# Eyewitness Identification of a "Familiar Stranger": The Influence of Personal Knowledge and Previous Interaction on Recall and Recognition Accuracy

by

Jennifer L. Pettalia

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Carleton University
Ottawa, Ontario

**Abstract**

The purpose of the present program of research was to examine the influence of a previous interaction and personal knowledge of a culprit on eyewitness recall and recognition accuracy. In the main study, participants (*N*=184) watched a video of a crime wherein the culprit was someone who: (1) they previously interacted with and, during that interaction, the culprit shared personal information about herself (more familiar condition), (2) they previously interacted with, although they received no personal information about the culprit (less familiar condition), or (3) they had no previous exposure to (stranger condition). There were no significant differences in participants' recall accuracy across the conditions. However, participants in the more and less familiar conditions were significantly more likely to correctly identify the culprit from a target-present lineup; there were no significant differences in the target-absent lineup. Also, more and less familiar participants were significantly more confident in their ability to correctly identify the culprit prior to the lineup task and during the lineup task than those in the stranger condition. Moreover, the confidence-accuracy relationship was strongest for those in the more familiar condition. A follow up study, completed approximately 14 months after the original study, was conducted to examine the influence of familiarity and delay on eyewitnesses who initially provided a correct lineup response. Delay was associated with a significant decline in accuracy rates and confidence for those in all conditions. Overall, results indicate that previous interaction, regardless of whether personal information is shared, may enhance eyewitness identification accuracy only when the culprit is in the identification lineup (i.e., when the suspect is guilty). Also,

more familiar eyewitnesses may have a better sense of their identification accuracy, as indicated by their confidence judgements.

**Acknowledgements**

This project would not have been completed without the support and assistance of a few very important people. Thank you to my mentor and supervisor, Dr. Joanna Pozzulo. Your insight, guidance, support, and patience have meant the world to me. I feel very lucky to have had you to guide me through my doctoral degree and I am appreciative of all of the time and "pep talks" you have given me throughout the years. I would also like to thank my amazing committee members, Dr. Craig Bennell and Dr. Adelle Forth. I feel privileged to have had your expert guidance throughout this project and I am very much appreciative of all of your time and support. Also, Carleton is very lucky to have Etelle Bourassa as a Graduate Studies Administrator and I am very lucky to have been touched by her kind and compassionate spirit – thank you Etelle for all your assistance.

Data collection would not have been possible without the assistance of my lab-mate, Emily Pica. Her patience, flexibility, and all-star confederate skills are greatly appreciated. I am also thankful for the support and friendship of my lab-mates Julie Dempsey and Jenny Reed - your advice and encouragement was immeasurable.

Lastly, I would like to thank my best friend and husband, AJ Pettalia. Your unconditional support throughout these many years (more than we want to count!) has allowed me to devote a lot of time to this project and I am forever appreciative for that.

**Table of Contents**

## List of Tables

## List of Appendices

# Eyewitness Identification of a "Familiar Stranger": The Influence of Personal Knowledge and Previous Interaction on Recall and Recognition Accuracy

## Introduction

Eyewitness evidence is frequent, significant, and persuasive in the prosecution of defendants in many criminal cases (e.g., Flowe, Mehta, & Ebbesen, 2011; Goldstein, Chance, & Schneller, 1989; Wells & Quinlivan, 2009). However, since the 1990's the shortcomings of a witness' memory have increasingly been highlighted by researchers, organizations, and the criminal justice system (Gross, Jacoby, Matheson, Montgomery, & Patel, 2005; Wells & Olson, 2003). For example, the Innocence Project, an organization dedicated to the exoneration of individuals who have been wrongfully convicted in the United States, reports that erroneous eyewitness evidence is the foremost reason for wrongful convictions in 72% of the over 325 exonerations they have processed since their inception in 1992 (Innocence Project, 2015). The grave consequences of faulty eyewitness testimony, such as convicting an innocent person, make this issue imperative. As a result, numerous studies utilizing "never before seen" targets (i.e., strangers) have been conducted and have informed the best practices in collecting and preserving eyewitness evidence (Technical Working Group for Eyewitness Evidence, 1999, 2003; Wells, Malpass, Lindsay, Fisher, Turtle, & Fulero, 2000).

Notably, though, many crimes (e.g., violent crimes, personal crimes, sexual assaults) in which eyewitnesses are utilized for conviction are not committed by strangers, rather, by individuals who are known, or familiar, to the victims or bystanders (Flowe et al., 2011; Memon, Havard, Clifford, Gabbert, & Watt, 2011; Valentine,

Pickering, & Darling, 2003). For example, Flowe and colleagues (2011) randomly sampled 725 felony cases from a District Attorney's office in California. They found that the majority (67%) of cases involved one or more eyewitnesses who were acquainted with the defendant. What's more, eyewitness evidence significantly increased the probability that a case would be prosecuted when the witness was acquainted with the defendant, whereas the influence of eyewitness evidence was not significant when the eyewitness was a stranger. Similarly, Gross and colleagues (2005) examined all known exonerations that occurred in the United States between 1989 and 2003, finding 328 individual exonerations in total. They reported that, while 72% of rape exonerations involved victims or witnesses who were strangers to the defendant, 86% of the exonerated murder defendants were acquainted with one or more of the witnesses who provided eyewitness evidence in their case.

Despite these statistics, alarmingly, the differences between these two distinct groups, stranger and familiar others, is rarely acknowledged in eyewitness evidence research or protocols. In particular, eyewitness recall and recognition accuracy of *familiar-strangers* – individuals who are not known to the eyewitness but with whom the eyewitness has come into contact with prior to the crime – is neglected in research thus far. Consequently, academics and lawyers may be tempted to estimate accuracy rates of the arguably distinct *familiar-stranger* from research results utilizing *strangers* (never before seen targets) or, in contrast, they may assume that identification of a familiar-stranger is similar to an identification of a known other. The gap in literature in this area

is great and provoked the present program of research concerning the overlooked, yet often encountered, eyewitness identification of the familiar-stranger.

The present studies will inform researchers and authorities as to the potential differences in eyewitness memory for strangers in comparison to familiar (but unknown) culprits. These studies will be the first to use an eyewitness paradigm to examine the influence of a previous interaction with the sharing of personal knowledge on eyewitness’s recall and recognition accuracy. In this dissertation, individuals were familiarized with a target via a scripted interview that involved or did not involve the disclosure of personal information. Five to ten days later they witnessed the familiar target commit a theft and were asked to provide a description and complete an identification lineup. Furthermore, approximately a year later correct witnesses completed the identification task again in order to examine their memory trace over an extended period of time.

The following sections of this paper will begin with an overview of eyewitness research and methodology, highlighting best practices and procedures. The extant literature regarding witness recall and recognition of strangers will be contrasted to the dearth of research on memory for familiar strangers. In particular, available literature regarding eyewitness recall accuracy, recognition accuracy, and confidence will be reviewed in order to provide context to the current study results. Also, theoretical perspectives for memory of familiar-strangers will be considered. Results from previous research and theory will be extended to the current program of research in order to

provide predictions and provide insight into the implications that such research holds in actual criminal investigations.

## Indicators of Eyewitness Accuracy

In many criminal investigations in which there is an eyewitness to the crime, eyewitness evidence may be collected. This "evidence" is essentially the witness's memory of the crime and the culprit (recall) and often also includes their ability to choose the culprit from an identification lineup (recognition). The eyewitness's confidence in his or her identification decision is often assessed and has been found to be an influential measure of eyewitness accuracy as well (e.g., Brewer & Burke, 2002). Each of these measures relating to eyewitness accuracy (i.e., recall, recognition, and confidence) will be discussed next.

### Recall

Eyewitness recall involves an account of what the eyewitness saw during the criminal event. Depending on the crime and the witness's involvement, recollections can be quite varied; however, in all cases it is recommended that accounts are taken as soon after the event as possible (Mackay & Paterson, 2014; Wang, Paterson, & Kemp, 2014), prior to or without interaction with other witnesses (e.g., Paterson, Kemp, & Ng, 2011), and in an open-ended format (i.e., with limited use of suggestive questioning; e.g., Lipton, 1977). A good deal of research has been conducted on best practices in eliciting recall from witnesses (e.g., Fisher, Brennan, & McCauley, 2002).

Recalled information collected from eyewitnesses can be very influential in determining the path of a criminal investigation (Meissner, Sporer, & Schooler, 2007).

For example, should witnesses report that the culprit drove off in a green pick-up truck, a green pick-up truck will likely become an important clue to the identity/location of the culprit. However, a review of research on an eyewitness’s ability to describe a once-seen culprit suggests that witnesses often provide vague or general details that can be applied to many people therefore offering little assistance in narrowing the scope of the investigation (Lindsay, Martin, & Webber, 1994; Meissner et al., 2007; Odinot & Wolters, 2006).

For example, Kuehn (1974) examined a random sample of victim person descriptions obtained from violent crime reports to the Seattle Police Department in 1967. The average number of descriptors provided was 7.2 and over 85% of victims provided six or more descriptors. The most commonly reported person descriptors were sex, age, height, build, race, weight, complexion, and hair colour (Kuehn, 1974). These results are consistent with similar studies assessing the number of details provided by actual crime victims and witnesses (e.g., van Koppen & Lochun, 1997).

Lindsay and colleagues (1994) conducted a study in order to determine how real world descriptions compared to descriptions typically obtained in the laboratory. The authors examined person recall from 100 laboratory witnesses and 105 real-life witnesses (obtained from newspaper articles). On average, laboratory witnesses reported 7.35 descriptors, most commonly regarding clothing, hair colour, and height. On the other hand, real-life witnesses reported significantly fewer details ($M = 3.94$), although they were similar in nature to the lab witnesses: sex, clothing, age, and height (Lindsay et al., 1994, Study 1). These rates may be underestimated due to the authors’ decision to count

any and all clothing descriptors as one item; nonetheless, results suggest that real-life witnesses provide similar information to laboratory witnesses, albeit at a diminished rate.

An advantage of laboratory witnesses is that recall accuracy can be measured. Unlike real-world cases, the culprit is known in laboratory experiments, thus eyewitness descriptions of the culprit can be verified. Results from laboratory studies indicate that recall accuracy rates are quite varied between and within studies (e.g., Yarmey & Yarmey, 1997). For example, Yarmey and Yarmey (1997) examined 603 community members' recall for a women who approached them on the street for assistance in finding jewelry or directions; the interaction time was approximately 15 seconds. Half of the participants were asked to recall the women using an interrogative approach (i.e., cued-recall, 15 questions) while that other half recalled the women using a narrative approach (i.e., open-ended recall, similar to that used in the present study). Participants who used a narrative recall approach provided significantly more accurate responses than those who use the interrogative recall approach. Of those who used the better performing narrative approach, 92% correctly reported type of lower clothing, 87% correctly reported design on lower clothing, 85% correctly reported hair colour, 75% correctly reported height, and 69% correctly described the design on the lower clothing. The least accurately reported items for those who used a narrative recall approach were age (23% correct), weight (29%), complexion (45%), colour of top clothing (45%), and hair style (49%).

**Recall of familiar-strangers.** The studies described above all involve the recall of "strangers"; to date, no known studies have examined eyewitness recall accuracy for familiar-strangers. In an effort to provide some insight into familiar-stranger recall

accuracy, though, we may look to research examining event repetition and extended exposure duration.

***Event repetition.*** Repetition priming involves the repeated presentation of a stimulus (in this case, a face or person) and is often found to increase recognition accuracy (Bruce, Burton, & Hancock, 2007). With repeated exposure to the culprit, witnesses may have the opportunity to view the culprit from different angles, with varying facial expressions, in numerous contexts, and so on. Face recognition researchers indicate that such varying experiences with the same face are likely to improve an individual's recognition accuracy, possibly because it strengthens their memory trace (e.g., Burton, Wilson, Cowan, & Bruce, 1999). In contrast, relatively little research has been conducted on the influence of event repetition, and in particular repeated viewing of a person, on recall accuracy. Of the research available, most has been conducted with child samples repeatedly exposed to events.

For example, McNichol, Shute and Tucker (1999) studied 6- and 7-year-old's recall of an event that they experienced once or repeatedly. The children visited a caravan on school property where a guest played games with them and dressed them up in costumes. The games, costumes, and physical items in the caravan changed for those who experienced the event repeatedly. The authors reported that children in the repeated-event group recalled significantly more items than the single-event group. Furthermore, the children in the repeated-event group where more accurate in recalling details that were consistent across the events than items that varied from event to event (McNichol et al., 1999).

Odegard, Cooper, Lampinen, Reyna and Brainerd (2009) conducted a similar experiment; however, all participants were repeatedly exposed to the events (i.e., there was no control group). In their study, 40 children (5–12 years old) were invited to four birthday parties over four weeks. The birthday parties followed a general pattern, but each had a theme and activities that varied from event to event. Ten days following the last birthday party, children’s recall and recognition memory were tested. Recall was tested using a face-to-face interview that began with open-ended questions and transitioned to cued-recall. There was an effect of age on recall accuracy such that younger children were significantly less accurate than older children. To measure their recognition accuracy, the researchers showed the children pictures of events at the party within a spread of filler pictures (i.e., photos of events that were not at the parties), and asked the children if they experienced the event. Children were found to be more conservative in their choices as they got older; however, no significant effect of age was found with regard to recognition accuracy (Odegard et al., 2009). Taken together, these results on children’s recall accuracy for repeated events suggest that repeated exposure may increase recall accuracy, particularly when there is little change from one event to another and that older children may perform better than younger children, when open-ended questioning is used. Although there is no known comparable research with adults at this time, it may be predicted that adults, like older children, will recall more information when repeatedly exposed (or familiar) to someone than with only one viewing of that person.

***Exposure duration.*** An eyewitness may be exposed to a culprit for a very limited time (such as a few seconds during a purse snatching) or a lengthy time (such as a kidnapping that can occur over a few hours). Similar to repeated exposure, being exposed to a culprit for a longer period of time may allow the eyewitness to become more familiar with the individual. As a result, their accuracy in identifying an individual with whom they had a longer exposure may be better than an individual with whom they had a limited exposure.

Yarmey, Jacob, and Porter (2002) tested the recall accuracy of 120 community members who were approached by a confederate who asked for directions or for assistance finding jewelry (the same procedure as Yarmey & Yarmey, 1997). The duration of the interaction between the participant and the confederate was either 5 or 30 seconds. Two minutes post-interaction participants completed a 14-item cued recall test regarding the confederates' physical characteristics and clothing. The authors found that participants who had more time to view the target were more accurate in their descriptions of eye colour and clothing – there were no significant differences for other physical characteristic descriptors. From this, it seems as though extended exposure duration, like repeated exposure, may enhance eyewitness recall to some degree, although the research base is currently too limited to draw firm conclusions. Subsequently, theoretical developments may provide some further insight.

**Theoretical implications.** As of late, a cognitive theory referred to as Fuzzy Trace Theory (FTT) has increasingly been used to explain some factors frequently found to influence eyewitness recall accuracy (e.g., Brainerd & Reyna, 2002). The theory asserts

that we actively encode and retrieve information within the context of previously encoded material such that we develop gist-like representations (called fuzzy traces) as well as more detailed accounts (called verbatim). FTT asserts that these two categories of memory trace (i.e., gist and verbatim) may be considered the extremities of a continuum. At one end of the continuum are verbatim traces which are detailed, precise and, as such, quite accurate memories, whereas at the other end are fuzzy-traces (gist) which are vague memories of the sense or pattern of an event that may be more meaningful but also less accurate. Verbatim and gist traces are parallel, but separate, processes that each have advantages and disadvantages for memory recall (Reyna, Mills, Estrada, & Brainerd, 2007). As previously indicated, verbatim traces are more factual and exact; therefore, accessing verbatim traces is regarded as an objective reproductive process that may be superior to gist in some ways. However, meaningful gist traces are found to be more accessible, stable over time, and less effortful than verbatim traces, leading to a fuzzy-processing preference whereby we prefer to remember by processing fuzzy rather than verbatim traces (Brainerd & Reyna, 1990). Furthermore, gist is generally all that is required throughout one’s daily activities.

How does this all factor into eyewitness recall and, in particular, the present study? Using FTT, one may predict that eyewitnesses’ recall for never before seen culprits may be driven by gist-traces. In lieu of multiple sources of information or experience with a person, eyewitnesses may rely on their general sense of what occurred during the criminal event and what the culprit looked like. If this is the case, they may be tempted to report items vaguely or “fill-in-the-blanks” with schematic information. On the other

hand, eyewitness recall for familiar-strangers may be retrieved from verbatim traces. These eyewitnesses may access factual and detailed content regarding the culprit that they have encoded from their previous encounters with him or her. If this is the case, one can assume that recall for familiar-strangers will be more detailed and accurate than recall elicited from never before seen witnesses.

While seemingly sufficient in providing a theoretical framework for recall accuracy, FTT is not without its critics; it is unable to account for social and emotional factors that influence event processing (e.g., Davies, 1995). In consideration of eyewitness studies, recall performance is not simply a result of the cognitive processing (and subsequent memory trace) at the time of exposure. A seemingly unlimited number of physical (e.g., lighting, distracters, view obstructions), social (e.g., demand characteristics, pressure), and emotional (e.g., stress, motivation) factors may be experienced before, during, and/or after exposure, and as a result, influence recall accuracy. However, given the absence of alternate theories to explain recall accuracy, FTT may be cautiously used in the present program of research.

In the present study, eyewitnesses were asked to provide a description of the crime and culprit using open-ended questions. There is very little previous research to inform predictions regarding recall accuracy for familiar-stranger witnesses. However, results from exposure repetition and extended exposure studies suggest that additional opportunities for viewing a culprit may enhance recall accuracy (e.g., McNichol et al., 1999; Yarmey et al., 2002). Moreover, should familiar-strangers witnesses access

verbatim traces, as outlined in FTT, they should be more likely to provide accurate information (e.g., Reyna et al., 2007).

**Recognition**

When a suspect is determined, investigators may call upon the witness once again to provide additional evidence: an eyewitness identification. An eyewitness identification is the primary eyewitness evidence used in an eyewitness case as it is the culminating declaration by the eyewitness that the defendant is the culprit. Identifications can be obtained in a variety of ways, such as live lineups or photo lineups. Photos are the most common method of presenting an identification lineup in Canada (Beaudry & Lindsay, 2006). As a result, photo lineups were used in the present studies.

**Target-present and target-absent lineups.** In a criminal investigation, eyewitnesses are typically shown 6 to 12 photographs displaying the face and upper body of the suspect amongst other people who share similar characteristics (i.e., fillers). If an eyewitness chooses the suspect from the alternatives, confidence that the suspect is guilty increases, whereas if an eyewitness rejects the lineup (does not choose anyone) confidence that the suspect is innocent may increase (Wells & Lindsay, 1980). However, whether the suspect is, in fact, guilty of the crime cannot be determined in real-world eyewitness identification lineups. Hence, the impetus for eyewitness research in order to assist in the prediction of an eyewitness’ accuracy based on various witness factors.

Eyewitness identification lineups used in research mimic real-life eyewitness lineups with the additional benefit of having a definitive guilty suspect. A lineup including a guilty suspect is often called a target-present (TP) lineup. An alternate lineup

is frequently used to simulate a situation in which the suspect is innocent; this type of lineup is called a target-absent (TA) lineup. In TP lineups (where the culprit appears in the lineup), participants may correctly choose the target, incorrectly choose a filler, or incorrectly dismiss the lineup. In TA lineups (where the culprit does not appear in the lineup), participants may correctly dismiss the lineup, incorrectly choose a filler, or incorrectly choose the innocent suspect. Taken together the five responses available to eyewitnesses in identification research are: (1) correct identification (choosing the target in a TP lineup); (2) correct rejection (dismissing a TA lineup); (3) false identification (choosing the innocent suspect); (4) filler identification (choosing a filler); or (5) false rejection (dismissing a TP lineup; Brewer & Palmer, 2010; Lindsay & Wells, 1985; Wells, 1984). In the present study, eyewitness accuracy will be assessed via participants' ability to correctly identify a target in a TP lineup or correctly reject a TA lineup; correct identification and correct rejection, respectively.

**Lineup procedures.** In addition to having TP or TA lineups, researchers and policing authorities have used both simultaneous and sequential lineup procedures[1] (Lindsay & Wells, 1985). In a simultaneous procedure eyewitnesses are presented with a photo array and asked to choose the target, if present (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997). Alternatively, in the sequential procedure eyewitnesses are shown one photograph at a time and must make a decision as to whether the person in the photo

[1] There are additional lineup procedures such as the elimination lineup procedure (Pozzulo & Lindsay, 1999) and the wildcard lineup procedure (Zajac & Karageorge, 2009); however, these procedures were primarily developed for the purpose of improving identification accuracy in children. Given that the present study uses an adult sample, these lineup procedures are not discussed here.

being shown is the target before they move on to the next photo (e.g., Steblay, Dysart, & Wells, 2011). Witnesses are not aware of the number of photos that they will be shown. A simultaneous lineup is said to require the eyewitness to make a relative decision because viewing the photo array in its entirety allows one to compare and contrast the various photos, thus being able to determine which one looks *most* like the target (e.g., Wells, 1984). On the other hand, a sequential lineup is thought to require an absolute decision because there is no opportunity to compare the photos to one another before deciding whether the image matches one’s memory of the target (e.g., Kneller, Memon, & Stevenage, 2001).

The debate amongst researchers as to which lineup method is most advantageous is ongoing. Initially, the sequential procedure was popularized due to its ability to significantly reduce the number of false identifications made by eyewitnesses compared to using the simultaneous lineup (e.g., Cutler & Penrod, 1988; Lindsay, Lea, & Fulford, 1991; Lindsay & Wells, 1985; Melara, DeWitt-Rickards, & O’Brien, 1989; Parker & Ryan, 1993; Sporer, 1993). However, after extended research including meta-analytic analyses it became clear that a more nuanced interpretation of data examining simultaneous versus sequential lineup procedures indicates that, in comparison to the simultaneous lineup, the sequential lineup elicits a more conservative choosing criterion from eyewitnesses (e.g., Gronlund, Wixted, & Mickes, 2014; Meissner, Tredoux, Parker, & MacLin, 2005). In other words, the sequential procedure results in eyewitnesses who are less likely to choose a lineup member from TA lineups (resulting in more accurate responses) as well as in TP lineups (resulting in fewer correct identifications; e.g.,

Lindsay & Wells, 1985; Lindsay et al., 1991; Meissner et al., 2005; Steblay, Dysart, Fulero, & Lindsay, 2001). Extrapolated to a real-world scenario, the simultaneous lineup may result in more innocent suspects being convicted whereas the sequential lineup may result in more guilty suspects remaining free (Wells, 2014). The question then becomes, which of these risks are we more willing to accept?

Wells (2014) argues that, due to the statistics obtained from archival data indicating that real-world eyewitnesses' decision criterion may in fact be too lax (i.e., they are making choices more frequently thereby increasing their likelihood of identifying a filler or innocent suspect) an adjustment of their criterion may be made through the use of the sequential procedure. Wells (2014) adds that there is asymmetry in the risks associated with the simultaneous and sequential lineups: the sequential lineup may (1) increase the likelihood that a guilty suspect remains free, whereas the simultaneous lineup may (1) increase the likelihood that an innocent suspect is convicted and (2) that the guilty suspect remains free. At this point, one procedure is not unanimously preferred over the other; however, due to the fact that the sequential lineup is most commonly reported to be used in Canada (Beaudry & Lindsay, 2006) the sequential procedure was used in the present study.

**Identification/recognition of familiar-strangers.** Eyewitness researchers commonly use an eyewitness paradigm, in which participants are exposed to the target (i.e., culprit) without knowledge that they will later have to remember the event or target. The target may be seen via a live exposure, videotaped re-enactment, or picture slides, and is often depicted engaging in a criminal activity such as a theft. After viewing the

target, participants are informed that they have witnessed a mock crime and are asked to provide descriptive information and to identify the target amongst similar looking foils in a lineup procedure. The eyewitness paradigm is meant to closely resemble that which occurs in actual criminal investigations and therefore provides results that may be applied to judicial processes and procedures. The few studies that have utilized the eyewitness paradigm to examine identification accuracy for familiar others will be reviewed below. However, given that there are very few studies that have used an eyewitness paradigm to study familiar-stranger identification, a review of results from face recognition research also will be provided.

Face recognition studies largely take a cognitive approach to understanding our ability to process and recognize faces. In contrast to eyewitness research, face recognition studies generally involve exposing participants to many (e.g., 15 or more) faces over a short period of time (i.e., seconds; e.g., Burton et al., 1999). These faces are typically photographs or still shots of the face that are utilized in both the exposure stage as well as the recognition stage of the experiment (e.g., Clutterbuck & Johnston, 2005). Immediately following the learning stage, or sometimes after a short delay, participants are typically presented with one photograph at a time and asked to indicate whether the face is one that was learned or not learned (e.g., Campbell et al., 1999). Another frequently used methodology is a type of matching task where the participant is simultaneously shown two faces, one in full view or an unblemished state and the other with some type of obstruction, and asked to identify whether the two photos are of the same person (e.g., Bonner, Burton, & Bruce, 2003). These types of face recognition

studies are very controlled and are often utilized to inform memory theory; however, they lack ecological validity (Pozzulo, 2007; Shapiro & Penrod, 1986). Nevertheless, these types of studies may provide insight into what one may expect to find in studies examining the influence of familiar-stranger recognition.

***Face recognition studies.*** Bruce and colleagues' (e.g., Bruce, 1982; Burton et al., 1999) have conducted extensive work in the area of recognition of stranger and familiar faces. In their review of the face recognition literature, Hancock, Bruce, and Burton (2000) concluded that never before seen face recognition is poorer than chance levels, even under ideal conditions. On the other hand, familiar faces are found to be more accurately identified (compared to never before seen faces), particularly when they are more obscure (Burton et al., 1999) or when the viewpoint is changed (Bruce, 1982). For example, Burton and colleague (1999) exposed undergraduates to poor-quality images of professors by whom some of the students had been taught. All students were then shown 20 high-quality photographs and told that half of the photos were that of individuals who appeared in the previous images; they were asked to rate the likelihood (on a 7-point scale) that each photograph matched a person with whom they had previously seen. As predicted, students who were identifying known professors performed significantly better than students who were identifying professors who were new to them. Results of this study, and ones similar to it, suggest that adults' recognition of familiar faces is significantly greater than their recognition of never before seen faces (e.g., Burton et al., 1999). However, such findings lead one to wonder about where the transition is from never before seen to familiar and, subsequently, inaccurate to accurate.

Clutterbuck and Johnston’s (2005) study provides some insight into the transition from “never before seen” to “familiar” face. They measured undergraduates’ reaction time and accuracy at matching face pairs of famous (known), familiarized, and never before seen faces. In the first stage, experimenters familiarized students with a set (12) of faces by presenting them 5 times at 4 seconds (i.e., the 5@4 condition) and a second set of faces were presented 10 times at 2 seconds (i.e.,10@2 condition). In the testing stage of the experiment, students were provided with 96 face pairings: Each face pairing was composed of either a famous face, a familiarized face using the 5@4 method, a familiarized face using the 10@2 method, or a never before seen face presented beside a photo with the internal or external face features removed. Half of the face pairings were matched while the other half were mismatched. Participants were asked to push a button indicating whether the photos were the same or different as quickly as possible. Results suggested that famous/known faces are matched more rapidly than familiarized/learned and never before seen faces, and familiarized faces (only the 10@2 condition) more quickly than never before seen faces; although, these findings are only apparent in the case of making a mismatch decision and opposite results are found when measuring match rates. Overall, these results reveal a complex distinction between never before seen, familiarized, and known faces.

Bonner and colleagues (2003) systematically studied undergraduates’ recognition of familiarized faces. Participants were exposed to three randomly dispersed, short (30 seconds) video clips of 24 faces, half familiar and the other half never before seen, for three consecutive days. Face recognition was measured daily using a face matching task,

similar to the one used in Clutterbuck and Johnston’s (2005) study. Participants were simultaneously shown two photos: a still image from one of the videos previously viewed with the internal (eyes, nose, mouth) or external (hair, chin, ears) features of the face of the target removed and a portrait photograph of the whole face of the matching target or a distracter. The task of the participant was to decide whether the photos were of the same or different people as quickly and accurately as possible. The authors report that, for the familiarized faces, matching accuracy based on external face features remained relatively constant over the three days; however, accuracy based on internal face features improved over time (and familiarization) and, in fact, was as accurate as external face feature matching by the third day (Bonner et al., 2003). Many face recognition researchers have attributed the enhanced accuracy of familiar face recognition to a shift in reliance on internal compared to external face features (e.g., Campbell, Walker, & Baron-Cohen, 1995; Ellis, Shepherd, & Davies, 1979). Internal face features such as eyes, nose and mouth may be a more reliable and robust source of information as they do not generally change over time and situation; however, external features such as hair, hairline, and ears may more easily be disguised or changed from time to time (Ellis et al., 1979). Previous exposure to a face may allow someone to view inner face features under a variety of conditions and thus have a robust memorial representation of the familiar person’s face.

A meta-analysis by Johnston and Edmonds (2009) highlights the need to consider important differences in accuracy of face recognition when the face is unfamiliar as compared to when it is “familiar”. From a cognitive standpoint, the ability to recognize a familiar face is impervious to such factors as context and expression but is just as

vulnerable as a never before seen face to poor lighting conditions (Johnston & Edmonds, 2009). These findings have some relevancy to the eyewitness situation, although the differences in familiar and unfamiliar face identification accuracy has been vastly ignored in eyewitness research.

***Theoretical implications.*** Burton and colleagues (1999) proposed the Interactive Activation and Competition (IAC) model of face recognition. This model is comprised of Face Recognition Units (FRU) which house view-independent information regarding any view of a recognizable face. There also are Person Identity Nodes (PIN) which form a more global domain for person information. Burton et al. (1999) assert that person familiarity is a function of the PIN; when the PIN reaches a threshold, familiarity is signaled. PINs are linked to Semantic Information Units which house person information, such as their occupation and Name Recognition Units, which is their name. Lastly, there are Lexical Outputs which contain what would be produced during retrieval. This semantic-priming model aptly explains the impact of previous exposure on face recognition – the more frequently a face is seen the more readily its FRU-PIN link will be accessed (Burton et al., 1999). Thus, a familiar face will be recognized more quickly and semantic information associated with it will be more readily accessible.

***Eyewitness research.*** Even though the face recognition literature provides some insightful information regarding adults' ability to identify a never-before-seen face versus a familiar face, such literature may not directly inform eyewitness research. To re-enforce this notion, Shapiro and Penrod (1986) analyzed over 190 studies, 80% of which were facial recognition studies and the remaining 20% eyewitness identification studies,

and concluded that eyewitness identification studies have reliably lower correct identification rates (or “hits” in face recognition lingo) and higher false identification rates (“misses”) than the average face recognition study. However there are few studies that have used an eyewitness paradigm to examine familiar-stranger recognition accuracy. Below these few studies will be discussed along with studies that have considered extended exposure duration.

Steblay, Dietrich, Ryan, Raczynski, and James (2011) examined the influence of viewing a sequential lineup more than once on recognition accuracy. While not the primary purpose of their study (Experiment 1), Steblay and colleagues (2011) asked participants to report their familiarity with lineup members after they completed the lineup. Their confederate was a previous student at the school from which some participants were recruited and, therefore, some participants may have had previous exposure to the target. In their initial attempts at completing the sequential lineup, 8% of stranger witnesses correctly identified the target and 68% correctly rejected the lineup whereas, 60% of familiar witnesses correctly identified the target and 88% correctly rejected the lineup. In this study, participants were permitted to view the sequential lineup again and modify their responses if they so wished (of note, only 5% of familiar witnesses chose to view the sequential lineup for more than one time). Once participants had viewed the lineup for as many times as they wished, stranger witnesses correctly identified the target at a rate of 14% and correctly rejected the lineup at a rate of 45%. Alternatively, 66% of familiar witnesses correctly identified the target and 88% correctly rejected the lineup (Steblay et al., 2011). Familiar witnesses were significantly more

likely to correctly identify the culprit from a TP lineup and were significantly more likely to correctly reject a TA lineup. The authors also examined participants’ choosing rates (i.e., the rate in which any lineup member was chosen, target or filler) with respect to familiarity with the culprit. Familiar witnesses who viewed a TP lineup had the highest choosing rate and familiar witnesses who viewed the TA lineup had the lowest choosing rate. Moreover, diagnosticity of lineup decisions was significantly higher for familiar witnesses than for stranger witnesses. Finally, familiar witnesses were significantly more confident in their identification decisions than stranger witnesses (Steblay et al., 2011). This rare study provides a great deal of insight into familiar witness identification accuracy; however, given that measuring familiarity was post-hoc, usability of the results are limited. For example, the authors were unable to determine the extent of the familiar witnesses’ exposure to the culprit or to verify participants self-report claims of previous exposure. Pezdek and Stolzenberg’s (2014) recent research highlights that subjective ratings of familiarity with a person may not be wholly reliable.

Pezdek and Stolzenberg (2014) examined high school students’ ability to identify students who graduated from their school one year previously. The participants would have been in their first year of high school when the to-be-recognized students were completing their fourth year of high school. The study was conducted at two small (approximate enrolment of 750) all-girl schools. Each participant was presented with 40 yearbook photos, half of which pictured a student who graduated from their school (presumably familiar). Only 42% of schoolmate photos were correctly classified as familiar, whereas 23% of stranger photos were incorrectly labelled as familiar. Further

analysis indicated that familiar photos were significantly more likely to be classified as familiar than stranger photos; however, the authors note that the diagnostic error is high enough to warrant caution in the forensic utility of familiarity judgements. Similar to the Steblay et al. (2011) study, while this study provides insightful information regarding adults’ ability to accurately determine person familiarity, given that familiarity was not controlled or objectively measured, this study does not contribute fully to our understanding of the influence of familiarity on identification accuracy.

*Exposure duration.* Alternatively, there are studies using an eyewitness paradigm that have been conducted on the influence of extended exposure duration on recognition accuracy, many of which have been conducted with children. For example, in order to mimic the physical contact of sexual assault cases, Leippe, Romanczyk, and Manion (1991) had a confederate administer a skin sensitivity test that involved the confederate touching the participant. After a brief delay (less than 30 minutes), participants who were 5-6 years, 9-10 years, or college students were asked to identify the confederate (longer and more interactive exposure) as well as an individual who briefly interrupted the skin test (shorter and less interactive exposure). Twenty-five percent of the child groups inaccurately identified the confederate (long exposure duration); they were significantly less accurate than the college students. Both children’s groups were also less accurate in identifying the intruder (short exposure duration). However, the age effect was more pronounced in the case of a shorter exposure to the target (Leippe et al., 1991).

Similarly, Gross and Hayne (1996) used a live event to assess children’s eyewitness identification of targets of varying exposure durations. Five and 6-year-olds from New

Zealand went on a trip to the fire station where they were exposed to two tour guides for roughly an hour as well as two intruders who they viewed only briefly (approximately 30 seconds each). They completed an identification task for each target a few days later. Consistent with previous eyewitness identification studies, the children performed more accurately when they were given TP lineups (in comparison to TA lineups). Additionally, participants were more accurate in identifying the tour guide with whom they had a longer exposure than the intruder. A notable finding of this study is that exposure duration had a significant effect on TP lineups, not TA; in other words longer exposure to the target significantly increased correct identifications, however significant differences were not found for correct rejections (Gross & Hayne, 1996).

One of the few studies examining the effect of exposure duration on eyewitness accuracy of adults was published by Memon, Hope, and Bull (2003). Approximately half of their participants were undergraduate students (*M*age = 19 years). Participants were exposed to the target in a robbery videotape for 12 or 45 seconds and they were asked to identify the target using either a TP or TA lineup. Student participants who saw the target for 45 seconds were significantly more likely to correctly identify the target from a TP lineup (95% correct IDs) than participants who saw the target for 12 seconds (29% correct IDs). Also, the longer exposure duration (41% incorrect IDs) elicited significantly less incorrect identifications when student participants were provided a TA lineup than the shorter duration (90% incorrect IDs; Memon et al., 2003).

Horry, Halford, Brewer, Milne, and Bull (2014) examined 295 case files from a large police force in England. In total there were 833 eyewitness identification lineups

administered to stranger eyewitnesses (witnesses with previous exposure to culprits were excluded from the dataset), the majority of which were live lineups. The rate of suspect identification was significantly higher for cases involving longer exposure durations (more than 60 seconds) compared to shorter exposure durations (less than 60 seconds).

These archival results are consistent with a meta-analysis conducted by Bornstein, Deffenbacher, Penrod, and McGorty (2012) on laboratory face recognition and eyewitness studies. The authors examined the influence of exposure duration across 33 independent effect sizes reported in 25 articles. Overall, they found longer exposure duration elicited significantly greater accuracy than shorter exposure duration which is consistent with a previously published meta-analysis by Shapiro and Penrod (1986). However, Bornstein and colleagues (2012) maintain that increasing exposure duration may influence accuracy more so on the shorter end of the spectrum. For example, greatest effects appear when durations are less than 30 seconds, whereas beyond 30 seconds a more substantial difference is required to produce equivalent effects (Bornstein et al., 2012)

Overall, results from face recognition and eyewitness identification studies suggest that longer exposure duration may result in increased identification accuracy (Gross & Hayne, 1996; Leippe et al., 1991). Moreover, Burton and colleagues’ IAC model (1999), although typically applied to cognitive research, implies that face recognition for familiar faces is more quickly and accurately accomplished due to semantic priming. From this previous research, it may be predicted that familiar-stranger eyewitnesses would exhibit

higher correct identification rates (in TP lineups) and higher correct rejections rates (in TA lineups) than stranger eyewitnesses.

**Confidence**

The utility of confidence as an indicator of eyewitness accuracy is an ongoing and hotly debated topic for eyewitness researchers. The present study will not resolve these issues; however, eyewitness confidence will be measured in order to provide insight to the potential variability in confidence across stranger and familiar-stranger eyewitnesses.

There is no doubt that a confident eyewitness is persuasive to a jury (e.g., Brewer & Burke, 2002). However, the relationship between confidence and accuracy is inconsistent in the literature. Some researchers have found a significant confidence-accuracy (CA) relationship (e.g., Brewer & Wells, 2006) whereas others have not (Deffenbacher, 1980). Published meta-analyses or reviews of the literature thus far, generally, find an overall small to moderate relationship between confidence and identification accuracy (e.g., Bothwell, Deffenbacher, & Brigham, 1987; Sporer, Penrod, Read, & Cutler, 1995; Deffenbacher, 1980). Although, more recent studies examining the CA relationship have produced more optimistic results (i.e., higher *r*'s; Wells, Olson, & Charman, 2002). For example, in a recent review of stranger eyewitness studies by Leippe and Eisenstadt (2007), these authors estimate that in an ideal witnessing situation the CA *r* is approximately 0.40 and can be as high as 0.55 to 0.60.

Sporer and colleagues (1995) meta-analysis highlights a potential moderator of previous findings. Similar to previous researchers they found that the overall CA relationship yielded an $r = 0.29$. However, when they considered the CA relationship for

eyewitnesses who chose someone from a lineup (i.e., choosers) and those who did not choose someone from the lineup (i.e., nonchoosers), differences in the CA relationship arose. The CA relationship for choosers was $r = 0.41$ whereas the CA relationship for nonchoosers was $r = 0.12$, indicating that confidence may be a better predictor of accuracy for those who chose someone from a lineup than was once thought. Sporer et al. (1995) further partitioned their data to eliminate any choosers who picked a filler from the lineup; as filler choices are known errors, it can be reasonably assumed that these lineups will not be further processed in an investigation. Choosers (with filler choosers removed) yielded a CA relationship of $r = 0.39$.

**Familiar-stranger witnesses.** There are very few studies that have considered the confidence or the CA relationship across levels of eyewitnesses' familiarity with a culprit; however, research on the influence of exposure duration may provide some insight into predicted findings. Bothwell and colleagues (1987) conducted a meta-analysis similar to those reported previously; however, they considered exposure duration as a moderator variable. The overall weighted average CA relationship effect size for the 35 eyewitness studies considered was $r = 0.25$. Additionally, the correlation between CA and exposure duration was $r = 0.36$. The authors assert that these findings support the optimality hypothesis; that is, that ideal information-processing conditions increase the correlation between eyewitness confidence and identification accuracy. According to Deffenbacher (1980) such ideal conditions include: awareness of an impending memory test, optimal stress, a short amount of time between exposure and test, similar conditions

at encoding and test, extended exposure to the target (as found in Bothwell et al., 1987’s study), and (most notable for the present research) familiarity with the target, and so on.

On the other hand, results from Read, Vokey, and Hammersley’s (Experiment 1, 1990) study suggest a more complex relationship between exposure duration, accuracy, and confidence. The authors had 24 undergraduate students study 36 unique yearbook photos for varying durations (5, 12, or 20 seconds). Approximately 10 minutes later they were presented with 76 photos, 36 of which were the targets previously studied (however, the test pictures were from two years after the study pictures), and 40 distractors (never before seen photos). Results indicated that when study and test photos were similar, increases in exposure duration had a positive effect on accuracy; however, there was no significant effect when faces were average in similarity and, in fact, there was a negative effect of exposure duration when the study-test faces were dissimilar. However, remaining consistent with Deffenbacher’s (1980) optimality hypothesis, the magnitude of the CA relationship increased with extended exposure duration. For example, $r = -0.21$ at 5 seconds, -0.20 at 12 seconds, and 0.06 at 20 seconds (Read et al., 1990, Experiment 1).

Read (1995) conducted a series of experiments to further analyze these interesting results from the 1990 study. In Study One, 112 retail store clerks interacted with a target for either 30 to 60 seconds or four to 12 minutes. Two to four days later they completed either a TP or a TA lineup with distractors who were either of high similarity to the target or low similarity to the target. Overall, longer exposure duration elicited more correct identifications; however, also more false positives (of note, it appears that a false positive identification was an identification of *any* filler).

In Study Two (Read, 1995), 211 undergraduates enrolled in three classes were exposed to a male target posing as their teaching assistant; he was introduced to the class and a description of his role as a teaching assistant was provided. He assisted with the distribution of exams and left. In two of the classes, the target was mentioned in class on four occasions subsequent to the initial exposure (e.g., the professor indicated that the teaching assistant lost some exams). Altogether there were three conditions: 1) Control Condition: the target was only mentioned/seen during the introduction; 2) Retrieval Condition: the target was seen during the introduction and discussed in class on four occasions; 3) Description Condition: the target was seen during the introduction and discussed in class on four occasions, with participants in this condition being asked to describe the target before completing the identification task. Participants in all conditions made an identification decision three weeks post exposure. Two lineups were presented to all participants; the first was a TA lineup, then a TP lineup. Participants in the Retrieval Condition were significantly more likely to choose someone from the initial TA lineup than those in the Control Condition, although, no significant differences in accuracy were found for the TP lineup. The authors computed CA correlations. The CA relation for those in the Control Condition was highest ($r = 0.35$), followed by those in the Description Condition ($r = 0.31$), and the correlation for those in the Retrieval Condition was much lower ($r = 0.05$). Read suggests that his results may be explained by looking toward the availability hypothesis; that is, the notion that information that is more readily available in one's mind is more likely to influence their judgement or decision (Tversky & Kahneman, 1973). Due to the additional time participants in the Retrieval

Condition spent thinking about the target, information regarding him may have easily come to mind. Participants may have interpreted the ease with which they could recall the target as an indicator that they would be able to accurately recognize him. Moreover, Read differentiates between contextual knowledge (semantic information) and perceptual knowledge (visual information), and suggests that, in Study Two, participants in the Retrieval Condition misattributed their contextual knowledge as an indicator for perceptual knowledge. In his third study, Read explored the influence of contextual versus perceptual knowledge further.

The design of Study Three (Read, 1995) was a 3 (type of information rehearsed: none, perceptual, contextual) x 2 (lineup type: TP vs TA) x 2 (photo similarity: high vs. low) x 2 (target sex: male vs. female) between-subjects design. One hundred and ninety-two retail store clerks participated in a ten-minute interview followed by three telephone calls over the span of two weeks. The telephone calls were used to manipulate the type of information rehearsed. For example, for those in the Perceptual Condition the telephone calls consisted of the participant describing the target’s physical appearance, a discussion of an interaction with a customer, and describing the target’s physical characteristics again. Participants in the Contextual Condition received a phone call from the target him- or herself and they were asked to describe their interaction with the target (excluding physical characteristics of the target), to discuss topical events with the target (in order to develop rapport), and again to describe their interaction with the target. Lastly, participants in the Control Condition described customer interactions and appearances more generally. Participants were presented with a TP or a TA photo lineup

two weeks after the initial interview. Results indicated that participants in the Perceptual (67%) and Contextual (69%) Conditions performed similarly and at a better rate than the Control (58%) participants; although, these rates did not significantly differ from one another. Confidence in correct decisions was significantly higher than confidence in incorrect decisions, and confidence in TP lineup decisions was higher than confidence for TA lineup decisions. There was no overall effect of confidence across the rehearsal conditions. CA correlations also were computed. The overall CA correlation was $r$ = 0.47; for those in the Contextual Condition $r$ = 0.47, Perceptual Condition $r$ = 0.47, and Control Condition $r$ = 0.30. In consideration of the findings of Sporer et al. (1995), the CA relationship for choosers only was reported as well. When only considering the CA relationship for choosers the overall $r$ increased to 0.73; for choosers in the Contextual Condition $r$ = 0.60, in the Perceptual Condition $r$ = 0.85, and in the Control Condition $r$ = 0.79.

Overall there are no known studies that have examined identification confidence in familiar-stranger eyewitnesses. With stranger eyewitnesses, however, identification confidence generally shows a moderate relation to identification accuracy, suggesting that it may be used as somewhat of an indicator of eyewitness identification accuracy. When considering the influence of exposure duration and consequently, the impact of additional encoding opportunity, extended exposure duration seems to increase the CA relationship. In other words, the confidence ratings of eyewitnesses who have had more opportunity to view the culprit better align with their identification decisions. Extending this limited previous research to the current program of study, one may predict that

familiar-stranger eyewitnesses may be more confident in their lineup decisions and that their confidence judgements may be more highly correlated with their lineup decision than stranger eyewitnesses.

## Overview of Current Program of Research

In summation, there are relatively few research studies available concerning the accuracy of eyewitness identification of familiar-strangers. In fact, to date, there is no known research that has used the eyewitness paradigm to examine whether eyewitnesses' familiarity with the culprit influences their recall or recognition accuracy, or confidence.

This lack of empirical research may be because the notion of person "familiarity" is not a well-studied or understood concept (Mandler, 2008). Dual-process theorists use it to describe one of two parallel processes which represents an impoverished and less exact memory of something or someone. On the other hand, face recognition researchers refer to "familiar" faces as those that are known (e.g., famous people) or have been learned through repeated exposure. Social psychologists have defined familiarity as the frequency with which one is exposed to a person (Moreland & Zajonc, 1982). Eyewitness researchers have not formally adopted a concept of familiarity, although it seems as though a familiar eyewitness broadly refers to a witness who has had *any* previous exposure to the culprit. A broad definition such as this does not allow for research findings that would have a great deal of forensic utility in a courtroom.

For the purpose of the present program of research, the goal was to operationally define "familiarity" in a manner that would align with a witness's subjective experience of feeling familiar with a person. Given that the underlying goal of the present studies

was to apply the findings to a judicial context, such a definition of familiarity seems fitting. Eyewitnesses will have a subjective awareness of their familiarity with the culprit and/or defendant; however, they may not always have information regarding their previous exposure (i.e., they may not be able to relate how many times and for how long they had previously seen the culprit). In the present studies, an attempt was made to manipulate subjective feelings of familiarity with the target in order to determine whether familiarity influences eyewitness recall and recognition accuracy. In particular, the purpose of the present research was to garner insight into the identification accuracy of familiar-stranger eyewitnesses – witnesses who would fall somewhere between those who never before saw the culprit and those who know the culprit well (e.g., relative, close friend, etc.).

**Hypotheses**

The following predictions were made in regards to the influence of familiarity on recall accuracy, recognition accuracy, and confidence:

1. Eyewitnesses who are more familiar with the culprit will provide a greater overall number of descriptors and proportionately more correct descriptors, than eyewitness who are less familiar with the culprit, followed by stranger eyewitnesses.

2. a) The rate of correct identification (TP lineups) will be higher than the rate of correct rejection (TA lineups) regardless of level of familiarity.

2. b) For both types of lineups, participants will be more likely to provide a correct decision (i.e., correct identification in TP lineups or correct rejection in TA lineups)

when they are more familiar with the culprit compared to those who are less familiar, and those who are strangers to the culprit will be least likely to provide a correct response.

3. a) Eyewitnesses in the more familiar condition will rate their confidence highest across all confidence measures (i.e., pre-lineup confidence measure, lineup confidence, and post-lineup confidence measure), followed by those who are less familiar with the culprit, and lastly eyewitnesses for whom the culprit is a stranger.

3. b) The CA relationship (i.e., the correlation between confidence ratings and accuracy of lineup decision) will be highest for those in the more familiar condition, followed by those in the less familiar condition, and lastly those in the stranger condition. Moreover, these correlations will be higher when only considering participants who choose someone from the identification lineup.

## Pilot Studies

Prior to conducting the Main Study, a series of pilot studies were conducted to test the manipulations to be used in the Main Study (i.e., more familiar versus less familiar). It was of particular importance to ensure that attempts to manipulate person familiarity effectively produced changes in participants’ feelings of familiarity with the target.

### Pilot Study One

In an initial pilot study, videotapes of the target were used to manipulate exposure duration (one minute vs. nine minutes) and number of repeated exposures (one time vs. three times) in order to determine whether person exposure duration and/or repeated exposure effectively influenced participants’ perceptions of familiarity with the target.

The limited previous eyewitness research and some face recognition studies have found that extended exposure duration and exposure repetition increase recall and recognition accuracy, and confidence (e.g., McNichol et al., 1999; Read, 1995). However, none of these studies systematically examined whether extended exposure duration and exposure repetition effectively produced differences in participants' perceptions of familiarity. Regardless, given that previous research has been published using these manipulations it was a desirable starting point.

## Method

### Participants

Eighty-eight undergraduate students from Carleton University participated in Pilot Study One. Their ages ranged from 17 to 63 years ($M$ = 19.98, $SD$ = 5.06) and 76% ($N$ = 67) were women. Forty-four participants (50%) self-identified as being White (e.g., European) and 18 (21%) described themselves as Black (e.g., African Canadian). All participants received 1.5% credit towards a psychology course for taking part in this 75-minute study session.

**Study design.** Three levels of exposure were manipulated using videotaped exposure; participants were exposed to the target for one minute on one occasion (one minute duration/one time exposure), nine minutes on one occasion (nine minute duration/one time exposure) or for three minutes on three occasions (nine minute duration/three time exposure). The primary dependent measure was perception of familiarity.

**Materials**

**Demographics.** Each participant completed a demographics form that requested them to report their age, sex, primary language, and ethnicity. See Appendix A.

**Videos.** All participants were shown a series of videos (in total 27 minutes in length) of a female target reading children's story books (i.e., by Dr. Seuss and Robert Munsch). In the videos, the target was seated on a chair with the story book in her hands. No other stimuli or people were visible on the video. Exposure to the target was manipulated by having the target visible or not visible on the videotape. For example, in the nine-minute/one-time exposure condition participants saw the target on the video on one occasion for nine minutes; whereas, in the nine minute/three-time exposure condition they saw the target on the video on three separate occasions for a total of nine minutes. During the time when the target was not visible a still shot of a stack of books was shown on the screen.

**Familiarity.** Participants were asked to rate their familiarity with the storybook reader on a 100-point scale (0 *not at all familiar* to 100 *very familiar*).

**Description.** Participants were asked to describe the target in the video using an open-ended question (i.e., please write down everything that you can remember about the appearance of the person who read the storybooks on the video). Also, participants were asked to complete 16 multiple choice questions regarding the appearance of the target (e.g., the person who narrated the storybooks was (a) male, or (b) female). These activities were used as filler tasks.

**Photo array.** Participants were sequentially presented with six colour photographs, 4 x 6 inches, of the head and upper body of individuals who resembled the target in the videotape. The photos were presented to participants one at a time. The participants were instructed to: "Think back to the video. Think back to what the person reading the storybooks looks like. I am going to present you with some photos. The narrator you saw on the video may or may not be in the photos. You will only be allowed to look at each photo once. Please look at each photo and decide if it is or is not a picture of the narrator. If the picture is of the narrator, please place a check mark next to 'yes.' If it is not a picture of the narrator please place a check mark next to 'no'. You will not be able to move ahead in the sequence until you make a decision. Once you make a decision, you will not be allowed to re-examine the photo. After each decision, please rate your confidence using the scale below." The scale below was a 100-point confidence measure (0 *not at all sure* to 100 *absolutely sure*). This task was used to simulate a real-life criminal investigation; however, data obtained from this task was not analyzed for the purpose of this pilot study.

**Manipulation check form.** Participants were asked to verify that they did not know about the memory portion of the study prior to completing the photo array task and to state whether or not they personally knew any of the individuals shown in the photo array. No participants indicated that they were aware that the study would require an eyewitness identification and no participants indicated that they knew a lineup member.

**Procedure**

This pilot study took place in the Laboratory for Child Forensic Psychology at Carleton University. It was completed in one testing session lasting approximately 75 minutes. Upon entering the lab, participants were asked to read the informed consent form explaining that they would be participating in several activities that children typically perform. Students were told that they would be read storybooks via videotape. They were asked to pay attention to the story books so that they may later answer questions in regards to them. A breakdown of participants’ activities is available in Table 1. Various activities (i.e., word search and computer game) were offered to participants between video clips as filler tasks and distractors. Upon completion of the study, students were thanked for their time and given a debriefing form with information on the purpose of the study and available resources should they be required.

Table 1

*Pilot Study One procedural breakdown*

| Time | One-time Exposure (1 minutes) | One-time Exposure (9 minutes) | Three-time Exposure (9 minutes) |
|---|---|---|---|
| 0:00 | Informed Consent | Informed Consent | Informed Consent |
| 0:02 | Video 1 (9 minutes) *target not visible* | Video 1 (9 minutes) *target not visible* | Video 1 (9 minutes) *target visible for 3mins** |
| 0:11 | Demographic Form Word Search | Demographic Form Word Search | Demographic Form Word Search |
| 0:16 | Video 2 (9 minutes) *target not visible* | Video 2 (9 minutes) *target not visible* | Video 2 (9 minutes) *target visible for 3mins** |
| 0:25 | Word Search | Word Search | Word Search |
| 0:30 | Video 3 (9 minutes) *target visible for 1mins** | Video 3 (9 minutes) *target visible for 9mins* | Video 3 (9 minutes) *target visible for 3mins** |
| 0:39 | Familiarity Form Description Form | Familiarity Form Description Form | Familiarity Form Description Form |
| 0:50 | Computer Game | Computer Game | Computer Game |
| 1:05 | Identification Lineup Deception Awareness Debriefing | Identification Lineup Deception Awareness Debriefing | Identification Lineup Deception Awareness Debriefing |

*The timing of the exposure was counterbalanced across the 9-minute videotape. For example, some participants saw the target within the first 3 minutes of the 9-minute tape, others saw the target in the middle 3 minutes of the tape, and others saw the target in the last 3 minutes of the 9-minute video.

## Results

### Exposure Duration

An independent samples *t*-test was used to assess whether exposure duration (one minute versus nine minutes) influenced perceptions of familiarity. There were no significant differences in participants perceived familiarity when they viewed the target for one minute or nine minutes, $t(57) = 0.14$, $p = .89$, $d = 0.04$, 95% CI [-0.48, 0.55]. Descriptive results appear in Table 2.

Table 2

*Descriptive results from Pilot Study One*

| | One-time Exposure (1 minute) | One-time Exposure (9 minutes) | Three-time Exposure (9 minutes) |
|---|---|---|---|
| Familiarity | 31.03 (30.04) | 30.00 (26.78) | 31.25 (29.80) |

### Exposure Repetition

An independent samples *t*-test was used to assess whether exposure repetition (one time versus three times) influenced perceptions of familiarity. There were no significant differences in participants' perceived familiarity when they viewed the target on one occasion or three occasions, $t(56) = -0.17$, $p = .87$, $d = 0.04$, 95% CI [-0.47, 0.56]. Descriptive results appear in Table 2.

## Discussion

Results from this initial pilot study revealed that neither manipulation significantly influenced participants' perceptions of familiarity. This suggests that participants were

not describing person familiarity in terms of exposure duration or repeated exposure, at least not when the target was presented via videotape. At this point it became apparent that the best approach in understanding subjective experiences of familiarity may be to examine participants' definitions of person familiarity.

**Qualitative Analysis of "Familiarity"**

A subset of participants who completed pilot studies for this dissertation were asked to answer the question, "When someone says 'I am familiar with him,' what does that mean to you – how would you define 'familiar'?" Participants responded to this question in written format. In total, 75 undergraduate students answered this question. Their mean age was 21.77 years (*SD* = 5.32) and 71% (*N* = 53) were women.

All responses were entered into the qualitative research tool, NVivo. A word frequency was run in order to assist with determining themes. The most frequent relevant words (i.e., excluding words like "the") were: know (count = 72), person (52), familiar (44), seen (18), met (11), and friends (9). Next, all responses were coded for themes. Overall, 20 unique (i.e., did not overlap) themes were determined. The most common of which were: knowing something about the person (26 coded items), just knowing the person (22), having seen the person (16), interaction with the person (14), and ability to recognize the person (8).

Overall participants' responses were quite vague, which is a further indicator that the concept of person familiarity is difficult to articulate. Many participants simply responded that person familiarity is knowing someone. For example, one participant stated that, *"To me, it just means that you know that person."* Similarly, another

participant said, *"Personally, I'd define familiar as someone I know. Not necessarily a friend but not a stranger."*

However, there were some common themes in participants' responses that provide some insight into how to manipulate subjective feelings of familiarity. The most common theme was the indication that familiarity was knowing something about the person. For example, one participant stated, *"Familiar means being aware of someone and something about them – the more you know the more familiar you are."* Another participant indicated, *"To be familiar with someone means to have knowledge of a few identifying factors, interests, beliefs, and goals of someone."* Other common themes were suggestions that the person must be previously seen and there must have been an interaction. For example, one participant stated, *"As in knows basic information about that person (name, nationality, occupation) and had at least a 30 min long conversation."* whereas another responded, *"Familiar means that you have seen that person before and that you might have talked to him/her before, and that you know things about him/her."*

Based on these qualitative responses, subjective familiarity may be elicited via personal interaction and personal knowledge of a person. Pilot Study Two was conducted to examine whether manipulating these variables effectively influenced feelings of familiarity.

## Pilot Study Two

Pilot Study Two was conducted in order to determine whether person familiarity may be influenced by personal interaction and/or knowledge of personal information, as

was indicated by participants' definitions of familiarity. In order to provide control over the personal interaction and the sharing of personal information a timed face-to-face interview with a target was used.[2]

## Method

### Participants

Undergraduate students from Carleton University participated in Pilot Study Two ($N$ = 69). Participants ranged in age from 17 to 44 years old ($M$ = 21.81, $SD$ = 5.48). Majority of the participants were women (78%) and many self-identified as being White (e.g., European, 46%).

**Study design.** Two levels of interaction with the target were manipulated: interaction without the sharing of personal information and interaction with the sharing of personal information. The dependent measures included a series of ratings to assess perceptions of familiarity with the target.

### Materials

**Demographics form.** Participants completed a demographics form to report their age, sex, primary language, and ethnicity. See Appendix A.

---

[2] The influence of exposure to a target *without interaction* was examined in a small ($N$ = 29) but separate study. Briefly, participants of this study were exposed to a target without interaction, with interaction, or with interaction and sharing of personal information for 5 structured minutes. Although the sample size was quite small, trends in the data suggested no differences in perceptions of familiarity between those in the no interaction and interaction without the sharing of personal information groups. In order to narrow the scope of this program of research, a decision was made to concentrate only on the two manipulations that were eliciting significantly different perceptions; that is, (1) personal interaction and (2) personal interaction with sharing of personal information.

**Familiarity interview.** Participants took part in a face-to-face structured interview (see Appendix B for script) in which they were asked to report information on their schooling, interests, and hobbies. The interview administrator was the target. The purpose of this task was to create an opportunity for the controlled sharing of personal information by the target; for participants in the personal information condition the confederate provided personal information about herself whereas for participants in the no personal information condition the target shared information attributed to the average person (i.e., non-personal).

**Word game.** Participants played a word game with the target. They were instructed as follows: "Next, I invite you to play a word game with me. I will start with a word such as elephant and then you will follow with any word that comes to your mind that begins with the last consonant or vowel from my word. For example, I would say elephant which ends with a *t* and then you would say a word that starts with a *t*, such as together. The words do not have to be related and we cannot use the same word more than once. We will continue back and forth like this until we reach 100 words or until the timer goes off." Clarification was provided as necessary and the words were recorded. The purpose of this task was to extend the amount of time the participant was exposed to the target; the word game was stopped once the combined time of the interview and the word game reached 30 minutes. For example, if a participant's interview was 15 minutes long, the length of time permitted to play the word game was 15 minutes.

**Questionnaire.** Participants were asked a series of open- and close-ended questions regarding their notion of familiarity and in particular their perceptions of familiarity with the target. See Appendix C for the questionnaire form.

**Procedure**

Upon entering the lab, participants completed an informed consent form and demographics form independently. Next, the target conducted the familiarity interview and played the word game with the participant. The combined duration of the interview and word game was 30 minutes. Participants were then provided the questionnaire to complete independently (the target was not present). Once complete, participants were thanked for their time and provided a verbal debriefing and a debriefing form.

## Results

A multivariate analysis of variance (MANOVA) was conducted in order to determine whether the sharing of personal information significantly influenced participants' perceptions of familiarity with the target. However, there were nine instances of missing data in nine participants that caused these cases to be deleted from the MANOVA. As a result, these nine instances were replaced by the item mean. Given that these nine instances represent less than 0.01% of the data analyzed, it is anticipated that using mean replacement did not bias the results. The omnibus $F$ test indicated that there is at least one significant difference in the perceptions of participants who were administered the personal versus the non-personal interview, Hotelling's $T = 0.74$, $F(15, 53) = 2.61$, $p < .01$, partial $\eta^2 = 0.42$, 95% CI [0.09, 0.43]. All post-hoc results are available in Table 3. As may be expected, participants who interacted with a target who

shared personal information rated their perceptions of having personal information about the target significantly higher than participants who interacted with the target who did not share personal information. Additionally, participants in the personal information sharing condition were significantly more likely to yield higher ratings of knowing the target, feeling acquainted with the target, and having things in common with the target. Lastly, there was a trend ($p = .06$) toward higher ratings of familiarity for those in the personal information condition compared to those in the no personal information condition.

Table 3

*Descriptive results from Pilot Study Two*

| | Non-Personal | Personal | *F* | *df* | *p* | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| How **familiar** are you..? | 23.09 (22.76) | 33.86 (24.32) | 3.60 | 1, 67 | .06 | 0.05 |
| How well do you **know**..? | 13.91 (20.54) | 25.46 (21.98) | 5.08 | 1, 67 | .03 | 0.07 |
| How much **exposure** have you had..? | 28.93 (24.79) | 28.50 (17.17) | 0.01 | 1, 67 | .93 | 0.00 |
| How many **minutes** would you estimate you saw..? | 30.59 (14.11) | 32.12 (23.43) | 0.11 | 1, 67 | .74 | 0.00 |
| How much **information** could you provide about..? | 26.18 (24.37) | 34.87 (15.97) | 3.09 | 1, 67 | .08 | 0.04 |
| How much **personal information** do you know..? | 9.12 (12.15) | 32.57 (21.05) | 31.87 | 1, 67 | .00 | 0.32 |
| How likely are you to **forget** that you met..? | 33.53 (31.03) | 23.43 (24.25) | 2.28 | 1, 67 | .14 | 0.03 |
| Do you think the researcher is a **stranger**? | 60.88 (29.06) | 54.29 (29.13) | 0.89 | 1, 67 | .35 | 0.01 |
| How much do you **not know** about..? | 71.32 (38.52) | 72.57 (27.58) | 0.02 | 1, 67 | .88 | 0.00 |
| How accurately do you think you could **describe**..? | 41.47 (26.42) | 48.86 (22.59) | 1.56 | 1, 67 | .22 | 0.02 |
| How likely would you be able to **recognize**..? | 55.59 (31.35) | 61.53 (28.20) | 0.69 | 1, 67 | .41 | 0.01 |
| How **confident** would you be in your recognition..? | 58.82 (30.43) | 59.43 (27.11) | 0.01 | 1, 67 | .93 | 0.00 |
| How **friendly** are you with the researcher? | 77.50 (25.11) | 76.00 (20.18) | 0.08 | 1, 67 | .79 | 0.00 |
| How well **acquainted** are you with the researcher? | 27.21 (26.20) | 44.86 (25.48) | 8.05 | 1, 67 | .01 | 0.11 |
| How much do you have in **common**..? | 26.20 (23.90) | 39.43 (21.69) | 5.80 | 1, 67 | .02 | 0.08 |

**Discussion**

When asked in open-ended format, participants indicated that familiarity with a person is attained by knowing something about the person, interacting with the person, and seeing the person. As a result, this pilot study was conducted to align with these perceptions: participants in Pilot Study Two interacted with a confederate for 30 minutes in which the confederate either shared or did not share personal information. Results from this study suggest that the sharing of personal information significantly influences subjective perceptions of familiarity (trend), knowing someone, and feeling acquainted with someone.

These results may be interpreted using Mandler’s (1981) explanation of the process of familiarity and recognition. Mandler (1981) asserts that an impression of familiarity is a result of the way that we organize intraevent features (called integration). Intraevent features are items found within the event. This is in contrast to how one would organize interevent information which is information that can be related to other events (called elaboration). Simply interacting with someone may elicit the process of integration as characteristics of that person (intraevent information) can be categorized in one’s mind. However, when the interaction involves the sharing of personal information, the opportunity for elaborative processes arises; this personal information can incite interevent processing, for example, the categorization of the individual based on their employment or preferences. Interevent processing or elaboration results in the ability to retrieve memories from a variety of sources and hence, know or feel acquainted with

someone. This is very similar to Read's (1995) distinction between perceptual (intraevent) and contextual (interevent) information.

More recently, Cloutier, Kelley, and Heatherton (2011) distinguished between perceptual-based familiarity and knowledge-based familiarity. Perceptually familiar faces are those which have been previously seen, although no personal or distinguishing information is known about them. On the other hand, personal-knowledge faces are faces with whom some personal information (e.g., social status) is known or learned.

## Main Study

Given the results from this series of pilot studies, for the Main Study familiarity was manipulated using personal interaction and personal knowledge of a target/culprit. For the main study, three levels of familiarity with the culprit were manipulated: (1) previous interaction with the culprit sharing personal information (more familiar condition); (2) previous interaction with the culprit without disclosure of personal information (less familiar condition); or (3) no previous exposure to the culprit (stranger condition).

## Method

### Participants

The original sample was comprised of 213 participants. However, 19 participants did not complete the study (i.e., did not return for Session Two within the time limit; yielding an 8.92% attrition rate) therefore their data were incomplete and removed from analyses. Data from an additional 10 participants were removed from analyses due to irregularities during data collection (e.g., the confederate was seen for more than the

prescribed 30 minute exposure time). The final sample consisted of 184 participants ranging in age from 17 to 60 years old ($M$ = 20.98, $SD$ = 5.86). Seventy-one percent ($N$ = 131) of the sample were women and 52% ($N$ = 95) self-identified as being White (e.g., European). The number of participants, and their demographic characteristics were evenly distributed across conditions[3].

**Study Design**

A 3 (familiarity with the culprit: more familiar vs. less familiar vs. stranger) x 2 (photographic lineup type: target-present vs. target-absent) between-subjects factorial design was used. There were six conditions with 30 to 32 participants per condition. The primary dependent measures were recall accuracy (i.e., total number and proportion of correct descriptors), recognition accuracy (i.e., proportion of correct identifications in target-present lineups and proportion of correct rejections in target-absent lineups), and confidence prior to completing the lineup task, during the lineup task, and upon completion of the lineup task.

---

[3] There were no significant differences in participants' age across exposure ($F(2, 181) = 0.19$, $p = .83$, partial $\eta^2 < .01$, 95% CI [0.00, 0.01]) and lineup ($t(182) = 0.41$, $p = .68$, $d = 0.06$, 95% CI [-0.23, 0.35]) conditions. There were no significant differences in participants' sex across exposure ($\chi^2(2, N = 184) = 0.98$, $p = .61$, Cramer's $V = 0.07$) and lineup ($\chi^2(1, N = 184) = 0.24$, $p = .63$, $\phi = 0.04$) conditions. Lastly, there were no significant differences in participants' race (White vs. Non-White) across exposure ($\chi^2(2, N = 184) = 0.70$, $p = .71$, Cramer's $V = 0.06$) and lineup ($\chi^2(1, N = 184) = 0.20$, $p = .66$, $\phi = 0.03$) conditions.

**Materials**

**Demographics form.** Participants completed a demographics form to report their age, sex, primary language, and ethnicity. See Appendix A.

**Familiarity interview.** Participants took part in a face-to-face structured interview (see Appendix B for script) in which they were asked to report information on their schooling, interests, and hobbies. The interview administrator was the confederate (for those in the more familiar or less familiar conditions) or the researcher (for those in the stranger condition). The purpose of this task was to create an opportunity for the controlled sharing of personal information by the confederate; for participants in the more familiar condition the confederate provided personal information about herself whereas for participants in the less familiar condition the confederate shared information attributed to the average person (i.e., non-personal). For participants in the stranger condition the interview was administered by the researcher and the sharing of personal or non-personal information during the interview was counterbalanced. The average time it took to complete the interview was 15.48 minutes[4] ($SD$ = 3.58) and the maximum amount of time permitted for the interview was 30 minutes. Participants' responses to the interview questions were written out by the interview administrator.

**Word game.** Participants played a word game with the confederate or the researcher, depending on the assigned condition. They were instructed as follows: "Next, I invite you to play a word game with me. I will start with a word such as elephant and

[4] The amount of time to complete the interview did not significantly differ across the three exposure conditions, $F(2, 179) = 0.73$, $p = .49$, partial $\eta^2 < 0.01$, 95% CI [0.00, 0.03].

then you will follow with any word that comes to your mind that begins with the last consonant or vowel from my word. For example, I would say elephant which ends with a *t* and then you would say a word that starts with a *t*, such as together. The words do not have to be related and we cannot use the same word more than once. We will continue back and forth like this until we reach 100 words or until the timer goes off." Clarification was provided as necessary and the words were recorded. The purpose of this task was to extend the amount of time the participant was exposed to the confederate; the word game was stopped once the combined time of the interview and the word game reached 30 minutes. For example, if a participant's interview was 15 minutes long, the length of time permitted to play the word game was 15 minutes but if he/she took 20 minutes for the interview, he/she played the word game for 10 minutes.

**Crime video.** Participants watched a video (52 seconds in duration) of the confederate committing a theft. The video depicted the female confederate (i.e., culprit) filling out a form while seated on a chair in a waiting room. No other persons are visible on the video. The confederate looks suspiciously toward an unattended laptop computer on three occasions and, finally, places the form that she is filling out to the side, gets up from her chair, places the laptop computer into her backpack and leaves the waiting room.

**Pre-lineup confidence form.** Participants responded to four questions using a 100-point rating scale (0 *not confident*; 100 *very confident*) in order to assess confidence in their ability to accurately describe and identify the culprit, prior to attempting the identification task. See Appendix D.

**Description form.** Participants were asked two open-ended questions requiring them to provide a description of the crime and a description of the culprit from the video. Also, participants were asked whether they had previously encountered the culprit and, if so, to provide any additional information that they may have obtained during previous interactions with the culprit. See Appendix E.

**Computer game.** As a filler task, participants played *Bejewelled*, a Tetris-like computer game where the goal is to match jewels in sets of three or more in order to move from one level to the next. This game was played on a laptop computer for 20 minutes.

**Photographic lineup**. Six colour photographs of the head and upper body of individuals who resembled the confederate were chosen to make up a fair photographic lineup.

***Creation of the photographic lineup.*** The photographic lineup was created using Doob and Kirshenbaum's (1973) mock witness paradigm. Twelve student volunteers independently watched the videotape of the mock crime and provided a written description of the target's appearance. These descriptions were collated and a composite description of the target was created. This description read, "The culprit is a female of Caucasian or European ethnicity. She has dark brown hair. She is average height with a slim to medium build. She was wearing blue jeans and a blue t-shirt." Next, nine photos that matched this description were chosen from the Laboratory for Child Forensic Psychology's database of lineup photos (this database consists of Carleton University students who have volunteered to have their photos taken for the purpose of eyewitness

studies and have consented to having their photos used). These nine photos were rank ordered in terms of their match to the description by three graduate students/professors. The top ranked six photos as well as the target's photo created the photo lineup to be tested for fairness.

Next, 30 student volunteers were shown the seven (six fillers and target) photos simultaneously. They were provided the composite description and the photo lineup[5] and instructed, "This description was provided by an eyewitness to a crime. Using this description, please chose which photo looks most like the culprit." The frequency of lineup photos chosen is displayed in Table 4. Using Tredoux's (1998) method, the lineup was found to have an acceptable effective size, E' = 5.06, 95% CI [3.89, 7.23]. Moreover, the target was not chosen at a rate higher than expected by chance ($p$ = .21), indicating that the lineup is fair (Malpass & Lindsay, 1999).

Table 4

*Proportion (frequency) of lineup members chosen from the lineup fairness test*

| Lineup Member | 1 | 2 | 3 | 4 | 5 | 6 | Target |
|---|---|---|---|---|---|---|---|
| Choosing Rate | 0.07 (2) | 0.03 (1) | 0.20 (6) | 0.30 (9) | 0.20 (6) | 0.07 (2) | 0.13 (4) |

***Administration of the photographic lineup.*** In TP lineups five fillers and the target (presented in a random position) appeared in the lineup, whereas in TA lineups the same five fillers and one designated replacement filler (i.e., innocent suspect;

[5] The photos were displayed in a random order for each participant.

presented in a random position) appeared in the lineup. Participants were presented with the photographic lineup using a sequential lineup procedure on a laptop computer via the software program Qualtrics. First, participants were presented with the following instructions: "Please think back to the video. Think back to what the culprit looks like. You will be presented with some photos. The culprit's picture may or may not be in the photos. Please take a look at each photo and decide if it is or is not a picture of the culprit. If the photo is a picture of the culprit, please click on 'yes.' If the photo is not a picture of the culprit, please click on 'no.' You will only be presented with one photo at a time and you will not be able to move ahead to the next photo until you have made a decision about the photo that is in front of you. Once you have made a decision about a photo you will not be able to re-examine the photo again. You will not know how many photos there are and you will continue through the sequence of photos even if you have said 'yes' to one of them." The photos then appeared on the screen one at a time, measuring approximately 4 x 6 inches in size. Along with each photo, participants were requested to click a button to indicate whether or not the picture on the screen was that of the culprit. The amount of time to make this decision was recorded. Following each decision participants were directed to a second screen where they were asked to rate how confident they were that their decision was correct (i.e., How confident are you that you made a correct decision?) using a 100-point scale (0 *not confident*; 100 *very confident*).

**Willingness to testify.** Upon completion of the sequential lineup, participants were asked to rate their willingness to testify to their identification decision in a court of law

using a 100-point scale (0 *not willing*; 100 *very willing*). This measure was used as a proxy measure for overall confidence in identification decision.

**Familiarity lineup response form.** Participants rated their familiarity with six persons using a sequential procedure. Photos, sized approximately 4 x 6 inches, of Angelina Jolie (actress/philanthropist), Hilary Duff (actress/singer), McKayla Maroney (American Olympic gymnast), the culprit, a filler from the identification lineup, and a student volunteer (a stranger) were presented to participants one at a time via a laptop computer. Participants were provided with the following definition: "For the purpose of the next task we are going to define *familiarity* as having some acquaintance or knowledge of someone. I am going to show you six photos of people who you may or may not know. After each photo, please rate how familiar you are with the person pictured in the photo." Following each photo, participants were requested to rate how familiar that person in the photo was to them (0 *not at all familiar*; 100 *absolutely familiar*).

**Manipulation check.** Participants were asked to report whether they were aware of the true purpose of the study prior to Session Two (in other words, whether they knew they were about to view a crime prior to the video) as well as their experience and/or knowledge of eyewitness identification research and procedures. No participants were aware of the true purpose of the study prior to watching the crime video. Overall, 22% reported that they had taken a class on eyewitness evidence and 10% of participants had previously completed an identification lineup. However, these participants did not

significantly differ from the rest of the participants in terms of their identification accuracy[6].

**Procedure**

After ethical approval was obtained from Carleton University's Research Ethics Board, a recruitment notice was posted on an online recruitment tool (i.e., SONA) accessible to undergraduate students registered in introductory psychology and statistics courses. Via SONA, interested students scheduled two appointments, five to ten days apart, to attend to the Laboratory for Child Forensic Psychology to complete a study titled "Goals, Dreams, and Vocabulary."

**Session one.** Upon entering the lab, participants completed an informed consent form and demographics form independently. Then, the researcher or confederate (depending on the condition) administered the familiarity interview (Appendix B) and played the word game with the participant. The combined duration of the interview and word game was 30 minutes. At the completion of this portion of the study, participants were thanked for their time and reminded of their follow-up appointment.

**Session two.** Participants returned to the Lab five to ten days after they completed Session One ($M$ = 7.06 days, $SD$ = 1.34) to participate in Session Two[7] of the study.

---

[6] Having taken a class on eyewitness accuracy did not significantly influence identification accuracy, $\chi^2(1, N = 183) = 0.31$, $p = .57$, $\phi = 0.04$. Also, previously completing an identification lineup did not significantly influence identification accuracy, $\chi^2(1, N = 183) = 0.30$, $p = .59$, $\phi = -0.04$.

[7] An independent samples $t$-test was used to determine whether the number of days between Session 1 (where the culprit was first seen by those in the familiar conditions) and Session 2 (where the crime video was shown and identification lineup completed) influenced identification accuracy. The number of days between Session One and Session Two did not significantly differ for those who made a correct lineup decision ($M$

Participants completed a second informed consent form followed by the viewing of the crime video. They were not told the nature of the video prior to watching it; instead, they were simply requested to watch a quick video on the laptop computer. The video was 52 seconds in duration. Following the viewing of the video, it was explained to participants that the video depicted a crime and that the remaining study session tasks would pertain to their role as an eyewitness to the videotaped crime. They completed the pre-lineup confidence form (Appendix D) followed by the description form (Appendix E). In order to implement a delay between the description task and the recognition task, participants played a computer game for 20 minutes. Once this time had expired they were provided with another laptop computer that prompted them through the photographic lineup, confidence measures, and familiarity lineup. Next, a paper copy of the manipulation check form as well as a consent to keep data form[8] was provided to the participants to complete. Participants were verbally debriefed as to the nature of the study and provided the opportunity to ask questions. Also, they were provided with a debriefing form and some contact information to take with them. Participants received 1% in Session One and 1.5% in Session Two towards their final grade in select psychology courses at Carleton University.

---

= 6.96 days, *SD* = 1.34) and those who made an incorrect lineup decision (*M* = 7.28, *SD* = 1.31), $t(159) = -1.35$, $p = .18$, $d = -0.24$, 95% CI [-0.10, 0.52].

[8] Participants viewed the crime video under the guise that the study they were participating in was titled, "Goals, Dreams, and Vocabulary." Given that this title was misleading and that details regarding the eyewitness aspect of the study were withheld until after viewing the video, a Consent to Keep Data Form was provided to ensure that participants were willing to have their data included in the study after they were provided all details of the study purposes.

Both sessions were completed in a private room within a lab. Study tasks were completed individually without the presence of other participants or researchers (aside from the interview and word game in Session One). See Table 5 for design details within the procedure.

Table 5

*Study procedure per condition*

| | | More Familiar | Less Familiar | Stranger |
|---|---|---|---|---|
| Session One | *Administrator* | *Confederate (later culprit)* | *Confederate (later culprit)* | *Researcher* |
| | Step One | Consent, Demographics | Consent, Demographics | Consent, Demographics |
| | *Step Two* | *Interview (administrator sharing personal information)* with Word Game | *Interview (administrator not sharing personal information)* with Word Game | *Interview (sharing/not sharing personal information was counterbalanced)* with Word Game |
| Session Two | Administrator | Researcher | Researcher | Researcher |
| | Step One | Consent, Crime Video | Consent, Crime Video | Consent, Crime Video |
| | Step Two | Confidence, Description | Confidence, Description | Confidence, Description |
| | Step Three | Computer Game | Computer Game | Computer Game |
| | *Step Four* | *Lineup (TP or TA)* | *Lineup (TP or TA)* | *Lineup (TP or TA)* |
| | Step Five | Confidence, Familiarity | Confidence, Familiarity | Confidence, Familiarity |
| | Step Six | Manipulation Check, Consent, Debriefing | Manipulation Check, Consent, Debriefing | Manipulation Check, Consent, Debriefing |

## Results

### Manipulation Check

The goal in this study was to manipulate familiarity in such a manner that participants' subjective perceptions of familiarity with the target would vary across three levels: more familiar, less familiar, stranger. These three levels are thought to occur in the lower end of a known person-unknown person continuum. In order to determine whether the culprit was in fact perceived to be more or less familiar based on the study manipulations, participants completed a familiarity lineup in which they rated six photos on their level of familiarity. This task was always completed near the end of the study so as not to alert participants as to the study variables of interest. The photos included well-known celebrities such as Angelina Jolie (mean familiarity rating = 84.23, *SD* = 26.67) and Hilary Duff (*M* = 77.88, *SD* = 33.63), as well as the target (*M* = 69.97, *SD* = 35.15), a lineup filler (*M* = 22.34, *SD* = 32.15), a U.S. gymnast (*M* = 12.84, *SD* = 24.55), and a complete stranger for whom the participants had not previously seen (*M* = 5.02, *SD* = 15.37). As intended, overall the target fell somewhere between known persons (e.g., Angelina Jolie) and complete strangers (e.g., the never before seen person). Mean ratings across conditions are shown in Table 6.

Table 6

*Mean (SD) familiarity ratings of lineup members, across conditions*

| | Angelina Jolie | | Lineup Member | | Stranger | | Hilary Duff | | Target | | Gymnast | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Condition | TP | TA | TP | TA | TP | TA | TP | TA | TP | TA | TP | TA |
| More Familiar | 77.30 (29.91) | 88.47 (19.23) | 11.43 (24.38) | 24.34 (36.29) | 3.93 (14.99) | 1.63 (3.07) | 63.97 (41.16) | 87.69 (26.11) | 71.87 (29.19) | 88.16 (21.92) | 7.90 (17.17) | 14.53 (27.10) |
| Less Familiar | 84.67 (29.71) | 88.47 (22.85) | 15.37 (24.15) | 28.73 (37.59) | 5.50 (13.98) | 6.60 (17.09) | 81.87 (32.27) | 78.47 (33.87) | 81.03 (29.60) | 86.83 (25.34) | 13.03 (26.40) | 16.13 (28.50) |
| Stranger | 82.53 (31.01) | 83.17 (26.56) | 27.90 (34.17) | 26.90 (32.08) | 4.30 (13.43) | 5.23 (15.84) | 76.60 (34.77) | 78.03 (30.23) | 44.20 (36.74) | 46.50 (37.82) | 15.27 (26.59) | 10.07 (20.29) |

Familiarity ratings of the culprit across exposure conditions were examined in order to determine whether three levels of subjective familiarity were accomplished through the present study manipulations. Familiarity ratings of the culprit significantly differed across exposure conditions, $F(2, 179) = 28.95$, $p < .001$, partial $\eta^2 = 0.24$, 95% CI [0.15, 0.32]. Post hoc Tukey tests within the ANOVA indicate that the more familiar witnesses ($M = 80.27$, $SD = 26.77$) rated their perceptions of familiarity significantly higher than the stranger witnesses ($M = 45.35$, $SD = 35.15$, $p < .001$) and the less familiar witnesses ($M = 83.93$, $SD = 27.47$) rated their perceptions of familiarity significantly higher than the stranger witnesses, $p < .001$. There was no significant difference in the familiarity ratings of more and less familiar eyewitnesses, $p = .79$. Therefore these results suggest that, unlike the pilot study findings, participants in the main study did not significantly differ in their perceptions of familiarity across sharing vs. not sharing of personal information groups. It is noted that, while the conditions are labelled more familiar vs. less familiar in the following results section, the difference between these groups occur only in the fact that more familiar participants received personal information about the target whereas less familiar participants did not; the more/less familiarity distinction cannot be attributed to participants subjective experiences of familiarity.

**Hypothesis Testing**

Two types of memory were assessed in this study: recall memory and recognition memory. Recall of the culprit was assessed by asking participants to report, using an open-ended question, everything that they can remember about the culprit. Whereas recognition memory was assessed by having participants complete a photo identification

lineup. Recall memory will be discussed first followed by recognition memory. Moreover, a third variable that is often considered in a criminal investigation, confidence, was measured in the present study and will be discussed last.

**Recall Memory**

**Coding data.** Participants' responses to an open-ended statement requesting them to, "Please describe everything that you can remember regarding the culprit's appearance." were used to assess witness recall accuracy. Responses were coded for accuracy (1 = correct, 2 = incorrect) across 46 categories[9] of potential descriptors. Descriptions for age, height, and weight were deemed correct if they fell within a range (i.e., age = +/- 3 years, height = +/- 2 inches, weight = +/- 10 pounds). Intrusions also were recorded. A descriptor was considered an intrusion if an item was reported that was not present on the crime video, for example if the culprit was reported to have been wearing a scarf when no scarf was worn. All 184 response sets were coded by the primary researcher. In order to assess inter-rater reliability, forty randomly selected response sets (22%) were coded a second time by a trained undergraduate psychology student from Carleton University. In total, 537 comparisons were made and nine of these comparisons did not match, yielding 98% agreement. Cohen's kappa statistics were

[9] The 46 categories included: sex, race, skin colour, age, height, weight, build, hair colour, hair shade, hair length, hair style, hair thickness, presence of school bag, colours of school bag (3), pattern on school bag, presence of running shoes, brand of running shoes, colour of running shoes, pattern on running shoes, presence of jeans, colour of jeans, shade of jeans, condition of jeans, presence of t-shirt, colour of t-shirt, style of t-shirt, shade of t-shirt, eye colour, eye shape, presence of bracelets (2), colour of bracelets (2), location of bracelets (2), shape of nose, presence of necklace, shape of ears, shape of face, placement of cheek bones, shape of mouth, thickness of eyebrows, shape of eyebrows, and shape of chin.

computed to assess the inter-rater reliability for the two aggregate measures of recall accuracy: number of correct descriptors and number of incorrect descriptors. The resulting kappa indicated almost perfect agreement between coders for number of correct descriptors ($\kappa = 0.86$, $N = 40$) as well as number of incorrect descriptors ($\kappa = 0.84$, $N = 40$). All inconsistencies between coders were discussed and resolved.

**Cleaning the data**. The aggregate measure of number of descriptors was computed by summing the number of descriptors in each category. Overall the average number of descriptors was 12.22 items (*SD* = 3.99); the minimum number of descriptors was 3 and the maximum was 26. The proportion of correct descriptors was computed by dividing the number of correct descriptors by the total number of descriptors. Overall the average proportion of correct descriptors was 92% (*SD* = 9.02). Proportions ranged from 57% to 100%. The number of intrusions was determined by the number of descriptors that were reported that were not present in the crime video. Only 12 participants reported intrusions. As a result, any statistical comparisons using this dependent measure would involve a very small sample size; therefore, this measure of accuracy was not considered further.

Given that multivariate analysis of variance (MANOVA) was planned to be used to compare number of descriptors and proportion of correct descriptors across the three exposure conditions, univariate normality was assessed for the two dependent variables. For number of descriptors, skewness (skewness = 0.53, *SE* = 0.18) and kurtosis (kurtosis = 0.27, *SE* = 0.36) were in the normal range. However, for the proportion of correct descriptors, the distribution was negatively skewed (skewness = -1.40, *SE* = 0.18) and

leptokurtic (kurtosis = 2.03, *SE* = 0.36). *Z*-scores were computed and three outliers were identified (i.e., z-scores = -3.30, -3.58, -3.89) and removed. Normality was assessed for the new data set and the distribution was still negatively skewed (skewness = -1.09, *SE* = 0.18), although kurtosis was now in the normal range (kurtosis = 0.70, *SE* = 0.36). Examination of new *z*-scores indicated no outliers. It was determined that the cause of the negative skew was due to the high degree of proportional accuracy in participants' recall responses. Given that this natural pattern of results is logical and that MANOVA is robust to a skewed distribution such as this, the dataset was left as is. The new distribution ranged from 67% to 100% ($M = 92.81$, $SD = 8.05$).

**Recall Accuracy**

A one-way MANOVA was used to examine the influence of exposure condition on the number of culprit descriptors and the proportion of correct descriptors recalled by participants. It was hypothesized that witnesses in the more familiar condition would report more culprit descriptors and that their descriptions would be more accurate than those in the less familiar condition, and that both of the familiar conditions would report a greater number of descriptors and would be more accurate than witnesses in the stranger condition. The omnibus *F*-test indicated no significant differences between groups, Wilk's $\lambda = 0.99$, $F(4, 352) = 0.47$, $p = .76$, partial $\eta^2 = 0.01$, 95% CI [0.00, 0.01]. See Table 7 for descriptive statistics.

Table 7

*Mean (SD) number/proportion of descriptors across three exposure conditions*

| | More Familiar *N* = 62 | Less Familiar *N* = 60 | Stranger *N* = 58 |
|---|---|---|---|
| Total Descriptors | 12.40 (4.42) | 12.45 (3.47) | 11.91 (4.04) |
| Proportion Correct | 92.88 (7.61) | 93.58 (7.63) | 91.93 (8.93) |

Before moving on to recognition memory results, a logistic regression was run to determine whether recall accuracy predicts recognition accuracy. The dependent variable was recognition accuracy (correct or incorrect) and the predictors were total number of descriptors and proportion of correct descriptors. The model was significant, $\chi^2(2) = 6.62$, $p = .04$, Nagelkerke $R^2 = 0.05$. Results indicated that a greater number of person descriptors predicted recognition accuracy, Wald = 5.72, $p = .02$, OR = 1.12, 95% CI [1.02, 1.23] whereas the proportion of correct descriptors did not, Wald = 0.41, $p = .52$.

**Recognition Memory**

**Partitioning the data.** Half of the participants viewed a lineup with the target present while the other half viewed a lineup with the target absent. TP lineups mimic real-life situations in which the police include the guilty suspect in the lineup. In this case, the eyewitness's task is to choose this person from the lineup. On the other hand, TA lineups mimic real-life situations in which the police do not include the guilty suspect in the lineup (i.e., their suspect is innocent). In this case the eyewitness's task is to not choose any one from the lineup and to reject the lineup altogether. Since a TP versus a

TA lineup require different tasks from the eyewitness, that arguably vary in terms of decision making and social demands, accuracy rates for TP and TA lineups will first be compared and then they will be analyzed separately throughout the following analyses.

**Assumption checks.** Pearson chi-square tests were used to examine identification accuracy. There were no outliers as the data were categorical and there were no missing values. However, there are a few assumptions underlying the use of chi-square analysis. Use of chi-square requires data to be independent (e.g., the data cannot be repeated measures) and when contingency tables are greater than 2 x 2 (which they were for some analyses in this study), expected frequencies for each cell should exceed one and no less than 20% of expected cell frequencies should be less than five. These assumptions were met in all chi-square analyses discussed below.

**Spontaneous Recognition**

While completing the recall task (i.e., describing the culprit) participants were asked whether they met the culprit prior to having seen her in the crime video (i.e., yes or no). This question was posed prior to the completion of the identification lineup. A chi-square analysis was used to determine whether exposure condition influenced accuracy in spontaneous recognition of the culprit. Participants' responses were correct if they indicated "yes" when in the more or less familiar conditions or "no" when in the stranger condition. Alternatively, participants' responses were incorrect if they indicated "yes" when in the stranger condition or "no" when in the more or less familiar condition. Results indicated no significant differences across conditions for those who correctly relayed their previous exposure to the culprit (more familiar = 82% correct; less familiar

= 83% correct; stranger = 76% correct), $\chi^2(2, N = 180) = 1.04$, $p = .59$, Cramer's $V = 0.08$.

However, participants who were correct in stating whether they had or had not previously met the culprit were significantly more likely to accurately identify the culprit in the identification lineup (84%) than those who incorrectly stated whether they had or had not previously met the culprit (71%), $\chi^2(1, N = 180) = 3.95$, $p = .05$, $\phi = 0.15$.

**Recognition Accuracy**

A 2 (TP vs. TA) x 2 (correct vs. incorrect lineup decision) Pearson chi-square was conducted. Contradictory to our study hypothesis, participants in the target-present condition were not significantly more likely to correctly make an identification decision than participants in the target-absent condition, $\chi^2(1, N = 184) = 1.33$, $p = .25$, $\phi = -0.09$. Approximately 58% of eyewitnesses who were provided a TP lineup correctly identified the culprit, whereas approximately 64% of eyewitnesses who were provided a TA lineup did not chose anyone from the lineup (i.e., correctly rejected the lineup).

**Correct identifications (TP lineup).** A Pearson chi-square was used to compare correct identifications and incorrect identifications (i.e., filler identifications or incorrect rejections) across the three levels of target exposure (i.e., more familiar, less familiar, and stranger) for those who completed the TP lineup only. It was hypothesized that witnesses in the more familiar condition would be more likely to correctly identify the target than participants in the less familiar condition, and that participants in either of these conditions would be more likely to correctly identify the target than participants in the stranger condition.

Results of the Pearson chi-square suggest that identification accuracy differed across exposure conditions, $\chi^2(2, N = 92) = 26.15$, $p < .001$, Cramer's $V = 0.53$, see Table 8 for descriptive results. Follow up chi-squares on each pairwise comparison were performed with a Bonferroni correction ($\alpha = .017$). Contrary to study predictions, witnesses in the more familiar condition did not produce significantly more correct identifications than witnesses in the less familiar condition, $\chi^2(1, N = 62) = 1.17$, $p = .28$, OR = 0.45, 95% CI [0.10, 1.98]. However, participants in the more familiar condition did produce significantly greater correct identifications than participants in the stranger condition, $\chi^2(1, N = 61) = 13.95$, $p < .001$, OR = 8.33, 95% CI [2.58, 26.86] and participants in the less familiar condition produced significantly greater correct identifications than participants in the stranger condition, $\chi^2(1, N = 61) = 21.08$, $p < .001$, OR = 18.67, 95% CI [4.55, 76.62].

Table 8

*Proportion (n) of identification decisions across conditions for TP and TA lineups*

| | More Familiar | Less Familiar | Stranger |
|---|---|---|---|
| Target Present | | | |
| Correct Identification | 80.65 (25) | 90.32 (28) | 33.33 (10) |
| Foil Identification | 12.90 (4) | 3.23 (1) | 10.00 (3) |
| Incorrect Rejection | 6.45 (2) | 6.45 (2) | 56.67 (17) |
| Target Absent | | | |
| Correct Rejection | 78.13 (25) | 83.33 (25) | 66.67 (20) |
| Foil Identification | 6.25 (2) | 0.00 (0) | 26.67 (8) |
| False Identification | 15.63 (5) | 16.67 (5) | 6.67 (2) |

**Correct rejections (TA lineup).** Pearson chi-square also was used to examine correct rejections (compared to filler and false identifications) across the three levels of target exposure for those who completed the TA lineup. It was hypothesized that witnesses in the more familiar condition would be significantly more likely to produce correct rejections than witnesses in the less familiar condition, and that witnesses in either of these conditions would be significantly more likely to correctly reject a lineup than witnesses in the stranger condition.

Contrary to study predictions, results indicated that correct rejection rates did not significantly differ across exposure conditions, $\chi^2(2, N = 92) = 2.40$, $p = .30$, Cramer's $V = 0.16$ (see Table 8 for descriptive statistics).

**Diagnosticity ratios.** Wells and Lindsay (1980) introduced the diagnosticity ratio (DR) as a measure of the degree to which a lineup procedure is diagnostic of the suspect's guilt. The DR is computed by dividing the probability that a guilty suspect is chosen from a TP lineup (i.e. correct identification) by the probability that an innocent suspect is chosen from a TA lineup (i.e. false identification) (i.e., p(IDS/(S = C))/p(IDS/(S ≠ C)). In the context of the present study, the DR may be used to determine whether familiar witnesses' identification decisions are more indicative of guilt than stranger witnesses' identification decisions. Results indicated that witnesses in the more familiar condition were five times more likely to correctly identify the guilty suspect than to identify the innocent suspect (DR = 5.16). Witnesses in the less familiar condition also were five times more likely to correctly identify the guilty suspect than to identify the innocent suspect (DR = 5.42). Lastly, witnesses in the stranger condition were five times more likely to identify the guilty suspect than to identify the innocent suspect (DR = 5.00).

**Choosers.** A Pearson chi-square was conducted to determine whether exposure condition influenced choosing behaviour. Results indicated a trend toward significant differences in choosing rates across exposure conditions, $\chi^2(2, N = 184) = 5.80$, $p = .06$, Cramer's $V = 0.18$. Fifty-seven percent of participants in the more familiar condition chose someone from the lineup and 57% of participants in the less familiar condition chose someone from the lineup, whereas 38% of participants in the stranger condition made a lineup choice.

**Confidence**

Confidence was measured on three occasions throughout the study. Immediately after watching the crime video, during the identification lineup, and after completing the identification lineup.

The first measure of confidence was taken immediately after the participant viewed the crime video (in Session 2) and prior to completing any other forms. Participants were asked to rate their confidence in their ability to describe the culprit’s appearance, the accuracy of their description of the culprit, their ability to recognize the culprit, and the accuracy of their recognition of the culprit. The ratings were done using a 100-point scale (0 *not at all confident;* 100 *very confident*). A bivariate correlation was run on all four measures of confidence (see Table 9). Confidence in ability and accuracy of description of the culprit were highly correlated ($r = 0.82$) and confidence in ability and accuracy of recognition of the culprit were highly correlated ($r = 0.93$). Therefore, the four measures of confidence were collapsed into two by averaging across the correlated measures.

Table 9

*Correlations between confidence measures*

| | Confidence in description of culprit | Confidence in accuracy of description of culprit | Confidence in recognition of culprit | Confidence in accuracy of recognition of culprit |
|---|---|---|---|---|
| Confidence in description of culprit | 1.00 | 0.82* | 0.64* | 0.67* |
| Confidence in accuracy of description of culprit | --- | 1.00 | 0.64* | 0.68* |
| Confidence in recognition of culprit | --- | --- | 1.00 | 0.93* |
| Confidence in accuracy of recognition of culprit | --- | --- | --- | 1.00 |

*$p < .001$

**Data cleaning.** The new composite variables for description confidence and recognition confidence were examined for normality. Description confidence ratings ranged from 0 to 100 with a mean of 76.17 and standard deviation of 15.63. The distribution was negatively skewed (skewness = -1.13, $SE = 0.18$, $p < .001$) and leptokurtic (kurtosis = 2.76, $SE = 0.36$, $p < .001$). *Z*-scores were computed and analyzed. An outlier was identified with a *z*-score of -4.87. The description confidence score for this participant was zero. This data point was deleted and normality statistics were re-run. The new mean was 76.59 ($SD = 14.61$) with scores ranging from 30 to 100. The

skewness (skewness = 0.69, *SE* = 0.18) and kurtosis (kurtosis = 0.46, *SE* = 0.36) were now in the normal range.

The same procedure was conducted with the new composite variable for recognition confidence. Scores for this variable ranged from 0 to 100 and the mean was 80.11 (*SD* = 18.88). The distribution was negatively skewed (skewness = -1.30, *SE* = 0.18, $p < .001$) and leptokurtic (kurtosis = 2.00, *SE* = 0.36, $p < .001$). *Z*-scores indicated two outliers; with *z*-scores of -4.24 and -3.98 and raw scores of 0 and 5 respectively. These scores were deleted and tests were re-run. The new data set ranged from 35 to 100 (*M* = 80.97, *SD* = 17.11; skewness = 0.87, *SE* = 0.18; kurtosis = 0.21, *SE* = 0.36).

The second measure of confidence (i.e., lineup confidence) was taken during the sequential photographic lineup. Participants were required to make a decision about each lineup member one at a time; they were not permitted to move on from the photo until they decided whether the lineup member was the culprit. Once they made their decision by clicking yes or no, they were directed to a follow up page asking them how confident they were in their decision. This was done for every lineup member. An average lineup confidence rating was computed to assess participants overall lineup confidence. TP lineups and TA lineups were considered separately.

For TP lineups, confidence ratings ranged from 0 to 100 (*M* = 84.92, *SD* = 18.49) and the distribution was negatively skewed (skewness = -2.53, *SE* = 0.25, $p < .001$) and leptokurtic (kurtosis = 8.59, *SE* = 0.50, $p < .001$). *Z*-scores were computed and two outliers were identified with z-scores of -4.59 and -4.56 and raw scores of 0.00 and 0.67 respectively. These scores were deleted and *z*-scores were computed again. A hidden

outlier was identified with a z-score of -4.05 and raw score of 31.67. Once this outlier was deleted the distribution was found to be within the normal range (17.17 to 100; $M$ = 87.42, $SD$ = 12.35; skewness = -1.02, $SE$ = 0.26; kurtosis = 0.64, $SE$ = 0.51).

For TA lineups, confidence ratings ranged from 17 to 100 ($M$ = 85.17, $SD$ = 14.71) and the distribution was negatively skewed (skewness = -1.54, $SE$ = 0.25, $p < .001$) and leptokurtic (kurtosis = 4.06, $SE$ = 0.50, $p < .001$). $Z$-scores indicated that there was one outlier (z-score = -4.64, raw score = 17.00); this score was deleted. The new distribution was within normal range (48.67 to 100; $M$ = 85.92, $SD$ = 12.90; skewness = -0.81, $SE$ = 0.25; kurtosis = 0.04, $SE$ = 0.50).

Lastly, immediately after completing the photographic lineup participants were asked to rate their willingness to testify to their lineup decision in a court of law (0 *not at all*; 100 *absolutely*). For TP lineups, willingness to testify ratings ranged from 0 to 100 ($M$ = 77.17, $SD$ = 28.38) and the distribution was negatively skewed (skewness = -1.38, $SE$ = 0.25, $p < .001$; kurtosis statistic = 0.96, $SE$ = 0.50). Z-scores were examined; however, no outliers were identified. For TA lineups, willingness to testify ratings ranged from 0 to 100 ($M$ = 65.98, $SD$ = 30.65) and the distribution was normal (skewness = -0.73, $SE$ = 0.25; kurtosis statistic = -0.54, $SE$ = 0.50).

**Pre-confidence.** Eyewitnesses' confidence in their ability to describe the culprit did not significantly differ by exposure condition $F(2, 179) = 2.60$, $p = .08$, partial $\eta^2 = 0.03$, 95% CI [0.00, 0.07]. However, eyewitnesses' confidence in their ability to recognize the culprit did significantly differ by exposure condition, $F(2, 178) = 9.76$, $p < .001$, partial $\eta^2 = 0.10$, 95% CI [0.10, 0.17]. Post hoc Tukey tests run within the ANOVA

indicated that eyewitnesses in the more familiar condition were significantly more confident than eyewitnesses in the stranger condition, $p < .01$, and that eyewitnesses in the less familiar condition were significantly more confident than eyewitnesses in the stranger condition, $p < .001$. See Table 10 for descriptive results.

Table 10

*Mean (SD) confidence in ability to describe and recognize the culprit across exposure conditions*

| Exposure Condition | Confidence in Description | Confidence in Recognition |
|---|---|---|
| More Familiar | 77.13 (14.87) | 82.46 (16.34) |
| Less Familiar | 79.26 (13.10) | 86.56 (14.33) |
| Stranger | 73.33 (15.39) | 73.64 (18.17) |

**Lineup confidence.** Eyewitnesses' average confidence rating in their lineup decisions significantly differed across exposure conditions. In TP lineups ($F(2, 86) = 5.55$, $p < .01$, partial $\eta^2 = 0.11$, 95% CI [0.02, 0.21]) more familiar eyewitnesses were significantly more confident than stranger eyewitnesses ($p = .03$) and less familiar eyewitnesses were significantly more confident in their lineup decision than stranger eyewitnesses ($p < .01$). Similarly, in the TA lineups ($F(2, 88) = 8.57$, $p < .001$, partial $\eta^2 = 0.16$, 95% CI [0.05, 0.27]) more familiar eyewitnesses were significantly more confident than stranger eyewitnesses ($p < .01$) and less familiar eyewitnesses were significantly more confident than stranger eyewitnesses ($p < .01$). See Table 11 for descriptive results.

Table 11

*Mean (SD) confidence in lineup decision across exposure conditions*

| Condition | Confidence in TP Lineup ID | Confidence in TA Lineup ID |
|---|---|---|
| More Familiar | 89.52 (11.06) | 89.21 (11.96) |
| Less Familiar | 90.87 (10.06) | 89.90 (9.54) |
| Stranger | 81.35 (14.01) | 78.54 (13.90) |

**Willingness to testify.** ANOVA was used to examine whether participant willingness to testify was influenced by participants' exposure to the culprit (more familiar, less familiar, stranger), identification lineup (TP, TA), and identification decision accuracy (correct, incorrect). There was no main effect of exposure condition, $F(2, 172) = 1.63$, $p = .20$, partial $\eta^2 = 0.02$, 95% CI [0.00, 0.06], or lineup condition, $F(1, 172) = 0.05$, $p = .82$, partial $\eta^2 < 0.01$, 95% CI [0.00, 0.01]. However, there was a main effect of identification accuracy, $F(1, 172) = 5.80$, $p = .02$, partial $\eta^2 = 0.03$, 95% CI [0.00, 0.09]. Participants who made a correct identification decision were significantly more willing to testify ($M = 75.26$, $SD = 28.65$) than participants who made an incorrect identification decision ($M = 61.98$, $SD = 31.51$), $t(182) = 2.74$, $p < .01$, $d = 0.44$, 95% CI [0.11, 0.70].

This main effect is qualified by a significant interaction between exposure and identification accuracy, $F(2, 172) = 3.48$, $p = .03$, partial $\eta^2 = 0.04$, 95% CI [0.00, 0.09].

Post hoc *t*-tests with a Bonferroni correction ($\alpha = .017$) indicate more familiar eyewitness were significantly more willing to testify when their identification decision was correct ($M = 83.98$, $SD = 22.80$) than when their decision was incorrect ($M = 55.15$, $SD = 34.29$), $t(61) = 3.64$, $p < .01$, $d = 0.99$, 95% CI [0.39, 1.43]. On the other hand, there were no significant differences when eyewitnesses were less familiar with the culprit (correct ($M = 76.06$, $SD = 29.90$), incorrect ($M = 70.00$, $SD = 33.25$), $t(59) = 0.53$, $p = .60$, $d = 0.14$, 95% CI [-0.37, 0.64]) or strangers to the culprit (correct ($M = 59.30$, $SD = 29.33$), incorrect ($M = 62.80$, $SD = 30.31$), $t(58) = -0.46$, $p = .65$, $d = 0.12$, 95% CI [-0.39, 0.62]). Also, there was a significant interaction between lineup condition and accuracy, $F(2, 172) = 7.19$, $p < .01$, partial $\eta^2 = 0.04$, 95% CI [0.02, 0.14]. Post hoc *t*-tests with a Bonferroni correction ($\alpha = .025$) indicated that participants who received a TP lineup were significantly more willing to testify when they correctly identified the culprit ($M = 85.06$, $SD = 21.02$) than when they made an incorrect decision (i.e., identified a filler or incorrectly rejected the lineup, $M = 60.03$, $SD = 34.56$), $t(90) = 4.29$, $p < .001$, $d = 0.88$, 95% CI [0.46, 1.32]. However, there were no significant differences when participants were presented with a TA lineup, correct $M = 66.43$, $SD = 31.73$, incorrect $M = 64.55$, $SD = 27.55$, $t(90) = 0.25$, $p = .80$, $d = 0.05$, 95% CI [-0.36, 0.46]. There was no significant interaction between exposure condition and lineup condition, $F(2, 172) = 0.87$, $p = .42$, partial $\eta^2 = 0.01$, 95% CI [0.00, 0.04]. The three-way interaction was not significant, $F(2, 172) = 2.41$, $p = .09$, partial $\eta^2 = 0.03$, 95% CI [0.00, 0.07]. Descriptive statistics are available in Table 12.

Table 12

*Participants' mean (SD) willingness to testify in a court of law regarding their lineup decision, across exposure conditions*

| | Target Present | | Target Absent | |
|---|---|---|---|---|
| | Correct | Incorrect | Correct | Incorrect |
| More Familiar | 91.92 (11.30) | 41.50 (41.20) | 76.04 (28.33) | 66.86 (24.29) |
| *N* | 25 | 6 | 25 | 7 |
| Less Familiar | 86.04 (20.45) | 55.00 (39.05) | 64.88 (34.92) | 79.00 (30.03) |
| *N* | 28 | 3 | 25 | 5 |
| Stranger | 65.20 (29.60) | 66.35 (31.60) | 56.35 (29.50) | 55.70 (27.75) |
| *N* | 10 | 19 | 19 | 9 |

**Confidence-accuracy relationship.** The "willingness to testify" confidence measure was used to compute confidence-accuracy relationships given that this measure most closely resembles confidence measures used by previous researchers (i.e., a request for confidence immediately upon completion of the identification lineup). It was predicted that participants in the more familiar condition would have a higher CA relationship than those in the less familiar condition, and that participants in the stranger condition would have the smallest CA relationship.

All CA relationships were computed using point biserial correlations. The overall CA correlation was significant, $r(184) = 0.20$, $p < .01$, 95% CI [0.06, 0.33]. However, the CA relationship for witnesses in the more familiar condition was $r(63) = 0.42$, $p <$

.01, 95% CI [0.20, 0.60], in the less familiar condition was $r(61) = 0.07$, $p = .60$, 95% CI [-0.18, 0.31], and in the stranger condition was $r(60) = -0.06$, $p = .65$, 95% CI [-0.31, 0.19]. Fisher's *z* transformations were conducted to determine whether these correlations were significantly different. The CA relationship for more familiar witnesses was significantly greater than the CA relationship for less familiar witnesses ($p = .04$) and stranger witnesses ($p = .04$). There was no significant difference between the CA relationship for less familiar witnesses and stranger witnesses ($p = .96$)

Further analyses were conducted with choosers only. The overall CA relationship for those who picked someone from the identification lineup was $r(94) = 0.36$, $p < .001$, 95% CI [0.17, 0.52]. Similar to the overall CA relationship results, more familiar eyewitnesses' CA relationship was $r(36) = 0.60$, $p < .001$, 95% CI [0.35, 0.77], less familiar witnesses was $r(35) = 0.14$, $p = .41$, 95% CI [-0.19, 0.44], and stranger witnesses $r(23) = 0.07$, $p = .74$, 95% CI [-0.33, 0.45]. The CA relationship for more familiar witnesses was significantly greater than the CA relationship for less familiar witnesses ($p = .02$) and stranger witnesses ($p = .03$). There was no significant difference between the CA relationship for less familiar and stranger witnesses ($p = .80$).

## Follow Up Study

Given that the Main Study yielded very high identification accuracy rates for more and less familiar witnesses, a follow up study was conducted. The purpose of the follow up study was to determine whether an extended period of time between an initial identification lineup and a second identification lineup would decrease accuracy and, if

so, whether differences would occur across more familiar witnesses, less familiar witnesses, and stranger witnesses.

Eyewitness research examining the influence of repeated or extended exposure duration and retention interval on identification accuracy has produced mixed results. For example, Cain, Baker-Ward, and Eaton (2005) introduced preschoolers (19 mos. to 5-years-old) to a caregiver for seven weeks. The children’s identification accuracy was then assessed utilizing a TP simultaneous lineup both one week following the exposure as well as three months later. While older children were more accurate than younger children, there was no significant effect of delay in identification accuracy. Alternatively, a non-significant pattern in the data suggests that accuracy improved over time. The authors caution, though, that this result may be a lineup practice effect, given that the design was repeated measures. However, a study by Cutler, Penrod, and Martens (1987) examining varying retention intervals produced similar counterintuitive results. In their study, participants watched a videotaped staged robbery where they were exposed to the target for 30 or 75 seconds. While varying exposure durations did not produce significant results, first year psychology students were better able to identify the target after a one week delay than they did 30 minutes following exposure (Cutler et al., 1987).

On the other hand, there are many researchers who have reported an opposite and, some would say, more expected finding; that recognition accuracy decreases with time (e.g., Deffenbacher, Bornstein, McGorty, & Penrod, 2008). In archival studies, witnesses have been found to be less likely to make an identification as the delay between the time of the crime and the identification lineup increases (e.g., Tollestrup, Turtle & Yuille,

1994). For example, in Tollestrup et al.’s (1994) examination of robbery cases, they found that a suspect was more likely to be chosen from the lineup 0-1 days after the crime (78% of the time), followed by after 4 days (46%), after 19 days (33%), and after 120 days (14%). These findings coincide with results of a meta-analysis conducted by Shapiro and Penrod (1986). In their analysis of 18 face recognition and eyewitness studies in which retention interval (i.e., the amount of time between exposure/study and identification) was examined, they found that shorter retention intervals are associated with greater accuracy than longer retention intervals. A more recent meta-analysis by Deffenbacher and colleagues (2008) conducted specifically on the effect of retention interval and identification accuracy found similar results.

These findings, along with others that indicate a negative correlation of identification accuracy over time (e.g., Shapiro & Penrod, 1986), are in-line with Ebbinghaus’ (1949) well documented forgetting curve. The forgetting curve encompasses the idea that things are forgotten over time, exponentially. While the forgetting curve itself is well-known, the exact fit of the forgetting curve with various stimuli is unknown (Wells & Quinlivan, 2009). For example, we do not know how many days it takes for a face to be completely forgotten. Furthermore, does the forgetting curve shift when the face is familiar?

**Hypotheses**

The following hypotheses were made concerning the influence of familiarity on recognition accuracy and confidence, after an extended delay:

F4. Participants in the more familiar condition will be more likely to correctly identify the target than participants in the less familiar conditions, and both the more/less familiar condition participants will be more likely to correctly identify the target than participants in the stranger condition.

F5. a) Eyewitnesses in the more familiar condition will rate their confidence highest across all confidence measures (i.e., pre-lineup confidence measure, lineup confidence, and post-lineup confidence measure), followed by those who are less familiar with the culprit, and lastly eyewitnesses for whom the culprit is a stranger.

F5. b) The CA relationship will be highest for those in the more familiar condition, followed by those in the less familiar condition, and lastly those in the stranger condition.

## Method

### Participants

In the main study, 132 participants made a correct identification decision (i.e., correctly identified the culprit from a TP lineup or correctly rejected a TA lineup). These 132 participants were contacted approximately 14 months later ($M$ = 14.06, $SD$ = 2.85) to examine their identification accuracy after a long delay. Of the original 132, six participants were not reachable due to an incorrect email address. As a result, 126 participants were invited to participate in the follow up study and 85 participated; yielding a 67% response rate.

Each participant was asked to report whether they had any exposure to the culprit (confederate) in the time between the Main Study and the Follow Up Study. Ten

participants responded yes or did not provide an answer to this question; their data were removed due to the fact that their responses may be influenced by their additional (not controlled) exposure to the culprit. A secondary manipulation check took place given that the confederate ran a study (with potentially overlapping participants) within the year and a half time delay. Consent forms from the confederates study were cross-checked with the participants of the follow up study to ensure that there was no overlap (and, therefore, no exposure with the confederate). The overlap that did occur was accounted for by participants who reported having seen the culprit during the delay; therefore, no additional participants were removed from the data analysis. The final sample was 75 participants (*M*age = 20.47, *SD* = 5.39, 75% Women, 54% White).

**Study design.** A 3 (familiarity with the culprit: more familiar vs. less familiar vs. stranger) x 2 (first photographic lineup type: target-present vs. target-absent) between-subjects factorial design was used; participants were considered to hold membership in the same condition in which they were randomly assigned during the Main Study. The primary dependent measures were recognition accuracy (i.e., proportion of correct identifications in target-present lineups) and confidence prior to completing the lineup task, during the lineup task, and upon completion of the lineup task.

**Materials**

**Pre-lineup confidence form.** Participants were asked to rate how confident they were in their ability to recognize the culprit who they saw in a videotaped crime approximately one year ago using a 100-point rating scale (0 *not confident*; 100 *very confident*).

**Photographic lineup**. The photo lineup used in the Main Study was used in the follow up. The sequential lineup was always target-present and the location of the target was randomly determined. The amount of time participants' took to make their lineup decision was recorded and they were asked to rate how confident they were in their decision following each lineup decision (0 *not confident*; 100 *very confident*).

**Willingness to testify.** Upon completion of the sequential lineup, participants were asked to rate their willingness to testify to their identification decision in a court of law using a 100-point scale (0 *not willing*; 100 *very willing*). This measure was used as a proxy measure for overall confidence in identification decision.

**Manipulation check.** Participants were asked to report whether they had any exposure to the culprit /confederate since participating in the Main Study.

**Procedure**

After ethical approval was obtained from Carleton University, students were recruited to participate via email. Participants wishing to complete the follow up study were prompted to click on a link to the online study. They were asked to complete an informed consent form. Then, they were invited to complete the confidence form, the identification task, post-identification form, and manipulation check. Lastly, participants were directed to a debriefing form that provided them details on the follow up study's purpose and contact information should they have any questions or concerns. Participants received a $5 gift card for participation in this study.

## Results

### Recognition Memory

**Assumption checks.** Pearson chi-square tests were used to examine identification accuracy. There were no outliers as the data were categorical and there were no missing values. However, there are a few assumptions underlying the use of chi-square analysis. Use of chi-square requires data to be independent (e.g., the data cannot be repeated measures) and when contingency tables are greater than 2 x 2 (which they are for some analyses in this study), expected frequencies for each cell should exceed one and no less than 20% of expected cell frequencies should be less than five. These assumptions were met in all chi-square analyses discussed below.

### Recognition Accuracy

A 2 (TP vs. TA) x 2 (correct vs. incorrect lineup decision) Pearson chi-square was conducted. Participants who were previously provided a TP lineup were not significantly more likely to correctly make an identification decision than participants who previously received a TA lineup, $\chi^2(1, N = 75) = 1.83$, $p = .18$, $\phi = 0.16$. Approximately 61% of eyewitnesses who were previously provided a TP lineup correctly identified the culprit, whereas approximately 45% of eyewitnesses who were previously provided a TA lineup correctly identified the culprit in this follow up study. Given that there were no significant differences here and the sample size was small, participants who were initially provided a TP or a TA lineup in the Main Study were combined for future analyses.

**Correct identifications.** A Pearson chi-square was used to compare correct identifications and incorrect identifications (i.e., foil identifications or incorrect

rejections) across the three levels of target exposure (i.e., more familiar, less familiar, and stranger). Results of the Pearson chi-square suggest that identification accuracy did not significantly differ across exposure conditions, $\chi^2(2, N = 75) = 3.93$, $p = .14$, Cramer’s $V = 0.23$, see Table 13 for descriptive results. Given that all of these participants would have provided a correct response when presented with the identification lineup in the Main Study, probabilities presented in Table 13 represent the percent decrease in accuracy from the initial lineup and the delayed lineup. For example, 48% of more familiar witnesses who initially provided a correct lineup response provided an incorrect lineup response approximately 14 months later, 38% of the less familiar witnesses who initially provided a correct lineup response provided an incorrect lineup response approximately 14 months later, and 67% of the stranger witnesses who initially provided a correct lineup response provided an incorrect lineup response approximately 14 months later. Inferential analysis of these results would be the same as the initial chi-square, indicating that identification accuracy following a 14 month delay did not significantly differ amongst exposure conditions.

Table 13

*Proportion (n) of identification decisions across exposure conditions*

| | More Familiar *N* = 25 | Less Familiar *N* = 32 | Stranger *N* =18 |
|---|---|---|---|
| Correct Identification | 52.00 (13) | 62.50 (20) | 33.33 (6) |
| Filler Identification | 36.00 (9) | 18.75 (6) | 38.89 (7) |
| Incorrect Rejection | 12.00 (3) | 18.75 (6) | 27.78 (5) |

**Confidence**

Confidence was measured on three occasions throughout the study. Before completing the identification lineup, during the identification lineup, and after completing the identification lineup.

**Pre-confidence.** Eyewitnesses' confidence in their ability to recognize the culprit did not significantly differ by exposure condition $F(2, 67) = 0.61$, $p = .55$, partial $\eta^2 = 0.02$, 95% CI [0.00, 0.08]. Mean confidence ratings were similar across more familiar ($M = 49.42$, $SD = 31.98$), less familiar ($M = 47.17$, $SD = 31.62$), and stranger ($M = 38.88$, $SD = 26.08$) conditions.

Pre-confidence reported in the Main Study was compared to pre-confidence reported in the Follow Up study using a paired-samples *t*-test. Results indicated that pre-lineup confidence ratings in the Main Study ($M = 86.52$, $SD = 15.61$) were significantly higher than pre-lineup confidence ratings in the Follow Up Study ($M = 46.28$, $SD = 30.58$, $t(68) = 11.03$, $p < .001$, $d = 1.33$, 95% CI [1.00, 1.65]. Also, a mixed-model

ANOVA was performed to determine whether changes in confidence (within subjects factor) differed across exposure conditions (between subjects factor). Results indicated no significant difference in changes in pre-lineup confidence across exposure conditions, Wilk's $\lambda = 0.99$, $F(2, 66) = 0.46$, $p = .64$, partial $\eta^2 = .01$, 95% CI [0.00, 0.07]. Descriptive results are available in Table 14.

Table 14

*Mean (SD) pre-lineup confidence ratings from Main Study and Follow Up*

| Pre-lineup Confidence Rating | More Familiar *N* = 23 | Less Familiar *N* = 30 | Stranger *N* = 16 |
|---|---|---|---|
| Main Study | 84.35 (18.05) | 87.50 (13.57) | 81.88 (15.90) |
| Follow Up | 50.26 (32.43) | 47.17 (31.62) | 38.88 (26.08) |

*Note*. Means (*SD*) reported in this table are based on the means obtained from the repeated measures analysis. They may differ slightly from means reported previously or in text due to the need to "match" data in the repeated design (i.e., some data points may have been removed for this analysis due to missing data).

**Lineup confidence.** Eyewitnesses' average confidence rating in their lineup decisions did not significantly differ across exposure conditions, $F(2, 72) = 2.57$, $p = .08$, partial $\eta^2 = 0.07$, 95% CI [0.00, 0.16]. Participants in the more familiar ($M = 70.48$, $SD = 27.68$), less familiar ($M = 70.72$, $SD = 27.98$), and stranger ($M = 53.35$, $SD = 29.19$) conditions yielded similar lineup confidence ratings.

Lineup confidence reported in the Main Study was compared to lineup confidence reported in the Follow Up Study using a paired-samples *t*-test. Results indicated that eyewitnesses' lineup confidence in the Main Study ($M = 89.02$, $SD = 11.25$) was

significantly higher than eyewitnesses' lineup confidence in the Follow Up Study ($M$ = 58.46, $SD$ = 32.98), $t(73) = 8.74$, $p < .001$, $d = 1.02$, 95% CI [0.73, 1.29]. Also, a mixed-model ANOVA was used to determine whether the decrease in confidence (within subjects) differed across exposure conditions (between subjects). Results indicated a trend in the data, Wilk's $\lambda = 0.93$, $F(2, 71) = 2.50$, $p = .09$, partial $\eta^2 = 0.07$, 95% CI [0.00, 0.16]. Post hoc Tukey tests within the ANOVA indicate that there was a significantly greater decrease in lineup confidence for more familiar eyewitnesses than stranger eyewitnesses, $p = .01$, as well as a significantly greater decrease in lineup confidence for less familiar eyewitnesses than stranger eyewitnesses, $p = .03$. However, there was no significant difference between more and less familiar eyewitnesses, $p = .80$. Descriptive results are reported in Table 15.

Table 15

*Mean (SD) lineup confidence in the Main Study and Follow Up, across exposure conditions*

| | Main Study Lineup Confidence | Follow Up Study Lineup Confidence |
|---|---|---|
| More Familiar | 91.15 (11.56) | 67.50 (28.90) |
| Less Familiar | 90.32 (10.21) | 61.91 (33.59) |
| Stranger | 83.87 (11.60) | 40.28 (31.56) |

*Note*. Means (*SD*) reported in this table are based on the means obtained from the repeated measures analysis. They may differ slightly from means reported previously or in text due to the need to "match" data in the repeated design (i.e., some data points may have been removed for this analysis due to missing data).

**Willingness to testify.** An ANOVA was used to examine whether participant willingness to testify was influenced by their exposure to the culprit (more familiar, less familiar, stranger) and accuracy (correct, incorrect). There was a main effect of identification accuracy, $F(1, 64) = 25.83$, $p < .001$, partial $\eta^2 = 0.29$, 95% CI [0.14, 0.42]. Participants who made a correct identification decision were significantly more willing to testify ($M = 63.53$, $SD = 34.88$) than participants who made an incorrect identification decision ($M = 22.38$, $SD = 23.90$). There was no main effect of exposure condition, $F(2, 64) = 0.31$, $p = .74$, partial $\eta^2 = 0.01$, 95% CI [0.00, 0.06], or interaction, $F(2, 64) = 1.02$, $p = .37$, partial $\eta^2 = 0.03$, 95% CI [0.00, 0.11]. Descriptive results are available in Table 16.

Table 16

*Participants' mean (SD) willingness to testify in a court of law regarding their lineup decision, across exposure conditions*

| | Correct | Incorrect |
|---|---|---|
| More Familiar | 73.08 (30.96) | 18.36 (22.72) |
| *N* | 13 | 11 |
| Less Familiar | 60.58 (37.79) | 25.45 (26.01) |
| *N* | 19 | 11 |
| Stranger | 52.17 (33.68) | 23.40 (24.70) |
| *N* | 6 | 10 |

A paired-samples *t*-test was used to compare willingness to testify ratings from the Main Study to the Follow Up Study. Results indicated that participants' willingness to testify was significantly higher in the Main Study ($M = 76.44$, $SD = 28.19$) than it was in the Follow Up Study ($M = 44.71$, $SD = 36.54$), $t(69) = 7.09$, $p = .00$, $d = 0.85$, 95% CI [0.57, 1.12]. A mixed-model ANOVA was run with exposure condition (between-subjects) and willingness to testify (repeated measures). Results indicated no significant interaction between exposure condition and willingness to testify, Wilk's $\lambda = 0.96$, $F(2, 67) = 1.39$, $p = .26$, partial $\eta^2 = 0.04$, 95% CI [0.00, 0.12].

***Confidence-accuracy relationship.*** Point-biserial correlations were computed overall, and across exposure conditions. The overall CA relationship was $r(70) = 0.57$, $p < .001$, 95% CI [0.39, 0.71]. The CA relationship for witnesses in the more familiar condition was $r(24) = 0.72$, $p < .001$, 95% CI [0.46, 0.87], in less familiar condition was $r(30) = .46$, $p = .01$, 95% CI [0.13, 0.70], and in the stranger condition was $r(16) = 0.47$, $p = .07$, 95% CI [0.01, 0.77]. However, when examining the differences between these correlations using Fisher's *z* transformations, correlations were not significantly different between exposure groups ($p > .05$).

When considering only participants who chose someone from the lineup, the overall CA relationship was $r(57) = 0.53$, $p < .001$, 95% CI [0.32, 0.69]. The CA relationship for those in the more familiar condition was $r(21) = 0.75$, $p < .001$, 95% CI [0.49, 0.89], the less familiar condition was $r(25) = 0.31$, $p = .13$, 95% CI [-0.08, 0.62] and in the stranger condition was $r(11) = 0.44$, $p = .18$, 95% CI [-0.15, 0.80]. The CA relationship was significantly greater in the more familiar witnesses than it was in the less

familiar witness ($p$ = .04); however, there were no significant differences in the CA relationship for those in the more familiar and stranger conditions ($p$ = .24), or those in the less familiar and stranger conditions ($p$ = .71).

## Discussion

The present program of research provides a novel and preliminary examination of "familiar-stranger" eyewitness accuracy in order to fill a gap in eyewitness literature and to provide some much needed insight for criminal investigators. The bulk of empirical studies to date have examined the recall and recognition accuracy of eyewitnesses who have not previously seen the culprit, i.e., stranger eyewitnesses. This extant research has informed the criminal justice system and has resulted in many advances to procedures and protocols used when handling eyewitnesses. The present study results do not discount the value of this previous literature; however, it does draw attention to the risks of generalizing research conducted with stranger eyewitness to *all* eyewitnesses. Archival research suggests that witnesses who are previously acquainted with the defendant are frequently encountered in the criminal justice system and, in fact, they may be more frequently called upon in criminal prosecutions than stranger eyewitnesses (e.g., Flowe et al., 2011). In the interest of continuing the progressive and thorough nature of eyewitness research to date, it is of utmost importance to consider the similarities and differences between stranger and familiar-stranger eyewitnesses in order to ensure that we are providing eyewitnesses and criminal investigators with the best approaches to collecting accurate and reliable evidence.

In this first known attempt to explore the recall and recognition accuracy of familiar-stranger eyewitnesses, three levels of familiarity were examined: 1) a more familiar stranger was operationally defined as a witness who had a previous 30-minute interaction with the culprit and was provided personal information about the culprit, 2) a less familiar stranger was operationally defined as a witness who had a previous 30-minute interaction with the culprit, and 3) these familiar-stranger groups were compared to a control (stranger) group who did not have any previous contact with the culprit. The purpose of the studies was to examine whether recall accuracy measured in an open-ended description, recognition accuracy using a sequential photo lineup, and confidence measured before, during, and after the photo identification lineup differed across these groups. Results from the present studies suggest complex findings in regards to familiarity influences on eyewitness recall accuracy, recognition accuracy, and confidence.

**The Concept of Familiarity**

When embarking on the present program of research, it quickly became apparent that witness familiarity with the culprit was a pervasive yet widely varied concept. This may be one of the reasons that it is currently neglected in the eyewitness field of research. Previous eyewitness literature that have made reference to familiar eyewitnesses often describe this phenomenon with vague and, at times, contradictory definitions (e.g., Clutterbuck & Johnston, 2005). For example, face recognition researchers often describe a familiar face as one that was previously seen (through learning exercises) or is known (i.e., someone famous; e.g., Clutterbuck & Johnston, 2005; Gardiner, Ramponi, &

Richardson-Klavehn, 2002), whereas proponents of the know-remember-guess paradigm describe a participant’s sense of familiarity with a face as occurring when the participant remembers having seen the face but is not sure of the context, or even whether they have actually seen the face before (e.g., Gardiner et al., 2002). These varying definitions of person familiarity are fundamentally different within the context of a criminal investigation. Whether a witness responds, “I think I have met that person before but I do not know where” or “I have seen that person on previous occasions at the store” or “I know that person, she is in a movie”, may lead investigators to lend more or less credibility to the witness. Determining a clear and uniformly accepted definition of “familiarity” within eyewitness research is essential to a greater understanding of familiar witness accuracy. The present program of research serves as a foundation for an operational definition of a familiar witness. Given the applied focus of the research, it was most desirable to define familiarity according to participants’ perceptions, rather than from theoretical or cognitive points of view. In the real-world, investigators are likely going to have to rely on witnesses’ perceptions. Taking this approach, it is clear that participants’ perceptions of familiarity is a function of previous interaction with the culprit. In open-ended responses as well as when previous interaction was experimentally manipulated, participants perceived previous interaction with the culprit to significantly increase their familiarity with them. An additional factor, personal knowledge, was examined in the present studies as well. A group of participants received personal information about the culprit in addition to the previous interaction. In a pilot study this factor was found to significantly increase perceptions of familiarity beyond that

which occurred with just previous interaction, although this result did not occur in the main study. The reason for this inconsistency may have been due to a power issue; the pilot study had a smaller sample size than the main study. Alternatively, methodological differences may have been a factor; in the pilot study familiarity rankings were taken during the same testing session as the familiarity manipulation, whereas in the main study familiarity rankings were taken in a second session after all other measures were completed. It is recommended that future research delve further into the notion of familiarity and in particular in the context of familiar eyewitnesses. It is speculated that witness familiarity may be considered on a continuum, with stranger witnesses on one end and known witnesses (e.g., frequently seen relatives) on the other. A manipulation check in the main study indicated that the more familiar (i.e., interaction with personal information) and less familiar (i.e., interaction without personal information) groups manipulated in the present studies fell into a familiar-stranger category – i.e., participants in these groups rated the target as more familiar than a stranger but less familiar than a well-known celebrity. As a result, the present study findings may be considered a starting point to understanding the influence of the lower echelons of the familiarity (stranger vs. familiar-stranger) continuum on eyewitness accuracy.

**The Familiar-stranger Eyewitness**

The primary goal of the main study was to determine whether familiar-stranger eyewitnesses differed in terms of recall and recognition accuracy as compared to stranger eyewitnesses. Results suggest a complex relationship between familiarity and accuracy measures. Despite having additional exposure time, familiar eyewitnesses did not

provide significantly more (or significantly more accurate) descriptive information about the culprit. However, they were significantly more likely to correctly identify the culprit from a target-present lineup. Interestingly, there were no significant differences in ability to correctly reject a target-absent lineup. Results from a follow up study conducted more than a year after the initial identification lineup suggest that a very long delay before a second identification does not change the pattern of results for target-present lineups, although accuracy significantly declined.

**Recall accuracy.** Contrary to study hypotheses, more and less familiar eyewitnesses provided a similar number of culprit descriptors as stranger eyewitnesses and stranger eyewitnesses were just as accurate in the their descriptions of the culprit as were more and less familiar eyewitnesses. These results suggest that eyewitnesses who are familiar with the culprit may be no more informative than stranger eyewitnesses during the primary stages of an investigation (i.e., describing the culprit). This is the first known study to have examined recall accuracy across stranger versus familiar-stranger eyewitnesses; therefore, there is no previous research to compare. Future research is strongly recommended to determine whether the present study results may be generalized to real-world investigations.

In the meantime, a few caveats to these results should be noted. Qualitative reports from the eyewitnesses in the present study were analyzed for physical descriptors only. While the culprit was viewed an additional time by more and less familiar eyewitnesses, the appearance of the culprit changed slightly (i.e., her clothing) from their additional meeting to the crime video. As a result, recall of clothing viewed during the crime would

not be expected to differ amongst familiar and stranger eyewitnesses. Moreover, previous research examining the influence of exposure repetition on recall accuracy in children (there are no known studies conducted with adults) suggests that recall accuracy is more accurate when there is little to no change across repeated exposures (McNichol et al., 1999).

Moreover, in the main study results both familiar and stranger eyewitnesses provided, on average, a greater number of physical descriptors (i.e., 12 descriptors) of the culprit compared to that which is reported in previous research (e.g., 7 descriptors; Lindsay et al., 1994). Perhaps the approximately one-minute viewing time during the crime was sufficient to encode just as much information as may be encoded during a 30-minute interaction. As in many eyewitness research studies, the participants in the present study were tested under ideal conditions (e.g., little to no distractions, clear view of the culprit, etc.). A familiar-stranger advantage may be viewed under less than ideal conditions; however, future research is necessary to make this determination.

Even though, according to the present study results, familiarity with the culprit may not be an indicator of recall accuracy, the amount of information (but not the accuracy of information) reported in the recall task predicted recognition accuracy (correct or incorrect). A result such as this may suggest that witnesses who provide more information about the culprit would be better able to later recognize the culprit from an identification lineup. Regardless of their familiarity with the culprit, witnesses who reported a greater number of descriptors were more likely to correctly complete the identification lineup (i.e., correct identification in target-present lineup or correct

rejection in a target-absent lineup). This result is in contrast to previous research which generally finds no correlation between eyewitness recall performance and recognition performance (e.g., Bothwell et al., 1987; Meissner et al., 2007; van Koppen & Lochun, 1997).

Given the preliminary results of this study, more research is needed to make any firm conclusions; however, the results of this study warrant caution in giving more credence to more familiar eyewitnesses’ culprit descriptions. Familiar eyewitnesses may not be able to provide any additional or more accurate information regarding a culprit than stranger eyewitnesses. On the other hand, results from the present study suggest that for both stranger and familiar-stranger eyewitnesses, the amount of information that an eyewitness provides may be indicative of the witness’s ability to make a correct lineup decision.

**Recognition accuracy.** One of the most important pieces of evidence that may be obtained from an eyewitness is his or her lineup identification decision. It was hypothesized that more familiar eyewitnesses would be more accurate in their lineup decision, followed by less familiar eyewitnesses and that stranger eyewitnesses would be least accurate. Results from the present study partially supported this hypothesis. More and less familiar eyewitnesses were significantly more likely to make a correct identification in a target-present lineup; however, there were no significant differences across exposure conditions in the target-absent lineup. Extrapolated to a real-life situation, these results suggest that familiar-stranger witnesses may be more likely to correctly identify the culprit when the police have detained a guilty suspect but not when

the suspect is innocent. In other words, familiar-stranger witnesses are not significantly less likely to make a false identification than stranger witnesses. These results are partially consistent with the limited previous research (i.e., Steblay et al., 2011).

Face recognition studies that have used a target-present ranking system (rather than a yes or no) to identify previously seen persons, found that familiar persons were recognized more frequently than stranger persons (Burton et al., 1999). Also, Bornstein and colleagues (2012) reported similar results from their meta-analysis on exposure duration and identification accuracy. Although there were only three studies in which TP and TA lineups were considered separately, taken together results from these studies indicated that longer exposure time is more beneficial for TP than TA lineups (i.e., effect sizes were larger for TP lineups than TA lineups; Bornstein et al., 2012). In one of the only other eyewitness studies known to have measured familiarity, albeit post hoc, Steblay and colleagues (2011) found similar results, although their target-absent lineup also was significant across exposure conditions. The primary goal of Steblay et al.’s (2011) study was to assess the use of sequential lineup laps – i.e., allowing participants to view the lineup on more than one occasion; however, given that a portion of the study participants had previous exposure to the identification target, the influence of self-reported familiarity with the culprit was measured as a secondary variable. The pattern of results in this study were very similar to the present study. For example, Steblay et al.’s (2011) stranger participants correctly identified the target at a rate of 8% and correctly rejected the target at a rate of 68% on the first sequential lineup lap, compared to rates of 33% and 67% respectively in the present study. Steblay et al.’s (2011) familiar

participants correctly identified the target at a rate of 60% and correctly rejected the target at a rate of 88%, compared to the more familiar participants in the present study who correctly rejected at a rate of 81% and 78% respectively.

When considering the results of the present study more broadly, using diagnosticity ratios, it appears that familiar-stranger witnesses (participants from both the more familiar-stranger group and less familiar-stranger group) are just as likely as stranger witnesses to identify the guilty over the innocent suspect. This result is possibly due to stranger witnesses taking on a more conservative choosing criterion than familiar-stranger eyewitnesses; stranger eyewitnesses were significantly less likely to choose someone from a lineup than familiar-stranger eyewitnesses. This more stringent choosing approach also would explain the differences in accuracy across exposure condition when analyzing the TP and TA lineup responses separately. For example, stranger eyewitnesses were less likely to choose someone from the TP lineup, resulting in stranger eyewitnesses being less likely to make a correct decision. However, stranger eyewitnesses also were less likely to choose someone from a TA lineup resulting in them being just as likely to make a correct decision as familiar-stranger eyewitnesses. Previous research examining choosing rates for sequential lineups indicate that, overall, stranger eyewitnesses tend to adopt a more conservative criterion when presented with a sequential lineup (in comparison to a simultaneous lineup; e.g., Meissner et al., 2005). Given that this is the first study to systematically examine eyewitness accuracy of familiar-stranger eyewitnesses (Steblay et al., 2011 used a post hoc approach) there is no previous research to compare; however, it seems as though familiar-stranger eyewitnesses

may have a more lenient choosing criteria, being that they are more likely to choose someone from the lineup. Archival studies examining the influence of extended exposure duration finds some support for these results, in that witnesses with longer exposure to the culprit are more likely to make a culprit identification (Horry et al., 2014; Memon et al., 2011). Also, an eyewitness study by Read (Study One; 1995) found that witnesses with longer exposure to the culprit were more likely to correctly identify the target as well as make a false identification.

Given that witnesses may be asked to recount or even complete another identification lineup after a lengthy delay, the present study participants completed a second identification lineup approximately 14 months after their participation in the main study. Accuracy rates dropped substantially over the delay; however, the rates at which they declined were not significantly different across exposure conditions. This may indicate that the familiar-stranger eyewitness advantage found in an initial TP lineup disappears with time. However, due to the fact that the follow up study was post hoc and, therefore, sample size was quite low, power may have been an issue.

Another notable finding with respect to familiarity and recognition accuracy is the lack of differences between the more familiar and less familiar conditions. The difference between these two groups was the presentation of personal knowledge about the culprit in the more familiar condition. As noted earlier, a manipulation check indicated no significant differences in participants’ perceptions of familiarity across these two groups and that is likely one of the reasons that no differences were found between these groups in terms of recognition accuracy. However, further insight into why these

two groups did not differ may be offered by the interactive activation and competition (IAC) model of face recognition (Burton et al., 1999). The IAC model suggests that face recognition is a function of Face Recognition Units that are linked to Person Identity Nodes (PIN). Person familiarity is a threshold that is reached at the PIN that subsequently triggers Semantic Information Units that house meaningful information about the person. The present study did not directly examine this model; however, it may be surmised that participants in the more familiar condition who had personal information about the culprit would be more likely to recognize the culprit than a less familiar person or stranger because they did not have personal information about the culprit. The knowledge of personal information for those in the more familiar group may result in more information being stored in the Semantic Information Units producing a greater likelihood of triggering a familiarity judgment. However, contrary to this, the present study results suggest that there were no significant differences in the recognition accuracy of more familiar and less familiar eyewitnesses. This may be explained by the order in which information is processed according to the IAC model. Encoded face information is first assessed at the Face Recognition Unit (view independent information) and then the PIN. It is at this stage that the model suggests a determination of familiarity is made. If that is the case, then witnesses who are making an identification decision may not go further to retrieve Semantic Information Units because it is not necessary for the task at hand. All they are being requested to do is recognize the person, not recollect any personal information about them. If this is the case, it may be that personal information was not used in the recognition task by more familiar eyewitnesses thus making their

decisions very similar to less familiar eyewitnesses. On this note, it is important to consider the identification task of eyewitnesses and in what ways familiarity may or may not influence this decision. Ultimately, recognizing a photograph amongst alternatives is a very visual task and previous knowledge of the culprit may not provide any advantages in this task.

Lastly, when interpreting these results it is important to make note of emerging research being conducted by Wells and colleagues (e.g., Charman & Wells, 2007). The design of the present study, similar to the design of most eyewitness research studies, makes an assumption that the base rate of any given real-life police lineup would have a 50% likelihood of having a guilty suspect and a 50% likelihood of having an innocent suspect. With this assumption, diagnosticity ratios suggests little variability in the likelihood of stranger versus familiar-stranger eyewitnesses’ recognition accuracy. However, changes to this base rate will impact the results and the subsequent interpretation of familiar-stranger recognition accuracy. For example, given that familiar-stranger witnesses were significantly more likely to correctly identify the culprit from the target-present lineup, a shift in the base rate to a 90% probability of having a lineup with a guilty suspect would likely result in more distinctions between stranger and familiar-stranger recognition accuracy (e.g., Wells & Olson, 2003). Overall, while the results of the present study provide insight into the relationship between accuracy rates between familiar-stranger and stranger eyewitnesses, the absolute accuracy rate values may provide little meaning to a real-life case. As Wells and Luus (1990) have asserted, lineup identification decisions are only one of many factors that may be considered in

determining the accuracy of eyewitness evidence. There are many other forms of evidence that may be available in a given case that will skew the base rate and/or contribute to the diagnosticity of guilt.

**Confidence of the Familiar-stranger Eyewitness**

Eyewitness confidence is an influential factor in a criminal case and for many years researchers argued that its influence was unfounded; however, more recent research suggests that there may be value in using confidence as an indicator of eyewitness accuracy. Confidence was measured on three occasions in the present study: before completion of the identification lineup, during the identification lineup, and a confidence proxy (i.e., willingness to testify to identification decision) was measured immediately after completion of the identification lineup. More and less familiar witnesses were more confident in the pre-lineup recognition of the culprit than stranger witnesses. Moreover, for both the target-present and target-absent lineups, more and less familiar witnesses were, on average, more confident in their lineup decisions than stranger eyewitnesses. Lastly, an interaction occurred whereby more familiar eyewitnesses with correct identification decisions were more willing to testify than more familiar eyewitnesses who were incorrect. Taken together, these results suggest that the timing of confidence judgements may be particularly important in the case of the familiar-stranger eyewitnesses.

Familiar-stranger eyewitnesses went into the identification lineup task feeling more confident than stranger eyewitnesses. This may be due to their additional experience with the culprit; they may assume that the task will be relatively easy given that they have

had a good deal of previous exposure to the culprit. They also may be overestimating the strength of their memory trace or, because they were able to recognize the culprit as someone who they had previously seen, they may believe that they will easily recognize the culprit again in an identification lineup. Given that witnesses have not yet made an identification decision, pre-lineup confidence may not be as valuable as confidence statements provided during the decision-making process in the identification lineup.

The present study results suggest that the confidence of familiar-stranger eyewitnesses during the lineup may need to be interpreted with caution. Familiar-stranger eyewitness confidence remained higher than stranger eyewitness confidence overall even though more and less familiar eyewitnesses were more accurate in correctly identifying the target from the TP lineup but were not significantly more likely to correctly reject a TA lineup, compared to stranger eyewitnesses. Familiar-stranger eyewitness identification confidence in the TA lineups may be mistakenly inflated. Albeit, correct witnesses were more confident than incorrect witnesses overall.

A proxy measure of confidence, participant willingness to testify to their lineup decision in a court of law, was assessed upon completion of the identification lineup. This may be a more realistic measure of confidence, as criminal investigators may not request confidence until the completion of the identification lineup. Also, making salient that the participant’s lineup decision would hold weight in a court of law may prompt them to adopt a more stringent confidence criterion. There is evidence that more familiar eyewitnesses did adjust their confidence with this last question because in this last measure of confidence, more familiar eyewitnesses who were correct were significantly

more confident than incorrect familiar eyewitnesses; there were no significant differences with the less familiar eyewitnesses and stranger eyewitnesses. These results suggest that more familiar eyewitnesses were better able to discriminate between correct and incorrect responses than less familiar and stranger eyewitnesses.

Moreover, overall, participants who correctly identified the culprit from the TP lineup were significantly more willing to testify than those who incorrectly responded to the TP lineup. This result may be due to a delayed relative judgement that may be possible when completing a TP sequential lineup. Participants may have initially committed to a lineup member but after having seen all lineup members, and potentially the target later in the lineup, may realize that their initial identification decision was incorrect.

**Confidence-accuracy relationship**. The primary purpose for measuring confidence in the present study was to determine whether familiar-stranger eyewitnesses’ confidence ratings are more indicative of accuracy than stranger eyewitnesses’ confidence ratings. The final confidence measurement, willingness to testify, was used to determine confidence-accuracy correlations.

Overall, the confidence-accuracy (CA) relationship in the present study was significant, $r = 0.20$. This CA relationship is similar to previous studies (e.g., Bothwell et al., 1987; Sporer et al., 1995), and somewhat lower than what may be predicted by Leippe and Eisenstadt (2007). However, the CA relationship across groups was quite varied. The CA relationship for those in the more familiar condition was moderate ($r = 0.42$) and significant; whereas the CA relationship for those in the less familiar condition

was near zero ($r$ = 0.07) and even negative in the stranger condition ($r$ = -0.06). When considering that previous studies are most comparable to the stranger condition, the present study results are quite lower than what may be typically found in research. This may suggest that, for participants in the stranger condition, the identification decision in the present study was difficult.

Given that it may be more ecologically valid to consider only witnesses who made a choice from a lineup because someone who made a non-identification (i.e., rejected a lineup) will likely not be asked to testify to their decision in a court of law, the CA relationship for choosers only was examined. The CA relationship for choosers only remained significant and was slightly higher ($r$ = 0.36) than the overall CA relationship. The same increase in CA relationship occurred when Sporer et al. (1995) isolated out choosers ($r$ = 0.41), identifying choosing as a moderating factor in the CA relationship. Similar to the overall CA relationship results, more familiar eyewitnesses' CA relationship was strongest ($r$ = 0.60), followed by less familiar witnesses ($r$ = 0.14) and stranger witnesses ($r$ = 0.07). The more familiar witness CA relationship was the only significant correlation.

On the other hand, these results are in opposition to those reported by Pezdek and Stolzenberg (2014). Participants in their study rated the familiarity of previous school colleagues and strangers. They were more confident in their familiar responses than they were in their non-familiar responses; however, the overall CA relationship was $r$ = 0.16. Considering familiar and non-familiar separately the CA relationship for familiar was $r$ = 0.12 and for unfamiliar was $r$ = 0.25. It is difficult to discern the reason for the

inconsistency between the present study and Pezdek and Stolzenberg’s (2014); however, it is notable that familiarity was manipulated in the present study, whereas Pezdek and Stolzenberg (2014) simply measured subjective ratings of it. Regardless, the present study results are consistent with the optimality hypothesis; more exposure and better quality exposure was related to confidence ratings that were more consistent with accuracy (Bothwell et al., 1987).

Overall CA relationships increased from the main study to the follow up study approximately 14 months later. For example, the choosers-only CA $r$ was 0.36 in the main study and increased to $r = 0.53$ in the follow up. A similar pattern was found for all relationships. This may indicate that, with time, eyewitnesses become more reserved in their confidence judgments which bodes well for the use of confidence as a measure of accuracy.

A final note regarding the CA relationship described in the present studies. The analysis used to assess the CA relationship was a point-biserial correlation. This type of analysis has been frequently used in previous studies and is suitable for determining differences between groups (e.g., Olsson, 2000). In addition, given that the experimental conditions in the present study were stringent and time intensive a limited sample size was obtained; relative to what is typically required for other types of confidence-accuracy analyses (e.g., a calibration approach; Juslin, Olsson, & Winman, 1996). As a result, the point-biserial correlation was deemed most appropriate for the present study; however, there are alternate confidence-accuracy analyses that may be used, such as a calibration approach, with familiar-stranger eyewitnesses in future research that may provide

additional information regarding this relationship. For example, using a calibration approach one may come to the conclusion that witnesses with 90% confidence have an 80% likelihood of having made an accurate lineup decision (e.g., Brewer & Wells, 2006). Moreover, it may be the case that familiar-stranger eyewitness confidence may calibrate differently than stranger eyewitness confidence. Future research in this area is highly recommended.

**Strengths and Limitations**

Results from the present study suggest that witnesses who are familiar with the culprit are similar to stranger witnesses in some respects (i.e., their ability to provide descriptive information about the culprit and reject a photo lineup that does not contain the culprit) whereas they are more accurate in others (i.e., their ability to identify the culprit from a TP lineup and provide confidence judgements that correlate with the accuracy of their lineup decision). The lack of research conducted to date on familiar eyewitnesses suggests that these witnesses are being treated the same as stranger eyewitnesses which, as suggested by the present study results, may be reasonable in some ways but could be detrimental in others.

Face recognition researchers have studied familiar face recognition for some time and their research has developed cognitive theories regarding the recognition of stranger versus familiar faces. According to these cognitive researchers familiar faces are often recognized more quickly and accurately than stranger faces in most circumstances. This research can be used to inform eyewitness research to some extent; however, it must be recognized that the face recognition paradigm neglects some key procedures found in an

eyewitness situation such as the presence/exposure to the entire person (not just the face) and the need to choose the culprit amongst alternatives. Results from the present Pilot Study One, where manipulations used in face recognition research (i.e., exposure repetition and exposure duration via videotape) did not produce the subjective familiarity that is the focus of eyewitness familiarity judgements, further supports the notion that studies using the eyewitness paradigm may be best suited to inform the criminal justice system. There are only two known eyewitness studies that have measured familiarity. In 2011, Steblay and colleagues examined familiarity as a secondary variable when looking at the use of sequential lineup laps and in 2014 Pezdek and Stolzenberg had participants make familiarity judgements of previous school mates or strangers. Along with the present study, these previous studies are an important contribution to the limited knowledge that is currently available regarding familiar witness recognition accuracy. However, neither of the previous studies controlled the previous exposure participants had with the target; in both Steblay et al. (2011) and Pezdek and Stolzenberg (2014) the “familiarity” that participants had with the target was varied and, quite frankly, is unknown due to their use of self-report familiarity judgements (Steblay et al., 2011) and the assumption that school mates will have frequent exposure to one another (Pezdek & Stolzenberg, 2014). On the other hand, the use of strict control measures used in the present study allows for firm conclusions to be made regarding participants’ previous exposure with the culprit, hence their objective familiarity with the culprit. With this definitive manipulation of familiarity, the present study supersedes the limited previous research in providing conclusions regarding the influence of familiarity on eyewitness

accuracy. As such, results from the present study provide novel and much needed insight to the field of eyewitness identification accuracy and provides impetus for similar research in the future.

There are a few ways that the present research may be improved upon in future research in order to further strengthen the results and generalizability to real-world situations. For example, an eyewitness paradigm was used in the present study as it is considered more ecologically valid than face recognition studies. However, there are some criticisms to the applicability of an experimental method to real-life criminal cases. For example, in the present study witnesses viewed a video-tape of a mock crime – the crime was not live and not real. There were no dangers to the eyewitnesses and the stress levels of participants are suspected to have been fairly low. These are factors that may have influenced the memory and identification decisions of real-life witnesses that were not factored into the present study results. Having acknowledged all of this, the purpose of the present study was to examine differences in familiarity and, as such, it was important to keep these variables controlled in order to be able to adequately determine that differences amongst participants were due to the primary independent variable, witness familiarity, alone.

Relatedly, it is acknowledged that results obtained in the present study may represent a best-case scenario in that participant eyewitnesses were provided an ideal witnessing condition. The culprit was viewed on the videotape for nearly one minute, the participant was purposefully provided no distractions, and they were asked to watch the video. Moreover, while all participants completed the recall and recognition tasks in the

same room, for more and less familiar eyewitnesses this was also the same room with which they had a previous exposure with the culprit. Consequently encoding specificity may have enhanced their memory (Tulving & Thomson, 1973). There may have been cues in the room that triggered their memory for their previous encounter with the culprit.

The sample used in the present study was undergraduate psychology students. Using a homogenous group such as this who are generally younger and perhaps more educated than the general population may limit generalizability. However, researchers in eyewitness and juror decision making fields have examined the differences between student samples and community samples and have found little to no differences (e.g., O’Rourke, Penrod, Cutler, & Stuve, 1989).

Nonetheless, due to the novelty of the present study it serves as a great starting point for future research on familiar eyewitnesses and, at present, the only source of much needed information regarding the performance of objectively familiar eyewitnesses’ recall and recognition performance and confidence.

**Real-world Implications and Future Directions**

There are a few important insights that the present program of research may provide criminal investigators. Primarily, the present study’s first look into the familiar-stranger suggests that familiar-stranger witnesses may be just as susceptible to making a false positive identification as a stranger eyewitness. If this is the case, investigators should be wary in giving more credit to witnesses who have claimed to have had previous exposure to the culprit. While this result is preliminary, there are a number of examples of so-called familiar-stranger eyewitnesses who provided evidence that resulted in the

wrongful conviction of an innocent suspect in the Innocence Project archives (Innocent Project, 2015). For example, in May 1991, a man broke into the home of two teenage girls in North Carolina. The intruder held the girls at knife-point and raped them. Both girls separately identified Joseph Lamont Abbitt as the culprit; a man who they had seen on a few occasions prior to the crime because he lived in their neighbourhood and had visited their home. Based on the girls’ testimony, Abbitt was convicted and sentenced to two consecutive life sentences and 110 years imprisonment. However, 14 years after the conviction, DNA tests were conducted that absolved Abbitt from the crime. Despite having previous exposure to the innocent suspect as well as lengthy and close contact with the culprit, the young girls were unable to distinguish this familiar innocent suspect from the true culprit.

While confidence has been measured and studied in the eyewitness field for many years, more recent interest and innovative approaches are rapidly advancing its utility in the criminal justice system. Results from the present study suggest that findings may be unique when considering the confidence-accuracy relationship in familiar-stranger eyewitnesses. Familiar-stranger eyewitnesses’ confidence estimates may be more indicative of the accuracy of their identification decision than stranger eyewitnesses.

Overall, familiarity between an eyewitnesses and a culprit is an important factor that is currently neglected in empirical research. This variable must be explored further in order to inform criminal investigators of its impact on witness accuracy and confidence.

**References**

Beaudry, J.L., & Lindsay, R.C.L. (2006). Current identification procedure practices: A survey of Ontario Police Officers. *The Canadian Journal of Police and Security Services, 4*, 178-183.

Bonner, L., Burton, A.M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition, 10*(5), 527-536. doi:10.1080/13506280244000168

Bornstein, B.H., Deffenbacher, K.A., Penrod, S.D., & McGorty, E.K. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime & Law, 18*(5), 473-490.

Bothwell, R.K., Deffenbacher, K.A., & Brigham, J.C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*(4), 691-695.

Brainerd, C.J., & Reyna, V.F. (1990). Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review, 10*(4), 365-403.

Brainerd, C.J., & Reyna, V.F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*(5), 164-169.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26*(3), 353-364.

Brewer, N. & Palmer, M.A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*(1), 77-96. doi:10.1348/135532509X414765

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

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*(1), 105-116.

Bruce, V., Burton, M., & Hancock, P. (2007). Remembering faces. In R.C.L. Lindsay, D.F. Ross, J.D. Read, & M.P. Toglia (Eds.), *The handbook of eyewitness psychology, volume II: Memory for people* (pp. 87-100). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Burton, A.M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*(3), 243-248. doi:10.1111/1467-9280.00144

Cain, W.J., Baker-Ward, L., & Eaton, K.L. (2005). A face in the crowd: The influences of familiarity and delay on preschoolers’ recognition. *Psychology, Crime & Law, 11*(3), 315-327. doi: 10.1080/10683160418331294835

Campbell, R., Coleman, M., Walker, J., Benson, P.J., Wallace, S., Michelotti, J., & Baron-Cohen, S. (1999). When does the inner-face advantage in familiar face recognition arise and why? *Visual Cognition, 6*(2), 197-216. doi:10.1080/713756807

Campbell, R., Walker, J., & Baron-Cohen, S. (1995). The development of differential use of inner and outer face features in familiar face identification. *Journal of Experimental Child Psychology, 59*(2), 196-210. doi:10.1006/jecp.1995.1009

Charman, S., & Wells, G.L. (2007). Applied lineup theory. In R.C.L. Lindsay, D.F. Ross, J.D. Read, & M.P. Toglia (Eds.), *The handbook of eyewitness psychology, volume II: Memory for people* (pp. 219-254). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Cloutier, J., Kelley, W.M., & Heatherton, T.F. (2011). The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. *Social Neuroscience, 6*(1), 63-75. doi: 10.1080/17470911003693622

Clutterbuck, R., & Johnston, R.A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17*(1), 97-116. doi:10.1080/09541440340000439

Cutler, B.L., & Penrod, S.D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73*(2), 281-290. doi:10.1037/0021-9010.73.2.281

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior, 11*(3), 233-258. doi: 10.1007/BF01044644

Davies, G.M. (1995). Fuzzy-trace theory and eyewitness memory. *Learning and Individual Differences, 7*(2), 111-114.

Deffenbacher, K.A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*(4), 243-260.

Deffenbacher, K.A., Bornstein, B.H., McGorty, E.K., & Penrod, S.D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness’s memory

representation. *Journal of Experimental Psychology: Applied, 14*(2), 139-150. doi: 10.1037/1076-898X.14.2.139

Doob, A.N., & Kirshenbaum, H.M. (1973). The effects on arousal of frustration and aggressive films. *Journal of Experimental Social Psychology, 9*(1), 57-64. doi: 10.1016/0022-1031(73)90062-0

Ebbinghaus, H. (1949). Experiments in memory. In W. Dennis (Ed.), *Readings in general psychology* (pp. 225-230). New York, NY: Prentice-Hall Inc.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*(4), 431-439.

Fisher, R.P., Brennan, K.H., & McCauley, M.R. (2002). The cognitive interview method to enhance eyewitness recall. In M.L. Eisen, J.A. Quas, & G.S. Goodman, *Memory and suggestibility in the forensic interview* (pp. 265-286). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Flowe, H.D., Mehta, A., & Ebbesen, E.B. (2011). The role of eyewitness identification evidence in felony case dispositions. *Psychology, Public Policy, and Law, 17*(1), 140-159. doi:10.1037/a0021311

Gardiner, J.M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory, 10*(2), 83-98, doi: 10.1080/09658210143000281

Goldstein, A.G., Chance, J.E., & Schneller, G.R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Bulletin of the Psychonomic Society, 27*(1), 71-74.

Gronlund, S.D., Wixted, J.T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Direction in Psychology Science, 23*(1), 3-10.

Gross, G., & Hayne, H. (1996). Eyewitness identification by 5- to 6-year-old children. *Law and Human Behavior, 20*(3), 359-373. doi:10.1007/BF01499028

Hancock, P.J.B., Bruce, V., & Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*(9), 330-337. doi:10.1016/S1364-6613(00)01519-9

Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law and Human Behavior, 38*(1), 94-108.

Innocence Project (2015). In *Innocence Project*. Retrieved from www.innocenceproject.org

Johnston, R.A., & Edmonds, A.J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*(5), 577-596. doi: 10.1080/09658210902976969

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnositicity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1304-1316. doi: 10.1037/0278-7393.22.5.1304

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

Kuehn, L.L. (1974). Looking down a gun barrel: Person perception and violent crime. *Perceptual and Motor Skills, 39*, 1159-1164.

Leippe, M.R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R.C.L. Lindsay, D.F. Ross, D.J. Read, & M.P. Toglia (Eds.), *The handbook of eyewitness psychology, volume II: Memory for people* (pp. 377-425). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Leippe, M.R., Romanczyk, A., & Manion, A.P. (1991). Eyewitness memory for a touching experience: Accuracy differences between child and adult witnesses. *Journal of Applied Psychology, 76*(3), 367-379. doi:10.1037/0021-9010.76.3.367

Lindsay, R.C.L., Lea, J.A., & Fulford, J.A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*(5), 741-745. doi:10.1037/0021-9010.76.5.741

Lindsay, R.C.L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18*(5), 527-541. doi: 10.1007/BF01499172

Lindsay, R.C.L., Pozzulo, J.D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of

adults and children. *Law and Human Behavior, 21*(4), 391-404. doi:10.1023/A: 1024807202926

Lindsay, R.C.L., & Wells, G.L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*(3), 556-564. doi:10.1037/0021-9010.70.3.556

Lipton, J.P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology, 62*(1), 90-95.

Mackay, T.L., & Paterson, H.M. (2014). How does timing of recall affect eyewitness memory and psychological distress? *Journal of Police and Criminal Psychology*. Advanced online publication. doi: 10.1007/511896-014-9156-2

Malpass, R.S., & Lindsay, R.C.L. (1999). Measuring line-up fairness. *Applied Cognitive Psychology, 13*, S1-S7. doi: 10.1002/(SICI)1099-0720(199911)13:1

Mandler, G. (1981). The recognition of previous encounters. *American Scientist, 69*(2), 211-218.

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychology Science, 3*(5), 390-399. doi: 10.1111/j.1745-6924.2008.00087.x

McNichol, S., Shute, R., & Tucker, A. (1999). Children’s eyewitness memory for a repeated event. *Child Abuse & Neglect, 23*(11), 1127-1139. doi:10.1016/S0145-2134(99)00084-8

Meissner, C.A., Sporer, S.L., & Schooler, J.W. (2007). Person descriptions as eyewitness evidence. In R.C.L. Lindsay, D.F. Ross, J.D. Read, & M.P. Toglia (Eds.), *The*

*handbook of eyewitness psychology, volume II: Memory for people (pp. 3-34)*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

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

Melara, R.D., DeWitt-Rickards, T.S., & O’Brien, T.P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology, 74*(5), 706-713. doi:10.1037/0021-9010.74.5.706

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime, & Law, 17*(8), 711-729. doi: 10.1080/10683160903524333

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*(3), 339-354. doi: 10.1348/000712603767876262

Moreland, R.L., & Zajonc, R.B. (1982). Exposure effects in person perception: Familiarity, similarity, and attraction. *Journal of Experimental Social Psychology, 18*(5), 395-415. doi:10.1016/0022-1031(82)90062-2

Odegard, T.N., Cooper, C.M., Lampinen, J.M., Reyna, V.F., & Brainerd, C.J. (2009). Children’s eyewitness memory for multiple real-life events. *Child Development, 80*(6), 1877-1890. doi:10.1111/j.1467-8624.2009.01373.x

Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology, 20*(7), 973-985. doi: 10.1002/acp.1263

Olsson, N. (2000). A comparison of correlation, calibration, and diagnositicy as measures of the confidence-accuracy relationship in witness identification. *Journal of Applied Psychology, 85*(4), 504-511. doi: 10.1037/0021-9010.85.4.504

O’Rourke, T.E., Penrod, S.D., Cutler, B.L., & Stuve, T.E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior, 13*(4), 385-395. doi: 10.1007/BF01056410

Parker, J.E., & Ryan, V. (1993). An attempt to reduce guessing behaviour in children’s and adults’ eyewitness identifications. *Law and Human Behavior, 17*(1), 11-26. doi:10.1007/BF01044534

Paterson, H.M., Kemp, R.I., & Ng, J.R. (2011). Combating co-witness contamination: Attempting to decrease the negative effects of discussion on eyewitness memory. *Applied Cognitive Psychology, 25*(1), 43-52.

Pezdek, K., & Stolzenberg, S. (2014). Are individuals’ familiarity judgments diagnostic of prior contact? *Psychology, Crime & Law, 20*(4), 302-314. doi: 10.1080/1068316X.2013.772181

Pozzulo, J.D. (2007). Person description and identification by child witnesses. In R.C.L. Lindsay, D.F. Ross, J.D. Read, & M.P. Toglia (Eds.), *The handbook of eyewitness psychology, volume II: Memory for people* (pp. 283-307). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Pozzulo, J.D., & Lindsay, R.C.L. (1999). Elimination lineups: An improved eyewitness procedure for child eyewitnesses. *Journal of Applied Psychology, 84,* 167-176. doi:10.1037/0021-9010.84.2.167

Read, J.D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology, 9*(2), 91-121. doi: 10.1002/acp.2350090202

Read, J.D., Vokey, J.R., & Hammersley, R. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy-confidence relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 870-882.

Reyna, V.F., Mills, B., Estrada, S., & Brainerd, C.J. (2007). False memory in children: Data, theory, and legal implication. In M.P. Toglia, J.D. Read, & D.F. Ross (Eds.), *The handbook of eyewitness psychology, volume 1: Memory for events* (pp. 479-507). Mahwah, NJ: Lawrence Erlbaum Associates.

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*(2), 139-156. doi:10.1037/0033-2909.100.2.139

Sporer, S.L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*(1), 22-33. doi:10.1037/0021-9010.78.1.22

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

Steblay, N.K., Dietrich, H.L., Ryan, S.L., Raczynski, J.L., & James, K.A. (2011). Sequential lineup laps and eyewitness accuracy. *Law and Human Behavior, 35*(4), 262-274.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R.C.L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*(5), 459-473. doi:10.1023/A:1012888715007

Steblay, N.K., Dysart, J.E., & Wells, G.L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17*(1), 99-139. doi:10.1037/a0021650

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* National Institute of Justice, U.S. Department of Justice.

Technical Working Group for Eyewitness Evidence. (2003). *Eyewitness evidence: A trainer’s manual for law enforcement.* National Institute of Justice, U.S. Department of Justice.

Tollestrup, P.A., Turtle, J.W., & Yuille, J.C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D.F. Ross, J.D. Read, & M.P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-160). New York, NY: Cambridge University Press.

Tredoux, C.G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22*(2), 217-237. doi:10.1023/A:1025746220886

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373. doi: 10.1037/h0020071

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232. doi: 10.1016/0010-0285(73)90033-9

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17*, 969-993. doi: 10.1002/acp.939

van Koppen, P.J., & Lochun, S.K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21*(6), 661-685. doi: 0147-7307/97/1200-0661

Wang, E., Paterson, H., & Kemp, R. (2014). The effects of immediate recall on eyewitness accuracy and susceptibility to misinformation. *Psychology, Crime, & Law, 20*(7), 619-634. doi: 10.1080/1068316x.2013.854788

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14*(2), 89-103. doi:10.1111/j.1559-1816.1984.tb02223.x

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

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

Wells, G.L., & Luus, C.E. (1990). The diagnosticity of a lineup should not be confused with the diagnostic value of nonlineup evidence. *Journal of Applied Psychology, 75*(5), 511-516. doi: 10.1037/0021-9010.75.5.511

Wells, G.L., Malpass, R.S., Lindsay, R.C.L., Fisher, R.P., Turtle, J.W., & Fulero, S.M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist, 55*(6), 581-598. doi:10.1037/0003-066X.55.6.581

Wells, G.L., & Olson, E.A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295. doi: 10.1146/annurev.psych.54.101601.145028

Wells, G.L., Olson, E.A., & Charman, S.D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science, 11*(5), 151-154. doi: 10.1111/1467-8721.00189

Wells, G.L., & Quinlivan, D.S. (2009). Suggestive eyewitness identification procedures and the Supreme Court’s reliability test in light of eyewitness science: 30 years later. *Law and Human Behavior, 33*(1), 1-24. doi:10.1007/s10979-008-9130-3

Yarmey, A.D., Jacob, J., & Porter, A. (2002). Person recall in field settings. *Journal of Applied Social Psychology, 32*(11), 2354-2367.

Yarmey, A.D., & Yarmey, M.J. (1997). Eyewitness recall and duration estimates in field settings. *Journal of Applied Social Psychology, 27*(4), 330-344. doi: 10.1111/j.1559-1816.1997.tb00635.x

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children’s target-absent lineup performance. *Applied Cognitive Psychology, 23*, 358-368. doi: 10.1002/acp.1511

**Appendix A**

Demographics Form

Your Age: _____

Your Sex: _____

Your Primary Language: _____

Please indicate which ethnic group you would consider yourself to belong to:

☐ White (e.g., European)

☐ Black (e.g. African, African American, African Canadian, Caribbean)

☐ East Asian (e.g. Chinese, Japanese, Korean, Polynesian)

☐ South Asian (e.g. Indian, Pakistani, Sri Lankan, Bangladeshi)

☐ Southeast Asian (e.g. Burmese, Cambodian, Filipino, Laotian, Malaysian, Thai, Vietnamese)

☐ West Asian (e.g. Arabian, Armenian, Iranian, Israeli, Lebanese, Palestinian, Syrian, Turkish)

☐ Latin American (e.g. Mexican, Indigenous Central and South American)

☐ Aboriginal Canadian/Native Canadian/First Nations

☐ Mixed origin, please specify:_____

☐ Other: _____

## Appendix B

Structured Interview for Pilot Study Two/Main Study

**No Personal Information – Less Familiar Condition**

To start off with, I am going to give you a little more information about *typical students that come to the lab to do this study*. *Students are usually undergraduates* here at Carleton University. *They are usually in their first year of studies*; most of them taking the Bachelor of Arts program, majoring in psychology. *The average age of participants is 20*-years-old. *Most students are not married and do not have children*. *Many are* attending Carleton on a full-time basis but also have a part-time job. *I estimate that about 50% of students do not consider Ottawa to be their home town*; *those who are not from Ottawa will likely move back home or to another community after they graduate*. *When asked, students report coming from families of all different sizes, but many are only children or only have one sibling*. Lastly, *the types of activities that students most commonly report engaging in in their free time are: playing sports, hanging out with friends, or playing video games*.

Now that you know a little bit more about *the typical student that comes into the lab*, I am going to ask you a few questions about yourself. These questions mostly pertain to your opinions, your interests, and your hobbies. They are standard questions and none of them are meant to be personal. Please know that you are not obligated to answer any of these questions and you are able to say, "I don't know" or "I don't want to answer that question" to any one of the following questions.

The first few questions are going to be about university, and in particular about Carleton University.

When asked, a lot of students say they chose their university for a lot of reasons. For example, *some people come to Carleton because of the location, the staff or professors, or because of the programs that it offers*. What would you say was the main reason you chose to come to Carleton?

*Some Bachelor degree programs are 3 years while others are 4 years*. How many years do you plan to be at Carleton?

Most of the students who complete psychology studies for credit are majoring in psychology, but some students take intro psych or stats just for interest. What program are you majoring in?

*The largest population of students at Carleton is in their 1st year of studies*. What year are you in?

*Most students at Carleton choose to* attend university on a full-time basis. Are you a full-time or part-time student?

*One of the more popular classes to take at Carleton is a Biology class by Michael Runtz*. What has been your favourite course at Carleton thus far?

In many introductory psychology courses students are graded based on their performance on tests, but sometimes there are opportunities to write an essay or make a presentation for the class instead. *Many students prefer presentations*. If you had to choose between taking a multiple choice test, writing an essay, or doing a presentation for marks, which one would you chose?

University courses used to require a lot of independent work; however, more recently efforts have been made to implement more activities that students can do in groups. *Many students indicate that they prefer group work*. If you had to choose to work on a project by yourself or with a group, which one would you choose?

*Many students indicate that they* would prefer night classes over morning classes. Would you consider yourself to be a morning person or a night person?

In the next set of questions, I am going to ask you about your thoughts and experiences of Ottawa.

Students attending Carleton come from all over the world. Are you from Ottawa?

In a poll, people living in Ottawa reported their favourite things about living in Ottawa. *Some people reported that they liked the hockey games and concerts and others enjoyed all of the museums*. What would you say is your favourite thing to do in Ottawa?

Ottawa and the surrounding area host many festivals and events *such as Bluesfest, Canada Day celebrations, and Ribfest*. What would you say is your favourite festival or event?

Ottawa has many different types of restaurants where patrons can try different kinds of ethnic foods *such as Vietnamese, Italian, Thai, or Greek*. What is your favourite kind of food or restaurant?

In Ottawa, there are many sports or recreational activities that you can participate in professionally or recreationally. For example, *there is rowing, hockey, or soccer teams that you can be a part of, if you wish*. Do you participate in any sports?

Ottawa also has a great artistic community and there are opportunities to express your creative side. *There are opportunities for acting, painting, sculpting, and playing music, to name a few*. Do you play any musical instruments?

This next set of questions I like to call the "favourites" section. I am going to ask you to think of some of your favourite things.

Many people like country music but *there are also quite a few students who prefer pop*. What is your favourite kind of music or musician?

Empire has released a list of the top 500 movies ever made. *The top 3 are: The Godfather, Raiders of the Lost Ark, and Star Wars Episode 5.* What is your favourite movie of all time?

The television network Fox has many popular shows. *The most watched shows on Fox over the past 25 years are: American Idol, Glee, and The Simpsons*. What is your favourite television show?

*Many people choose to watch TV during their free time*. If you had to choose between watching TV, reading a book, or going for a walk, which one would you choose?

Next, I will ask you some questions regarding your goals and aspirations in terms of a career.

*A lot of students choose to work while they are attending university*. Do you have a part-time or full-time job?

*Many students are attending university in order to further their education and increase the chances that they will get a job in the field that they enjoy*. What job are you planning on getting after you graduate from university?

When determining which career path you want to take you may consider various factors *like salary, job location, and hours in order* to help you decide which type of employment you want. What factor do you think is most important to you when considering a job opportunity?

*The most common age to retire is approximately 60 years old.* What age would you like to retire at?

The last set of questions are a lot more general and may make you think a little bit – but in a fun, no pressure kind of way.

Flash can move at lightning fast speed, Superman can fly, and the Hulk is abundantly strong. *If given the choice most people say they would like to fly*. If you could have any superhuman power, which one would you want?

*Some people play the Lottery in hopes of winning big and being able to quit their jobs or purchase luxury items*. If you won a million dollars what would be the first thing that you would do or buy?

National Geographic recently released a list of the top places in the world to visit. *At the top of the list were India, Portugal, and Argentina*. If you could visit anywhere in the world where would you go?

*Many people go on vacation to relax, so they prefer a warm location on the beach*, but others may prefer a more education-oriented vacation visiting museums and landmarks, while others would like to spend their vacations partying in a busy and bright city. If you had to choose between vacationing at a beach resort country like Jamaica, a museum haven like Washington, or a big bustling city like Las Vegas, which one would you choose?

Okay that is all the questions I have for you for right now. Thank you for bearing with me for those questions!

**Personal Information – More Familiar Condition**

To start off with, I am going to give you a little more information about *myself*. *I am a grad student* here at Carleton University. *I am in my 4th year of studies in the PhD program*, majoring in psychology. *I am 25*-years-old. *I am married and do not have any children*. *I am* attending Carleton on a full-time basis but also have a part-time job. *I am not originally from Ottawa, I moved here from the United States to complete the PhD program here at Carleton*. *I originally had intentions to move back home after finishing school but now I have fallen in love with Ottawa and will likely stay*. *I come from a large extended family but I actually only have one older brother*. Lastly, *in my free time I am usually playing sports or watching TV; I also like to travel when I have the money and time to do so*.

Now that you know a little bit more about *me*, I am going to ask you a few questions about yourself. These questions mostly pertain to your opinions, your interests, and your hobbies. They are standard questions and none of them are meant to be personal. Please know that you are not obligated to answer any of these questions and you are able to say, "I don't know" or "I don't want to answer that question" to any one of the following questions.

The first few questions are going to be about university, and in particular about Carleton University.

When asked, a lot of students say they chose their university for a lot of reasons. For example, *I came to Carleton because I heard that they had a great PhD program and I*

*wanted the chance to work with my supervisor*. What would you say was the main reason you chose to come to Carleton?

*I am planning to be at Carleton for 5 years*. How many years do you plan to be at Carleton?

Most of the students who complete psychology studies for credit are majoring in psychology *like me*, but some students take intro psych or stats just for interest. What program are you majoring in?

*I am in my 4th year of studies*. What year are you in?

*I* attend university on a full-time basis. Are you a full-time or part-time student?

*One of my favourite classes was a Social Psychology class*. What has been your favourite course at Carleton thus far?

In many introductory psychology courses students are graded based on their performance on tests, but sometimes there are opportunities to write an essay or make a presentation for the class instead. *I prefer writing tests*. If you had to choose between taking a multiple choice test, writing an essay, or doing a presentation for marks, which one would you chose?

University courses used to require a lot of independent work; however, more recently efforts have been made to implement more activities that students can do in groups. *Personally, I prefer individual work*. If you had to choose to work on a project by yourself or with a group, which one would you choose?

*I* would prefer night classes over morning classes. Would you consider yourself to be a morning person or a night person?

In the next set of questions, I am going to ask you about your thoughts and experiences of Ottawa.

Students attending Carleton come from all over the world. Are you from Ottawa?

In a poll, people living in Ottawa reported their favourite things about living in Ottawa. *One of my favourite things to do in Ottawa is to walk around the Byward Market or Parliament Hill*. What would you say is your favourite thing to do in Ottawa?

Ottawa and the surrounding area host many festivals and events; *my favourites being the Canada Day celebrations and Ribfest*. What would you say is your favourite festival or event?

Ottawa has many different types of restaurants where patrons can try different kinds of ethnic foods. *One type of food that I like the most is Japanese at the restaurant Japanese Village*. What is your favourite kind of food or restaurant?

In Ottawa, there are many sports or recreational activities that you can participate in professionally or recreationally. For example, *I've tried out recreational rowing and baseball*. Do you participate in any sports?

Ottawa also has a great artistic community and there are opportunities to express your creative side. *Personally, I do not act, paint, sculpt, or play an instrument*. Do you play any musical instruments?

This next set of questions I like to call the "favourites" section. I am going to ask you to think of some of your favourite things.

Many people like country music but *I prefer pop*. What is your favourite kind of music or musician?

Empire has released a list of the top 500 movies ever made. *My top 3 are: The Lion King, The Wizard of Oz, and E.T*. What is your favourite movie of all time?

The television network Fox has many popular shows. *The shows that I watch most frequently are: Survivor, The Big Bang Theory, and Criminal Minds*. What is your favourite television show?

*I usually watch TV during my free time*. If you had to choose between watching TV, reading a book, or going for a walk, which one would you choose?

Next, I will ask you some questions regarding your goals and aspirations in terms of a career.

*I am working part-time as a teaching assistant while I am attending university*. Do you have a part-time or full-time job?

*I am attending university in order to further my education and hopefully increase the chances that I will get a job as a teacher when I graduate*. What job are you planning on getting after you graduate from university?

When determining which career path you want to take you may consider various factors to help you decide which type of employment you want. *For example, it is important to me what type of hours I would be working*. What factor do you think is most important to you when considering a job opportunity?

*I think I would want to retire at about 65 years of age.* What age would you like to retire at?

The last set of questions are a lot more general and may make you think a little bit – but in a fun, no pressure kind of way.

Flash can move at lightning fast speed, Superman can fly, and the Hulk is abundantly strong. *I would like to have superhuman intelligence*. If you could have any superhuman power, which one would you want?

*If I won the lottery one of the first things I would do is buy a house*. If you won a million dollars what would be the first thing that you would do or buy?

National Geographic recently released a list of the top places in the world to visit. *If I could go anywhere in the world I would probably go to New Zealand*. If you could visit anywhere in the world where would you go?

*I go on vacation to relax so I prefer a warm location on the beach*, but others may prefer a more education-oriented vacation visiting museums and landmarks, while others would like to spend their vacations partying in a busy and bright city. If you had to choose between vacationing at a beach resort country like Jamaica, a museum haven like Washington, or a big bustling city like Las Vegas, which one would you choose?

Okay that is all the questions I have for you for right now. Thank you for bearing with me for those questions!

**Appendix C**

Pilot Study Two Questionnaire Form

These questions were presented in a random order for each participant.

1. How familiar are you with the researcher who you met when you started this study session? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not at all familiar Very familiar

2. How well do you think you know the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not well Very well

3. How much exposure do you feel that you had with the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
None A lot

4. How many minutes would you estimate you saw the researcher?_____

5. How accurately do you think you could describe the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not accurately Very accurately

6. How likely would you be able to recognize the researcher in public? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not likely Very likely

7. How confident would be in the accuracy of your recognition of the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not confident Very confident

8. How much information could you provide someone about the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
None A lot

9. How much personal information do you know about the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
None A lot

10. How likely are you to forget that you met the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not likely Very likely

11. To what degree do you think the researcher is a stranger? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not at all a stranger Very much so a stranger

12. How much do you <u>not</u> know about the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Nothing A lot

13. How friendly are you with the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not friendly Very friendly

14. How well acquainted are you with the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not well acquainted Very well acquainted

15. How much do you have in common with the researcher? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Nothing A lot

## Appendix D

Confidence Form

1. How confident are you that you can describe the culprit’s appearance? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not confident Very confident

2. How confident are you that your description of the culprit will be correct? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not confident Very confident

3. How confident are you that you can recognize the culprit? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not confident Very confident

4. How confident are you that your recognition of the culprit will be correct? Please circle your response:

0……10…….20…….30…….40…….50…….60……..70…….80…….90…….100
Not confident Very confident

## Appendix E

Eyewitness Description Form

You have now become the sole eyewitness to a crime. The police are requesting your assistance in capturing the culprit. Please answer the following questions as thoroughly and accurately as possible.

Please describe everything that you can remember about the crime that occurred in the video.

_____

_____

_____

_____

_____

_____

_____

_____

Please describe everything that you can remember regarding the culprit.

_____

_____

_____

_____

_____

_____

_____

Sometimes crimes are committed by individuals with whom the eyewitness has seen or met before. Have you seen or met the culprit before?

☐ Yes ☐ No

If yes, is there any additional information that you may provide police based on your previous encounter with the culprit?

_____

_____

_____

_____

_____

_____

_____

_____

_____