A REPEATED FORCED-CHOICE LINEUP PROCEDURE: EXAMINING THE IMPACT ON CHILD AND ADULT EYEWITNESSES

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy
in
Experimental and Applied Psychology
University of Regina

By
Kaila Chantal Bruer
Regina, Saskatchewan
March, 2017

© 2017: K. C. Bruer
Kaila Chantal Bruer, candidate for the degree of Doctor of Philosophy in Experimental and Applied Psychology, has presented a thesis titled, *A Repeated Forced-Choice Lineup Procedure: Examining the Impact on Child and Adult Eyewitnesses*, in an oral examination held on March 24, 2017. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner: *Dr. Tammy Marche, University of Saskatchewan*

Co-Supervisors: Dr. Heather Price, Department of Psychology

Dr. Kristi Wright, Department of Psychology

Committee Member: Dr. Chris Oriet, Department of Psychology

Committee Member: Dr. Jeff Loucks, Department of Psychology

Committee Member: Dr. Rick Ruddell, Department of Justice Studies

Chair of Defense: Dr. Marc Spooner, Faculty of Education

*Via teleconference*
Abstract

In two experiments and one follow-up analysis, I examined the impact of using a repeated force-choice (RFC) lineup procedure with child and adult eyewitnesses. The RFC procedure divides the identification task into a series of exhaustive binary comparisons (i.e., round-robin design) and, in doing so, provides information about (a) who the witness believes is the suspect (if any) and, (b) additional information about how each face in the lineup matches the witness’ memory of a target, relative to every other face. Results from Experiment 1 indicate that younger children (6-to-8-year-olds) struggled with the RFC procedure, while older children (9-to-11-year-olds) performed at least as well with the RFC procedure as with a simultaneous procedure. In Experiment 2, the comparable performance in the simultaneous and RFC procedures was replicated with adult eyewitnesses. Follow-up analyses examined the additional information provided by the RFC in Experiments 1 and 2 and found evidence that witnesses’ patterns of responding during the RFC procedure can be used to estimate selection bias or memory strength associated with an individual witness’ lineup decision.
Acknowledgements

I would like to thank Social Science and Humanities Research Council of Canada (SSHRC), the American Psychology-Law Society (AP-LS), and the Faculty of Graduate Studies and Research at the University of Regina for their funding of this research and/or my studies. I would especially like to take a moment to express my gratitude for my supervisor, Dr. Heather Price. I wish to express my deepest gratitude for your training, advice, and wisdom. You are a mentor in the truest sense. Thank you for the long hours of discussion about my dissertation, research, and career. Without you, this would not have been possible.

I wish to thank my committee members, Dr. Jeff Loucks, Dr. Chris Oriet, and Dr. Rick Ruddell, for the helpful and suggestions made through this process. Their guidance helped to make this research what it is today. I also wish to thank the research assistants involved in data collection for this project, the wonderful staff of EYES camp, and all the parents and children who helped make this research idea a reality. To my family, thank you for your never-ending support and encouragement. You mean the world to me. I would also like to extend a special thank-you to Adam. Thank you for your patience and support and for listening to me talk about my research for many, many hours.

Post-Defence Acknowledgement

I would like to thank my external examiner, Dr. Tammy Marche, for her helpful discussion and comments. It was an honour to have my work reviewed by you.
Table of Contents

Abstract .......................................................................................................................... i

Acknowledgements ....................................................................................................... ii

Table of Contents .......................................................................................................... iii

List of Tables ................................................................................................................... viii

List of Figures ................................................................................................................ ix

List of Appendices ....................................................................................................... x

1.0 General Introduction ............................................................................................... 1

1.1 The Eyewitness Problem ....................................................................................... 3

1.2 The Eyewitness Identification Task ....................................................................... 3

1.3 Why Do Eyewitnesses Make Mistakes? ................................................................. 4

1.4 Using a Forced-Choice Model of Face Recognition ............................................... 5

1.5 Repeated Forced-Choice Procedure ..................................................................... 8

1.5.1 Stimulus set size ............................................................................................... 9

1.5.2 The repetition advantage ................................................................................ 9

1.6 Overview of Dissertation Research ...................................................................... 10

2.0 Experiment 1. The Repeated Forced-Choice Procedure with Children ............. 11

2.1 Experiment 1 Introduction .................................................................................... 11

2.2 An Historical Approach to Understanding Children as Witnesses ..................... 13

2.3 Why Do Child Eyewitnesses Make Mistakes? .................................................... 15

2.3.1 Pressure to choose. ......................................................................................... 15

2.3.2 Inhibitory control ......................................................................................... 16

2.3.3 Decision criteria ......................................................................................... 20
6.4 Present Experiment .................................................................83
7.0 Method ..................................................................................84
  7.1 Participants and Design .........................................................84
  7.2 Procedure .............................................................................84
8.0 Results ....................................................................................85
  8.1 Lineup Identification Choice .................................................85
    8.1.1 Target-present .................................................................85
    8.1.2 Target-absent .................................................................86
  8.2 Survival Statistics .................................................................87
  8.3 Lineup Comparison ..............................................................88
    8.3.1 Signal Detection Theory .................................................88
    8.3.2 Diagnosticity ratios .........................................................88
    8.3.3 Information gain .............................................................90
    8.3.4 BREE curves .................................................................91
  8.4 Supplementary Memorial Information ..................................91
    8.4.1 Confidence-Accuracy relationship ................................91
    8.4.2 Response latency ............................................................92
    8.4.3 Decision patterns and Z-scores ......................................93
    8.4.4 Predictive utility of Z-scores ..........................................94
  8.5 Individual Witness Discrimination .......................................94
    8.5.1 Index of discrimination ...............................................94
    8.5.2 Profile analysis .............................................................95
9.0 Experiment 2 Discussion ......................................................95
10.0 Additional Analyses. A Round Robin Model to Assess Recognition Memory

Strength with the RFC Procedure .................................................................................. 100

10.1 Additional Analyses Introduction ........................................................................... 100

10.2 Problems with Applying Traditional Analyses ..................................................... 101

11.0 Method and Results .............................................................................................. 102

11.1 A Round Robin Model using the Probability of Responses ................................. 102

11.2 Model Development ............................................................................................ 103

11.3 Application of Model Data to Experiment Data .................................................... 103

11.3.1 Model fit ........................................................................................................ 103

11.4 Using the Model to Interpret Witness Responding ............................................... 108

12.0 Additional Analyses Discussion ........................................................................... 112

12.1 Future Directions ................................................................................................. 112

13.0 General Discussion ............................................................................................... 114

13.1 Applied Implications ........................................................................................... 116

13.2 Theoretical Contribution ..................................................................................... 117

13.3 Future Research .................................................................................................. 120

13.5 Conclusion ........................................................................................................... 121

14.0 References ............................................................................................................ 122
List of Tables

Table 1. Identification Responses for Experiment 1 ......................................................... 44
Table 2. Suspect Survival Rates in Experiment 1 ................................................................. 48
Table 3. Signal Detection Theory Information for Experiment 1 ........................................ 51
Table 4. Diagnosticity Ratios for Experiment 1 ................................................................. 53
Table 5. Correlations between Confidence and Accuracy - Experiment 1 ....................... 60
Table 6. Frequency of Z-scores for each Decision Type for Experiment 1 ..................... 64
Table 7. Logistic Regression Model Information-Experiment 1 ..................................... 67
Table 8. Profile Analysis for Experiment 1 ....................................................................... 72
Table 9. Identification Responses for Experiment 2 ......................................................... 87
Table 10. Suspect Survival Rates in Experiment 2 ............................................................... 87
Table 11. Signal Detection Theory Information for Experiment 2 .................................... 88
Table 12. Diagnosticity Ratios for Experiment 2 ............................................................... 88
Table 13. Correlations between Confidence and Accuracy - Experiment 2 ................. 92
Table 14. Frequency of Z-scores for each Decision Type for Experiment 2 ................. 94
Table 15. Profile Analysis for Experiment 2 .................................................................... 95
Table 16. Model Fit Statistics for Experiments 1 and 2 ..................................................... 105
Table 17. Model Simulation Odds Ratio – Z-score Look-up Table ................................. 108
List of Figures

*Figure 1.* Images of the target-present and target-absent simultaneous lineups .......... 38

*Figure 2.* Sample of face-off stimuli during Rounds 1-3 and RFC during Rounds 1-2 ... 40

*Figure 3.* Sample of face-off stimuli during Round 4 and RFC during Round 3 .......... 40

*Figure 4.* Information gain in Experiment 1 .................................................. 55

*Figure 5.* Base-rate effect-equivalency (BREE) curve in Experiment 1. .................. 57

*Figure 6.* Mean response latency for decision made with the suspect ...................... 62

*Figure 7.* Plotted probabilities of response accuracy at each level of confidence ......... 69

*Figure 8.* Predictive probability of Z-scores with confidence ratings ....................... 70

*Figure 9.* Information gain in Experiment 2 ................................................. 90

*Figure 10.* Base-rate effect-equivalency (BREE) curve in Experiment 2 ................. 91

*Figure 11.* Mean response latency for Decisions made when the suspect ................. 93

*Figure 12.* P-values plotted for children and adults ..................................... 107
List of Appendices

Appendix A. Repeated Forced-Choice Program Design .................................................. 142
Appendix B. Pilot Testing .................................................................................................. 144
Appendix C. Parent Consent Form .................................................................................. 146
Appendix D. Child Assent and Debriefing ..................................................................... 148
Appendix E. Ethics Approval Certificate .......................................................................... 149
Appendix F. Lineup Decisions for each Target ................................................................. 150
Appendix G. Analyses of Guessing Behaviour in Children .............................................. 151
Appendix H. Z-score Calculation ..................................................................................... 153
**A Repeated Forced-Choice Lineup Procedure-Examining the Impact on Child and Adult Eyewitnesses**

**1.0 General Introduction**

“The spectre of erroneous convictions based on honest and convincing, but mistaken, eyewitness identification haunts the criminal law” (Justice David Doherty, *R v. Quercia*, 1990). As Justice Doherty notes, even with the best of intentions, eyewitness identifications can be inaccurate. Adults and particularly children struggle with lineup identifications (Fitzgerald & Price, 2015; Parker & Ryan, 1993; Pozzulo & Lindsay, 1998; Smith & Cutler, 2013). Misidentifications are problematic because they can implicate an innocent suspect in a crime and result in a wrongful conviction (Connors, Lundregan, Miller, & McEwan, 1996; Gross & Shaffer, 2012; Innocence Project, 2011; Wells et al., 1998). There is a strong reliance on eyewitness evidence when determining fact in criminal cases (e.g., Bradfield & Wells, 2000). Due to this reliance, a great deal of effort has been made to learn as much information as possible about a witness’ memory when assessing the likelihood of guilt.

Traditional lineup procedures involve the presentation of one (e.g., sequential) or more (e.g., simultaneous) photos at a time and typically provide one key piece of information: who in the lineup, if any, the witness believes to be the perpetrator. This information is obtained by asking a witness to make a single, categorical decision when presented with a suspect placed amongst fillers, or known innocents. This piece of information can often be supported using the witness’ confidence rating made following their decision. Given the high value placed on eyewitness evidence in the justice system (e.g., Beaudry, Lindsay, Leach, Mansour, Bertrand & Kalmet, 2015; Bradfield & Wells,
In this dissertation, I continue this exploration of alternative approaches to understanding eyewitness performance. Specifically, I explored a new lineup procedure—the repeated force-choice (RFC) procedure. The RFC procedure divides the identification task into a series of exhaustive binary comparisons and was designed to produce the same key piece of information that traditional lineups do (i.e., a categorical decision) but also to provide additional memorial information that can be used to understand more about a witness’ lineup behaviour. In Experiment 1, I compared how this new procedure works with child witnesses, relative to the face-off procedure (i.e., a paired-presentation style; Price & Fitzgerald, 2016) and the simultaneous procedure in which all pictures are shown at once. In Experiment 2, I compared adult witness performance using the RFC to the simultaneous procedure. Finally, in the additional
analyses, I developed a theoretical model to learn more about how responding during the RFC procedure can be used to better understand an individual witnesses’ memory for a suspect.

1.1 The Eyewitness Problem

The Canadian criminal justice system relies heavily on eyewitness evidence when determining fact in criminal cases. Since the year 2000, more than 2100 criminal bench cases have included testimony of an eyewitness (Canadian Legal Information Institute, 2014). Despite the value placed on eyewitness evidence, eyewitnesses of all ages have demonstrated a propensity for making errors, including misidentifications (Smith & Cutler, 2013). Misidentifications are problematic in that they can implicate an innocent suspect for a crime and result in wrongful convictions (Gross & Shaffer, 2012; Innocence Project, 2011) or they can contribute to a guilty person going free.

1.2 The Eyewitness Identification Task

The main task of an eyewitness is to help police identify a guilty perpetrator. Identification evidence is typically collected using a lineup task in which an eyewitness is presented with a series of photographs. A lineup usually contains a single suspect as well as a number of fillers (also referred to as foils) who are known innocents that are included to reduce undue focus on the suspect (Lindsay & Wells, 1985). In Canada, police agencies typically employ a 12- (Levi & Lindsay, 2001; Pozzulo, 2014) or 8-person lineup (Department of Justice, 2011). The benefit of including fillers in a lineup is that police and researchers gain additional information about the witness’ memory capacity and potential susceptibility to reliance on social cues. For example, an identification of a
filler is a known error and may suggest a weak memory for the suspect (Penrod, 2003; Wells & Olson, 2002; Wells et al., 2006).

When a lineup is compiled by police, two scenarios are possible. First, police may have caught the guilty perpetrator and, therefore, the suspect in the lineup is guilty (i.e., culprit- or target-present lineup). On the other hand, the police may have apprehended the wrong person and the suspect in the lineup is innocent (i.e., culprit- or target-absent lineup). An eyewitness is tasked with determining which picture (if any) is of the perpetrator. Witnesses can make one of three choices: identify the suspect (innocent or guilty), identify a filler (a known innocent), or decide the suspect is not present and reject the lineup (Charman & Wells, 2006; Lindsay & Wells, 1980). In applied legal settings, police do not know if a suspect included in a lineup is guilty or innocent. However, in a research setting, researchers have control over the task and are able to assess the frequency with which eyewitnesses choose innocent relative to guilty suspects. Legal professionals can use this type of information to assess the success of lineup procedures developed by applied researchers and thus, to inform lineup policies.

1.3 Why Do Eyewitnesses Make Mistakes?

Even under the most ideal conditions (e.g., strong memory, unbiased lineup instructions), eyewitness identifications can be inaccurate. Mistaken eyewitness identification is the leading cause of confirmed wrongful convictions in Canada (Association in Defence of the Wrongly Convicted, 2014) and in the United States (Innocence Project, 2014). Eyewitness researchers have revealed numerous factors that influence eyewitness identification. These factors have been divided into two categories (Wells, 1978). The first category, termed system variables, are factors the justice system
has control over, such as how police ask questions, how lineups are administered, or the time delay between witnessing an event and memory retrieval.

The other category, estimator variables, are factors that are out of the system’s control, such as how long the witness saw the perpetrator, how much lighting was present, and how tired or distracted the eyewitness was. As suggested by the name, the impact of estimator variables on a witness’ memory can only be estimated. A child’s cognitive development, for example, has the potential to influence how he/she is able to remember an event and understand instructions provided by police; therefore, level of cognitive development can be classified as an estimator variable. Although it is important to understand how both system and estimator variables impact memory, methods of improving eyewitness memory should target those system variables that the legal system has control over (Wells, 1988). As such, this research focused on a system variable, the lineup procedure, and how lineup administration processes can be adapted to overcome limitations faced by child and adult eyewitnesses and to learn more about recognition memory.

1.4 Using a Forced-Choice Model of Face Recognition

Accuracy of face recognition outside of the eyewitness context is typically assessed using two methods: the old/new (i.e., yes/no) format and the two-alternative forced-choice format (Jang, Wixted, & Huber, 2009). The old/new format involves serially presenting previously seen faces (multiple targets) among a series of new faces (lures) after which participants indicate whether they have seen the face before (“old”) or not (“new”). On the other hand, the two-alternative forced choice (2AFC) method tests how accurately a participant can discriminate between two visual stimuli that are
presented simultaneously (Bogacz, Brown, Moehlish, Holmes, & Cohen, 2006; Rakover & Cahlon, 2001). In this task, participants are shown a single target and then, during the testing phase, they are shown two faces and asked to decide which of the two is, or is most similar to, the target face (e.g., Jenkins, Lavie, & Driver, 2005).

The two-alternative forced choice model postulates that a witness integrates evidence favouring each option over time and, ultimately, he or she will make a final decision when a sufficient level of evidence has been accumulated favoring one of the options (Bogacz et al., 2006). Several models explain how a person integrates evidence over time. For example, the Drift Diffusion Model (Smith, 2000) argues that evidence is stochastically (randomly) accumulated for and against the two choices during each exposure until a decision threshold is met. In another model, the mutual inhibition model (Usher & McClelland, 2001), the accumulation of evidence for each option has an inhibiting effect on the other, such that as evidence favouring one increases, it mitigates the accumulation of evidence for the other. While the present research does not speak to the mechanisms behind integration of evidence, the aim was to examine whether an adapted two-alternative forced choice recognition model can be used to gather recognition memory evidence from eyewitnesses.

In the recognition memory literature the forced-choice recognition paradigm frequently results in more accurate recognition decision-making when compared to the old/new method for both adults (e.g., Deffenbacher, Leu, & Brown, 1981; Green & Moses, 1966; Jesteadt & Bilger, 1974; Jung et al., 2009) and children aged 4 through 11 (e.g., Deruelle, Rondan, Gepner, & Tardif, 2004; Hood, Macrae, Cole-Davies & Dais, 2003). This superiority of forced-choice methods is especially observed in situations with
efficient encoding conditions (good view, warned about memory testing; Deffenbacher et al., 1981).

However, the methodology used in this facial recognition research is difficult to apply to lineup research. Facial recognition research using an alternative forced choice method typically presents multiple targets to a participant (e.g., Deffenbacher et al., 1981). In contrast, a typical eyewitness context only shows one target amongst a series of distracters. Moreover, a critical factor that researchers propose as the driver behind the superiority of alternative forced choice methods over old/new methods is that the observer knows that one of the two items presented is a target. The observer can then simply choose the option that has a stronger match to their memory of the face previously seen. In contrast, in a lineup context witnesses may be shown a target-absent photo array.

Because of these substantial differences between the forced-choice and eyewitness paradigms, lineup research has focused primarily on adapting the old/new approach of face recognition to measuring eyewitness accuracy. Prior to development of the recent face-off procedure that involved showing two lineup members at a time to school-aged children (Price & Fitzgerald, 2016), no lineup researcher to my knowledge has utilized a forced-choice model to measure eyewitness accuracy.

In the present work, I developed a new lineup procedure that also utilized a forced-choice assessment approach. However, to adapt the forced-choice method to an eyewitness context, I introduced an element of repetition that resulted in a repeated, forced-choice procedure. With the repeated forced-choice procedure, witnesses are shown several pairs of lineup members and are told that, out of all of the pairs they will be shown, only one person may (or may not be) the target person. When shown each pair,
witnesses are asked, “Which of these two looks MOST like the person you saw?” This contrasts with some traditional alternative-forced choice models in which each pair contains one of multiple targets and participants are asked, “Which of these two faces is the one you saw.” Though the task is similar, the focus for participants is quite different in these two task (which is most similar versus which is the one you saw) which may produce rather different results than what has been observed in the facial recognition literature (i.e., superior recognition accuracy with forced-choice).

Different retrieval techniques may engage different types of memory processing. Dual-processing theories of recognition postulate that memory is the result of two distinct memory processes: recollection and familiarity (Jacoby, 1991; see Yonelinas, 2002 for a review). Recollection involves conscious, intentional remembering of episodic details. Familiarity, on the other hand, involves a much faster, unconscious form of memory processing that is more based on semantic (not episodic) details of the event/stimuli. It is possible that different retrieval techniques will activate different forms of memory processing. For instance, the question of, ‘which is MOST like the person you saw,’ may be more likely to engage familiarity processing. Alternatively, ‘which IS the person you saw,’ may be more likely to engage recollection processing. Although the present research does not directly assess evidence in support of a dual-processing theory of memory, it is important to consider how different questioning techniques may affect the accuracy of memory reports.

1.5 Repeated Forced-Choice Procedure

The repeated forced-choice (RFC) design is similar to a round robin design (see Fienberg & Wasserman, 1981; Wong, 1984) because witnesses are presented with all
possible pairings of lineup members and asked, for each pair, to choose who looks *most* like the target (i.e., ‘most-similar’ judgments). Once all pairings are decided, witnesses are provided with the two lineup members that were selected most often and are then asked to make another ‘most-similar’ decision between them. Lastly, the final remaining lineup member is compared against a blank silhouette picture (salient rejection option, or wildcard, indicating the absence of the target; Zajac & Karageorge, 2009) and witnesses decide if the remaining picture is, in fact, the target. There are two features of the RFC procedure that were expected to impact lineup performance: reduced stimulus set size (i.e., showing only two pictures at once) and repetition.

1.5.1 Stimulus set size. Simultaneous, sequential (Lindsay & Wells, 1985), and other adapted lineups procedures (e.g., elimination procedure; Pozzulo & Lindsay, 1999) all require witnesses to compress a substantial amount of complex information into a single decision. For example, when presented with a simultaneous lineup, a witness is required to compare all faces to their memory of a target, evaluate the strength of those comparisons, and make a decision. In visual search tasks, such as a lineup, attention attributed to the task is impacted by the set size (Palmer, 1994). Presenting witnesses, especially children, with a task that requires them to visually assess multiple options likely depletes their cognitive resources (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Pozzulo & Lindsay, 1999). Thus, reducing the sample size to include only two lineup members at once was expected to reduce the strain on cognitive resources that, in turn, may improve lineup accuracy for all children.

1.5.2 The repetition advantage. In the RFC procedure, a witness sees every lineup member paired off with every other member. In an eight-person lineup, this results
in a witness seeing each lineup member seven times\(^1\). Critically, the use of an alternative forced-choice model, like the RFC procedure, allows for the acquisition of additional information about a witness’ behaviour that would normally not be available (i.e., a quantitative assessment of recognition memory; Jang et al., 2009). That is, the repetitive nature of the RFC procedure allowed me to collect information about: (a) an eyewitness’ final lineup decision; (b) response latency for each decision; and, (c) each witness’ selection patterns of the suspect relative to the other faces in the lineup. Thus, the repetition involved in the RFC procedure provides an opportunity to collect supplementary information about a witness’ memory, without solely relying on his/her final lineup decision (and a possible accompanying confidence assessment). This information can be used to better understand a witness’ decision and may help investigators attribute weight to a witness’ decision.

1.6 Overview of Dissertation Research

The purpose of this dissertation was to examine the utility of a repeated-forced choice lineup procedure to measure recognition memory. This was done through conducting two experiments and one follow-up analysis. Experiment 1 compared children’s (aged 6- to 8- and 9- to 11-years-old) performance on the RFC procedure with two other lineup procedures: the well-established simultaneous procedure with wildcard (showing all lineup members at once) and the recently developed face-off procedure (showing lineups members in pairs; Price & Fitzgerald, 2016). Experiment 2 examined the impact of the RFC procedure on adult eyewitnesses’ lineup performance, relative to the simultaneous (with wildcard) procedure. In the follow-up analyses, a model was

\(^1\) The formula to determine the number of decisions based on lineup size is \(n^* (n-1)/2\). For example, a lineup containing six members would include 15 decisions \((6^*5/2)\).
developed and applied to data from Experiments 1 and 2 to provide additional memorial information that can be used to understand a witness’ responding during the RFC procedure. The final section provides a general discussion of this body of research.

2.0 Experiment 1. The Repeated Forced-Choice Procedure with Children

2.1 Experiment 1 Introduction

Children are frequently involved in the criminal justice system as victims or as witnesses and, in some cases, such as a sexual assault or abuse, a child may be the sole witness to a crime (Cain, Baker-Ward, & Eaton, 2005; Carter, Weithorn, & Behrman, 1999). While children are capable of accurately reporting personally experienced events (Ceci & Bruck, 1993), they are more likely than adults are to identify an innocent person as a suspect when the perpetrator is absent from a police lineup (Parker & Ryan, 1993; Pozzulo & Lindsay, 1998; Zajac & Karageorge, 2009). For example, on August 17, 1992, 11-year-old Holly Staker was sexually assaulted and murdered while babysitting two children. The only witness to this crime was Taylor, who was just two years old at the time. Taylor identified a man, Juan Rivera, who police had previously targeted as a suspect for the crime (National Registry of Exonerations, 2014). Over a decade later, DNA evidence was used to exonerate Rivera. Another case involving a child eyewitness is the wrongful conviction of Danny Brown. In 1981, Brown’s girlfriend was murdered and the only eyewitnesses were her three small children (6-year-old son and twin 3-year-old daughters). The 6-year-old identified Brown as the assailant (National Registry of Exonerations, 2014). Despite concerns raised about the reliability of a 6-year-old child’s identification, his testimony was used to convict Brown. Eighteen years after his conviction, Brown was exonerated using DNA evidence.
The cause of children’s poorer identification performance is not fully understood, but research suggests that it is likely the result of a convergence of memory, cognitive development (Simpson et al., 2012; Humphries, Holliday, & Flowe, 2012), and social-influence (Beal et al., 1995; Pozzulo & Lindsay, 1999) factors that differentially impact children and adults. Many studies have explored methods to improve how children perform on lineup tasks (e.g., Harvard & Memon, 2013; Price & Fitzgerald, 2016; Pozzulo & Lindsay, 1999; Zajac & Karageorge, 2009). This body of research has found that, with the proper supports (e.g., including a salient rejection option in a lineup; Zajac & Karageorge, 2009), children’s performance can improve. However, we still lack an intervention that can reliably reduce children’s problematic choosing (Brewer, Keast, & Sauer, 2010; Fitzgerald & Price, 2015). Child eyewitnesses present a unique challenge and, therefore, more work is needed to understand why children perform more poorly on lineup tasks than adults.

The goal of this experiment is to explore the utility of a novel lineup procedure for use with child eyewitnesses. This lineup procedure, the repeated forced-choice method, was designed as a supportive tool to improve children’s identification performance. To understand the effectiveness of this new procedure with children, I compared identification accuracy associated with the repeated forced-choice procedure with accuracy from two other procedures: the well-established simultaneous procedure (showing all lineup members at once) that has become a standard procedure in the child eyewitness literature against which new procedures are compared (e.g., Beresford & Blades, 2006; Pozzulo & Balfour, 2006), and the recent face-off procedure (showing lineups members in pairs; Price & Fitzgerald, 2016).
2.2 An Historical Approach to Understanding Children as Witnesses

A primary concern regarding children as witnesses pertains to the reliability of the information they report. Reliability refers to a child’s ability to observe, store, retrieve and communicate accurate information about a witnessed event (Ceci & Bruck, 1993). Researchers from both cognitive and social areas of psychology have sought to explore this problem. Prior to the 1980s, much of what we knew about children’s potential as witnesses came from cognitive research on children’s face recognition. As discussed earlier, accuracy of face recognition is typically assessed using one of two methods: the old/new (i.e., yes/no) format and the two-alternative forced-choice format (Jang et al., 2009). The forced-choice recognition paradigm has been used a great deal with children in the face recognition literature and has been able to establish basic developmental differences in face recognition (e.g., Deruelle, Rondan, Gepner, & Tardif, 2004; Hood et al., 2003). This face recognition research has generally found that recognition accuracy improves with age, especially between the ages of 5 and 12 (e.g., Carey, Diamond, & Woods, 1980; Chung & Thomson, 1995; Diamond & Carey, 1977; Ellis, Shepherd, & Bruce, 1973). These results contributed to a common perception that young children make poor witnesses because of poorer face recognition abilities (Ceci, Ross, & Toglia, 1987).

However, in the 1980s, a series of high-profile American cases involving children as witnesses increased interest in researching children as eyewitnesses (Pozzulo, 2007). With this increased focus on children as witnesses, face recognition research came under scrutiny for its lack of ecological validity and application in the legal system. In response to this criticism of face recognition literature, social and cognitive researchers began
conducting lineup experiments in which children would be unknowingly exposed to an event (either live or video) during which they would see a target person. Then, using a variation of the old/new format, eyewitnesses were provided with a lineup and asked to identify the target among a series of new faces. Lineup experiments are arguably much less taxing on a child’s working memory than recognition studies because they are required to remember only one target—rather than upwards of 20 (Pozzulo, 2007).

An influx of research with children revealed that child witnesses were capable of encoding and retrieving accurate information, given favourable circumstances, such as clear, unbiased instructions or good encoding conditions (e.g., Burgwyn-Bailes, Baker-Ward, Gordon, & Ornstein, 2001; Ceci & Bruck, 1993; Peterson & Whalen, 2001; Pozzulo & Lindsay, 1998; Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998). Similar conclusions began to emerge from research on children’s ability to remember autobiographical events (e.g., Fivush, 1993; Hudson & Fivush, 1991; Nelson & Greundel, 1981).

However, despite the evidence that children are capable of retrieving accurate information about experienced events, lineup research also revealed that children do not always perform at comparable levels with adult eyewitnesses (Fitzgerald & Price, 2015). Fitzgerald and Price meta-analyzed 62 eyewitness studies that examined children’s performance on lineup tasks and concluded that children are less likely than adults to correctly identify a guilty suspect and more likely to misidentify an innocent lineup member when no guilty suspect is present. In short, the literature shows that compared to adults, child eyewitnesses are more likely to choose from a lineup (Fitzgerald & Price, 2015; Parker & Ryan, 1993; Pozzulo & Lindsay, 1998). This high rate of choosing, of
course, can be problematic when memory is poor or when the suspect is innocent. The cause of children’s high rate of choosing is not fully understood, but there has been a great deal of research attempting to understand why this robust age difference exists and to develop strategies to reduce or eliminate children’s problematic choosing.

2.3 Why Do Child Eyewitnesses Make Mistakes?

In exploring the social, cognitive, and developmental memory literature, three key factors emerge as likely contributors to children’s problematic choosing in target absent lineups: perceived social pressure to choose, underdeveloped impulse control, and use of overly lenient decision criteria.

2.3.1 Pressure to choose. The first factor relates to children’s vulnerability to demand characteristics or social pressure to choose from a lineup (Beal et al., 1995; Parker & Ryan, 1993; Pozzulo, 2007; Pozzulo & Lindsay, 1998). There is a level of implicit pressure to make a choice when presented with a lineup. For instance, Wells and Luus (1990) argued that, when presented with a lineup task, participants might have the impression that not making a selection is akin to being unwilling to complete the task. While eyewitnesses of all ages may experience pressure to choose, adults may be better able to resist such pressure (Beal et al., 1995). Children, on the other hand, may be particularly susceptible to pressure because the person asking them to make a decision is perceived as an authority figure (e.g., they want to make the adult/police officer happy; Pozzulo & Lindsay, 1998). Moreover, Dunlevy and Cherryman (2013) argued that because children are not able to fully consider an interviewer’s motive for asking them to view the lineup (e.g., to see if the target is there or not), they only focus on the outcome
of the identification task. That is, children may consider making a choice from a lineup as more helpful than not making a choice.

Some research has found evidence that social pressure is related to children’s problematic choosing. Pozzulo and Lindsay (1992), for example, found that when witnesses are given the option to say, “I don’t know,” children do so far less often than adults do. More recently, Pozzulo, Dempsey, Bruer, and Sheahan (2012) designed an experiment to reduce the cognitive load placed on children during a lineup task in order to examine the influence of social factors. To do this, the authors used two cartoon targets that were well known to the children and argued that near-ceiling correct identification rates could be expected in target-present lineups. The results confirmed this, with near-perfect identification rates when the two targets were present (100% and 97% for each target). If choosing in a target-absent lineup was primarily based on memory of the familiar target, children would be expected to also have a 100% correct rejection rate (clearly the well-known target is not there). However, correct rejection rates for the two cartoon targets were much lower than expected (67% and 80%). Thus, even though the children were familiar with the targets, some still did not indicate that the target was missing from the pictures. These results suggest that children’s inaccurate responding in target-absent conditions is, at least in part, due to social factors. As such, there is a need for child-friendly identification procedures to reduce the perception that they have to choose from a lineup.

2.3.2 Inhibitory control. The second factor that contributes to children’s problematic choosing is impulsivity in responding that results from underdeveloped inhibitory control (Simpson et al., 2012). Inhibitory control can be understood as the
ability to ignore task-irrelevant information, processes, and automatic or prepotent responses (Levy & Anderson, 2002; Roberts & Powell, 2005). When children are presented with a task that requires the use of executive functions, they need to be able to inhibit a prepotent response in order to engage in active computation and determine a correct response (Gerstadt, Hong, & Diamond, 1994). However, decades of executive control research have established that, compared to adults, children have less developed inhibitory control (e.g., Davidson, Amso, Cruess, Anderson, & Diamond, 2006) and thus are not able to easily reach the active computation stage. Notable improvements in inhibitory control have been documented between the ages of 4 and 7 (Dempster, 1993) and between 6 and 9 (Williams et al., 1999), wherein children begin to demonstrate improved suppression of automatic or prepotent responses (Bruck & Ceci, 1997; Diamond, Kirkham, & Amso, 2002; Williams et al., 1999).

Inhibitory control is a fundamental component of accurate performance during an identification task. Roberts and Powell (2005) argued that accurately reporting target information is related to the ability to inhibit irrelevant (i.e., non-target) information. Inhibition likely plays a greater role in children’s responding when children are presented with a target-absent lineup than with a target-present lineup. More specifically, a target-absent lineup presents a situation in which there is no direct match between the lineup members and memory (because the target is not in the lineup) but there are, however, many similar faces presented. Due to the implicit social pressure to choose during a lineup task, a child’s natural inclination (or prepotent response) is arguably to select an individual from a lineup, even if they do not see the target. However, children are
typically instructed to do just the opposite when they do not see a target—say the target is “not here.”

Similar to the finding that increased inhibitory control is needed to prevent retrieval of non-targets when targets and non-targets are high in similarity (Anderson, Bjork & Bjork, 1994; Roberts & Powell, 2005), a target-absent lineup may increase reliance on response inhibition to prevent selection of these similar, but non-target faces. However, because children have limited ability to inhibit prepotent responses, they tend to make impulsive decisions (Brewer & Day, 2005; Diamond et al., 2002). For example, Roberts and Powell (2005) explored the relationship between inhibitory control and children’s suggestibility. In this study, children (5- to 7-year-olds) witnessed a staged event and were asked if they recognized target (correct) and non-target (inaccurate) information. They found that children who demonstrated high levels of response inhibition were less susceptible to affirming the presence of suggested details. Similarly, Humphries and colleagues (2012) found that one third of children selected the first of nine pictures presented when lineup members were presented sequentially. This tendency to choose implies that this behaviour is a result of impulsivity and may be an expression of pre-potent responses.

Despite children’s impulsivity and underdeveloped inhibitory control, there is evidence suggesting that, with support, children’s performance on inhibition tasks can be improved. For example, strategies that introduce delays in responding (increasing response latency) have been found to help children reach the active computation phase because it gives time for prepotent responses to pass (Diamond et al., 2002; Simpson et al., 2012). Diamond and colleagues (2002) for example, used the day/night task (see
Gerstadt et al., 1994 for a description) and found that when provided with a cognitive crutch (i.e., presenting instructions as a jingle) children took more time to think about instructions and, as a result, they were better able to accurately apply complex rules. There is also evidence that shows slowing response time improves young eyewitness’ performance during lineup tasks (Beal et al., 1995; Bruer & Pozzulo, 2014), especially in target-absent conditions (Dunlevy, 2005; Dunlevy & Cherryman, 2013). Bruer and Pozzulo (2014), for example, explored response latency in young children (4- to -6-year-olds) and adults when making lineup decisions. The researchers found that, unlike adults where speed led to improved accuracy, children were more accurate when they took more time to make a decision. Beal and colleagues (1995) similarly found that young children (5-to-6-year-olds) were more accurate when they made decisions more slowly. Dunlevy and Cherryman (2013) found that increased response latency was related to children’s higher accuracy rates only in target-absent lineups. Children may have faster response times when a target is present because decisions are based on recognition memory. When the target is absent, a child must engage more deliberate, time-consuming cognitive processes to correctly reject a lineup (e.g., inhibiting of prepotent responses).

In short, a lack of fully developed inhibitory control likely contributes to children’s tendency to misidentify innocent people from lineups. Therefore, innovations in lineup procedures would benefit from a focus on overcoming children’s executive control limitations. Modifying a lineup procedure to slow down responding should have several benefits for children. Slowing down decision latency provides children much needed time that allows them to think about the task and, in turn, gives more opportunity to overcome their natural inclination to choose. Increased time is especially important in
target-absent conditions in which a child’s inclination to choose may be inflated due to a lack of a strong match between their memory of a perpetrator and the lineup stimuli.

2.3.3 Decision criteria. The third factor that likely contributes to children’s problematic choosing is their use of overly lenient decision criteria. The ability to accurately monitor and regulate decision criteria is dependent upon a child’s metacognitive abilities. Metacognition generally refers to the awareness, monitoring, and regulating of one’s cognitive processes, including memory (Flavell & Wellman, 1977; Haller, Child, & Walberg, 1988). Research has demonstrated that children lack the metacognitive capacity to effectively regulate their recognition memory reporting (Bryce & Whitebread, 2012; Haller et al., 1988; Keast, Brewer, & Wells, 2007; Roebers, 2002; Roebers & Howie, 2003).

Without the ability to monitor and regulate their cognitive processes, are children able to recognize when they do not recognize a face and adjust their decision criterion accordingly? According to Dunlevy and Cherryman (2013), children do adjust their decision criteria—but not in the anticipated direction. Instead of using more conservative decision criteria, children use more lenient decision criteria when there is no close match or when their memory for a perpetrator is weak. Evidence of children’s use of overly lenient decision criteria can be seen by examining differences in correct identification rates in different lineups. Humphries and colleagues (2012), for example, examined the impact of simultaneous and sequential lineups (individual presentation of lineup members), and a type of lineup specifically developed for children, the elimination video lineup, on young (5- to 6-year-olds) and older (9- to 10-year-olds) children’s identification performance. Compared to the other lineups, sequential lineups produced
the lowest rate of correct identifications. The authors argued that this reduction in correct identifications can be explained by the use of more conservative decision criteria (more incorrect rejections of the lineup) for older children; however, the pattern of responding for the young children was not as easily explained as they were more likely to misidentify a lineup filler. To explain this, the authors argued that young children’s responding was driven by a lack of understanding of the task. It is also possible that high incorrect choosing in young children could be explained by the use of more lenient decision criteria (causing them to choose more) that reflects poor understanding of the consequences of the lineup task.

Under Signal Detection Theory, a witness’ choice of criterion may depend upon the perceived consequences of outcomes. For example, if consequences for saying a suspect is in a lineup when, in fact, he is not present are perceived by a witness to be costly, then a witness may be less willing to make that choice. However, if a witness perceives a high cost to not making a selection from a lineup, then he/she may be more likely to make an identification. Children may not be fully able to understand an investigator’s motive for asking them to complete a lineup task (Dunlevy & Cherryman, 2013). Instead, children may focus on egocentric consequences of a lineup decision (e.g., “I want to do a good job” or “I want to please the adult”), rather than broader consequences of their lineup decision (“I want to make sure the correct person gets into trouble”).

Brewer and colleagues (2010) argued the key to reducing children’s positive identification in target-absent conditions is to encourage children to use stricter decision criteria. However, considering that children may not possess the metacognitive abilities
to easily adjust decision criteria (Keast et al., 2007) this can pose a large obstacle to overcome. In order to mitigate problems associated with using overly lenient decision criteria, particularly when no suspect is present, it would be beneficial to conduct lineup procedures that acquires as much information about a child’s memory as possible, with minimal reliance on a child’s final decision (that is dependent upon decision criterion). The repeated forced-choice lineup procedure was designed to capture information about a child’s final lineup decision, but also additional supplementary information that may be used to assess children’s memory without relying solely on their final decision.

2.4 What We Know About System Variable Changes

Eyewitnesses are usually provided with the opportunity to provide a verbal description of a suspect as well as the opportunity to identify the suspect from a lineup. Wells (1993) characterized the lineup as a way of collecting information about a particular component of an eyewitness’ memory that cannot be articulated—recognition memory. The lineup is a critical indicator of eyewitness performance and, because it is something the legal system has control over, variations to lineup identification procedures are one of the most researched techniques to improve performance of child eyewitnesses.

2.5 Changing Lineup Instructions and Contents

Children’s problematic choosing behaviour is thought to be related, at least in part, to a lack of awareness of rejection as an option and the guessing that stems from perceived pressure to make a selection from the lineup (Gross & Hayne, 1996). One method that has shown to be very successful in reducing misidentification with adult witnesses, and is now considered best practice (National Research Council of the National Academies, 2014), is to include non-biased instructions that tell the witness that
the guilty perpetrator ‘may or may not be here’ (Steblay, 1997). However, these instructions alone do not have the same impact on children’s identification performance as they do for adults. Thus, researchers have looked to other methods, such as changing the composition of a lineup, to make rejection a more salient option for children.

Some researchers have found that adding a visual image in the lineup to represent the absence of the perpetrator provides children with a non-verbal method to reject a lineup and reduces misidentifications (Davies, Tarrant, & Flin, 1989; Dunlevy & Cherryman, 2013; Harvard & Memon, 2012; Zajac & Karageorge, 2009). The introduction of these salient rejection options has been generally successful at reducing children’s problematic choosing. For example, Zajac and Karageorge (2009) were able to reduce inaccurate identifications made by children by 25% in target-absent lineups (at no cost to target-present performance) by including a silhouette shape with a question mark in the lineup (termed the ‘wildcard’). The success of the wildcard demonstrates how, with proper supports, children’s performance on lineup tasks can be improved. To determine if the current investigation demonstrates improvement beyond that of salient rejection options, Zajac and Karageorge’s (2009) wildcard was included in all lineups.

2.6 Changing the Way Lineups are Presented

Another method that has shown some success in reducing children’s problematic choosing has been to adjust the presentation of images in a photo lineup. There are two traditional techniques of presenting photographic lineups to witnesses. The simultaneous procedure, involves showing all lineup members (suspect and fillers) at the same time and the witness identifies which individual—if any—is the perpetrator. Presenting the images all at once is argued to evoke relative judgment (i.e., comparative decision-
making; Wells, 1984; Wells et al., 1998). The use of a relative judgement strategy is thought to work well when the target is present among the photographs, but is potentially problematic when the target is absent as there will always be someone in the lineup who more closely matched the witness’ memory of the perpetrator than other lineup members (Wells et al., 2006). Alternatively, the sequential lineup procedure involves presenting one lineup member at a time and the witness is required to decide whether the image is of the perpetrator before seeing the next image. Serial presentation of images is argued to evoke an absolute judgment (Wells, 1984; Wells et al., 1998) in which the witness is required to compare the image with only their memory of the perpetrator and, as a result, to use more conservative decision-making (Lindsay, Mansour, Beaudry, Leach, & Bertrand, 2009; Palmer & Brewer, 2012; Steblay, Dysart & Wells, 2011; Wells et al., 1998). Although children may benefit from a procedure that encourages more conservative responding (e.g., Dunlevy & Cherryman, 2013), use of the sequential lineup procedure does not eliminate problematic choosing for children (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997; Parker & Ryan, 1993; Pozzulo & Lindsay, 1998; Steblay et al., 2011).

Given that eyewitness identification errors can result in wrongful convictions, these traditional approaches to administering lineups have been carefully scrutinized by scholars (Cutler & Penrod, 1995; Sauer et al., 2008; Wells et al., 1998). An argument has recently been put forth suggesting that eyewitness researchers have been working within boundaries placed on them via the legal system and that there is a need to develop alternative approaches to understanding how to improve eyewitness performance. To quote Wells, Memon and Penrod, (2006):
It could be argued that research has been profoundly conservative in its approach to the eyewitness-identification problem. Specifically, researchers have tended to operate within the confines of the traditional lineup, in which a suspect is placed among fillers and the eyewitness makes a verbal identification. But what if the lineup had never existed and the legal system turned to psychology to determine how information could be extracted from eyewitnesses’ memories? (pp. 68)

The problem with high misidentification rates may be, at least in part, the product of how these traditional lineup tasks are conducted (Hiller & Weber, 2013). Moreover, these traditional lineup procedures were not designed with children in mind and, as such, do not meet the special developmental needs of children. Both simultaneous and sequential lineup procedures require witnesses to compress a significant amount of complex information into a single decision (Sauer et al., 2008). For example, when presented with a simultaneous lineup, a child is required to compare all faces to their memory of a target, evaluate the strength of those comparisons, and make a decision. Presenting children with a task that requires them to track multiple options (i.e., cost of choice or search cost; Vohs, Baumeister, Schmeichel, Twenge, Nelson, & Tice, 2008) likely depletes their cognitive resources (Pozzulo & Lindsay, 1999). The sequential lineup, and to a lesser extent, the simultaneous lineup, requires the eyewitness to use a conservative decision criterion when making an identification in order avoid identifying an innocent person—something that children tend not to do (Dunlevy & Cherryman, 2013). In response to these concerns, some researchers have designed innovative, developmentally appropriate lineup procedures to overcome these issues.
2.6.1 Elimination procedure. The elimination lineup procedure was developed by Pozzulo and Lindsay (1999) as a means to improve children’s accuracy by overcoming developmentally based cognitive limitations. Pozzulo and Lindsay argued that a simultaneous lineup task was too difficult for a child and, in response, divided the task into two judgements, rather than one, in order to reduce the cognitive load. The first judgement involves presenting the photographs simultaneously and requesting that the child pick the one that looks MOST like the guilty person (i.e., relative judgement). Then, all but the selected picture is removed and the child is left with one picture and asked to decide if this is, in fact, the guilty person (i.e., absolute judgement). Removing all the unselected pictures is argued to reduce the cognitive load, allowing children to focus on just one face rather than several. When comparing the elimination to the simultaneous procedures, the results were promising. Using the elimination procedure with children increased the correct rejection rate with no cost to correct identification rates (Pozzulo & Lindsay, 1999). This effect has been replicated by Pozzulo and colleagues (3-6-year-olds; Pozzulo, Dempsey, & Crescini, 2009; 8-13 year olds; Pozzulo & Balfour, 2006).

Although the elimination procedure is a promising method to improve children’s performance, there is room for improvement. For example, the first task still places a large cognitive load on children. It still requires children to compare all faces to their memory of a target, evaluate the strength of those comparisons, and make a ‘most similar’ decision. Children are required to condense a lot of complex information into a single decision of who looks most like their memory of the target. Moreover, there is new evidence that suggests that the nature of the elimination task may further encourage choosing in children. A recent meta-analysis (Fitzgerald & Price, 2015) revealed that the
use of the elimination procedure with children appears to produce correct identification rates comparable with those observed in adults; however, the authors caution that the comparable correct identification may be the result of the elimination procedure increasing overall choosing behaviour in children (54%) relative to adults (38%).

2.6.2 Face-off procedure. To further reduce the cognitive load placed on children, Price and Fitzgerald (2016) tested a new lineup procedure with children, called the face-off procedure. Unlike previous methods that focused on the absolute versus relative judgment premise of identification decisions, the theoretical foundation of the face-off method was built on the notion of reducing stimulus set size and was designed to increase the inhibition of prepotent responses by delaying an identification decision. The procedure involved showing witnesses two images. From those two, they must choose which one looks most similar to a perpetrator\(^2\). The ‘loser’ is eliminated from the procedure and the ‘winner’ is subsequently compared to another ‘winner’ image. This procedure continues until the final picture is presented to the witness, who then must decide whether it is a picture of the guilty perpetrator. By presenting the stimuli in pairs, the researchers argued that making a single judgement about a pair of photographs would reduce the strain on cognitive resources because children would not have to make multiple judgements about multiple pictures. Relative to simultaneous and elimination

\(^2\)This was not the first procedure that presented lineup members in pairs. Wells and Pozzulo (2006) examined adult eyewitness’ memory of multiple perpetrators using six, two-person lineups, presented serially. In this procedure, participants were shown pairs of lineup members and were told that each pair may or may not contain a picture of either the assailant or accomplice, but not both. For each pair, participants were required to make a decision before moving onto the next pair. Though not significant, Wells and Pozzulo found a trend suggesting that two-person serial lineups produced more correct rejections than a simultaneous or sequential lineup.
lineups, Price and Fitzgerald (2016) found that the face-off procedure reduced children’s problematic choosing when the target was not included in the lineup, and this benefit did not come at the cost of correct identifications. Price and Fitzgerald (2016) found that the face-off procedure reduced false suspect identifications by up to 25%, (Exp. 2) relative to simultaneous procedures.

One of the strengths of the face-off procedure is that it helps children inhibit prepotent responses by requesting several decisions, rather than one. This caters to children’s urge to choose, while slowing down decision-making, ultimately giving children more time to respond accurately. In addition, by ‘chunking’ the decisions into a series of simple two-person comparisons, this procedure presumably reduces the cognitive load on the child. Despite the clear advances both from an applied and theoretical standpoint, there are limitations that stem from the design of the face-off procedure. The procedure evokes repeated viewing of only some of the lineup members. For example, if correctly identified, a guilty suspect is viewed four times, whereas a lineup member eliminated in the first stage is only viewed once. It is possible that repetition of only few faces can create a commitment-based confound in children’s choosing. Commitment bias is a tendency to select a suspect from a lineup on the basis that he/she has seen the suspect on multiple occasions previously (e.g., in another lineup) and wants to maintain consistency in their choosing (Lampinen, Neuschatz, & Cling, 2012). One way to maintain the advantage of repeated choosing but mitigate the risk of a commitment effect is by ensuring that all lineup members are presented an equal number of times.
2.7 Changing the Type of Information Collected from a Lineup Decision

Typically, eyewitness identification performance is measured using a witness’ categorical lineup decision (i.e., suspect identification, filler identification, rejection). The problem with relying on categorical decisions is that they are dependent upon a witness’ decision criterion. This is especially problematic with child witnesses who, due to various factors (e.g., limited metacognitive ability, social pressure), appear to use overly lenient decision criteria. In order to reduce children’s high choosing rates, some research has focused on mitigating the problems that cause witnesses to lower their decision criteria and choose more often (Bruer, Fitzgerald, Price, & Sauer, 2017; Sauer et al., 2008). For example, Sauer and colleagues (2008) mitigated the impact of the potential problem of decision criteria with adult witnesses by using a confidence procedure that removed the need for witnesses to use one decision criterion to make a single decision.

Sauer and colleagues asked adult witnesses to make a confidence judgement for each lineup member and, using the confidence ratings, inferred lineup decisions by using a criterion that maximized the combined proportion of correct identifications and rejections (fit ratio; see Sauer et al., 2008). Bruer and colleagues (2017) tested this procedure with child eyewitnesses ($M_{age} = 10$) and compared the accuracy of responses to a traditional categorical decision. The researchers found that using children’s confidence ratings led to more accurate identification responses than traditional lineup procedures. That is, relying on children’s confidence ratings allowed the children to indicate the degree of match between each picture and his/her memory of the perpetrator, without relying on their ability to condense complex information into a single, categorical response. Given that the use of an appropriate decision criterion appears to be a challenge
for children (Brewer et al., 2010), a lineup procedure that does not solely rely on a child’s identification decisions when determining the strength of a witness’ memory helps to mitigate problems of decision criterion that may otherwise impact a child’s final decision.

2.8 Repeated Forced-Choice Procedure and Children

In the present work, I examined how a new lineup procedure that utilizes a forced-choice assessment approach would impact child eyewitness accuracy. Charman and Wells (2006) argued that theoretical considerations of a lineup procedure should consider the constraints and applied goals of those who use lineups in the legal system. When dealing with child witnesses, a constraint that needs to be addressed is children’s tendency to choose from a target-absent lineup (Pozzulo & Lindsay, 1998). As reviewed above, this problematic behaviour is likely impacted by developmental factors that uniquely impact children, such as immature inhibitory control, perceived social pressure to choose, and use of lenient decision criteria. For a lineup procedure to be effective with children, it should address three issues. First, a lineup procedure should be designed in such a way that it considers the social demand to choose. Second, it must be carried out in such a way that children will be given ample time and opportunity to inhibit prepotent responses to choose from a lineup. Third, a lineup procedure should take into account the metacognitive abilities of children that impact decision criteria. When designing the RFC procedure, consideration of these issues facing child eyewitnesses was important. As previously noted, the key benefits of the RFC were expected to stem from two factors: reduced stimulus set size (i.e., showing only two pictures at once) and repetition.

2.8.1 Stimulus set size. Children are thought to struggle with a traditional lineup task because children are required to track and compress a large amount of information
into a single, complex decision (who, if any is the guilty person) which depletes their cognitive resources (e.g., Hiller & Weber, 2013; Pozzulo & Lindsay, 1999). In a recent study introducing the face-off lineup procedure with school-aged children (aged 6- to 15-years old), Price and Fitzgerald (2016) found that reducing the stimulus set size to a series of pairs improved the accuracy of children’s responding in target absent lineups, relative to the simultaneous procedure. Price and Fitzgerald argued that showing two faces at once is advantageous because it provides children with a clear structure to guide their use of relative judgement during a lineup task.

In keeping with this argument, I also hypothesized that presenting eyewitnesses with pairs of faces will allow for the assessment of discriminability between previously seen and previously unseen faces using a presentation method that is less cognitively-taxing on witnesses. Presenting complex stimuli in smaller pieces is expected to reduce the amount of target irrelevant information presented to the child, thereby reducing the cognitive load placed on the child during the task. Additionally, presenting the stimuli in pairs ensures that the children will have a chance to look at all lineup members, without having to examine all lineup members at once. Thus, it was hypothesized that paired-presentation would translate into more accurate responding in lineup procedures with a smaller stimulus set size, including the RFC procedure.

The present study compared the RFC procedure with the face-off procedure (Price & Fitzgerald, 2016) and the simultaneous (with wildcard) procedure. Both the RFC and face-off procedures present lineup members in pairs and, as such, both procedures were expected to produce more accurate responding than the simultaneous procedure. However, I also expected to observe differences between the face-off and RFC
procedure. Although the face-off and RFC procedures utilize similar presentation formats (i.e., paired presentation of lineup members), the RFC is unique in that it introduces repetition.

2.8.2 Repetition. A key distinction between the RFC procedure and other child-friendly lineup procedures is the repetitive nature of the RFC task. In the RFC procedure, a witness sees every lineup member paired off with every other member. There are several potential benefits of introducing repetition into the procedure. First, repeated-choosing is expected to reduce problems associated with perceived social pressure to choose because the RFC procedure requires that children choose each time a pair of pictures is presented. Children choose at high rates from lineups (Brewer & Day, 2005) and this choosing is, in part, due to perceived social pressure (Beal et al., 1995). Therefore, explicitly instructing children to choose repeatedly between lineup members allows this choosing behaviour to happen with minimal impact on the final decision.

Second, repetition should help mitigate the negative impact of underdeveloped inhibitory control on children’s overall responding. By increasing the number of decisions (by repeating the faces), the exposure time to all the faces is increased as well as the total time taken before a final decision is made. If a child is asked to make a decision about a lineup member more than once, initial selection of that lineup member that may have resulted from prepotent responding can be corrected with the next several decisions. That is, increasing the number of decisions is expected to provide children with the time needed to engage in active computation to make a decision based on their recognition memory. Thus, it was hypothesized that the RFC procedure would result in improved
performance for children, relative to the other lineup procedures that do not involve repeated viewing of all lineup members (i.e., simultaneous and face-off procedures).

Lastly, the repetitive nature of the RFC method has the potential to provide an index of recognition memory that may be used to manage issues related to metacognitive development. As discussed earlier, choosing is problematic because children may not have the ability to engage in the metacognitive processes needed to adjust their decision criteria (e.g., Bryce & Whitebread, 2012; Keast et al., 2007). Using an adapted alternative forced-choice model like the RFC procedure allows for a quantitative assessment of recognition memory (Jang et al., 2009). In the RFC procedure, children are asked to compare all lineup members and, in doing so, they effectively ‘semi-rank’ the faces from most similar to their memory of the target to least similar. The pattern of children’s responding can provide an index of the degree of match between each picture and the children’s memory of the target. Critically, this information stemming from the repetitive nature of the RFC provides an opportunity to gather additional information about the strength of a witness’ memory, without solely relying on his/her final lineup decision. This method of assessing children’s recognition memory is then less susceptible to the problems that may otherwise cause witnesses to lower their decision criteria and choose more often. This information may then be used to contextualize a witness’ decision and help investigators attribute weight to a witness’ decision. Therefore, I hypothesized that the information about children’s behaviour (e.g., choosing patterns) collected during the RFC procedure would be predictive of identification accuracy.

### 2.8.3 Age as a factor

Social pressure, inhibitory control, and metacognitive development are key factors informing the hypotheses of this study. However, age is an
important consideration when predicting how these factors may influence performance on the RFC procedure. In this experiment, I examined age categorically, rather than continuously for a number of reasons. A recent meta-analysis reported that children demonstrate improvement and start performing more like adults on lineup tasks around the age of 9 years (Fitzgerald & Price, 2015). As children age past 9 years, a more gradual increase with an eventual plateauing of identification accuracy is expected. Thus, I expected to find better overall lineup identification performance with the older age group than with the younger age group.

Moreover, meta-cognition and executive functioning is thought to play a role in children’s identification performance (Hiller & Weber, 2013; Koriat, Goldsmith, Schneider & Nakash-Dura, 2001; Roebers & Howie, 2003). Previous research has found that around the age of 9 to 10, children begin to perform more like adults on meta-cognitive association tasks (e.g., Schneider, 1986; Wellman, 1978). For this reason also, I hypothesized differences between younger and older children’s performance. The above stated benefits of the RFC procedure were hypothesized to improve older children’s performance, relative to the standard, or simultaneous, lineup procedure. For younger children, the direction of the hypotheses was less clear. It was possible that use of the RFC procedure would help younger children overcome some of the more pronounced developmental limitations that normally interfere with a lineup decision (e.g., compressing all lineup information into one, single decision). However, it is also plausible that proper use of a decision-intensive procedure—like the RFC—requires more developed meta-cognitive and executive functioning levels. Thus, how the RFC procedure would work with young children was more exploratory.
2.9 Hypotheses

1. The smaller stimulus set size and repetition involved during the RFC procedure would produce higher identification accuracy than either the simultaneous or the face-off procedure.

2. The supplementary information provided by the RFC procedure was hypothesized to be indicative of identification accuracy.

3. The benefits of the RFC were hypothesized to translate into more accurate responding with the older children relative to the younger children.

3.0 Method

3.1 Design

This study was a 3 (lineup procedure: simultaneous, face-off, RFC) x 2 (target: present, absent) x 2 (age: younger, older children), between-subjects design.

3.2 Participants

Children aged 6- to 11-years-old ($M_{age} = 9.04, SD = 1.52$), $N = 451$, were recruited from a summer science camp hosted by the University of Regina. Both parental consent and participant assent was collected for all participants. The consent form requested basic demographic information (i.e., age, sex). The age distributions for each lineup procedure were highly similar: simultaneous ($M = 9.00, SD = 1.55$, range = 6-11), face-off ($M = 9.05, SD = 1.52$, range = 6-11), and RFC ($M = 9.10, SD = 1.48$, range = 6-11). Given the relatively wide age range, I explored age differences between the younger (6- to 8-year-olds) and older children (9- to 11-year olds) in response accuracy across the procedures.
3.3 Materials

3.3.1 Target event. Children participated in a live event in which a female and male research assistant (the targets, Dr. Jessica and Dr. Matthew), visited children in groups of approximately 15 children. After introducing themselves, the targets engaged the children to help find a missing piece of science equipment. To ensure the targets were memorable as Dr. Jessica and Dr. Matthew, they wore white lab coats. During the searching activity, the visitors directly interacted with the children to ensure children were actively involved in the search. The activity lasted about 10-12 minutes.

3.3.2 Lineups. All lineups contained eight members, including one suspect and seven fillers. In addition, lineups all contained a visual rejection option (the wildcard; Zajac & Karageorge, 2009). Fillers and innocent suspects (16 photographs) were selected from the Glasgow Unfamiliar Face Database (Burton, White, & McNeill, 2010). In order to select the pictures, 200 photographs that matched the two targets’ general description were selected (100 females and 100 males). Next, independent adult judges (n = 24) provided pairwise (subjective) similarity ratings between photographs of the target and 100 potential fillers (preselected on gender) on a 10-point Likert-type scale (1 = not at all similar, 10 = highly similar). Mean ratings were used to select fillers, ensuring fillers were neither too similar nor too distinct from the target (Fitzgerald, Price, Oriet, & Charman, 2013). The mean similarity rating for the female fillers was slightly lower (M = 3.16) than for the male target (M = 3.60). All lineup images (180H x 288W pixels) were displayed on an 11-inch touch screen tablet and shown using Eprime 2.0 software that

---

3 N = 9 independent adult raters were asked to rate similarity for the female target and an additional 15 independent adult raters were asked to rate similarity for both female and male targets.
recorded the participants’ responses. Lineup bias measures (Malpass, 1981; Wells, Leippe, & Ostrom, 1979) indicated that the lineups were appropriate.

Lineup presentation was dependent upon the lineup procedure manipulation, described below. Note that the sequential procedure was not included as a baseline or comparison procedure as children have been documented to perform poorly on this task (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997; Parker & Ryan, 1993; Pozzulo & Lindsay, 1998; Steblay et al., 2011). Moreover, I prioritized ensuring that procedures were used that all involved an element of direct relative judgement (i.e., seeing multiple faces at once) in this initial exploration in order to ensure some comparability across procedures. The order/location of the lineup images, including the suspect, was randomized across participants.

3.3.3 Target presence manipulation. The presence of the targets (i.e., Dr. Jessica/Dr. Matthew) was manipulated, with half of lineups target-present and half target-absent. Target-present lineups contained a target from the live event as well as seven fillers. In target-absent lineups, the targets were replaced with the most similar-looking innocent suspect that was rated higher than the mean similarity rating (similarity rating: $M = 3.32$ for female; $M = 5.86$ for male). The spatial positions of the targets were randomized across participants.

3.3.4 Lineup procedure manipulation. All lineup procedures were conducted electronically, using Eprime 2.0 software that recorded the child’s responses as well as their response latency.

Simultaneous. Simultaneous lineups were displayed in a randomized, 3x3 array with the wildcard in the centre of the array (see Figure 1). Participants were asked to
either select the one lineup member they believed was the target (i.e., Dr. Matthew/Jessica) or indicate that the target was not in the lineup (i.e., select the wildcard) with the following instructions:

_The computer is going to show you some pictures. I need you to help me figure out if Dr. Matthew/Jessica’s picture is one of the pictures. There may be a picture of Dr. Matthew/Jessica here or there may not be a picture of Dr. Matthew/Jessica. Think back to what Dr. Matthew/Jessica looked like. I want you to compare your memory of Dr. Matthew/Jessica to each of these pictures. If you see Dr. Matthew/Jessica’s picture, touch it. If you don’t see his/her picture, touch the shadow picture in the middle._

![Image of simultaneous lineups](image)

_Figure 1. Images of the target present (left) and target absent (right) simultaneous lineups. Lineups included the target person (left, top right), the innocent suspect (right, top right), and the seven fillers._
Face-off. Face-off lineups were presented to participants in pairs. There were four Rounds presented to participants and, for each round, participants were presented with several pairs. For each pair, participants were asked to identify the person who looked MOST like the target. For Round 1, two random faces (Pair 1) were compared (i.e., a face-off match; See Figure 2). The winner (chosen lineup member) for the first match was stored for later use and the loser (non-chosen lineup member) was removed from the lineup. This was repeated for the next three pairs (i.e., Pair 2, 3, and 4). During Round 2, participants were presented with the winners of Pair 1 and Pair 2 and the winner of this match was stored for the next round, while the loser was removed. Next, participants were presented with winners of Pair 3 and Pair 4. During Round 3, participants were presented with the two winners from Round 2. Round 4 involved showing participants the overall winner and asking them to make a final decision as to whether this winner was the target (Figure 3). Participants were presented with the following instructions during rounds 1 through 4:

[Round 1-3] The computer is going to show you some pictures. I need you to help me figure out if Dr. Matthew/Jessica’s picture is shown. There may be a picture of Dr. Matthew/Jessica shown or there may not be a picture of Dr. Matthew/Jessica. The computer is going to show you two pictures at a time, in pairs. Each time you see a pair, I want you to touch the person who looks MOST like Dr. Matthew/Jessica. Even if you don’t think either picture looks like Dr. Matthew/Jessica, I need you to touch the one that looks MOST like him/her.

Which of these looks MOST like Dr. Matthew/Jessica?
Figure 2. Sample of face-off stimuli presented in pairs during Rounds 1-3 as well as sample of RFC stimuli presented during Rounds 1 and 2. Pair includes the female target person (left) and a filler (right).

[Round 4] The computer is going to show you the picture that you thought looked MOST like Dr. Matthew/Jessica, but that doesn’t necessarily mean it is Dr. Matthew/Jessica. Remember, his/her picture might not have even been shown at all, so this might be a picture of Dr. Matthew/Jessica or it might be a picture of someone else. If you think it is Dr. Matthew/Jessica’s picture, touch the picture. If you don’t think it is Dr. Matthew/Jessica’s picture, touch the shadow picture.

Figure 3. Sample of face-off stimuli presented in pairs during Round 4 as well as sample of RFC stimuli presented during Round 3. Pair includes the female target person (right) and the wildcard (left) that represents the absence of the target.

Repeated-Forced Choice. Participants in the Repeated-Forced Choice lineup condition were also presented with lineup members in pairs. This lineup contains three Rounds presented to participants and, for each round, participants were presented pairs of faces. For each pair, participants were asked to identify the one who looks MOST like the target (i.e., a most similar judgement). For the first round (Round 1), participants were asked to make a decision on 28 pairs—that is, they saw all possible combinations or
permutations of face pairs (Figure 2). During the second round (Round 2), participants were presented with the two faces they chose MOST often (Figure 2). Again, they were asked to choose the one who looks MOST like the target. For Round 2 to be successful, the computer program was designed to deal with situations of ‘tied’ lineup members by selecting the face chosen during the direct comparison of the ‘tied’ faces. See Appendix A for detailed information on how often lineup members ‘tied’ and how these ‘ties’ were dealt with in the program.

During the final round (Round 3), participants were presented with the face they selected from the previous round along with the wildcard. They were asked to make a final decision about whether or not the face was the target (Figure 3). Pilot testing was conducted to ensure the functionality of the program as well as to glean feedback from children on task length and comprehensibility (Appendix B). Round 1 of the RFC procedure included the same instructions listed for Rounds 1-3 of the face-off procedure. Instructions for Rounds 2 and 3 follow:

[Round 2 – see Figure 2] *Now the computer is going to show you the two pictures that you picked most often. I need you to pick the one that you think looks MOST like Dr. Matthew/Jessica.*

[Round 3 – see Figure 3] *The computer is going to show you the picture that you thought looked MOST like Dr. Matthew/Jessica, but that doesn’t necessarily mean it is Dr. Matthew/Jessica. Remember, his/her picture might not have even been shown at all, so this might be a picture of Dr. Matthew/Jessica or it might be a picture of someone else. If you think it is Dr. Matthew/Jessica’s picture, touch the*
picture. If you don’t think it is Dr. Matthew/Jessica’s picture, touch the shadow picture.

3.4 Procedure

On the first day, a male and a female research assistant visited children in small groups (about 10-15 children per group) and introduced themselves as scientists who had been working in the room before the children arrived. The targets informed the group that they lost something of value and needed the group’s help to find it. As a group, they searched for the object. During the search, the targets moved around the room and directly interacted with many children to ensure they were memorable. Once the male target found the missing object in a silly place (his pocket), they thanked the children and left. The interactions were audio recorded to ensure consistency in the interactions and each session lasted about 10 minutes.

The following day, different research assistants worked with children individually. Only children with parental consent were invited to participate (see Appendix C). After rapport-building and receiving assent from the children (Appendix D), children conducted the lineup tasks (one for the male target and one for the female target). A computer program was used to administer the lineup to limit administrator influences on the child eyewitnesses (see Wells & Bradfield, 1998). Children were told that they would be shown some pictures and would be asked to decide if the visiting scientists who interrupted their class were in one of the pictures. Children were asked to identify both targets and, as such, were shown two lineups (one for the male and one for the female target). The order of the lineups was counterbalanced (half saw the male first, half saw the female first), but children always completed the same procedure for both
targets (e.g., both simultaneous). The research assistant who launched the computer program did not know whether the suspect would be present in the lineup; this was randomly determined by the computer. In addition, administrators sat across from the children, holding the tablet computer in such a way that prevented viewing of the participant’s selection. Once each lineup was completed, children were asked to provide a confidence rating (i.e., how sure are you?) on a scale of 1 to 11. Once completed, children were thanked for their participation and given a prize. All interviews were audio recorded. Ethical approval was obtained prior to data collection (Appendix E).

4.0 Results

4.1 Identification Responses

Children’s identification responses were categorized into one of three responses: suspect identifications, filler selections, or lineup rejections. In an eight-person lineup, regardless of target presence, there is always one suspect, seven fillers, and a wildcard. Suspect identifications are either accurate or inaccurate, depending on the target-presence condition. A suspect identification represents a correct identification in target-present conditions, while a suspect identification in a target-absent condition represents a false identification of the innocent suspect. If the wildcard was selected (i.e., no lineup member was chosen), the response was classified as a rejection. A rejection is correct in a target-absent condition, but an error in the target-present condition. Filler selections were always considered errors in both target-present and target-absent conditions. See Table 1 for the identification response rates for the simultaneous, face-off, and RFC procedures across age groups.
Table 1

Identification Responses for Experiment 1

<table>
<thead>
<tr>
<th>Age</th>
<th>Target</th>
<th>Procedure</th>
<th>Identification Decision</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 9</td>
<td>Present</td>
<td>Simultaneous</td>
<td>Suspect</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face-off</td>
<td>Filler</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFC</td>
<td>Reject</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>Suspect</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face-off</td>
<td>Filler</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFC</td>
<td>Reject</td>
<td>0.20</td>
</tr>
<tr>
<td>≥ 9</td>
<td>Present</td>
<td>Simultaneous</td>
<td>Suspect</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face-off</td>
<td>Filler</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFC</td>
<td>Reject</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>Suspect</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face-off</td>
<td>Filler</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFC</td>
<td>Reject</td>
<td>0.06</td>
</tr>
<tr>
<td>Total</td>
<td>Present</td>
<td>Simultaneous</td>
<td>Suspect</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face-off</td>
<td>Filler</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFC</td>
<td>Reject</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>Suspect</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face-off</td>
<td>Filler</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFC</td>
<td>Reject</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Note. Given that each witness made two identifications, n represents the number of identifications made in each condition. Nine children withdrew mid-task and only completed one lineup (four did not complete the male target lineup, and five did not complete the female target lineup).
4.2 Lineup Identification Choice

To understand the influence of lineup procedure and eyewitness age on identification responses for the two targets, separate 3 (procedure: simultaneous, face-off, RFC) x 2 (actor: male, female), x 2 (age: < 9, ≥ 9) x 3 (lineup response: suspect, filler, reject) hierarchical log-linear analyses (HILOG) were conducted for target-present and target-absent lineups. For each HILOG, lineup response was the independent variable; therefore, only interactions with this variable were explored. In addition, odds ratios (OR) and associated 95% confidence intervals were computed as an effect size for significant differences between procedures. Confidence intervals that do not overlap with 1.00 represent a significant difference (α = .05).

4.2.1 Target-present. The HILOG with target-present lineups revealed that the highest order model (i.e., a 4-way interaction between all variables) was nonsignificant, $\chi^2(4) = 1.75, p = .78$, indicating that the model containing all four variables did not provide adequate fit with the data and, therefore, the association between all four variables together was not explored further. Similarly, the model revealed inadequate fit for a model containing three variables, $\chi^2(12) = 10.49, p = .57$, and two variables, $\chi^2(13) = 18.54, p = .14$. Because specific age-related hypotheses were made, the partial associations for age-related effects were explored. Results revealed a two-way interaction between age and lineup response, $\chi^2(2) = 8.76, p = .02$, such that the younger children picked the guilty suspect less often ($M = 0.41$ versus $M = 0.55$; $z = 2.83, p = .005$) and rejected the lineup more often ($M = 0.39$ versus $M = 0.27$; $z = 2.53, p = .01$) than older children. There were no differences in filler picks ($M = 0.21$ versus $M = 0.18$; $z = 0.75, p = .45$). There was no interaction between actor (target) and lineup response, $\chi^2(2) = 2.82,$
46

$p = .25$, indicating that responses were consistent across the two targets. In sum, when the
target was present lineup behaviour was similar across procedures.

4.2.2 Target-absent. The HILOG with target-absent lineups revealed that the
highest-order interaction (between age, lineup response, actor, and procedure) was not
significant, $\chi^2(4) = 5.96, p = .20$. The highest order model to retain that provided a
significant fit to the data was a 3-way interaction, $\chi^2(12) = 23.34, p = .03$. Partial
associations revealed a significant interaction between age, procedure, and response,
$\chi^2(4) = 12.07, p = .02$. To examine this three-way interaction, the effect of lineup
procedure on response was examined for each age group. Younger children picked the
innocent suspect more often ($z = 2.55, p = .01$) in the RFC procedure than in the face-off
procedure. The odds that the innocent suspect would be chosen by younger children from
a RFC lineup were seven times the odds for the face-off procedure, $OR = 6.53$ [95% CI =
1.34, 31.66]. However, the rate of filler selection ($z = 0.42, p = .67$) and correct rejection
($z = 1.24, p = .21$) of lineups did not differ between the RFC and the face-off procedure.
Innocent suspect identifications in the simultaneous procedure did not differ from either
the RFC or face-off procedures ($z = 0.79, p = .43; z = 1.78, p = .08$, respectively).
Likewise, children’s rate of rejecting the lineup in the simultaneous procedure did not
differ from the RFC or face-off procedures ($z = 1.40, p = .16; z = 0.20, p = .84$,
respectively). No other differences in responding were found across procedure types (all
$p$’s $> .05$). For older children, response type did not significantly differ across the three
lineup procedures (all $p$’s $> .05$). Thus, younger children incorrectly identified innocent
suspects more often in the RFC procedure than in the face-off procedure, and there were
no differences in target-absent lineup behaviour across the three procedures for older children.

There was also a two-way interaction between actor and response, \( \chi^2(4) = 19.86, p < .001 \); however, there was no significant three-way interaction between actor, response and procedure, \( \chi^2(4) = 4.04, p = .40 \). This indicates that the absolute performance was different across the two actors (targets)—however, this is not unexpected given the high level of variability between different faces. Follow-up analyses revealed that younger children performed generally more poorly on the female target lineup than the male target lineup\(^4\). These absolute differences in responses to targets, however, did not vary by lineup procedure. As such, the two targets are collapsed for all subsequent analyses.

4.3 Survival Statistics

Survival analysis is a method used to describe the proportion of cases surviving at various time points (Tabachnick & Fidell, 2007). Survival analyses have been used in previous eyewitness research to identify the likelihood that the suspect survived different levels of a lineup task (Pozzulo & Lindsay, 1999; Price & Fitzgerald, 2016). Survival statistics can only be reported for the face-off and RFC procedures because they contain multiple decisions, whereas the simultaneous procedure only contains one decision. Table 2 presents the survival statistics across four rounds of decisions. The face-off procedure included four rounds that contained eight decisions whereas the RFC procedure included three rounds containing 30 decisions. Round 1 includes the initial round of ‘most similar’ decisions for both procedures (four for face-off, 28 for RFC). Round 2 only applies to the

\(^4\)See Appendices F and G for a detailed description of differences between actors and the analyses that were used to examine for guessing behaviour in younger child witnesses.
face-off procedure and includes two additional ‘most similar’ decisions. For both lineup procedures, Round 3 (Final Comparison) represents the final ‘most similar’ decision between the two faces picked most and Round 4 represents the final decision in which the only remaining picture (the one chosen most frequently) was paired with a wildcard and the participant decided if the remaining picture was the target. For both lineup procedures, the Round 4 (Final Choice) in Table 2 indicates how many times the suspect (guilty or innocent) was selected over the wildcard.

Table 2

*Suspect Survival Rates in Experiment 1*

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>Survived R1</th>
<th>Survived R2</th>
<th>Survived R3 - Final Comparison</th>
<th>R4 - Final Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 9</td>
<td>RFC TP 46</td>
<td>.46</td>
<td>-</td>
<td>.30</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>TA 50</td>
<td>.34</td>
<td>-</td>
<td>.22</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>Face-off TP 49</td>
<td>.82</td>
<td>.63</td>
<td>.55</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>TA 57</td>
<td>.39</td>
<td>.16</td>
<td>.11</td>
<td>.04</td>
</tr>
<tr>
<td>≥ 9</td>
<td>RFC TP 108</td>
<td>.74</td>
<td>-</td>
<td>.61</td>
<td>.58</td>
</tr>
<tr>
<td></td>
<td>TA 83</td>
<td>.22</td>
<td>-</td>
<td>.13</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Face-off TP 81</td>
<td>.85</td>
<td>.68</td>
<td>.58</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>TA 81</td>
<td>.51</td>
<td>.25</td>
<td>.19</td>
<td>.12</td>
</tr>
</tbody>
</table>

*Note.* TA= Target-absent, TP= Target-present. R1-R4 = Round 1, 2, 3, or 4

When comparing the suspect survival rates for target-absent relative to target-present lineups, there is an expected decrease in how often the innocent suspect survived each round. With the exception of the RFC procedure used with younger children, the guilty suspect survived Round 1 better than chance (50%) and the innocent suspect
survived Round 1 less than or near chance. Table 2 provides evidence of young children performing better with the face-off procedure (higher suspect survival rates in target-present, lower suspect survival rates in target-absent) than the RFC procedure. The guilty suspect survival rate in target-present lineups was better in the face-off at Round 1, $z = 3.88, p < .001$, and Round 3, $z = 2.52, p = .01$. Likewise, the innocent suspect survival rate was better (i.e., lower) in the face-off at Round 4, $z = 2.55, p < .01$. Alternatively, with the older children, suspect survival rates are similar to the face-off for both target-present, $z = 0.55, p = .59$, and target-absent procedures, $z = 1.34, p = .18$, in Round 4.

### 4.4 Lineup Comparison

An eyewitness lineup task typically involves a participant looking at a series of faces and determining whether they have seen any of the faces before. Researchers commonly use two methods to compare eyewitness performance across lineup procedures: diagnosticity ratios (Wells & Lindsay, 1980) and the signal-detection statistic, $d'$ (Wixted & Mickes, 2014). Diagnosticity ratios provide a measurement of how likely participants are to identify a guilty suspect versus an innocent suspect. The larger the diagnosticity ratio is, the greater the likelihood that an eyewitness will identify a guilty suspect and not an innocent suspect. The $d'$ statistic is based on signal-detection theory (SDT) and the extent that the properties of a lineup can influence a person’s (specifically, a group of peoples’) ability to discriminate previously seen from previously unseen faces (i.e., a measure of sensitivity; Wixted & Mickes, 2014). To measure performance, SDT examines the relationship between hit (correct identification) and false alarm (innocent suspect identification) rates to assess discrimination ability and a witness’ criterion ($csuspect$; Mickes, Moreland, Clark & Wixted, 2014). A criterion is a
measure of how likely or willing a witness is to identify someone from a lineup—a response bias (Swets, 1996; Wixted & Mickes, 2014).

Both forms of lineup comparison information have advantages and limitations. As Palmer, Brewer and Weber (2010) point out, application of signal detection theory to an eyewitness identification decision is problematic because eyewitness decisions are not simple binary decisions (hit or miss for target-present stimulus; false alarm or correct reject for target-absent stimulus), but include an additional response type: filler selections. Using traditional SDT methods to calculate $d'$, these filler identification classifications pose a problem because they can be classified as either a false alarm or a miss, but are not genuine examples of either response type. Some researchers have opted to treat these filler identifications in target-present lineups as misses, thus, excluding the responses in the calculation of $d'$ (e.g., Meissner, Tredoux, Parker, & MacLin, 2005; Mickes et al., 2014). Alternatively, some researchers prefer signal detection theory information because diagnosticity is susceptible to influences on response criterion (e.g., Wixted & Mickes, 2012). In other words, when a procedure encourages more conservative responding, diagnosticity ratios can be artificially inflated. Given that both measures are used in the literature to compare performance across lineup procedures, both are presented in this dissertation.

4.4.1 Signal Detection Theory information. Discriminability ($d'$) can be conceptualized as an index of how well a group discriminates between guilty and innocent suspects, with a higher number more indicative of discrimination. To assess discriminability (or sensitivity), $d'$ was calculated using the formula $d' = zH - zFA$ (MacMillan & Creelman, 1991; Mickes et al., 2014), where $zH$ refers to the z
transformed hit rate (correct identifications in target-present conditions) and zFA refers to
the z transformed false alarm rate (innocent suspect selection in target-absent conditions).
Larger values of $d'$ indicate a higher level of discriminability and thus, the most useful
procedure. To assess response bias ($c_{\text{suspect}}$), the formula provided by MacMillan and
Creelman (1991) was used: $c = -0.5(zH + zFA)$. Thus, higher $c_{\text{suspect}}$ indicates bias against
identifying a suspect. Table 3 outlines $d'$ and response bias statistics calculated for the
different lineup procedures. Not surprisingly, the face-off procedure has the highest
discriminability with younger children, while the simultaneous was highest with older
children—however, this was closely followed by the RFC procedure.

Table 3

*Signal Detection Theory Information ($d'$ and $c_{\text{suspect}}$) for Lineup Comparison in
Experiment 1.*

<table>
<thead>
<tr>
<th>Age</th>
<th>Procedure</th>
<th>$d'$</th>
<th>$c_{\text{suspect}}$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 9</td>
<td>Simultaneous</td>
<td>0.94</td>
<td>0.39</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Face-off</td>
<td>1.59</td>
<td>0.77</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>RFC</td>
<td>0.45</td>
<td>0.26</td>
<td>96</td>
</tr>
<tr>
<td>≥ 9</td>
<td>Simultaneous</td>
<td>2.04</td>
<td>0.69</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td>Face-off</td>
<td>1.37</td>
<td>0.22</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>RFC</td>
<td>1.93</td>
<td>0.39</td>
<td>191</td>
</tr>
</tbody>
</table>
4.4.2 Diagnosticity ratios. Diagnosticity ratios and inferential confidence intervals (Palmer et al., 2010; Tryon, 2001) are presented in Table 4 and allow for comparison across the different lineup procedures (Wells & Lindsay, 1980). A diagnosticity ratio provides evidence regarding the likelihood that the identified suspect is the guilty suspect. A diagnosticity ratio of 1.0 indicates that the two events (i.e., identifying a guilty suspect versus identifying an innocent suspect) are equally likely. Departure from 1.0 indicates differences in the probability of these two events. For example, a ratio of 2.0 indicates that children were twice as likely to identify the culprit as the innocent suspect. In Table 4, procedures with confidence intervals that do not overlap with 1 can be considered diagnostic. Traditionally, diagnosticity is calculated using suspect identification rates. However, as argued by Price and Fitzgerald (2016), the diagnosticity of filler selection and rejection outcomes should be reported as these decisions can be diagnostic of a suspect’s innocence (Wells & Lindsay, 1980; Wells & Olson, 2002). Thus, diagnosticity was calculated for all three types of decision outcomes—suspect, filler, and rejections.

For younger children, the face-off procedure was significantly more diagnostic of suspect guilt than both the simultaneous and RFC procedures, which produced lower but similar diagnosticity ratios. For older children, the opposite pattern was found wherein the face-off procedure was least diagnostic of suspect guilt, while the simultaneous was most diagnostic, followed by the RFC procedure. Inferential confidence intervals indicate that the simultaneous procedure was significantly more diagnostic than the face-off procedure for older children, while the RFC procedure did not significantly differ from
either procedure. Thus, it appears that the RFC procedure worked well for older, but not younger children.

Table 4

*Diagnosticity Ratios and 95% Inferential Confidence Intervals for Experiment 1*

<table>
<thead>
<tr>
<th>Age</th>
<th>Procedure</th>
<th>Diagnostic of Guilt</th>
<th></th>
<th>Diagnostic of Innocence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Suspect Identification</td>
<td>Filler Selection</td>
<td>Rejection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DR</td>
<td>UL</td>
<td>LL</td>
<td>DR</td>
<td>UL</td>
</tr>
<tr>
<td>6-8.99</td>
<td>3.10</td>
<td>5.52</td>
<td>1.74</td>
<td>1.53</td>
<td>2.63</td>
</tr>
<tr>
<td></td>
<td>12.80</td>
<td>28.23</td>
<td>5.80</td>
<td>3.30</td>
<td>6.31</td>
</tr>
<tr>
<td></td>
<td>1.63</td>
<td>2.68</td>
<td>0.99</td>
<td>1.10</td>
<td>1.65</td>
</tr>
<tr>
<td>≥ 9</td>
<td>16.68</td>
<td>36.50</td>
<td>7.62</td>
<td>1.65</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>4.40</td>
<td>7.06</td>
<td>2.74</td>
<td>1.63</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td>9.68</td>
<td>19.06</td>
<td>4.92</td>
<td>1.69</td>
<td>2.51</td>
</tr>
</tbody>
</table>

*Note.* DR = diagnosticity ratio. LL = lower limit and UL = upper limit of 95% inferential confidence interval. Inferential confidence intervals were calculated (Tryon, 2001) using a bootstrapping procedure (Palmer et al., 2010) in which log scales (ln) were used to more closely approximate a normal distribution.

The confidence intervals in Table 4 have been converted from log back to their original unit.

For suspect identifications, diagnosticity was computed as the ratio of target-present/target-absent responses. For filler selections and rejections, diagnosticity was computed as the ratio of target-absent/target-present responses.

**4.4.3. Information gain.** Diagnosticity information can be interpreted as indications of guilt or innocence under the assumption that the guilty suspect is present in the lineup half of the time (50% base rate). While this is frequently the case for
experiments, as compellingly argued by Wells and colleagues (2015), it is not reflective of the ‘real world’ in which the precise base rates are unknown and constantly changing. To account for this unknown variable, it is important to examine diagnosticity ratios across all possible base rates using an information gain analysis. According to Wells and Olson (2002), the amount of information gained from a suspect identification is calculated by taking the absolute difference of the posterior probability of suspect guilt from the prior probability (base rate) of suspect guilt. The information gained from eyewitness decisions can be calculated for all possible prior probabilities and plotted as a curve (see Wells & Olson, 2002; Wells et al., 2015).

The information gain curves for simultaneous, face-off, and RFC procedures are shown in Figure 4. When focusing on suspect identifications made by younger children, more information about suspect guilt was gained from the face-off procedure, followed by the simultaneous and RFC procedures. For older children, the simultaneous and RFC procedures provided similar information regarding suspect guilt, followed by the face-off procedure. The simultaneous and RFC procedures were best able to provide information about suspect guilt at lower base rates. For filler identifications, the simultaneous and face-off procedures provided more information about suspect guilt than did RFC for both age groups; however, the simultaneous procedure provided the most information regarding innocence—particularly when the base rate was higher. For rejections, the face-off procedure produced the most information about suspect innocence for younger children—though all levels of information gain were small for this age group. For the older children, rejections made during the RFC procedure provided the most information about suspect guilt, particularly when the base rate was about 68%.
Figure 4. Information gain in Experiment 1.
To compare suspect diagnosticity (and information gain) across the three procedures, a z-value ratio was calculated using the arcsine transformation method (see Wells & Olson, 2002) and relative risk ratios (RRR) were calculated as an indicator of effect size (Bland & Altman, 2003). RRR ratios can also be used as an alternative method to test for significant differences in diagnosticity between procedures and, if this method is used, confidence intervals not overlapping with 1.00 indicate statistical significance.

For younger children, the RFC procedure was not significantly more diagnostic than the simultaneous procedure, \( z = -1.36, p = .09, RRR = 0.53, [0.19, 1.46] \), nor the face-off procedure, \( z = -2.85, p = .002, RRR = 0.13 [0.03, 0.61] \). Similarly, the simultaneous and face-off procedures did not differ, \( z = -1.54, p = .06, RRR = 0.24 [0.05, 1.18] \). For older children, the RFC procedure was not significantly more diagnostic than the simultaneous procedure, \( z = -0.08, p = .47, RRR = 0.58 [0.14, 2.41] \), nor was it significant different from the face-off procedure, \( z = 1.41, p = .08, RRR = 2.20 [0.76, 6.35] \). Similarly, the simultaneous was not more diagnostic than the face-off procedure, \( z = 1.48, p = .07, RRR = 3.79 [1.04, 13.72] \). Thus, it appears that the RFC procedure allowed older, but not younger children, to discriminate between guilty and innocent suspects at an equivalent rate to the simultaneous procedure.

**4.4.4. BREE curves.** Wells and colleagues (2015) proposed the use of base-rate effect-equivalency (BREE) curves to estimate the change in base rate needed to improve the diagnosticity from one procedure to match a more diagnostic procedure. BREE curves outlining the required change in base rates of suspect identification for the simultaneous, face-off, and RFC procedures are outlined in Figure 5. For the younger children, there is evidence that the face-off procedure performed better at lower base
rates. Assuming a base rate of 50% (half of the lineups include a guilty suspect), the base rate of the simultaneous procedure would need to increase by 31% and the RFC procedure would need to increase by 39% to produce suspect identifications that are as diagnostic of suspect guilt as the face-off procedure. For the older children, there is evidence of the simultaneous procedure performing better at lower base rates. As can be seen in Figure 5, the base rate of the face-off and RFC procedures would need to increase by 29% and 13%, respectively, to produce suspect identifications that are as diagnostic of suspect guilt as the simultaneous procedure.

![Figure 5. Base-rate effect-equivalency (BREE) curve in Experiment 1 for suspect identifications from simultaneous, face-off, and RFC procedures for each age group. For](image-url)
each curve, the first procedure listed represented the procedure with the higher diagnosticity.

4.7 Supplementary Memorial Information

Thus far, the results have focused on the traditional information collected from the three lineup procedures: lineup selection and what this selection can tell us about the likelihood of suspect guilt. From these metrics, the RFC procedure performs similar to the simultaneous lineup procedure with older children, but worse than the face-off procedure with younger children. To further explore the utility of the RFC procedure and to explain age differences in decision accuracy with the RFC procedure, it is important to consider supplementary information that is frequently used to contextualize a witness’ lineup choice from traditional procedures: post-identification confidence ratings and response latency.

4.7.1. Confidence-Accuracy relationship. Post-identification confidence ratings can be indicative of accuracy for those who choose from a lineup (e.g., Brewer & Wells, 2006; Sporer et al., 1995). To examine whether confidence was predictive of overall accuracy, logistic regressions were run for each lineup procedure. For younger children, confidence ratings did not significantly predict overall accuracy for the simultaneous, $B = .09, \text{Exp}(B) = 1.09$; Nagelkerke pseudo $R^2 = .02, \chi^2(1) = 1.83, p = .18$, the face-off, $B = .11, \text{Exp}(B) = 1.11$; Nagelkerke pseudo $R^2 = .04, \chi^2(1) = 2.87, p = .09$, or the RFC procedures, $B = -.06, \text{Exp}(B) = 0.94$; Nagelkerke pseudo $R^2 = .01, \chi^2(1) = 0.47, p = .50$. For the older children, confidence ratings did significantly predict overall accuracy for the simultaneous, $B = .18, \text{Exp}(B) = 1.19$; Nagelkerke pseudo $R^2 = .05, \chi^2(1) = 6.95, p = .008$, the face-off, $B = .23, \text{Exp}(B) = 1.26$; Nagelkerke pseudo $R^2 = .08, \chi^2(1) = 9.97, p = .001$. 
.002, and the RFC procedure, $B = .34$, $\exp(B) = 1.09$; Nagelkerke pseudo $R^2 = .14$, $\chi^2(1) = 20.76$, $p < .001$. These data demonstrated that post-identification confidence is predictive of accurate responding for older children. Follow-up examination using point biserial correlations (Weber & Brewer, 2004) were conducted to better understand which type of accurate responses (i.e., positive or choosing accuracy versus negative or rejection accuracy) was related to confidence. These analyses revealed that these effects with older children were primarily driven by confidence-accuracy relationships for positive responses (see Table 5). Although confidence was not predictive of overall accuracy for younger children, Table 5 suggests that the confidence ratings provided by children during the face-off procedure was significantly correlated with accurate, positive responding (i.e., correct choosing from the lineup).
Table 5

*Point bi-*serial correlations between confidence ratings and lineup accuracy for positive (correctly choosing from the guilty suspect from the lineup) and negative (correctly rejecting a target-absent lineup) responses in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Positive Decision</th>
<th>Negative Decision</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p</td>
<td>n</td>
</tr>
<tr>
<td>Younger Children</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simultaneous</td>
<td>.26</td>
<td>.06</td>
<td>56</td>
</tr>
<tr>
<td>Face-off</td>
<td>.46</td>
<td>&lt;.001</td>
<td>53</td>
</tr>
<tr>
<td>RFC</td>
<td>-.04</td>
<td>.75</td>
<td>58</td>
</tr>
<tr>
<td>Older Children</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simultaneous</td>
<td>.47</td>
<td>&lt;.001</td>
<td>95</td>
</tr>
<tr>
<td>Face-off</td>
<td>.39</td>
<td>&lt;.001</td>
<td>96</td>
</tr>
<tr>
<td>RFC</td>
<td>.46</td>
<td>&lt;.001</td>
<td>114</td>
</tr>
</tbody>
</table>

Note. Given that each witness made two identifications, n represents the number of identifications made in each condition. r refers to Pearson’s r or Pearson’s correlation.

**4.7.2. Response latency.** On average, it took children less than four minutes to complete the RFC procedure (younger children: $M = 3.87$ minutes, $SD = 1.76$; older children: $M = 3.46$, $SD = 1.10$). It took children less than two minutes to complete the face-off procedure (younger children: $M = 1.93$, $SD = 0.84$; older children: $M = 1.77$, $SD = 0.67$) and less than one minute to complete the simultaneous procedure (younger children: $M = 0.23$, $SD = 0.19$; older children: $M = 0.22$, $SD = 0.22$). Mean response latency did not differ between age groups for the simultaneous or face-off procedure (see Table 5; all $ps > .05$); however, young children took longer to complete the RFC procedure than older children, $t(304) = 2.50$, $p = .01$. 
To examine whether response latency was predictive of overall accuracy, logistic regressions were run for each lineup procedure. Response latency did not predict overall accuracy for any group—with the exception of the RFC procedure with older children, $B = .00$, $\text{Exp}(B) = 1.00$; Nagelkerke pseudo $R^2 = .03$, $\chi^2(1) = 4.08$, $p = .04$, in which correct decisions were made faster ($M = 3.33$ minutes, $SD = 1.01$) than incorrect decisions ($M = 3.66$ minutes, $SD = 1.20$), $t(189) = 2.04$, $p = .04$. However, it is important to note that this difference in latency may simply be an artifact of noise involved in making so many decisions during the RFC procedure.

The computer program for the RFC procedure tracked viewing duration for each pair of faces shown. Examination of these data revealed that children from both age groups spent more time looking at the suspect’s face at the first exposure ($M = 3.84$ seconds, $SD = 2.36$) and progressively decreased with each subsequent viewing, with the last pairing being viewed for less than two seconds ($M = 1.98$ seconds, $SD = 1.61$; see Figure 6). In addition, older children took more time to make a decision when viewing the guilty suspect ($M = 2.58$ seconds, $SD = 1.67$) than the innocent suspect ($M = 2.21$, $SD = 2.39$), $t(279) = 3.19$, $p < .001$. This effect was especially strong when the older children ultimately selected the suspect from the pair ($M = 2.32$ seconds, $SD = 1.52$; $M = 1.86$, $SD = 1.83$, respectively; $t(262) = -3.40$, $p < .001$; see Figure 6). On the other hand, younger children’s viewing time of the guilty suspect was much more irregular; children spend a similar amount of time viewing the guilty suspect ($M = 3.26$, $SD = 1.19$) as they did the innocent suspect ($M = 3.13$, $SD = 1.12$) before making any choice, $t(107) = 0.29$, $p = .39$, and when choosing the suspect, $t(92) = 0.02$, $p = .49$ (see Figure 6). Thus, the
speed of younger children’s decisions was not influenced by the presence of the guilty suspect.

Figure 6. Mean Response Latency for Decision Made when the Suspect was Shown and Selected

4.7.2. Patterns of responding: Z-scores. Overall, post-identification confidence ratings and response latency information collected during the RFC procedure were both predictive of overall accuracy for older but not younger children. The RFC procedure, however, also provides supplementary information beyond the metrics available in standard lineup approaches that can be used to further index recognition memory.
Specifically, the RFC procedure tracked how many times a witness selected the suspect (guilty or innocent) relative to how many times he/she selected a filler. This pattern of witness responding can be used to assess a witness’ bias toward selection of the suspect. For example, if the suspect is selected in only one out of seven pairings, it suggests a low level of suspect selection bias, while selecting five of seven pairings suggests a higher level of bias.

To simplify response patterns, Z-scores were calculated to summarize information from the entire distribution of responding for each witness. The formula to calculate Z-scores is found in Appendix H. Z-scores are reflective of how frequently the witness chose the suspect and the magnitude to which the suspect was favored, relative to other lineup members (i.e., individual measure of suspect selection bias). Thus, higher, positive Z-scores indicate frequent choosing of a suspect and a more random distribution of selections across other lineup members. Higher negative Z-scores reflect clear favoring of a filler. Z-scores were grouped to allow for comparison between Z-scores and witness’ final lineup outcome (i.e., selecting a suspect, filler, or rejection; see Table 6). In particular, I was interested in exploring whether Z-scores were reflective of accurate discrimination between guilty and innocent suspects.
Table 6

*Frequency of Z-Scores for each decision type for each age group*

<table>
<thead>
<tr>
<th>Z Statistics</th>
<th>TA</th>
<th></th>
<th></th>
<th>TP</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Suspect</td>
<td>Filler</td>
<td>Reject</td>
<td>Suspect</td>
<td>Filler</td>
<td>Reject</td>
</tr>
<tr>
<td>Younger Children</td>
<td>n=6</td>
<td>n=21</td>
<td>n=23</td>
<td>n=14</td>
<td>n=16</td>
<td>n=16</td>
</tr>
<tr>
<td>2.00-2.47</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>0.33</td>
<td>0.05</td>
<td>0.09</td>
<td>0.43</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>0.50</td>
<td>0.14</td>
<td>0.09</td>
<td>0.43</td>
<td>0.19</td>
<td>0.06</td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>0.17</td>
<td>0.05</td>
<td>0.13</td>
<td>0.07</td>
<td>0.44</td>
<td>0.19</td>
</tr>
<tr>
<td>-2.47-(-0.01)</td>
<td>0.00</td>
<td>0.76</td>
<td>0.65</td>
<td>0.00</td>
<td>0.38</td>
<td>0.69</td>
</tr>
<tr>
<td>Older Children</td>
<td>n=4</td>
<td>n=26</td>
<td>n=53</td>
<td>n=63</td>
<td>n=20</td>
<td>n=25</td>
</tr>
<tr>
<td>2.00-2.47</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>0.50</td>
<td>0.04</td>
<td>0.08</td>
<td>0.46</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>0.50</td>
<td>0.04</td>
<td>0.13</td>
<td>0.40</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>0.00</td>
<td>0.39</td>
<td>0.40</td>
<td>0.02</td>
<td>0.45</td>
<td>0.44</td>
</tr>
<tr>
<td>-2.47-(-0.01)</td>
<td>0.00</td>
<td>0.54</td>
<td>0.38</td>
<td>0.00</td>
<td>0.20</td>
<td>0.24</td>
</tr>
</tbody>
</table>

*Note.* Columns may not sum to one due to rounding. Unlike the bins used to examine model fit (i.e., grouping Z-scores based on third percentile; See Table 16), these bins or groups of Z-scores were established considering ease of interpretation for a user.

For both younger and older children, Z-scores were significantly higher when the witness selected the guilty suspect (younger: $M = 1.35, SD = 0.35$; older: $M = 1.53, SD = 0.22$) than when they selected the innocent suspect (younger: $M = 0.85, SD = 1.00$; older:
When children correctly chose from the lineup (i.e., positive decisions), younger children had significantly lower Z-scores than older children, $t(76) = -2.58, p = .01$, indicating that younger children’s response patterns can be used to discriminate guilty from innocent suspects, but to a lesser extent than for older children.

### 4.7.3. Predictive utility of Z-scores.

Logistic regressions were run in order to determine whether knowledge of a witness’ pattern of responding (Z-scores) throughout the RFC procedure allows for post-diction of accuracy due to the high correlation between number of suspect picks and Z-score ($r > .97, p < .001$ for both age groups). The model produces very comparable results when number of suspect picks are included in the model in place of Z-score. Given the demonstrated role of post-identification confidence ratings, it was important to learn whether Z-scores predicted accuracy beyond the confidence ratings. Table 7 outlines the model results.

When examining the overall accuracy of responding, inclusion of Z-scores was predictive of accuracy over and above the inclusion of confidence ratings for older children; however, no model was produced for younger children. Recall that post-decision confidence ratings accounted for about 14% of the variance in accuracy for the RFC procedure for older children (section 4.7.1, Nagelkerke pseudo $R^2$). Adding Z-scores into the model increased the explanatory power of the model by about 7%, accounting for about 21% of the variance. This 21% of accuracy variance accounted for in the RFC procedure is much higher than the 5% of variance accounted for by confidence ratings alone in the simultaneous lineup ($B = .18, \exp(B) = 1.19, p = .01$; Nagelkerke pseudo $R^2 = .05, \chi^2(1) = 6.95, p = .008$).
To further explore the predictive utility of Z-scores for older children, separate logistic regressions were run for positive (i.e., those who chose from the lineup) and negative (those who rejected the lineup) decisions for each age group. No model was produced for negative responses indicating that neither Z-scores nor confidence explained variance in negative response accuracy for either younger or older children. For positive response accuracy, only Z-scores were included in the model and accounted for 55% of the variance in accuracy for younger children. For older children, Z-scores accounted for 77% of the variance in accuracy and when combined with confidence ratings (Step 2), these two predictors account for 80% of the variance for older children. Thus, Z-scores are predictive of positive response accuracy.
Table 7

*Logistic Regression Model Information Demonstrating Predictive Utility of Confidence Ratings and Z-scores.*

<table>
<thead>
<tr>
<th></th>
<th>Step 1</th>
<th>Step 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive Response Accuracy: Z-scores</td>
<td>Positive Response Accuracy: Confidence</td>
</tr>
<tr>
<td></td>
<td>Total Accuracy: Confidence</td>
<td>Total Accuracy: Z-score</td>
</tr>
<tr>
<td></td>
<td>B(SE) Exp ( \chi^2 )</td>
<td>B(SE) Exp ( R^2 ) ( \chi^2 )</td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-score</td>
<td>3.41(1.34) 30.29 .55 24.66*</td>
<td></td>
</tr>
<tr>
<td>Older</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-score</td>
<td>5.41(1.18) 224.54 .77 96.75* 5.36(1.24) 213.22</td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.49(.20) 1.64 .80 104.16*</td>
<td></td>
</tr>
<tr>
<td>Total (Older)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>.34*(.08) 1.41 .14 20.76* 0.31*(.08) 1.36</td>
<td></td>
</tr>
<tr>
<td>Z-score</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.54*(.16) 1.71 .21 32.07*</td>
<td></td>
</tr>
</tbody>
</table>

Note. \( R^2 \) refers to Nagelkerke \( R^2 \), * indicates statistical significance at a level of \( p < .001 \). Degrees of freedom associated with chi-square tests are 1 for Step 1 and 2 for Step 2. No model for total accuracy was produced for younger children.

Next, using a predictive equation (Log(p/1-p) = \( b_0 + b_1 \times 1 + b_2 \times 2 \)), probability estimates of the model were calculated at each level of confidence, with and without Z-scores included in the model. These scores were plotted to demonstrate the predictive
utility of considering both the Z-scores and post-identification confidence ratings in the RFC procedure. As seen in Figure 7, witnesses who provided a confidence rating of 7 or higher had a higher than chance (.50) probability of an accurate decision. Between confidence ratings of 7-9, including the witness’ Z-score with the confidence rating increases the probability of accurate responding. Beyond a confidence rating of 9, Z-scores did not notably add to the predictive utility of confidence ratings. To further this interpretation, in examining Figure 7, it appears that Z-scores higher than 1.00 add to the predicted probability of accuracy for positive responses—particularly for the middle confidence ranges (e.g., confidence ratings of 4 through 7) where there may be some apprehension about interpreting the predictive utility of the confidence ratings.
Figure 7. Plotted probabilities of overall and positive response accuracy at each level of confidence.

To gain a better understanding of how confidence and Z-scores interact to predict accuracy, see Figure 8. Regardless of the level of confidence provided, Z-scores less than 1.00 are not predictive of accuracy, while Z-scores higher than 1.00 helped to predict accuracy when corresponding levels of confidence might otherwise be considered ambiguous (i.e., post-identification confidence ratings of 4 through 7).
Figure 8. Predictive probability of Z-scores in combination with confidence ratings for older children.

4.8 Individual Witness Discrimination

Criminal investigators will be interested in the accuracy of individual witness responses—not group averages. To provide information to guide interpretation of individual witness responses, two analyses were conducted using information about witness’ response patterns (Z-scores).

4.8.1. Index of discrimination. To examine how well a witness’ pattern of responding during Round 1 of the RFC procedure could be used by children to discriminate or distinguish a previously seen (target) face from unseen (fillers and innocent suspect) faces, the adjusted normalized discrimination index (ANDI) was calculated (for the formulae, see Yaniv, Yates, & Smith, 1991). ANDI is a measure of

\[ \text{ANDI} = \frac{\text{Z-score difference}}{\text{confidence deviation}} \]

Note that ANDI could not be calculated for the face-off procedure because each face was not seen the same number of times. Calculations require comparable responding opportunities for each stimulus in the lineup (e.g., how many times was each face picked out of the possible 7 times). Comparable calculations cannot be made consistently for the face-off procedure.
variance in accuracy accounted for by patterns of responding which ranges from 0 (no discrimination) to 1 (perfect discrimination). For example, an ANDI score of .30 indicates that the pattern of responding can explain 30% of the variability in outcomes. A bootstrapping procedure was used to compute 95% inferential confidence intervals\(^7\) around the ANDI scores. Younger children’s pattern of responding does not explain variance in outcomes—that is, their responses during the first 28 decisions of the RFC procedure do not discriminate the target from unseen faces (ANDI = 0.01, 95% CI [-0.01, 0.03]). Older children were significantly better able to discriminate using the RFC procedure (ANDI = 0.23, 95% CI [0.18, 0.28]).

4.8.2. Profile analysis. Lastly, to highlight which individual sets or patterns of selections was likely (versus not likely) to indicate accurate discrimination between previously seen and previously unseen faces, Brewer and colleagues’ (2012) profile analysis was adapted\(^8\) and applied to the RFC responding patterns (Z-scores). The Z-scores were compared with the accuracy of witnesses’ final decisions (Round 3—i.e., did they choose the suspect, a filler, or reject the lineup?). Table 8 provides accuracy information as a function of discrepancy. As seen in Table 8, there is a linear pattern such that as discrepancy in response patterns increased, so did the likelihood of correct

---

\(^7\) This procedure (see Palmer et al., 2010; Tryon, 2001) used the observed data as a sampling distribution and conducted 250 replications to estimate variance of ANDI. This estimated variance provided the distribution needed to calculate confidence intervals.

\(^8\) Brewer and colleagues used a discrepancy score between the maximum rating and the next-highest rating on a 100-point scale. However, the expected variability with the model (a forced-choice model containing a 7-point scale) is less than what Brewer and colleagues (2012) were able to observe. In addition, multiple versus single maximum ratings do not hold as much interpretive meaning in a forced-choice procedure as they did in Brewer et al., (2012) confidence procedure. To overcome problems associated with the small scale in the RFC procedure, I conducted an adjusted profile analysis using the Z-scores as indices of discrepancy.
decisions. These results indicate that higher Z-scores are related to higher accuracy in final choices.

Table 8

Proportion of Correct Decisions and Number of Responses for each Z-score Level for Experiment 1

<table>
<thead>
<tr>
<th>Discrepancy</th>
<th>Number of responses</th>
<th>Proportion correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
<td>-2.47-(-0.01)</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>0.00-0.99</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>1.00-1.49</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>1.50-1.74</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>1.75-1.99</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2.00-2.47</td>
<td>3</td>
</tr>
</tbody>
</table>

5.0 Experiment 1 Discussion

Experiment 1 examined the impact of the repeated forced-choice (RFC) lineup procedure on younger (6- to 8-year-olds) and older (9- to 11-year-olds) children’s lineup performance, relative to a simultaneous and face-off procedure. The objective in conducting Experiment 1 was to explore the utility of using the RFC with child eyewitnesses. This was done two ways. First, traditional identification performance measures were calculated and compared across the three procedures. Second, the additional, behavioural information provided by the RFC procedure was explored to try to understand more about children’s choosing behaviour.
5.1 Identification Accuracy

When considering the impact of the RFC procedure on lineup behaviour, age mattered. The RFC procedure was less effective for young than older children, as indicated by higher innocent suspect selections. The face-off procedure appears to have worked best for children in this younger age group, with that procedure showing higher diagnosticity and better suspect survival rates than the RFC procedure. In addition, the face-off procedure reduced young children’s overall levels of choosing when compared to the RFC, though this reduction did not appear to align with better discrimination of guilty from innocent suspects. It is particularly interesting that the face-off procedure produced more accurate responding than the simultaneous with wildcard procedure—a procedure that has been documented to work well with children in this younger age group (e.g., Karageorge & Zajac, 2011). The present data are the first replication of the success of the paired presentation of the face-off procedure reported by Price and Fitzgerald (2016). This consistent finding suggests that, for younger children, the paired-presentation method may be particularly advantageous.

However, not all paired-presentation methods work for younger children. Given the presentation similarities between the face-off and the RFC procedures (i.e., dual-presentation/forced-choice decisions, instructions), lower performance during the RFC suggests that younger children struggled with either, (a) the length of time to complete the RFC procedure, (b) the large number of decisions, or (c) the repeated exposure to faces. These differences are important to consider for future research. Given that the RFC procedure was completed quite quickly (3.87 minutes per lineup on average), younger children’s difficulties with the RFC likely stem from one or both of the latter two
possibilities. For example, when considering the lengthy and erratic time taken (see Figure 6) to make a choice from a pair of faces containing a suspect, younger children seemed to struggle with discriminating between the guilty and innocent suspect—right from the first exposure. Considering this, the low suspect survival rates, and the low ANDI scores, it is possible that repeated exposure to faces decreased younger children’s ability to discriminate between previously seen from unseen faces. However, this is an empirical question for future research.

Older children, conversely, performed well with the RFC procedure—at least as well as they performed with the simultaneous (with wildcard) procedure. Both the lineup identification decisions and diagnosticity were statistically comparable between the two procedures. Given the strong evidence to support the use of the simultaneous with wildcard procedure with 8- to 11-year-old children (e.g., Zajac & Karageorge, 2009), the observation that the RFC procedure produced similar responding with a similar age group is encouraging.

The observed age differences in performance associated with the RFC procedure indicates that, developmentally speaking, something important happens around the age of 9 that allowed older children to use the RFC procedure more effectively. Reducing the stimulus set size to two images per decision and asking children to repeat those decisions was hypothesized to reduce social pressure to choose, reduce the cognitive effort involved in a lineup decision, and help manage pre-potent responses. This, in turn, was hypothesized to benefit children of all ages. However, the accuracy of lineup decisions suggests that the use of the RFC procedure did not allow the younger children to overcome their developmental limitations. Given that the RFC procedure increased
younger children’s incorrect choosing and had no impact on older children’s choosing (cf. simultaneous), it appears that the RFC procedure did not alleviate social pressure to choose in the final decision for either age group. Moreover, contrary to expectations, these results provide evidence that the RFC procedure did not encourage the use of stricter decision criteria. Rather than increasing the conservativeness of their decisions, the RFC procedure made younger children more lenient. Thus, the RFC procedure did not reduce the leniency in children’s decision criterion that was expected.

Instead, the age differences reported in the results suggest that appropriate use of the RFC procedure requires more developed executive control and meta-cognitive skills that the younger age group (between 6 and 8 years old) have not yet achieved. For instance, notable improvements in inhibitory control have been documented between the ages of 6 and 9 (Williams et al., 1999), wherein children begin to demonstrate improved suppression of automatic or prepotent responses (Bruck & Ceci, 1997; Diamond et al., 2002; Williams et al., 1999). Similarly, there is evidence that around the age 9 years, children are better able than their younger counterparts to monitor and regulate the reporting of memories to improve accuracy (e.g., Hiller & Weber, 2013; Koriat et al., 2001; Roebers & Howie, 2003). The erratic latency in responding when shown a suspect in combination with lower discrimination on the RFC lineup task provides evidence that younger children may not have had the executive or meta-cognitive functioning ability needed to effectively regulate their responding over the large number of decisions.

5.2 Prediction of Accuracy

Traditional lineup procedures collect one key piece of information from an eyewitness: who, if anyone, is the culprit. If recorded, two other pieces of evidence can
be used to contextualize or support the utility of a witness’ identification decision: post-identification confidence ratings and response latency. Thus, to appropriately evaluate the RFC procedure, it is important to consider witness’ identification decisions within the larger context of confidence and latency. Post-identification confidence was not predictive of accuracy in any condition for younger children; however, confidence ratings provided by older children were predictive of identification accuracy in all procedures. The confidence ratings provided by older children in the RFC procedure explained more variance (14%) in accuracy of responding than did the confidence ratings in the simultaneous (5%) or face-off (8%) procedures. Similarly, response latency was only predictive of overall accuracy for older children in the RFC procedure, such that faster decisions across the 30 pairs of faces were related to more accurate responding. Thus, the predictive utility of confidence ratings and response latency provided by the RFC was especially informative with the present sample. These results provide evidence that the RFC procedure provides comparative, if not superior, information about suspect guilt relative to a traditional, simultaneous procedure when used with older children.

In addition to the post-identification confidence ratings and response latency information, the RFC also provided information about witnesses’ patterns of responding (i.e., Z-scores). Higher Z-scores were found to be indicative of discrimination between guilty and innocent suspects as well as identification accuracy—especially for older children. Moreover, older children’s patterns of choosing were predictive of identification accuracy over and above the predictive information provided by confidence ratings. The Z-scores were especially helpful in predicting accuracy in situations where confidence was more ambiguous (e.g., confidence ratings of 4 through 7). These results demonstrate
how the behavioural information only available in the repeated forced-choice can be used to predict accuracy and better understand children’s responding.

5.3 Mechanistic Information

The repeated forced-choice procedure is different from a traditional lineup procedure because it was designed to not only provide information about a witness’ lineup decision (who, if anyone, is the culprit), but also to provide behavioural information about how witnesses responded to the lineup procedure. Capturing this behavioural information (e.g., responding patterns, response latency) provides a unique opportunity to learn more the mechanisms that underlie children’s choosing decisions.

One type of behavioural information collected by the repeated forced-choice program is response latency (reaction time). Response latency is assumed to reflect internal cognitive processes, with longer response times indicating more extensive cognitive processing, including memory trace processing (e.g., Brewer & Weber, 2008). In the adult eyewitness literature, latency has been discussed as a sensitive indicator of memory trace strength—wherein the speed of a match between a suspect and a witness’ memory of a culprit should occur faster with accurate matches than inaccurate matches or rejections (Sporer, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004).

The RFC procedure is unique in that it offers two different measures of response latency. First, there is the response time taken to make a final lineup decision (e.g., suspect, filler, reject). The same information can be collected from the simultaneous and face-off procedures. However, this type of reaction time is limited in that we cannot separate the cognitive processing involved with recognition for each individual lineup member—rather, it provides insight into the processing involved in deciding among all of
the lineup members. As Brewer and Weber (2008) argue, this type of identification latency is likely noisy and, without enforcing a rule of speed during decisions, latency can vary greatly without relating to accuracy. Regardless, there seems to be consistency in the literature that increased response latency is related to accurate identification decisions (e.g., Sporer, 1993) and the present experiment found evidence to support this with older children who completed the RFC procedure. However, the total time taken to complete a RFC procedure does not capture the nuanced information that is recorded about a witness’ individual decisions.

The second type of latency measured in the RFC procedure is how long children took to select between pairs of faces during each of the 28 decisions. This type of response latency is particularly interesting because it gives insight into decision processes throughout the procedure (e.g., time to select the innocent versus guilty suspects; time to select the suspect at each of the seven rounds of presentation). Similar types of response latency cannot be observed in standard simultaneous lineup procedures. Examination of this data revealed that older children required more time viewing the guilty suspect before selecting him than when selecting the innocent suspect.

This tells a different story about the relationship between response time and accuracy than what is shown in the above analyses and what is frequently reported in the adult literature (e.g., Sporer, 1993). However, this inverse relationship with children is not unheard of in the child eyewitness literature (Bruer & Pozzulo, 2014; Dunlevy & Cherryman, 2013). Although accurate older children completed the RFC lineup faster, they required more time and, therefore, may have engaged more complex processing to make an accurate decision when viewing and selecting the guilty suspect’s picture. This
information supports that supposition that executive control, specifically inhibitory control, is playing a role in children’s decision-making (Simpson et al., 2012). Younger children’s responses to the guilty suspect appeared to be as random as they were to the innocent suspect (see Figure 6). Despite the fact that older children likely have more developed executive control (Williams et al., 1999), they still appeared to need more time to accurately assess whether a face was the target person.

The second type of behavioural information provided by the RFC procedure pertains to witness choosing patterns. These choosing patterns (summarized as Z-scores) were found to be indicative of individual witness’ ability to discriminate between faces. Profile analyses and examination of mean Z-scores across guilty and innocent suspect selections indicates that both younger and older children are capable of discriminating previously seen from unseen faces. Moreover, these responding patterns were found to be predictive of accurate responding for older children.

These results indicate two things. First, the RFC provided evidence that younger children may have the ability to discriminate previously seen from previously unseen faces—though perhaps not to the extent that older children can. However, this ability may not translate into accurate lineup performance. This is important because it supports existing research that has found that young children have the ability to monitor their recognition but are impaired by the complex design of the lineup task (see Hiller & Weber, 2013). Perhaps asking a child to make a final decision is what negatively impacts the accuracy of their recognition reporting. Second, these results provide evidence that older children’s responding during the first round of the RFC procedure can be used to assess a child’s memory for a suspect, without (completely) relying on a child’s final
lineup decision. By being able to remove (or focus less) on a final decision, we are less dependent on a child’s ability to monitor their decision criterion appropriately in order to make an accurate final lineup decision.

5.4 Conclusions

Overall, in Experiment 1, the RFC procedure emerged as a potentially viable method to extract recognition memory evidence from 9-to-11-year old children. Specifically, the results suggest that the RFC procedure performs as well as a simultaneous lineup procedure in terms of accuracy, and the additional information captured during the procedure adds to the predictive utility of the RFC procedure. The repetitive nature of the RFC method can provide an index of recognition memory that may be used to manage issues related to metacognitive development. Critically, this information stemming from the repetitive nature of the RFC provides an opportunity to gather additional information about the strength of a witness’ memory, without solely relying on his/her final lineup decision. This method of assessing children’s recognition memory is less susceptible to the problems that may otherwise cause witnesses to lower their decision criteria and choose more often (Bruer et al., revision requested; Sauer et al., 2008).

Given that children begin performing more like adults on lineup tasks at around the age of the older group in Experiment 1 (age 9- to- 11-years; Fitzgerald & Price, 2015), and clear evidence of differential effectiveness across development for the RFC was observed, Experiment 2 sought to extend the age range of witnesses into young adulthood.
6.0 Experiment 2. The Repeated Forced-Choice Procedure with Adults

6.1 Experiment 2 Introduction

When a new lineup procedure works well for children, researchers are often faced with questions about how well the same procedure might work for adults. Implementing procedural reforms in the justice system is difficult, and will be even more difficult if the recommendations are different for witnesses of varying ages. Thus, ideally, a procedure that is recommended for use with child witnesses will also be recommended for use with adult witnesses. Yet, we have not yet clearly discovered at what age children perform at adult-like levels in identifications, and whether or not there are ways in which we can close the gap between children and adults. The aim of Experiment 2 was to explore the effectiveness of the RFC procedure relative to a simultaneous procedure with young adults. In doing so, this Experiment was designed to explore two questions: (1) can the RFC procedure be used to extract memorial information from adult eyewitnesses and, if so, (2) how can the use of the RFC procedure inform the discussion regarding how children are different from adults?

6.2 Adult Witnesses: A Larger Need

Child eyewitnesses warrant special attention in the lineup literature because they pose a unique problem—they choose too much or too leniently. This can translate into problems for police investigators. However, the numbers of adult eyewitnesses that contribute to wrongful convictions in the legal system far outweigh the number of child eyewitnesses (Innocence Project, 2016). Although adults may be better able to discriminate an innocent from a guilty suspect than children, they still frequently make mistakes (Fitzgerald & Price, 2015; Steblay et al., 2011). Because adults are not immune
to inaccurate lineup decisions, procedural reforms are equally as important for adult witnesses – and perhaps more important due to the greater frequency with which adults testify as eyewitnesses. While there are unique situations to consider when dealing with a child witness, advancing the field of lineup research will be best addressed by finding a procedure that works for all ages.

6.3 When do Children Perform more “Like” Adults?

A recent meta-analysis examining the relationship between age and lineup identification accuracy revealed that young adults (i.e., university samples) demonstrated superior performance on lineup tasks when compared to children (aged 5 through 13 years) due to increased ability to discriminate guilty from innocent suspects (Fitzgerald & Price, 2015). The ability to discriminate between guilty and innocent suspects was found to increase in children with age. As noted in Experiment 1, the difference in ability between children and adults likely stems from some combination of developmental variables that differently impact children and adults, such as meta-cognition (Hiller & Weber, 2012), inhibitory control (Price & Fitzgerald, 2016), or social pressure (Beal et al., 1992). The repeated forced-choice procedure may be uniquely positioned to help address these issues. By providing a comparison sample of adult identification performance and lineup behaviour, we can learn more about factors that differently impact children and adults.

Several indices were used to draw comparisons between adult responses with data from Experiment 1, including identification accuracy, response or decision time, discrimination, and diagnosticity of responses. While there is no magical age cut-off, there is evidence to suggest that children start to perform more like adult witnesses
around the age of 12- to 13 years old (Fitzgerald & Price, 2015). Thus, the younger sample in Experiment 1 is likely to reflect children who are less ‘adult-like’ on the spectrum and the older sample is likely to reflect children who are more ‘adult-like.’ Key differences that emerge between these three age groups (younger children, older children, and adults) may help provide information about the factors that may contribute to children’s poorer performance.

6.4 Present Experiment

In Experiment 2, I compared young adults’ lineup behaviour when presented with the RFC procedure, relative to a simultaneous procedure. To maintain consistency with the child sample in Experiment 1, the simultaneous procedure included the salient rejection option, the wildcard (Zajac & Karageorge, 2009). Although the repeated forced-choice (RFC) lineup procedure was designed considering the unique situation of child witnesses, the features of the RFC were also expected to benefit adults. For example, reducing the stimulus set size and introducing repetition was expected to reduce the cognitive load associated with the task. Specifically, seeing faces in repeated pairs was expected to reduce the effort involved in making each decision (no longer concurrently comparing all faces with each other). In turn, this reduction in cognitive effort would allow for more reliance on recognition memory when making a decision.

Three main hypotheses were examined. First, it was hypothesized that the RFC procedure would be a useful method to extract information about adult witness recognition memory; that adults would be able to use the RFC procedure to discriminate the previously seen face (guilty suspect) from previously unseen faces. Evidence to support these hypotheses would stem from higher levels of accuracy, diagnosticity, and
discrimination ($d'$) scores (cf. simultaneous procedure). Second, given that young adults are found to perform better than other age groups (Fitzgerald & Price, 2015), it was hypothesized that adult performance with the RFC procedure would be superior to both younger and older children in Experiment 1. Third, the supplementary memorial information available from the RFC procedure (e.g., reaction time, confidence ratings, Z-scores) was expected to reveal differences in how children and adults use the RFC procedure. In turn, these differences are expected to help understand children’s relatively poorer identification performance.

7.0 Method

7.1 Participants and Design

One hundred and fifty-three undergraduate students ($M_{age} = 20.86$, $SD = 5.32$; 79% women; 64% White, 10% Black) were recruited from the University of Regina Psychology Department participant pool. All participants received course credit as compensation. The same lineup materials from Experiment 1 were used, with one exception—Experiment 2 did not include the face-off procedure, a procedure designed for young children that has not previously been used with adults. This study was a 2 (lineup procedure: simultaneous, RFC) × 2 (target: present, absent) between-subjects design.

7.2 Procedure

The study was advertised as an investigation of perceptions of personality. Participants individually viewed two target event videos (approximately 2 minutes in length each), each containing one of the same targets as in Experiment 1. On the first day, participants filled out a consent form where they were told that they would be watching
videos of people and, the following day, would later be asked to answer questions about their perceptions of the people’s personalities. In the female target video, the target reads a book, spills a glass of water, and cleans it up. In the male target video, the target is studying and does stretches/exercises. The next day, the experimenter explained the true purpose of the study prior to administering the lineup on a computer program. Participants were asked to complete a second consent form that outlined the true nature of the study. Once consent was given, the experimenter began the lineup procedure. The experimenter was responsible for launching the computer program and providing instructions but was not near the participant when s/he viewed the lineup or made a decision. The same programs used in Experiment 1 were used to administer the lineups to the adults, with some adjustments to the instructions to better reflect the context (e.g., any references to camp and the visiting doctors were removed; the targets were referred to as “Jessica” and “Matthew” in this experiment). A screening question was used to ensure that participants were unfamiliar with the targets. Experimenters were blind to target presence, but were responsible for ensuring counterbalancing of the order in which the target lineups were shown (i.e., male or female target first). Participants were told that the people from the videos may or may not be present in the lineups.

**8.0 Results**

**8.1 Lineup Identification Choice**

**8.1.1 Target-present.** To understand the influence of lineup procedure on identification responses, a hierarchical log-linear analysis (HILOG) was conducted that

---

9 Four witnesses did not complete the lineup for the male target as they indicated the target was familiar to them (via screening questions).
included the two manipulated variables and identification response as the dependent variable. The 2 (procedure: simultaneous, RFC) x 2 (actor: male, female) x 3 (lineup response: suspect, filler, reject) HILOG for target-present lineups revealed that the highest order interaction was a two-way effect, \( \chi^2(5) = 18.14, p = .003 \). Partial association analyses indicated a two-way interaction between actor and response, \( \chi^2(2) = 17.55, p < .001 \), such that more correct identifications and fewer incorrect rejections were made in the female than the male lineup. Similar to Experiment 1, the lack of a three-way interaction between actor, response and procedure suggests that the differences in responding across actors were consistent across the two lineup procedures. As such, targets were collapsed for the remaining analyses. For a complete breakdown of responses by actor, see Appendix F.

**8.1.2 Target-absent.** The 2 (procedure: simultaneous, RFC) x 2 (actor: male, female) x 3 (lineup response: suspect, filler, reject) HILOG for target-absent conditions revealed no higher order effect, \( \chi^2(2) = 3.26, p = .20 \), nor a higher order two-way effect, \( \chi^2(5) = 5.80, p = .33 \). See Table 9 for witnesses’ lineup decisions.
Table 9

Identification Responses for Experiment 2

<table>
<thead>
<tr>
<th>Target</th>
<th>Procedure</th>
<th>Identification Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Suspect</td>
</tr>
<tr>
<td>Present</td>
<td>RFC</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
<td>0.64</td>
</tr>
<tr>
<td>Absent</td>
<td>RFC</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Note. Given that each witness made two identifications, n represents the number of identifications made in each condition.

8.2 Survival Statistics

Survival statistics were calculated for the RFC procedure to help identify the proportion of suspects that survived multiple rounds of lineup procedures. As seen in Table 10, the guilty suspect survived to the final round more frequently than did the innocent suspect, suggesting that witnesses are able to discriminate the guilty from innocent suspect using the RFC procedure. In particular, the number of innocent suspect selections decreased at each Round, while the number of guilty suspect selections remained relatively consistent at each Round.

Table 10

Suspect Survival Rates in Experiment 2

<table>
<thead>
<tr>
<th>Target</th>
<th>n</th>
<th>Survived R1</th>
<th>Survived R2</th>
<th>Survived R3</th>
<th>R4-Final Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>75</td>
<td>.76</td>
<td>-</td>
<td>.76</td>
<td>.71</td>
</tr>
<tr>
<td>TA</td>
<td>73</td>
<td>.55</td>
<td>-</td>
<td>.32</td>
<td>.18</td>
</tr>
</tbody>
</table>

Note. TA= Target-absent, TP= Target-present. R1-R4 = Round 1, 2, 3, or 4
8.3 Lineup Comparison

8.3.1 Signal Detection Theory. Similar to Experiment 1, $d'$ and response bias ($c_{suspect}$) were calculated for witness decisions across the two lineup procedures used.

As seen in Table 11, discrimination was higher for RFC than for the simultaneous lineup.

Table 11

*Signal Detection Information ($d'$ and $c_{suspect}$) for Experiment 2*

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Identification Decision</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$d'$</td>
<td>$c_{suspect}$</td>
</tr>
<tr>
<td>Simultaneous</td>
<td></td>
<td>1.06</td>
<td>0.21</td>
</tr>
<tr>
<td>RFC</td>
<td></td>
<td>1.47</td>
<td>0.19</td>
</tr>
</tbody>
</table>

8.3.2 Diagnosticity ratios. Diagnosticity ratios were calculated for all three types of decision outcomes for the simultaneous and RFC procedures. As seen in Table 12, the simultaneous and RFC procedures produced similar diagnosticity ratios for all three lineup decisions. The wide inferential confidence intervals indicate that any differences did not reach statistical significance. This finding indicates that adults have similar response patterns across the two lineup procedures.

Table 12

*Diagnosticity Ratios in Experiment 2*

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Diagnostic of Guilt</th>
<th>Diagnostic of Innocence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Suspect Identification</td>
<td>Filler Selection</td>
</tr>
<tr>
<td></td>
<td>$DR$ $LL$ $UL$</td>
<td>$DR$ $LL$ $UL$</td>
</tr>
<tr>
<td>Simultaneous</td>
<td>2.76 1.96 3.90</td>
<td>1.91 1.16 3.14</td>
</tr>
</tbody>
</table>
Note. DR = diagnosticity ratio. LL = lower limit 95\% inferential confidence interval. UL = upper limit of 95\% inferential confidence interval. Inferential confidence intervals were calculated (Tryon, 2001) using a bootstrapping procedure (Palmer et al., 2010) in which log scales (ln) were used to more closely approximate a normal distribution. The confidence intervals in Table 12 have been converted from log back to their original unit. For suspect identifications, diagnosticity was computed as the ratio of target-present/target-absent responses. For filler selections and rejections, diagnosticity was computed as the ratio of target-absent/target-present responses.

<table>
<thead>
<tr>
<th>RFC</th>
<th>3.97</th>
<th>2.68</th>
<th>5.88</th>
<th>4.45</th>
<th>2.32</th>
<th>8.54</th>
<th>2.18</th>
<th>1.51</th>
<th>3.16</th>
</tr>
</thead>
</table>
**8.3.3 Information gain.** The information gain curves for simultaneous and RFC procedures are shown in Figure 9. When focusing on suspect identifications, the RFC procedure provided more information about guilt than did the simultaneous procedure—especially when the base rate is low. Similarly, for filler identifications, the RFC procedure provided the most information regarding innocence—particularly when the base rate was about 50%. When looking to rejections, the two procedures are not notably distinguishable.

![Information Gain Curves](image)

*Figure 9. Information gain in Experiment 2.*
8.3.4 BREE curves. A BREE curve (Wells et al., 2015) outlining the required change in base rate of suspect identification for the simultaneous and RFC procedures is shown in Figure 10. Assuming a base rate of 50% (half of the lineups include a guilty suspect), the simultaneous procedure would need to only increase the base rate by 9% to produce suspect identifications that are as diagnostic of suspect guilt as the RFC procedure.

![Graph showing BREE curve](image)

*Figure 10. Base-rate effect-equivalency (BREE) curve in Experiment 2 for suspect identifications from RFC and simultaneous. The first procedure listed represents the procedure with the higher diagnosticity.*

8.4 Supplementary Memorial Information

8.4.1 Confidence-Accuracy relationship. Logistic regressions revealed that confidence ratings following each identification significantly predicted overall accuracy for the simultaneous, $B = .31$, $\text{Exp}(B) = 1.37$; Nagelkerke pseudo $R^2 = .16$, $\chi^2 (1) = 18.98$, $p < .001$, and RFC procedure, $B = .30$, $\text{Exp}(B) = 1.35$; Nagelkerke pseudo $R^2 = .13$, $\chi^2 (1) = 15.55$, $p < .001$. When only considering the post-identification rating, the 13% of variance explained in the RFC procedure is very similar to the 16% explained by
confidence ratings in the simultaneous lineup. Using point bi-serial correlations (Weber & Brewer, 2004), a significant confidence-accuracy relationship was found for positive, but not negative, responses in both procedures (see Table 13).

Table 13

<table>
<thead>
<tr>
<th></th>
<th>Positive Decision</th>
<th>Negative Decision</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p</td>
<td>n</td>
</tr>
<tr>
<td>Simultaneous</td>
<td>.55</td>
<td>&lt;.001</td>
<td>96</td>
</tr>
<tr>
<td>RFC</td>
<td>.45</td>
<td>&lt;.001</td>
<td>98</td>
</tr>
</tbody>
</table>

Note. Given that each witness made two identifications, n represents the number of identifications made in each condition. \( r \) = Pearson r.

8.4.2 Response latency. On average, it took adults less than three minutes to complete the RFC procedure (\( M = 2.69 \) minutes, \( SD = 1.09 \)) and less than a minute to complete the simultaneous procedure (\( M = 0.30 \) minutes, \( SD = 0.23 \)). Response latency significantly predicted overall accuracy for positive responses from the simultaneous procedure, \( B = .00, \exp(B) = 1.00; \) Nagelkerke pseudo \( R^2 = .19, \chi^2(1) = 14.42, p < .001. \)

Response latency also significantly predicted overall accuracy for positive responses from the RFC procedure, \( B = .00, \exp(B) = 1.00; \) Nagelkerke pseudo \( R^2 = .05, \chi^2(1) = 4.05, p = .04. \) Response latency was not indicative of accuracy in any other conditions. Adults took less time to make a decision when viewing a guilty suspect (\( M = 2.21, SD = 1.23; \) see Figure 11) than when viewing an innocent suspect (\( M = 3.06, SD = 1.49, t(182) = 4.19, \ p < .001 \). This trend was also observed when adults ultimately selected the suspect from the pair of faces (\( M = 2.45 \) seconds, \( SD = 1.51; M = 3.12, SD = 1.40, \)
respectively; \( t(179) = 3.10, \ p = .001; \) see Figure 11). This suggests that when shown a suspect, adults took less time to evaluate and make their decisions when looking at the guilty suspect.

*Figure 11.* Mean response latency for decisions made when the suspect was shown and selected.

**8.4.3 Decision patterns and Z-scores.** Z-scores were calculated to compare how often each pattern of responding aligned with each lineup outcome (see Table 14). Unlike what was observed in Experiment 1, Z-scores were not different for guilty suspect identifications (\( M = 1.11, \ SD = 0.47 \)) and innocent suspect identifications (\( M = 1.12, \ SD = 0.64 \)), \( t(63) = 0.03, \ p = .25 \).
Table 14

*Frequency of Z-Scores for each Decision Type*

<table>
<thead>
<tr>
<th>Z Statistics</th>
<th>TA</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Suspect</td>
<td>Filler</td>
<td>Reject</td>
<td>Suspect</td>
<td>Filler</td>
<td>Reject</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>n=13</td>
<td>n=26</td>
<td>n=34</td>
<td>n=53</td>
<td>n=6</td>
<td>n=17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.00-2.47</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>0.23</td>
<td>0.12</td>
<td>0.09</td>
<td>0.23</td>
<td>0.00</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>0.54</td>
<td>0.38</td>
<td>0.29</td>
<td>0.47</td>
<td>0.33</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>0.23</td>
<td>0.31</td>
<td>0.35</td>
<td>0.15</td>
<td>0.50</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.47(-0.01)</td>
<td>0.00</td>
<td>0.19</td>
<td>0.24</td>
<td>0.11</td>
<td>0.17</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Columns may not sum to one due to rounding. Unlike the bins used to examine model fit (i.e., grouping Z-scores based on third percentile; See Table 16), these bins or groups of Z-scores were decided considering ease of interpretation for a user.

### 8.4.4 Predictive utility of Z-scores

A logistic regression revealed that Z-scores or patterns of responding were not predictive over and above the 13% of variance in accuracy explained by post-identification confidence ratings ($B = .31, \text{Exp}(B) = 1.36, p = .003; \text{Nagelkerke pseudo } R^2 = .13, \chi^2(1) = 13.96, p < .001$).

### 8.5 Individual Witness Discrimination

#### 8.5.1 Index of discrimination

ANDI scores indicated that adults were able to discriminate the target well using the RFC procedure ($\text{ANDI} = 0.15, 95\% \text{CI} [0.08, 0.21]$). Inferential confidence intervals allow for the comparison across the two experiments and indicate that the ANDI scores for the younger children were significantly lower than the older children and adults.
8.5.2 Profile analysis. Profile analyses were conducted to highlight which individual sets or patterns of selections was likely (versus not likely) to indicate accurate discrimination between previously seen and previously unseen faces. Table 15 provides accuracy information as a function of discrepancy. As seen in Table 15, higher Z-scores are related to higher accuracy in final choices.

Table 15

Proportion of Correct Decisions and Number of Responses for each Z-score Level for Experiment 2 data.

<table>
<thead>
<tr>
<th>Discrepancy</th>
<th>Number of responses</th>
<th>Proportion correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.47-(-0.01)</td>
<td>24</td>
<td>0.58</td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>40</td>
<td>0.50</td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>61</td>
<td>0.57</td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>25</td>
<td>0.60</td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>3</td>
<td>1.00</td>
</tr>
<tr>
<td>2.00-2.47</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

9.0 Experiment 2 Discussion

The purpose of Experiment 2 was to explore how the RFC procedure influenced adults’ identification accuracy. Similar to what was observed with older children in Experiment 1, the RFC procedure produced similar choosing and selection behaviour as the simultaneous lineup (with wildcard) in adults. Though there were no statistically significant differences, compared to the simultaneous procedure, the RFC procedure was numerically more diagnostic of suspect guilt and, when fillers were identified, more
informative about suspect innocence. The RFC procedure produced numerically more correct identifications and fewer innocent suspect identifications than the simultaneous procedure. Notably, when examining the survival rates of the suspect in target present and absent lineups, the RFC procedure performed well; the guilty suspect’s survival rate was relatively consistent (only 4% decrease at Round 3) across the three rounds of decisions. Alternatively, when the target was absent, the survival rate of the innocent suspect decreased considerably at each decision round. Moreover, post-identification confidence ratings were predictive of overall accuracy for both procedures, including for those who made a positive response (chose from the lineup). Similarly, response latency associated with the final lineup decision was predictive of positive response accuracy for both procedures. Thus, there was clearly no cost to administering the RFC, relative to the simultaneous lineup (with wildcard).

There were some observable changes in witnesses’ lineup behaviour when presented with the RFC relative to the simultaneous procedure. While these differences were not significant, discussion of these patterns is warranted as they provide a foundation from which to understand the impact of the results from the next section of this dissertation, Additional Analyses. Specifically, presentation of the RFC procedure appears to produce fewer filler identifications when the guilty suspect is present. However, when the guilty suspect is absent, the observed reduction in innocent suspect identifications appears to be driven by an increase in filler identifications. Thus, the RFC procedure may simply shift inaccurate decisions in target-absent lineups from one error type (innocent suspect identifications) to another error type—filler identifications (i.e., filler siphoning; Wells, Smalarz, & Smith, 2015).
Many have argued that in an applied setting when a lineup is properly administered, filler identifications errors are less problematic than innocent suspect identifications as they are known to be innocent individuals (Wells, 2006). However, when an eyewitness fails to reject a lineup and, instead, identifies a known innocent individual from a lineup—there are implications for his/her perceived credibility as a witness (i.e., seen as a poor witness/poor memory). This is especially likely in situations where the only information received by investigators is a single decision from a witness. However, as discussed in the Additional Analyses, use of the RFC procedure can provide additional information about a witness’ memory that may help investigators to better assess the credibility of witnesses who select a known-innocent (i.e., filler) from a lineup. Thus, considering that both the simultaneous and RFC procedures produced high filler identifications in target-absent lineups, it can be argued the RFC procedure can provide more useful information to investigating officers when a situation emerges in which a witness identifies a filler.

Although there are differences in how participants from Experiment 1 and 2 viewed the targets (live versus video), comparing the results across the two experiments provides some important insight. As expected, the diagnosticity of the RFC procedure was higher with adults (3.97) than it was for younger children (1.63). However, contrary to expectations, older children had a higher diagnosticity ratio (9.68) and higher discrimination scores than adults. Although post-identification confidence ratings predicted overall accuracy, unlike what was observed with older children, Z-scores (or choosing patterns) did not add to the predictive utility of the procedure in adults.
Comparison across experiments suggests that the RFC procedure was most useful in measuring recognition memory for children aged 9 through 11.

It is also interesting to note the response latency distinction across experiments. Unlike what was observed with older children in Experiment 1, adult witnesses made faster decisions when picking the guilty suspect than when picking the innocent suspect during the RFC procedure. This suggests that, for adults in this sample, accurate recognition memory appeared to involve more immediate cognitive processing than it did for children. From a developmental perspective, this supports previous claims that the development of executive functioning (ability to inhibit target-irrelevant information) is a mediating factor in children’s lineup performance (Price & Fitzgerald, 2016; Roberts & Powell, 2005; Simpson et al., 2012). Specifically, it suggests that children need more time than adults need to correctly decide to report that a lineup member is the guilty suspect.

It is important to highlight, however, that the simultaneous procedure used in Experiment 2 included a visual, salient rejection option (i.e., the wildcard; Zajac & Karageorge, 2009). Though the wildcard was designed for and is typically used in lineups with children, there is evidence that the simultaneous with wildcard procedure encourages conservative responding with adult witnesses as indicated through increased rejections for both target-absent and target-present lineups (Bruer, Fitzgerald, Therrien & Price, 2014). Two implications of this finding should be highlighted. First, considering the similar rejection rates across the two lineup procedures, it is plausible that the RFC procedure also encourages more conservative decision-making. Second, it is likely that the exclusion of the wildcard from the simultaneous procedure would have resulted in
differences in performance across the RFC and simultaneous procedure in adults. The next natural step in this research will be to explore how the RFC compares to more commonly used procedures with adult witnesses, such as the simultaneous procedure (without wildcard) and the sequential procedure.

Taken together with the accuracy rates, diagnosticity ratios, and discrimination scores (i.e., ANDI), these results support the hypothesis that adult witnesses can use the RFC procedure to discriminate between guilty and innocent suspects. Although I did not observe the expected improvement in identification accuracy, the results suggest that there is no cost to administering the RFC, relative to the simultaneous lineup (with wildcard). This is encouraging when considering that the RFC procedure was designed to capture more information about a witness’ memory than other lineup procedures. To explore the utility of this additional information for interpreting a witness’ lineup decision, additional analyses were conducted using data from both Experiment 1 and Experiment 2.
10.0 Additional Analyses. A Round Robin Model to Assess Recognition Memory Strength with the RFC Procedure

10.1 Additional Analyses Introduction

Eyewitness’ response patterns during the first round of the repeated forced-choice procedure can be summarized as a Z-score that reflects an individual witness’ level of suspect selection bias. In Experiment 1, and to a lesser extent Experiment 2, the utility of Z-scores in reflecting accurate discrimination and in predicting witness accuracy are outlined. That is, higher Z-scores reflect discrimination that is more accurate and can be used in combination with a child witness’ post-identification confidence rating to estimate the likelihood of memory accuracy. However, reducing a witness’ pattern of responding to a single Z-score may obscure valuable information about the witness’ memory. In a legal setting, if an investigator uses the RFC, he or she will have a pattern of responding from a single witness and will need to interpret that pattern. The repeated forced-choice model has only a finite number of possible patterns of responding across the eight lineup members. These finite patterns can be translated into a tool to help interpret what a witness’ Z-score may indicate about memory.

The goal of these additional analyses was to explore the additional memorial information about a witness’ target memory provided by witness’ choosing patterns during the RFC procedure. To do so, I developed a model that aids in the interpretation of an individual witness’ pattern of responding as an indication of selection bias, which, in turn, may represent a proxy for memory strength. Below, the rationale for requiring a model is described as well as how the model was developed, and how the model aids in the interpretation of lineup behaviour.
10.2 Problems with Applying Traditional Analyses

There are alternative lineup procedures that gather data about each lineup member, rather than a single final lineup decision (e.g., Sauer et al., 2008; Weber & Varga, 2014). However, the statistical analyses used with this granular lineup data (e.g., Brewer et al.’s profile analyses, 2012; Sauer et al.’s classification algorithms, 2008) cannot be applied to the repeated forced-choice design. In these procedures, witnesses provided a rating of their confidence that each lineup member was the culprit. Each rating of a lineup member is made independent from the ratings given previously. That is, a witness is free to give a rating of 100 to all lineup members, or a rating of 0 to all lineup members.

The RFC design, in contrast, can be compared with a round robin design (see Fienberg & Wasserman, 1981; Wong, 1984) in which lineup members are compared based on similarity to memory of target. Each lineup member is compared against every other lineup member, resulting in 28 decisions in an eight-person lineup. As these 28 decisions are made, each selection will impact the possible remaining selections. For example, consider the situation in which the first pair of lineup members shown to a witness includes lineup members 4 and 2. When the witness selects the winner (e.g., 2 is most similar to their memory for the target), the winning member has a ‘score’ of one, while the losing member has a ‘score’ of zero. Thus, going forward, the winning member has a potential to ‘score’ the maximum, while the losing member cannot exceed a score of six. Thus, ‘scores’ during the RFC procedure are limited by previous responses. This produces interdependent responses that limit the variability between lineup members—
making it difficult to find meaningful differences using previously applied approaches. Therefore, a new method of analysis was required to overcome these issues.

11.0 Method and Results

11.1 A Round Robin Model using the Probability of Responses

When a witness chooses from the RFC procedure, his/her pattern of choosing reflects a selection bias for the suspect that may be interpreted as a proxy for memory strength or discrimination for the suspect. In order to develop a model for use with a round-robin design such as the RFC procedure, an assumption must be made to explain how responding during the model would reflect varying levels of suspect selection bias (i.e., potentially a proxy for memory strength). A low suspect selection bias may result if the witness has a weak memory for the suspect (e.g., poor encoding or target is absent from the lineup). In this case, the corresponding response pattern should be relatively random across the eight lineup members and reflect little selection bias for any single lineup member. That is, it can be expected that the likelihood of any single lineup member being selected over any other is approximately 50%. With a stronger suspect selection bias, a witness’ pattern of responding should be less random across the eight lineup members. That is, the likelihood that the suspect will be selected over another lineup member should be greater than chance (i.e., 60%, 70%, 80%, or 90%). Importantly, a strong selection bias does not necessarily equate with memory strength for the guilty suspect, but rather could also be a result of mistaken recognition or commitment effect to the suspect, given that witnesses were instructed to ‘select the one that looks MOST like the target’.
11.2 Model Development

To create a tool for applied use, a model was developed that was not based on any single data set, including the present data, but rather based on this conceptual framework of suspect selection bias surrounding the RFC procedure. First, five levels of suspect selection bias were selected to be represented in the model\(^\text{10}\). The first level (50\%) reflects a random level of selection bias whereby any lineup member, including the suspect, has a 50\% chance of being selected over any other lineup member. For the second level, the model was designed such that the suspect would win over any other face 60\% of the time, while each filler continued to have a 50\% chance of winning when paired against any other filler. I continued to increase the rate at which the suspect would win for the third level (70\%), fourth level (80\%), and fifth level (90\%). For each of these five selection bias levels, 300,000 simulations were run to determine the frequency at which all-possible\(^\text{11}\) patterns of responding occurred. To simplify these model-generated patterns of responding, Z-scores were calculated the same way as described in Experiments 1 and 2.

11.3 Application of Model Data to Experiment Data

11.3.1 Model fit. Next, I examined how well the model fit the data from Experiments 1 and 2 using a G-test for goodness of fit (likelihood ratio test; McDonald, 1989). G-statistics for each of the model parameters (bias at 5\% increments between 50\% and 90\%) were calculated separately for target-absent and target-present conditions. A good-fitting model would be one that produces no significant difference \((p > .05)\)

\(^{10}\) Note that, due to computational intensity of the process, these model parameters were selected to ensure efficiency. With more computational power, one could optimize and test the model using 300,000 simulations at more granular increments (e.g., 81, 82, 83).

\(^{11}\) All possible patterns when disregarding the rankings/order of fillers wins.
between the observed (experiment) data and the expected (model) data. Model fit results are presented in Table 16.

Overall, the model provided adequate fit for Experiments 1 and 2 data. For all children (both younger and older children) combined, the results indicated that the 48%-50% parameter model was a good fit for Z-scores for target-absent conditions (See Table 16). This indicates that, when the target is absent, the patterns of responses are statistically similar to the model’s definition of random responding. Likewise, for target-present conditions, the results indicate a good fit between Experiment 1 data and the 70% parameter model data. For adults, there were similar patterns in target-present lineups, wherein the 70% parameter model provided the best fit for the Experiment 2 data. However, when the target was absent, adults’ patterns of choosing are best described by the 65% parameter model (cf. children’s 50% bias). Given that model fit data was similar for both age groups of children, data are collapsed across age for subsequent analysis of model data.
Table 16

Model Fit Statistics for Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Target-Absent</th>
<th>Target-Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G Statistic</td>
<td>df</td>
</tr>
<tr>
<td>Younger Children</td>
<td>90%</td>
<td>344.77</td>
</tr>
<tr>
<td>80%</td>
<td>172.25</td>
<td>3</td>
</tr>
<tr>
<td>75%</td>
<td>120.89</td>
<td>3</td>
</tr>
<tr>
<td>70%</td>
<td>80.05</td>
<td>3</td>
</tr>
<tr>
<td>65%</td>
<td>50.91</td>
<td>3</td>
</tr>
<tr>
<td>60%</td>
<td>28.18</td>
<td>3</td>
</tr>
<tr>
<td>50%</td>
<td>3.80</td>
<td>3</td>
</tr>
<tr>
<td>48%</td>
<td>1.48</td>
<td>3</td>
</tr>
<tr>
<td>Older Children</td>
<td>90%</td>
<td>572.87</td>
</tr>
<tr>
<td>80%</td>
<td>284.72</td>
<td>3</td>
</tr>
<tr>
<td>75%</td>
<td>199.32</td>
<td>3</td>
</tr>
<tr>
<td>70%</td>
<td>131.15</td>
<td>3</td>
</tr>
<tr>
<td>65%</td>
<td>83.01</td>
<td>3</td>
</tr>
<tr>
<td>60%</td>
<td>45.19</td>
<td>3</td>
</tr>
<tr>
<td>50%</td>
<td>5.36</td>
<td>3</td>
</tr>
<tr>
<td>48%</td>
<td>1.95</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Target-Absent</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>910.91</td>
</tr>
<tr>
<td>All</td>
<td>90%</td>
<td>230.36</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>452.26</td>
</tr>
<tr>
<td>Children</td>
<td>75%</td>
<td>316.41</td>
</tr>
<tr>
<td></td>
<td>70%</td>
<td>207.94</td>
</tr>
<tr>
<td></td>
<td>65%</td>
<td>131.45</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>71.35</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>8.28</td>
</tr>
<tr>
<td></td>
<td>48%&lt;sup&gt;2&lt;/sup&gt;</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>170.47</td>
</tr>
<tr>
<td>Adults</td>
<td>80%</td>
<td>46.24</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>19.77</td>
</tr>
<tr>
<td></td>
<td>70%</td>
<td>4.53</td>
</tr>
<tr>
<td></td>
<td>65%</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>4.83</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>36.37</td>
</tr>
</tbody>
</table>

<sup>1</sup> df is calculated as n-1+number of parameters varied (1).  <sup>2</sup>The 48% parameter was added in only for target-absent conditions for children to demonstrate which parameter was a better fit for the data. Bolded information and p-values with an asterisk (*) indicates good model fit. Model fit was based on a binning approach by dividing frequency Z-scores into thirds.
Taken together, these results suggest that, when the target is absent from the lineup, children’s behaviour reflects random or unbiased choosing. Adults, on the other hand, appeared to more closely follow (or were better able to follow) the instruction to *select the one that looks MOST like the target* and, in turn, they favoured the innocent suspect more than the fillers. See Figure 12 for a visual comparison of which level of selection fit best with each age group. There are several interesting observations to be made in Figure 12. Specifically, the peaks of the target-present curves are lower on the y-axis (lower \( p \)-values) than the target-absent curves. That is, the target-present curves do not reach a \( p \)-value of 1 (i.e., perfect fit). This suggests that the model provides a better fit for target-absent data.

*Figure 12.* P-values plotted at each model parameter for children (collapsed across age) and adults. Model data were examined using a bootstrapping procedure with 5000 simulations. Note that the data in Table 16 are based on results from a sample of 300,000 simulations. In this instance, the number of simulations were reduced (to a level that produced results similar to the model with 300,000 simulations) to allow for more
granular examination of each parameter. The dashed line indicates the level of significance ($p = .05$)

11.4 Using the Model to Interpret Witness Responding

Once model fit was established (and which parameter best fit) with Experiment 1 and 2 data, the next step was to apply the model data to help interpret individual witness responses from Experiment 1 and 2. I am interested in knowing what patterns of $Z$-scores are most common at each of the varying levels of suspect selection bias in the model data. To compare the frequency of responding across the model data, odds ratios (OR) were calculated to indicate the likelihood of strong or weak suspect selection bias associated with a range of $Z$-scores. Using the OR only from the model data, a look-up table was created (Table 17). These OR in the look-up table indicate the likelihood that the pattern came from a higher suspect selection bias simulation (e.g., 70%, 80%) rather than a lower suspect selection bias or random (i.e., 50%) simulation.

Table 17

*Model Simulation Odds Ratio – Z-score Look-up Table*

<table>
<thead>
<tr>
<th>Z Statistic</th>
<th>OR:50%</th>
<th>OR:60%</th>
<th>OR:70%</th>
<th>OR:80%</th>
<th>OR:90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.00-2.47</td>
<td>27.58</td>
<td>9.60</td>
<td>3.98</td>
<td>1.88</td>
<td>-</td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>17.90</td>
<td>7.06</td>
<td>3.27</td>
<td>1.72</td>
<td>-</td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>7.41</td>
<td>3.47</td>
<td>1.92</td>
<td>1.13</td>
<td>-</td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>2.11</td>
<td>1.18</td>
<td>0.83</td>
<td>0.75</td>
<td>-</td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>0.22</td>
<td>0.19</td>
<td>0.21</td>
<td>0.33</td>
<td>-</td>
</tr>
<tr>
<td>-2.47-(-0.01)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.08</td>
<td>-</td>
</tr>
<tr>
<td>Z Statistic</td>
<td>OR:50%</td>
<td>OR:60%</td>
<td>OR:70%</td>
<td>OR:80%</td>
<td>OR:90%</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>2.00-2.47</td>
<td>14.66</td>
<td>5.10</td>
<td>2.12</td>
<td>-</td>
<td>0.53</td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>10.43</td>
<td>4.12</td>
<td>1.91</td>
<td>-</td>
<td>0.58</td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>6.57</td>
<td>3.07</td>
<td>1.71</td>
<td>-</td>
<td>0.89</td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>2.82</td>
<td>1.58</td>
<td>1.11</td>
<td>-</td>
<td>1.34</td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>0.69</td>
<td>0.58</td>
<td>0.64</td>
<td>-</td>
<td>3.08</td>
</tr>
<tr>
<td>-2.47(-0.01)</td>
<td>0.07</td>
<td>0.12</td>
<td>0.27</td>
<td>-</td>
<td>12.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.00-2.47</td>
<td>6.93</td>
<td>2.41</td>
<td>-</td>
<td>0.47</td>
<td>0.25</td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>5.47</td>
<td>2.16</td>
<td>-</td>
<td>0.52</td>
<td>0.31</td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>3.85</td>
<td>1.80</td>
<td>-</td>
<td>0.59</td>
<td>0.52</td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>2.54</td>
<td>1.43</td>
<td>-</td>
<td>0.90</td>
<td>1.21</td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>1.07</td>
<td>0.91</td>
<td>-</td>
<td>1.55</td>
<td>4.78</td>
</tr>
<tr>
<td>-2.47(-0.01)</td>
<td>0.26</td>
<td>0.45</td>
<td>-</td>
<td>3.73</td>
<td>47.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.00-2.47</td>
<td>2.87</td>
<td>-</td>
<td>0.41</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>1.75-1.99</td>
<td>2.54</td>
<td>-</td>
<td>0.46</td>
<td>0.24</td>
<td>0.14</td>
</tr>
<tr>
<td>1.50-1.74</td>
<td>2.14</td>
<td>-</td>
<td>0.55</td>
<td>0.33</td>
<td>0.29</td>
</tr>
<tr>
<td>1.00-1.49</td>
<td>1.78</td>
<td>-</td>
<td>0.70</td>
<td>0.63</td>
<td>0.84</td>
</tr>
<tr>
<td>0.00-0.99</td>
<td>1.18</td>
<td>-</td>
<td>1.10</td>
<td>1.72</td>
<td>5.28</td>
</tr>
</tbody>
</table>
Note. Each row reflects groupings of Z-scores. Each column title (indicated by darker shading) indicates the suspect selection bias level that each subtitle (indicated by “OR:”) is compared with. That is, the numerator in the odds ratio is the column title value while the denominator is the subtitle value. For example, under the column title of 90% (90% shaded title), the odds ratio of 1.88 is calculated by the frequency of the Z-score in the model at 90% divided by the frequency at the 80% level. Interpretation of this odds ratio is that the Z-score range of 2.00-2.47 is 1.88 times more likely to emerge at the 90% bias level, relative to the 80% bias level.

Next, I compared individual witness’ Z-scores from my experimental data to the look-up table to determine the likely level of suspect selection bias demonstrated by particular witnesses. Note that this comparison allows only for consideration of suspect (innocent or guilty) bias, not for another lineup member that may have been ultimately chosen from the lineup.

Comparison of Z-scores from the experiments with the Z-score ORs produced from the model can help to discriminate between seemingly similar witnesses. For example, take two older child witnesses from Experiment 1 whose performance appears
very similar at first glance. Both witnesses correctly identified the guilty suspect, had similar response latency (just over 2 minutes), and provided a confidence rating of 10 (out of 11). The information about Z-scores and the probability of selection bias can help to differentiate these similar witnesses. The first witness’ pattern of responding translated into a Z-score of 1.82. Using the table, one can observe that this particular witness’ pattern is likely to reflect a selection bias value of 70% or higher (this witness’ Z-score or selection pattern is 5.47 times more likely in the 70% bias level than in the random bias level). Conversely, the second witness’ pattern of responding produced a Z-score of 0.94. From the look-up table, it can be seen that this witness’ pattern of responding reflects more random selection bias (this witness’ Z-score is just as likely to appear in the 50% bias level as the 60% and 70% bias levels, as indicated by OR close to 1). From this, I surmised that the first witness likely had a stronger selection bias towards the suspect (and perhaps stronger memory for the suspect) than the second witness.

This information allows us to understand more about a witness’ ability to discriminate (i.e., selection bias) a suspect from fillers. More crucially, these ORs can be used to understand a witness’ memory for the suspect, even if the final decision was to select a filler or reject the lineup. Take, for instance, a 6-year-old participant who rejected a target-present lineup but produced a Z-score of 1.69, or an 11-year-old participant who selected a filler from a target-present lineup and produced a Z-score of 1.50. The table suggests that these patterns are 3.85 times more likely in the 70% bias level than in the 50% level. Thus, despite rejecting the guilty suspect or selecting a filler, the selection patterns indicate non-random discrimination of the guilty suspect.
12.0 Additional Analyses Discussion

Using a theoretical framework outlined by the repeated forced-choice lineup procedure, a model was developed to assist with interpreting the complex output of the RFC procedure. Because the model was designed to reflect the RFC procedure, it allowed for the control over all factors. The data from Experiments 1 and 2 were found to have a good match with the model data which allowed me to use the selection bias model data to further understand witnesses’ responding during the RFC procedure. This model presents an opportunity to learn more about individual differences in witness behaviour and develop a greater understanding of what selection patterns during the RFC indicate about memory—for example, how witnesses respond when a target is present versus absent or how children respond to a RFC procedure versus to adult witnesses.

With the model, a witness’ choosing pattern during the RFC procedure can be used to help assess the diagnostic information contained in a witness’ identification decision. Specifically, a witness’ responding during a RFC procedure can be assessed for the likelihood of random versus non-random choosing towards the suspect. This is important as it provides a framework to help infer a witness’ memory strength for the suspect. Though not diagnostic of witness reliability, this information may be useful when considering the desire to learn how much weight an investigator or legal decision maker should place on a particular witness’ evidence.

12.1 Future Directions

Although these initial analyses are promising, there is much work to be done to further develop the model. As a starting point in assessing the value in such a model, I thought it important to begin using a simple, exploratory approach. That is, the model
was designed to vary only one parameter (how frequently the suspect is selected at each level of selection bias [90% bias—the suspect wins 90% of the time] and kept the filler selections at random or 50%). The advantage of using a simplistic model was that it allowed for both a basic examination of whether the model fit the data and for a simplistic interpretation of the results.

Of course, this model was designed as an ideal. In a real world setting, one or two fillers may be a stronger match to a witness’ memory of a target than others—as opposed to this model which assumed all fillers would be selected at random. The next natural step is to use more computing power and sophisticated programming to test a model that not only varies how often the suspect wins, but also how often each filler wins. For example, ad-hoc similarity ratings between the lineup members could be used to evaluate expected frequency of filler wins. This would allow for a more granular examination of the data.

Another important factor to consider when interpreting the data is the selection bias variable itself. Although the model was designed under the assumption that responding during the model and RFC procedure likely reflects some aspect of memory strength, responding may be better reflected by another latent variable. For example, is responding reflective of lineup member similarity, strength of recall, confidence, or some combination of these, or other variables? For instance, responding during the RFC procedure may be mediated or reflective of other, more general cognitive abilities that, in turn, may indirectly impact memory performance or memory reporting (e.g., inhibitory control, metacognition, intelligence). Although this dissertation did not include tests of individual differences of such abilities, exploring how these may impact responding
during the RFC procedure is an important research question going forward. This remains an empirical question to be addressed though experimentation and future modeling. For example, to help clarify whether a witness’ pattern of responding during the RFC procedure is reflective of memory strength, future research could experimentally examine the relationship between Z-scores and memory strength. As an example, one could experimentally manipulate encoding quality (thereby manipulating memory strength) and measure the effects on Z-score outcomes. If differing patterns were observed in Z-scores in response to different encoding qualities, it would provide further support for the use of a witness’ pattern of responding to evaluate his/her memory strength for the target. This, in turn, would provide further evidence to support the use of this model.

13.0 General Discussion

Relying solely on an eyewitness identification decision comes with many uncertainties about a witness’ memory strength and the likelihood of suspect guilt. In my dissertation research, I examined a new, alternative lineup procedure that was designed to provide the same identification decision as traditional lineups, but also additional information about individual witnesses’ selection bias. This dissertation examined how witnesses from three different age groups (younger children, older children, and adults) fared using the repeated forced-choice (RFC) procedure. Experiment 1 examined the impact of the RFC lineup procedure on younger (6- to- 8-year-olds) and older (9- to- 11-year-olds) children’s lineup performance, relative to a simultaneous and face-off procedure. The objective in conducting Experiment 1 was to explore the utility of the RFC with child eyewitnesses, both in terms of identification accuracy and providing insight into children’s choosing behaviour. Experiment 2 expanded this research by
exploring how the RFC procedure influenced adults’ identification accuracy. In turn, this created a comparison group that allowed me to use the information provided by the RFC procedure to understand more about what developmental factors may be driving children’s high rates of choosing (cf. adults).

Results from Experiment 1 and Experiment 2 indicate that, for the RFC procedure, age matters. Use of the RFC procedure with young children did not produce favourable results—even increasing inaccurate choosing relative to the other procedures used. However, for children aged 9-11 years and adults, the RFC procedure compared favourably with the simultaneous (with wildcard) procedure. Considering the comparable identification performance along with the predictive utility of supplementary information, Experiments 1 and 2 results indicate no disadvantage, and in the case of older children, some advantages in using the RFC procedure to collect identification evidence when compared to the simultaneous procedure.

In the additional analyses, I demonstrated how an eyewitness’ pattern of responding during the RFC procedure could be used to understand more about each individual witness. Specifically, I demonstrated how each witness’ pattern of responding (Z-scores) reflects an individual level of suspect selection bias that, in turn, may be used to understand the likelihood that a witness’ responding was based on strong or weak recognition memory for the suspect. This information can be used as a tool to assist investigators or researchers with assigning weight to an individual witness’ decision. For example, this information may serve as a tool to screen for reliable witnesses, similar to how the blank lineup (i.e., a known target-absent lineup) has been used to identify unreliable witnesses (those who are prone to picking) before providing a witness with a
suspect-present lineup (Wells et al., 1998). Thus, Z-scores and ORs provided by the model may be used to help investigators identify which witnesses to value, but also which to not value. Importantly, I demonstrate that interpretation of this behavioural information can be made independent of the witness’ final lineup decisions (i.e., whether they ultimately selected the suspect, a filler, or rejected to lineup).

13.1 Applied Implications

Although far too early to recommend applied implementation, if the RFC procedure was to be used in an applied setting, there are several conceivable advantages based on the current data. First, the RFC requires electronic administration\textsuperscript{12}, thereby increasing consistency and minimizing administrator influences. Second, the evidence suggests that the RFC works well (at least as well as the frequently used simultaneous procedure) for both older children (9-11 years) and adult witnesses, allowing for consistent lineup administration for both age groups.

Third, and perhaps most importantly, the RFC procedure was designed to provide investigators with additional meaningful information. The purpose of a police lineup task is not to test the eyewitness (categorical correct/incorrect) but rather to help investigators gather evidence as to the guilt of a possible suspect (Charman & Wells, 2007). Sauer and colleagues (2008) discussed the importance of providing a single eyewitness decision for use in an investigation and in court. Research has clearly established that hearing an eyewitness say ‘that’s the person I saw’ is a powerful and persuasive form of evidence (e.g., Boyce, Lindsay, & Brimacombe, 2008; Cutler, Penrod & Dexter, 1990). Because of

\textsuperscript{12}Although conducting the RFC procedure in person is possible, it may be vulnerable to human error and bias.
this, the use of such a clear-cut form of eyewitness evidence is not likely to disappear from the legal system anytime soon. There is, however, a growing movement in the lineup literature to support the use of probabilistic evidence of suspect guilt, rather than categorical (Sauer & Brewer, 2015). Recognition memory is conceived of as existing on a continuum and, therefore, it follows that use of recognition memory as evidence of suspect guilt should also be perceived as a graded or continuous piece of evidence. The RFC procedure is unique in that it is capable of providing a categorical lineup decision made by the suspect (i.e., suspect, filler, and rejection) but also provides probabilistic evidence of guilt that will help investigators and researchers to better understand a witness’ memory, with or without relying on that lineup identification decision.

13.2 Theoretical Contribution

This research contributes to a greater understanding of eyewitness memory. First, for those over the age of 9 years in the present work, it demonstrated that eyewitness recognition memory can be assessed using an adjusted forced-choice task. This is important because, prior to the introduction of the recent face-off procedure (Price & Fitzgerald, 2016), all lineup procedures tested have been based on the old/new method of recognition memory. When compared to a traditional lineup task, the simultaneous procedure, the repeated forced-choice (RFC) procedure produced similar accuracy. Though, obviously, an increase in accuracy is more desirable, there is a great deal of value in finding new procedures that produce comparable accuracy because it opens the door for further development and new avenues for exploring children’s capabilities. Thus, this dissertation provides evidence to support future investigation into adapting alternative forced-choice recognition assessment methods to an eyewitness paradigm.
Second, age differences in performance with the RFC procedure between the younger and older children indicate that there are important developmental changes that occur around the age of 9 that allows these children to use the RFC procedure more effectively. Younger children’s erratic latency in responding to stimuli containing a guilty and innocent suspect, in combination with low identification accuracy and lower discriminatory abilities, provides evidence that children 8-years and younger may not have the executive functioning ability needed to effectively regulate their responding over the large number of decisions.

Third, the additional information collected about witnesses while they completed the RFC procedure (i.e., confidence, response latency, responding patterns), provides insight that allows us to learn more about the cognitive processes that underlie a witness’ decision. Previous work with adult eyewitnesses using traditional procedures has established that, when an eyewitness selects someone from a lineup (i.e., positive response), decisions that are made quickly and that are accompanied by higher post-identification confidence ratings are related to higher accuracy (e.g., Brewer & Weber, 2008; Brewer & Wells, 2006; Sporer et al., 1995). Although post-identification confidence and response latency cannot be used to indisputably establish guilt or innocence, these are important indices of memory strength. For example, when a witness makes a positive identification decision slowly and reports low confidence, it is indicative of a witness making a decision based on limited memory trace strength. In the present work, I replicated the confidence-accuracy and response-accuracy relationship with adult witnesses (Experiment 2), wherein accurate lineup decisions were made faster than inaccurate decisions. However, different results were observed with older children,
such that children required more time to separate guilty from innocent suspects. These results suggest that, when shown a suspect, children engage in more complex cognitive processing than adults do in order to correctly pick the guilty suspect. These findings help to support previous claims that discriminating a previously seen face from a series of previously unseen faces is dependent upon executive functioning and requires more cognitive effort from children (e.g., Hiller & Weber, 2013; Keast et al., 2007; Parker & Ryan, 1993; Pozzulo & Lindsay, 1999; Price & Fitzgerald, 2016).

The repeated forced-choice procedure also provides individual witness patterns of responding (i.e., Z-scores) that allow investigators and researchers to explore individual differences across eyewitnesses. Having access to this individualistic type of data is rarely available using traditional lineup procedures. Like other alternative lineup procedures (e.g., Sauer et al., 2008), the RFC procedure allows the user to acquire additional information about a witness’ memory that can be used to distinguish between two, seemingly identical identification decisions. In Experiment 1 older children’s response patterns were predictive of accurate, positive responding—even more so than post-identification confidence ratings. Higher Z-scores helped to discriminate between accurate and inaccurate choosers—especially when witnesses provided ambiguous confidence ratings (e.g., confidence of 4-7). However, for adults, Z-scores did not predict accuracy better than confidence ratings. As demonstrated by the selection bias analyses in the Additional Analyses, the likely reason that Z-scores were predictive of accuracy for older children and not adults may result from differences in the ability to adhere to task instructions. Children responded with lower selection bias (more randomly) when shown a target-absent lineup than did adults. That is, adults treated the innocent and guilty
suspects similarly (i.e., selected the one that looks most like the target), whereas children’s choosing associated with the guilty suspect differed from their choosing of the innocent suspect. Overall, these results suggest that using the RFC procedure is most beneficial for use with children aged 9 through 11 years old, among the age groups studied.

13.3 Future Research

Although the results of this research are encouraging, further investigation on the potential utility of the RFC procedure is required. The most important priority for a new procedure is replication. It will be critical to independently replicate the finding of equivalent performance between the RFC and simultaneous (with wildcard) procedure. Likewise, it is important to further test the predictive utility of confidence, latency, and decision patterns of the RFC procedure with different samples to help establish external validity. Second, this research should be extended to explore how RFC performs relative to other lineup procedures—such as the sequential procedure. As children tend to struggle with sequential procedures (Lindsay et al., 1997; Pozzulo & Lindsay, 1998; Steblay et al., 2011), it is expected that the RFC procedure will outperform the sequential procedure for this age group. However, with adults, the sequential procedure has been thought to elicit more conservative responding than the simultaneous lineup (Lindsay, Mansour, Beaudry, Leach, & Bertrand, 2009; Palmer & Brewer, 2012; Steblay et al., 2011; Wells et al., 1998) and, as such, how the RFC would perform relative to the sequential procedure remains an empirical question.

Third, there is benefit in exploring other ways to examine what can be learned from witness’ decision patterns in the RFC procedure. After exploring several different
indices (e.g., Max, Mean, Variance, Max-Mean/SD), I settled on using a Z-score because it captured the information and variability contained in each witness’ suspect selections, relative to his/her filler selections. However, it would be ideal to directly link response patterns (Z-scores) with memory strength and future research could address this experimentally.

13.5 Conclusion

The repeated forced-choice (RFC) lineup procedure was designed as an alternative method to collect memorial evidence from eyewitnesses. This dissertation provides evidence that the RFC procedure produces similar lineup responding to the simultaneous (with wildcard) procedure for both older children and adult eyewitnesses. This is particularly promising when considering that the RFC procedure, by design, provides additional memorial information that can be used to understand the predictive utility of a witness’ decision as well as the underlying cognitive processing that may be involved in lineup decisions. This dissertation provides early evidence that demonstrates the utility of using a force-choice method to examine eyewitness memory.
14.0 References


Bruer, K. C., Fitzgerald, R. J., Price, H. L., & Sauer, J. D. (2017). How sure are you that this is the man you saw? Using confidence judgments to identify a target with child eyewitnesses. Manuscript accepted with minor revisions to *Law and Human Behavior*.


Dupuis, P. R., & Lindsay, R. C. L. (2007). Radical alternatives to traditional lineups. In R. C. L. Lindsay, D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *The handbook of*


Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute $d'$, not the


APPENDIX A: Repeated Forced-Choice Program Design

This lineup contains three *Rounds* presented to participants and, for each round, participants were presented pairs of faces. During Round 1, each lineup is paired off with every other lineup member and a witness is asked to identify the one who looks MOST like the target (i.e., a most similar judgement). This is conceptually similar to a round robin design. During Round 2, participants were presented with the two faces they chose MOST often and were asked to choose the one who looks MOST like the target. For Round 2 to be successful, the computer program was designed to deal with situations of ‘tied’ lineup members. For Experiment 1 (i.e., child witnesses) no ties emerged in more than half (54%) of cases, and the computer program could move forward. However, 16% of the time, two lineup members were tied for the most selected face (i.e., tied for ‘first place’). To resolve this issue, the program explored who won when these two lineup members were matched during Round 1. The winner of that match was determined to be the most selected face and the program could continue. In 22% of cases, there was a clear first place winner but multiple people tied for second place. For these situations, the program attempted to resolve these by referencing the individual matches and, if it was unsuccessful, the witness skipped Round 2 and progressed directly to Round 3. In 8% of the cases (*n* = 24), three faces were tied for first place. Of those, 3% (*n* = 9) were resolved using the similar approach described above, while the remaining 4% (*n* = 15) were deemed irresolvable and, thus, the lineup ended with no decision from that
witness\textsuperscript{13}. The irresolvable cases were event split between target present \((n = 8)\) and target absent \((n = 7)\) lineups.

For Experiment 2 (i.e., adult witnesses), 69\% of the cases saw no ties emerged in Round 1. In 26\% of cases, there was a resolvable tie between lineup members, while an additional 6\% \((n = 10)\) of cases were irresolvable ties and thus, no lineup decision was made. More irresolvable cases were found in target absent \((n = 8)\) than target present \((n = 2)\) lineups.

\textsuperscript{13} In the situations in which no clear winner emerged from Round 1, the witness’ behaviour is indicative of random choosing. Thus, this situation may reflect poor memory for any lineup member. While I did not classify the responses for the present experiment, it is possible that the 4\% of witnesses could be considered to have made a lineup rejection.
APPENDIX B: Pilot Testing

Two small pilot projects were conducted with children during preparation for this research. The first pilot project focused on the electronic presentation of the repeated forced-choice (RFC) lineup. Given the complexities involved in stimuli selection for the RFC lineup, I used example stimuli to work out any programming issues, to ensure children understood the task, and to determine the appropriate length of the task to avoid participant fatigue. Using Eprime 2.0 (Psychology Software Tools, 2014), two versions of the RFC procedure were created. First, a version with 6 person lineups was tested with 13 children ($M_{age} = 8.40, SD = 1.40$). I asked children for feedback. The six-person lineup required that children make 17 decisions (15 pairs, 1 final tie-breaker pair, and final decision). All children indicated that they understood the instructions, how to conduct the task, and indicated that the task was fun. Upon prompting about the length, all children indicated the length was good and that it could have been longer. As a result, the task was adjusted to include eight lineup members to better reflect lineups used in Canada (Levi & Lindsay, 2001; Yarmey, 2003). Again, the longer task (28 decisions, 1 final tie-breaker, and final decision) was tested with 19 different children ($M_{age} = 10.00, SD = 1.54$) and the feedback was similar. With the exception of two participants (who were distracted and wanted to rejoin the group), no participants had a problem with the length of the task. The eight-person lineup task took an average of less than 4 minutes to complete (measure electronically).

One concern with the RFC procedure was how the computer program will deal with ‘ties’ or faces that were picked the same amount. Of the 32 children, more than half (52%) of the cases results in no ties, and 32% of the time two faces were tied for first, but this tie was resolved easily by the program. 13% of the time there was a clear first place,
but two were tied for second—but again, these were resolved. Only 1 of 32 cases was problematic, with three lineup members being tied for second place. In that case, the final comparison was skipped and the child was simply asked to make a decision regarding whether the first place winner was or was not the target. In short, based on this sample, the computer program worked well and was able to deal with any issues that emerged from children’s responding.
APPENDIX C - Parental Consent Form

Project: Examining the impact of repeated choosing on eyewitness’s lineup identifications

Dear Parent/Guardian,

This letter requests permission to invite your child to participate in a research project approved by the University of Regina Research Ethics Board. The goal of this project is to understand more about how children remember personally experienced events.

On one occasion a University of Regina student researcher will come into your child’s summer camp to play games with some of the children and ask your children for help conducting a task. One day after the play activities, the children with parent/guardian permission will participate in a brief audio recorded interview about their knowledge of the activities. Children will also be asked to participate in a lineup identification task using a computer.

Before all components of participation, your child will be asked if he or she wants to participate and told that he or she can refuse to answer any question(s) or stop at any time without consequence or explanation. If you choose to allow your child to participate, we ask that you not tell him or her that memory will later be tested about their experiences, as this may influence what is remembered. If you decide not to allow your child to join us, it will have no effect on his or her status within the program.

If you allow your child to participate, and if your child also agrees to join us, we will maintain the strictest standards of confidentiality as required by the law. The audio recorded interviews will be transcribed and all identifying information will be removed (e.g., child’s name, location, etc.). We will identify children by a participant number only and all personal information will be stored in a secure location at the University of Regina. Any information obtained from this study will be kept confidential and only group results will be reported.

This project was approved by the Research Ethics Board, University of Regina. If research participants have any questions or concerns about their rights or treatment as participants, they may contact the Chair of the Research Ethics Board at 306-585-4775 or by e-mail: research.ethics@uregina.ca. Out of town participants may call collect. Questions about this research may also be directed to the principal investigator Kaila Bruer (306-337-3339; bruier20k@uregina.ca) or her faculty supervisor, Dr. Heather Price (306-585-4297; heather.price@uregina.ca) in the Psychology Department at the University of Regina. We sincerely appreciate your consideration of this request.

Sincerely,

Kaila Bruer
Doctoral Student in Experimental and Applied Psychology
Department of Psychology, University of Regina
Permission Form

**PART I:**
Do you agree to have your child participate in this memory study?

___ Yes, ALLOW my child to participate in the memory study

___ No, do NOT ALLOW my child to participate

Child’s Name: __________________________ Date of Birth: __________________________

Parent/Guardian name: __________________________

Signature: __________________________ Date: __________________________

**PART II:**
Would you like a summary of the research results? If so, please provide your e-mail address below.

_____________________________

**PART III:**
We would also like to request permission to use your child’s audio recorded interview in future research studies. If you agree, the audiofile of your child’s conversation would be played for future participants who may evaluate, for example, the quality of the questions and responses. Only anonymous versions of the recorded interview will be played - all identifying information (e.g., names, locations) will be removed from the recording.

___ ALLOW my child’s audio recorded interview to be used in future studies

___ Do NOT ALLOW my child’s audio recorded interview to be used in future studies

**PART IV:**
Would you like to be contacted in the future about participating in other research projects? If you select yes, you are granting us permission to contact you, provide you with information about an upcoming study, and invite you to participate.

___ YES, please contact me to invite my child to participate in future studies.

Email: __________________________

Phone: __________________________

___ NO, please do not contact me to invite my child to participate in future studies.
APPENDIX D – Assent and Debriefing

**RAPPORT BUILDING:**
As soon as you meet the child you’ll be interviewing, you should start building. A good way to do this is to ask them a couple questions (e.g., How is camp going? What kind of stuff are you doing today? Are you doing anything else this summer?). In most cases, you will be able to establish a sufficient level of rapport by the time you find a place to sit down, but take more time if child still appears uncomfortable.

**PARTICIPANT INFORMATION:**

“This is participant number _______. Alright, so like I said before, my name is ________, and I am here talking to ________ (child’s name).

**ASSENT:**

“I was wondering if you would be okay with answering a few questions for me today?”

If the child says *yes*…

“Okay great! Just so you know, I’ll be asking you a few questions and if you *don’t feel like answering them* that is okay. Or, if you *don’t want to talk to me anymore* that is okay too. So, if you want to stop at any time just say so, alright?

[wait for response]

To start things off, can you tell me how old you are ?”

If the child says *no*…

Thank them for their time and allow them to select a prize.

**DEBRIEFING:**

“Okay, that is all of the questions I have for you. Do you have any questions for me? [If yes, answer questions. If no, continue...] Thanks a lot for answering my questions and paying attention to my instructions. You really helped us out today. As a thank-you, we have a box of prizes and you can pick out your favourite one to keep.” [show child to treasure suitcase]

**Note:** If at any point the child seems uncomfortable, ask the child if he/she is okay with continuing on with the interview. If the child wishes to stop the interview, thank the child for their time and allow them to select a prize.
Appendix E: Research Ethics Approval Certificate

University of Regina

Research Ethics Board
Certificate of Approval

PRINCIPAL INVESTIGATOR
Kalia Bruer

DEPARTMENT
Psychology

REB #
2014-199

SUPERVISOR
Dr. Heather Price – Psychology

FUNDER(S)
Unfunded

TITLE
Examining the Impact of Repeated Choosing on Eyewitness’s Lineup Identifications

APPROVAL OF
Application for Behavioural Research Ethics Review
Appendix A – DNA recruitment information
Appendix B – Demographic information sheet
Appendix C – Complete (2") Consent form for Adults
Appendix D – Debriefing Form for Adults
Appendix E – Lineup Confidence Question
Appendix F – Consent Form for Children
Appendix G – Initial Consent form for Adults
Appendix H – Assent and Debriefing for Children
Appendix I – Similarity Rating Survey – Online Consent and Question

APPROVED ON
December 10, 2014

RENEWAL DATE
December 10, 2015

Full Board Meeting

Delegated Review

CERTIFICATION
The University of Regina Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol, consent process or documents.

Any significant changes to your proposed methods, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS
In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion.

Please refer to the following website for further instructions: http://www.uregina.ca/research/REB/main.shtml

Dr. David Senkow, Acting Chair
University of Regina
Research Ethics Board

Please send all correspondence to:
Office for Research, Innovation and Partnerships
University of Regina
Research and Innovation Centre 109
Regina, SK S4S OA2
Telephone: (306) 585-4775 Fax: (306) 585-4993
research.ethics@uregina.ca
APPENDIX F: Lineup Decision from Experiments 1 and 2 for each Target

<table>
<thead>
<tr>
<th>Age</th>
<th>Target</th>
<th>Procedure</th>
<th>Identification Procedure</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Suspect</td>
<td>Filler</td>
</tr>
<tr>
<td>Male</td>
<td>&lt; 9</td>
<td>Present</td>
<td>Simultaneous</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.36</td>
</tr>
<tr>
<td>≥ 9</td>
<td>Present</td>
<td>Simultaneous</td>
<td>.53</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.08</td>
</tr>
<tr>
<td>Adults</td>
<td>Present</td>
<td>Simultaneous</td>
<td>.51</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.26</td>
</tr>
<tr>
<td>Female</td>
<td>&lt; 9</td>
<td>Present</td>
<td>Simultaneous</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.04</td>
</tr>
<tr>
<td>≥ 9</td>
<td>Present</td>
<td>Simultaneous</td>
<td>.51</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Face-off</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.04</td>
</tr>
<tr>
<td>Adults</td>
<td>Present</td>
<td>Simultaneous</td>
<td>.76</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absent</td>
<td>Simultaneous</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFC</td>
<td>.10</td>
</tr>
</tbody>
</table>
APPENDIX G: Analysis to Examine Guessing Behaviour in Younger Child Witnesses

Younger children performed quite poorly across all lineup procedures, especially with the female target. To better understand why this may have been, follow-up analyses were conducted. When younger children did select the target, their selections indicated guessing behaviour as they were just as likely to select a filler as they were to select the target, across all three lineups (simultaneous: 0.23 vs. 0.21; face-off: 0.25 vs. 0.27; RFC: 0.26 vs. 0.27). Using a dispersion index (DI; MacLennan, 2003, unpublished manuscript) I was able to calculate how close to random (0-1.00, with 1.00 indicating randomness) children’s responses were. For younger children, target-absent responses were very close to random (0.92), which is expected given that no lineup member would strongly match a witness’ memory for the perpetrator. However, when the female target was present in the lineup, their responses were also close to random (0.72). For the male target, younger children’s responses produced randomness measures more similar to what would be expected, with target-present responses being further away from random (0.52) than the target-absent responses (0.83). For older children, target-present responses were much further away from randomness (female = 0.48; male = 0.47) than their target-absent responses (female = 0.90; male = 0.88).

In addition, a multinomial χ² test can statistically measure the distance of observed responses from random responding. When the female target was present, younger children’s suspect selections relative to filler selections were not statistically different from random, χ²(1) = 1.39, p = .24. This same pattern was not observed with the male target or with the older children, χ²(1) = 9.00, p = .003. For older children, suspect selections relative to filler selections were statistically different from random in a target-
present lineup for both female, $\chi^2(1) = 22.77, p = .001$, and male targets, $\chi^2(1) = 30.75, p = .001$. Thus, based on these analyses, I can conclude that younger children’s responses to the female target lineups were likely reflective of guessing behaviour.

I expect that younger children found the female target difficult to identify for two reasons. The first reason reflects differences in perceived similarity between children and the adults that were used to build the lineup. Post-hoc similarity ratings (scale of 1 being not at all similar to 10 being very similar) provided by independent children ($N = 92$) of a similar age (ranged from 6 through 11-years-old) revealed that the female target was perceived, on average, to be significantly more similar in appearance to her fillers (higher similarity; $M = 3.69, SD = 1.49$) than the male target (lower similarity; $M = 3.19, SD = 1.27$), $t(91) = 3.90, p < .001$—though level of similarity was not deliberately manipulated. That is, children rated the lineup members as more similar to the female target, indicating perceived difficulty with the lineup. This perceived level of difficulty may have proved to be more of a challenge for the younger children during the lineup procedure, however, no research has yet explored differences between how children and adults perceive similarity between faces. Second, during the live event, the female target did not attract the same level of attention as the male target due to a more reserved demeanor. Thus, target exposure to the female was likely less-salient for these children. For a complete breakdown on participants responding to each individual target, see Appendix F.
Appendix H: Z-score Calculation

(1) Identify the number of selections of the suspect.
(2) Calculate the mean ranking of all lineup members.

\[ M_a = \frac{\text{Suspect} + \text{Filler 1} + \text{Filler 2} + \text{Filler 3} + \text{Filler 4} + \text{Filler 5} + \text{Filler 6} + \text{Filler 7}}{8} \]

(3) Calculate the standard deviation (average distance from the mean) of all lineup members.

\[ SD = \sqrt{\frac{(M_a - \text{Suspect})^2 + (M_a - \text{Filler 1})^2 + \cdots + (M_a - \text{Filler 7})^2}{n-1}} \]

(4) Calculate a standardized score (Z-score) by subtracting the mean rankings from the suspect selections and dividing this by the standard deviation.

\[ Z - \text{Score} = \frac{(\text{Suspect} - M_a)}{SD} \]

(5) Match the Z-score to Table 17 to assess the likely suspect bias level.

**Example**

(1) Suspect was selected 7 of 7 possible times. Suspect = 7.
(2) Mean selection/ranking for all lineup members.
(3) Calculate the standard deviation.

Filler 1 = 6, Filler 2 = 5, Filler 3 = 3, Filler 4 = 3, Filler 5 = 2, Filler 6 = 2, Filler 7 = 0

\[ M_a = \frac{7 + 6 + 5 + 3 + 3 + 2 + 2 + 0}{8} = 3.50 \]

\[ SD = \sqrt{\frac{(3.50 - 7)^2 + (3.50 - 6)^2 + (3.50 - 5)^2 + (3.50 - 3)^2 + (3.50 - 3)^2 + (3.50 - 2)^2 + (3.50 - 2)^2 + (3.50 - 0)^2}{7}} = 2.33 \]

(4) Calculate the Z-score.

\[ Z - \text{Score} = \frac{7 - 3.50}{2.33} = 1.50 \]