DEVELOPMENTAL DIFFERENCES IN THE USE OF INTENTION AND OUTCOME IN FACE RECOGNITION IN AN EYEWITNESS IDENTIFICATION CONTEXT

by

TONI SPRING

A dissertation submitted to the Graduate Faculty in Developmental Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The Graduate Center of The City University of New York

2009

This manuscript has been read and accepted for the Graduate Faculty in Developmental Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Herbert D. Saltzstein

Date

Chair of Examining Committee

Maureen O'Connor

Date

Executive Officer

Steve Penrod

Margaret-Ellen Pipe

Jennifer Dysart

Maureen O'Connor

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

By

Toni Spring

Advisor: Professor Herbert D. Saltzstein

Young children's (ages, 7-9), older children's (ages, 10-12), adolescents' (ages, 13-15) and adults' (ages, 19-25) (total *N*= 122) use of intention and outcome in eyewitness identification was investigated in this study. The research strategy here was to study children's eyewitness identification tasks within the theoretical perspectives of moral development and decision-making theory, in particular how the intentionality and outcome of a transgression affected their eyewitness identification, both in its accuracy and in the kinds of errors made. In particular, it focused on the balance between false positives and false negatives. A false positive [F+] is saying that the person is guilty when he/she is, in fact, innocent. A false negative [F-] is saying the person is <u>not</u> guilty when he/she is, in fact, guilty. The identical act of setting a fire was framed in two different ways: (a) as a "neutral" intention act which resulted in a fire that ruins the restaurant; and, (b) as a "bad" intentional act of trying to set a fire that does not result in any damage to the restaurant. The developmental difference is precisely cast when the framings are mismatched between intent and outcome. Thus, the stimulus films were framed differently, by stressing the "neutral" or "bad" intentions of the actor and the greater or lesser severity of the outcome. The identification of the perpetrator was contrasted for the two acts (films) at four different age levels.

The major findings, of this study are: (1) that different moral framings of the act [transgression] affected the decisional criteria actually used by the 10-12 year-olds but not those used by the younger children [ages 7-9 year-olds] nor by the adolescents or adults. In particular, a) the younger children appear to use a general strategy, 'when in doubt, positively identify' regardless of the nature of the act; b) whereas the 10-12's appear to treat the neutral intent/ bad outcome (Fire) as more serious than the bad intent/neutral outcome (No Fire), and deserving of a more lenient decisional criterion that results in more false positive identifications. (2) However, both groups of children [7-9 and 10-12] explicitly state that false negatives [saying the person is <u>not</u> guilty when he/she is, in fact, guilty] are worse than or equally bad as false positives [saying that the person is guilty when he/she is, in fact, innocent] whereas adolescents and adults state that false positives [F+] are worse. This latter finding of stated beliefs appears not to differ according to the framing of the act, i.e., is not responsive to condition. Thus, there appears to be a discrepancy between performance and [stated] beliefs.

Acknowledgements

If a person's life was recorded in the format of a book the chapter on Achieving Academic Excellence would be one of the longest and certainly the most dynamic. One of the undeniable truths of this chapter on the quest for a Doctoral degree would be that all great journeys and all arduous undertakings are seldom accomplished alone. The support, encouragement and guidance we receive on our quest are paramount to our success. It is with the greatest respect and admiration that I acknowledge the following people.

I would like to gratefully and sincerely thank Dr. Herb Saltzstein for his guidance, enthusiasm, patience, and most importantly, his friendship during my doctoral studies at The Graduate Center. He encouraged me to not only grow as a student, but also as a researcher.

This dissertation would not be possible without the help of my doctoral committee; Herbert Saltzstein, Steve Penrod, and Mel Pipe. I am especially thankful for the committee's input, valuable discussions and accessibility. In addition, I'd like to thank Maureen O'Connor and Jennifer Dysart for taking the time to be my outside readers.

A special thank you goes to Teri (Yanxing) Cao for producing my film with such expertise. A loud thank you goes to David Barnard for giving a "voice" to my film as well as a face in the line-up. Thanks to Juraci da Silva for the use of her restaurant as the perfect setting for my film. A big hug goes to Yoko Takagi for the excellent help in formatting and for being a great friend. I would like to thank my Graduate School friends, Vienna, Rivka, Maria Helena, Anna, Felix, and Christina for their support and

encouragement throughout the years at the Graduate Center.

I thank my parents, Karin and Robbie, for encouraging me through the various stages of my graduate life. An enormous hug and big thanks go to my children, Samantha and Jake for their patience while I was writing this dissertation. Finally, and most importantly, I would like to thank my husband, Larry. His support has truly been the wind beneath my wings

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

Introduction

This research study illustrates the importance of the moral significance of an anti-social act for face recognition in the eyewitness context, specifically how different kinds of framing of the act affects decisional criteria. This approach may be particularly important when studying eyewitness identification by children and adolescence, whose moral framing of events is dramatically changing. The research strategy used here has rested on the combination of two very disparate approaches: a) decisional theory, in particular, signal detection theory, and b) moral development theory and research. The introduction is divided into four sections: first, a brief review of the main relevant findings from studies of children's eyewitness identification; second, a brief overview of Signal Detection Theory and how it might provide a way of testing hypotheses in this area, and third, a brief review of Piaget's theory of moral development and how it might help explicate some of the main developmental trends in eyewitness identification. The final section outlines the questions and hypotheses.

Overview of Research of Children's Eyewitness Identification

Criminal courts have traditionally relied heavily on eyewitness identification when trying suspects. In many criminal investigations an eyewitness's affirmative response in a line-up may be the only or the strongest evidence against a suspect. Recently, the use of DNA has confirmed the long suspected concern about the unreliability of much eyewitness identifications (Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998; Scheck, Neufeld & Dwyer, 2000). Before and since the introduction of such physical evidence, experimental studies on eyewitness identification

tasks have assisted researchers and practitioners in addressing some of the reasons for mistaken identifications (e.g., Cutler & Penrod, 1995; Loftus, 1979). In many studies, participants view a live staged event or watch a video of a staged crime and then are presented with a photo array, which may be presented simultaneously or sequentially. In these eyewitness identification studies participants are then asked to indicate whether they see the perpetrator in a yes-no recognition task and how confident they are of their response. Researchers have frequently found that eyewitnesses have poor accuracy although this poor performance is often accompanied by high levels of confidence (Meissner, 2002).

Many factors such as estimator and system variables (Narby, Cutler, & Penrod, 1996; Shapiro & Penrod, 1986; Wells, 1978; Wells and Olson, 2002) have been found to affect eyewitness performance. A system variable is one that is under control of the criminal justice system, while an estimator variable is not under the control of the criminal justice system. Some examples of the variables that can influence eyewitness lineup performance are weapon focus (e.g., Loftus, Loftus, Messo, 1987), stress and intoxication (e.g., Read, Yuille, & Tollestrup, 1992), post-identification feedback (e.g., Wells & Bradfield, 1998) and witness vs. perpetrator ethnicity (e.g., Meissner & Brigham, 2001). In their meta-analysis Meissner & Brigham (2001) report that race may play a role, such that participants are better able to recognize faces of their own race than faces of another race, although a recent dissertation study (da Silva, 2007) has suggested that the phenomenon may be more complicated than is usually claimed.

Accuracy of identification has also been shown to be related to age of the eyewitness. The legal requirements for child witness testimony have relaxed and

children as young as 3 have occasionally served as witnesses (Ceci & Bruck, 1993; Yuille & Wells, 1991). Some recent research has shown that under some circumstances, children as young as 3 can recount reliable narratives of past events (Butler, Gross, & Hayne, 1995; Fivush & Shukat, 1995; Goodman & Bottoms, 1993; Pipe, 1993; Salmon, Bidrose, & Pipe, 1995). Nonetheless, past research on the developmental pathways of [accuracy in] eyewitness identification has been often quite contradictory. Some of the questions that have been addressed in the research on the influence of age and development on children's eyewitness testimony are 1) Are young children generally less accurate in recalling events? ; 2) How does suggestibility affect the child's memory? ; 3) Does eyewitness identification follow a developmental trajectory? ; 4) What kind of information do children use in their eyewitness identifications? and, 5) Do they process the information differently than adults?

Many laboratory studies have focused on children's memories where the participants are asked in a standard sequential test, whether or not they have seen that face before (Chance and Goldstein, 1984; Cole and Loftus, 1987). Research studies that have used these paradigms have often shown that children's memory for unfamiliar faces improves with age (Carey, Diamond & Woods, 1980; Flin, 1980; Sophian & Stigler, 1981). Some research on face recognition abilities has shown a clear trajectory of improvement in face recognition accuracy from 5 years of age through 12 years of age (e.g., Chance & Goldstein, 1984). In Chance and Goldstein (1984) they report that for young preschool children accuracy typically ranges from 35% [and increases] up to 70% for 12- to 14-year-olds and adults. However, other research on the same phenomenon shows no developmental trajectory at all (e.g., Roebers, Moga & Schneider, 2001).

Children's performance on facial recognition tasks (as opposed to eyewitness identification tasks) under laboratory conditions may result in different findings than in children's performance that more closely resembles an eyewitness task that resembles a real-life situation. What most research suggests regarding children's memory is that free recall is more accurate, but when children are asked for a forced choice (i.e., yes or no) response, they are less accurate (Chance & Goldstein, 1984; Cole & Loftus, 1987). For example, recent research studies found that children's confidence judgments were found to discriminate correct responses from incorrect ones on open-ended unbiased recall questions (Roebers, 2002; Roebers & Howie, 2003).

Perhaps related to these developmental trends in accuracy and in particular, the tendency to make more false positive errors is the finding, well-summarized by Ceci and Bruck (1993) that there are highly "significant age differences in suggestibility, with pre-school aged children being disproportionately more vulnerable to suggestion than either school-aged children or adults and thus commit more false positive errors. The literature clearly reveals age differences in overall suggestibility, although the exact mechanisms involved in producing distortion in young children's reports are still being debated" (Ceci & Bruck , 1993, p.43). Thus, since suggestions are typically directed to saying that an individual is the perpetrator rather than the reverse; suggestions may exacerbate an already latent tendency to commit a false positive. However, it should be noted that some other researchers (e.g., Goodman, Rudy, Bottoms, & Aman, 1990) have found that when children are given misleading information regarding events that they have experienced as personally involving their own bodies, the children are less prone to suggestion. Thus, controversy characterizes the issue of the pervasiveness of suggestibility.

Decision-Making Strategies in Identification Lineups

Research studies that focus on eyewitness decision-making strategies have utilized two forms of lineups: sequential and simultaneous. In studies that use sequential lineups, photos are presented one at a time, and subjects are prompted to make a decision prior to seeing the next photo. In a simultaneous lineup, the photos are all shown at the same time and the subject is asked to pick out the perpetrator out of the line of photos. Lindsay and Wells (1985) have found in their study that there were fewer false positive errors made in the sequential lineups than in simultaneous lineups. Studies that have compared simultaneous lineups to sequential lineups, have found that sequential lineups reduce false identifications and increase correct rejections on target-absent lineups without substantially reducing the rate of correct identifications from target-present lineups (Cutler and Penrod, 1988; Kneller, Memon, & Stevenage, 2001; Lindsay & Bellinger, 1999; Sporer, 1993). This may be because simultaneous lineups encourage a relative judgment strategy to be used that is less accurate whereas a strategy of using absolutes judgment thus, encouraging judgments "based largely on evaluating each face individually" (Lindsay & Wells, 1985, p.558) which promotes greater accuracy. However, in a recent meta-analysis that examined the contrast between simultaneous and sequential lineups argues that sequential lineups can be preferred under certain conditions (Steblay, Dysart, Fulero, & Lindsay, 2001). Steblay et al. (2001) found that when the target-absent and target-present data are separated, the sequential lineups are better for minimizing false positive identifications (9% vs. 27%) and increases correct lineup rejections (72% vs. 49%). However, when the target is present, simultaneous lineups are better for achieving correct identifications (50% vs. 35%) and reducing false lineup

rejections (26% vs. 46%). Research studies that have used signal detection theory as the method of analysis also use the sequential lineup because accuracy and response criterion rely on multiple identification decisions.

However, reasonably well-established in the children's eyewitness testimony literature are the following: children's memories are often faulty and frequently include memories of events that never happened this is especially true of young children when they are asked yes and no questions perhaps closely related is young children's susceptibility to interviewers' suggestions; and finally witnesses' confidence in their own judgment is no guarantee of accuracy. Wells and Olson (2002) found that in many of the wrongful convictions "of innocent persons involved eyewitness identifications, and in every case, the mistaken eyewitnesses were extremely confident" (p. 155). A meta-analysis on children versus adult witnesses by Pozzulo and Lindsay (1998) found that very young children perform significantly worse than young adults on witness identification. However, when the perpetrator is in the line-up (target present) children perform nearly as well as young adults, but when the perpetrator is absent (target absent) from the lineup, children significantly commit more erroneous identifications than the young adults (Pozzulo & Lindsay, 1998; Cutler & Penrod, 1995). Again, this difference may be due to the young child's propensity to commit false positives.

Many of the current studies on children in eyewitness tasks suggest that young children are particularly prone to make more false positive than false negative errors and to over-guess in eyewitness identification tasks (e.g., Pozzulo & Lindsay, 1999; Cutler & Penrod, 1995). Pozzulo and Lindsay (1999) conclude that, an elevation in false positives

is a serious problem "reducing the credibility of children's identification evidence" (p. 175). Brewer and Day (2005) examined the eyewitness confidence-accuracy and decision latency relations in a sample of children aged 8-10 years and a group of adolescents and found "that young children were perhaps more impulsive in their identification responses" concluding that "neither confidence nor decision latency has promise for distinguishing accurate from inaccurate identification responses in children" (p.126). Thus, there is evidence to suggest that in young eyewitness' identification tasks poor accuracy may not just mean poor memory, but that they are employing a response strategy that *when in doubt – identify positively.*

A theoretical explanation for this apparent bias has remained elusive to date. One explanation may be how children develop an understanding of the consequences of committing false positives and false negatives. In other words, children may feel that the consequences of letting someone off, who might be guilty (false negative) are worse than the consequences of imprisoning someone who might be innocent (false positive) (Saltzstein & Peach, 2005). Although a rather extensive literature has now been developed concerning children's eyewitness testimony, with the exception of the few studies cited here, never to my knowledge has the focus of children's eyewitness testimony been studied through the lens of moral development. However, since eyewitness testimony has implications for morality and the practice of justice, the interest in studying children's eyewitness testimony may be particularly relevant. The research strategy here is to study children's eyewitness identification tasks within the theoretical perspectives of (a) decision-making theory and (b) moral development. Here, it is necessary to introduce the decision making perspectives, and, in particular,

Signal Detection Theory.

Basic Vocabulary of Signal Detection Theory

Signal detection decision-making theory (SDT) (Green & Swets, 1966) is a mathematical method of depicting the decision making process by which someone decides between different items (e.g., perpetrator and foils) and analyzes the process to its two underlying parameters. In decision making terms, there are four possible outcomes in any identification: (1) If a signal is present and a person correctly identifies the signal, then he/she has made a ‘positive hit’. (2) If a signal is not present and a person correctly does not identify a signal, then he/she has made a 'negative hit' or a correct rejection. (3) If the signal is absent and he/she says that the signal is present, then she/he has made a ‘false alarm’ or a false positive error. (4) If the signal is present and the person says it is absent then he/she has made a 'miss' or a false negative error.

The underlying model of SDT (Macmillan and Creelman, 1991; Swets, 1986) consists of two normal distributions, one representing a signal and another representing noise. How well a person can discriminate between ‘Signal Present’ and ‘Signal Absent’ trials is represented by the difference between the means of the two distributions, d'. The willingness of the person to say 'Signal Present' in response to an ambiguous stimulus is represented by the criterion (C). The logic of the SDT model is very similar to statistical hypothesis testing. The ‘Signal Absent’ distribution corresponds to the null hypothesized distribution, the ‘Signal Present’ is the alternative distribution, and the criterion is the alpha error rate. A research domain where SDT has been successfully applied is in the study of psychophysics and more recently in studies of memory (Van Zandt, 2000;

Vickers, 1979). The results of the experiment can be portrayed in what is called a decision matrix (Figure 1).

	Perpetrator	**Foil**
Signal	**Hit**	**False Positive**
Signal + Noise	**False Negative**	**Correct Rejection**

Figure 1.
Decision matrix

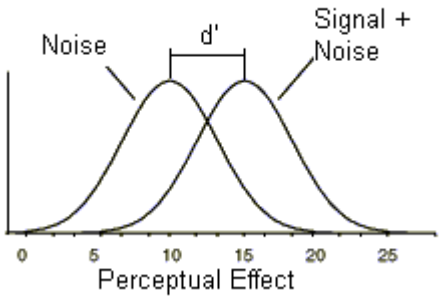


Figure 2.
Signal detection (*d'*) the ability to discriminate between Signal and Signal + Noise (Heeger, 1998)

A participant's sensitivity, as indexed by d', is how well the participant can differentiate items coming from the Signal Absent and Signal Present distributions (Figure 2). In other words, sensitivity measures the participant's ability to discriminate

between Signal and Signal + Noise. The greater the sensitivity score the more it translates to the higher rate that the individual is capable of discriminating between the perpetrator and the foils. Decisional criterion (i.e., response bias) represents the minimum level of internal certainty needed for the participant to decide that a signal was present (Figure 3).

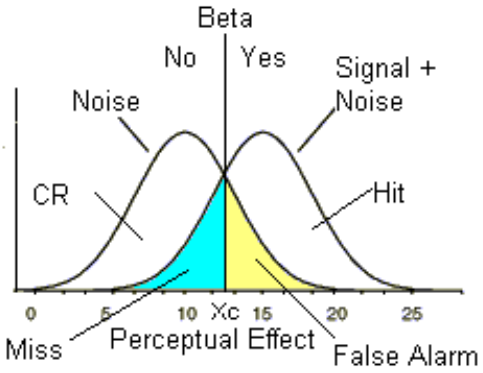


Figure 3.
Decisional criterion (The minimum level of internal certainty to decide that a signal was present) (Heeger, 1998)

Receiver Operating Characteristics (ROCs) represent the relationship between hits and false alarms, and can be used to describe performance in terms of d' (Figure 4). This is used to portray the equivalence of sensitivity across changing levels of biases. Each signal detection condition generates one point on the ROC. If the signal strength and the participant's sensitivity remain constant, then by changing Beta from one block of trials to another, a series of points are produced that move from a more conservative response to a more lax response. When connected these points make the ROC curve. As the response criterions are changed, the point on the individual curve changes, not the curve

itself.

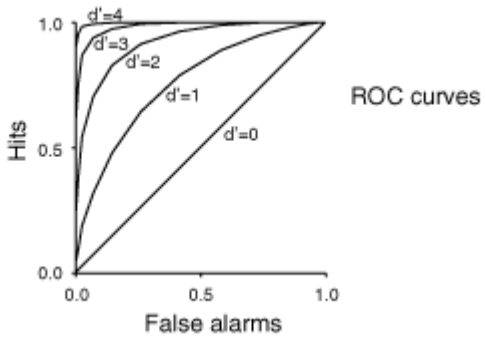


Figure 4.
ROC (the relationship between hits and false alarms in terms of d') (Heeger, 1998)

When d' is 0, the noise and the signal + noise distributions are the same and the number of false alarms and hits will be the same. That is represented by the diagonal in the figure. As d' gets larger, the curve bows away from the diagonal until at extreme values it is along the outer walls of graph. As d' gets larger, false positives decrease and hits increase. The most commonly used [by Green & Swets, 1966] SDT measure of sensitivity is d' (d prime), which is the standardized difference between the means of the 'Signal Present' and 'Signal Absent' distributions. The equation to calculate d' is as follows: d' = z(FA) - z(H) where FA and H are the False Alarm and Hit rates, respectively, that correspond to right-tail probabilities on the normal distribution. Thus, z(FA) and z(H) are the z scores that correspond to these right-tail p-values represented by FA and H. Larger absolute values of d' mean that a person is more sensitive to the difference between the 'Signal Present' and 'Signal Absent' distributions. Thus, d' values near zero indicate chance performance.

Bias (C), on the other hand, is the measure of the criterion used. A bias score of 0 indicates that the criterion is placed so as to equalize false positive and false negative

judgments. More positive bias scores indicate a more stringent or conservative bias score, indicating fewer false positive judgments relative to false negatives. This measure of decisional criterion measure where the eyewitness sets the decision (*very sure that it isn't the man, a little sure that it isn't the man, a little sure that it is the man, very sure that it is the man*) which indicates whether the eyewitness has a propensity to make one type of error over another, i.e., false positives over false negatives independent of accuracy. The typical eyewitness paradigm prevents a direct application of signal detection theory since in signal detection parameters of discrimination accuracy and response criterion rely on multiple identification decisions in which sometimes the target is present and other times when the target is absent. Meissner, Tredoux, Parker and MacLin (2005) introduced their novel lineup recognition paradigm in that participants were presented with multiple target faces from a series of target-present and target-absent lineups in order to apply the signal detection theory model. Their findings demonstrated "that individuals presented with a sequential lineup were significantly more conservative in their identification responses than individuals provided with a simultaneous lineup array" (p. 786). In the next section, will briefly outline the tenets of Piaget's theory of moral development and how it might be related to the changing or meaning of the event to be judged, and how this, in turn might affect how the child remembers the event and makes eyewitness identifications.

Piaget's Theory of Moral Development

Piaget (1932/1965) emphasized the close connection between the child's cognitive ability and his/her moral judgment. One of the key factors of his theory is that

moral judgment can be best understood from how the child views and understands rules. Initially, the rules he referred to were from how children played and reasoned about everyday games, such as marbles. He found that young children's reasoning can be characterized by: 1) Duty, as viewed by young children, is *heteronymous*. That is, an act that shows obedience to a rule or parental command is *ipso facto* good; any act which does not conform to rules is bad. 2) The letter of the law should be observed rather than the spirit of the law, and laws are immutable and cannot be changed by mutual agreement. 3) *Moral realism* is the belief that moral good depends on the outcome of an action rather than the intent behind the act. This *moral realism* produces a conception of responsibility in young children that rules should be followed rigidly without regard to their purpose. In other words, a young child will first evaluate acts not in accordance with the motive and intent [in law, the *mens rea*] that prompted them but in terms of their exact conformity with established rules/laws and their concrete outcome (e.g., damage done).

One of Piaget's (1932/1965) major findings in studying young children's responses to moral dilemma vignettes is that when judging culpability, the young child tends to focus on the outcome rather than on the intent of the action. That is, younger children's judgments differ from older ones and adults in that the younger children tend to give more weight to the amount of damage caused by a wrongful act than to the actor's intentions. For example, Piaget asked children which of two boys – well-intentioned John, who is trying to help his mother wipe the dishes, breaks 15 cups, or mal-intentioned Henry, who breaks 1 cup while stealing some jam- is naughtier and why. When children are questioned about the actor's intentions while paired with little to great damage young

children judged the actor John as being naughtier. While some researchers (Austin, Ruble, & Trabasso, 1997; Nummendal & Bass, 1976; Feldman, Frankel, & Sharp, 1971) have claimed that this developmental shift from outcome to intent is an artifact of method (e.g., order of presentation of the intent and outcome information) the developmental shift in emphasis for judgments appears to be a robust one, (e.g., Austin, Ruble, & Trabasso).

However, Piaget's original explanation of the shift has not been as consistently supported. He asserted that young children do not center or focus on the intent but rather center on the more concrete outcome. Several studies have shown that young children can describe action sequences as intentional or accidental quite accurately (e.g., Shultz, 1980; Berndt & Berndt, 1975) and yet still ignore intentions when judging the wrongness of the act. Baron, Granato, Spranca, & Teubal (1993) found that younger children use both intention and consequence information in their moral judgments in his study with second and seventh grade students. Yuill (1984) reported that 3-year-old children are able to distinguish accidental and intentional outcomes in their judgments of a character's emotional reaction to an outcome when they are given information about motive and outcome. Additional research has demonstrated that young children can use information about intentions in their moral judgments. For example, Armsby (1971) carried out investigations with moral dilemmas and found that, although younger children had some conception of intent, they still preferred to judge in terms of consequences because they found this easier. Thus, even though research has shown that children learn to distinguish intentional from unintentional acts already by the age of 3 to 4 years (i.e., Feinfield, Lee, Flavell, Green, & Flavell, 1999; Flavell, 1999), the children's ability to make distinctions

does not preclude the possibility that the relative weight of outcomes versus intentions may still shift with age, in favor of intentions, as originally claimed by Piaget (1932/1965). The relevance to forensic phenomena may already be apparent. In most criminal court proceedings the jury must infer the 'mens rea' or state of mind accompanying the act. In other words, the jury has to consider whether or not the defendant was acting with guilty intentions, malice, and foresight when the act was committed.

Piaget also reported that young children believe in the appropriateness of collective punishment (i.e., punishment of the group for the actions of the individual). He attributed this belief on the part of young children to their belief that infractions are, must, and should be automatically followed by punishment. He noted that older children sometimes approve of such group punishment but provide a rationale for it (e.g., that the group knows but would not help in punishing the guilty individual). A recent study by Tsukamoto and Saltzstein (2009) has confirmed this developmental trend with Japanese school children. The young one's preference for collective punishment, i.e., punishing the group for the act of an individual member, may render them also to incline to see guilt and punishment where the former is uncertain and the latter unjustified.

According to Piaget the general developmental changes from a focus on outcome to intent and to the appropriateness of collective punishment are largely due to increasing participation in the presence of the peer group and the need to establish social rules of fairness and cooperation in order to play and form peer groups. During this developmental phase of autonomy in the child, morality takes on a subjective character, in the sense that the rules that are to be followed are his/her own and maybe altered with

the consent of others and that intentions (in law, the '*mens rea'*) are key aspects to culpability. A sign of this transition to a more autonomous morality is the realization by the older child that the actors' intentions, not the act's outcomes, should serve as the basis of judging behavior. How might these shifts in moral orientation be reflected in children's eyewitness identification?

In most criminal proceedings the jury must consider the *mens rea*. In other words, the jury must consider whether the defendant was acting with guilty intentions, malice and foresight or in a similar fashion when he/she committed the act. The implications of the above are that young children should have difficulty in keeping focused on the actor's intentions, i.e., on the *mens rea*, and therefore judge more in terms of outcome of the act and the rules or laws apparently broken. Further, young children's belief on the focus of outcome versus intent, in collective punishment, and their preference for harsh, arbitrary or *expiative* justice punishment may incline them to seek punishment even under uncertain conditions. If children are inclined to seek punishment when there is a negative outcome, it may help to explain why young children are more willing to commit more false positive errors than adults.

Previous Research Using a Moral Decision-Making Framework

Developmental theory and previous research by Saltzstein and Peach (2005), suggest that younger children may believe that it is worse to let someone who is guilty go free than to convict an innocent person whereas older children believe the reverse. Thus, the overarching hypothesis would be that younger children prefer false positives to false negatives whereas older children would show the opposite preference, both in

performance and (stated) beliefs. This developmental difference was hypothesized to be *independent of differences in overall accuracy*. Saltzstein and Peach's results were that: (1) younger children are less sensitive (less discriminating between perp and foils) than are older children, (2) adolescents used a stricter decisional criterion than both older and younger children, independent of accuracy, (3) younger children reasoned in less moral terms when asked about the seriousness of false positives and false negatives than did older children or adolescents, and more of the older children chose false negatives as worse than false positives whereas adolescents said that both kinds of errors are equally bad.

Follow-up Study

A follow-up study by Spring (2006) tested the external validity of the findings by Saltzstein and Peach (2005) by repeating the study with children from two different communities and regional areas on Long Island, NY. The two communities differed in general socio-economic advantages, thus the labels "disadvantaged" and "advantaged". The poverty rate (percentage of students eligible for free or reduced lunches for the 2003-04 school year) in the former was 66.9% and in the latter, 7.9%. Participants were drawn from two age groups; younger children aged 7-9 years; older children aged 13-14 years. The second aim of this study was to investigate whether children's eyewitness performance and beliefs can be influenced by instructions, which emphasize the dangers of false positives or of false negatives. The method was essentially the same as in the study by Saltzstein and Peach (2005) except that biasing instructions were randomly introduced to bias the child towards favoring false positives [F+] or false negatives [F-]

and the moral framing of the act was depicted as stealing.

The study had a 2X3X2 factorial design with three between-subject factors: community (2), age of subject (3) and experimental biasing condition (2). There was computer administration of the task to allow for data collection by one experimenter and (b) the introduction of biasing instructions (See below). All the children were asked to view the film and to pretend that they were witnesses to the theft. Additionally, the children were told that their task was to identify the man who stole the bag.

All 51 participants, ages 7-9 (12 male, 13 female) and ages 13-15 (13 male, 13 female) were shown a short video depicting four men, all dressed in identical red t-shirts, practicing soccer in a locker-room. At the end of the video one of the men leaves carrying two sports bags, one of which had been brought into the locker room by one of the other men. The task was framed by asking children to pretend that they were witnesses to a crime and that their task was to identify the man who stole the bag. After watching the film participants were asked to describe what they had seen. One of two different biasing instructions was randomly read to the children prior to viewing the sequential line-up on a random basis. The children's stated beliefs about false positives and false negatives were elicited by the post-task questions. Each child was shown the last false positive and the last false negative (in counter-balanced order) and the interviewer asked the child to imagine that he or she had made an error. How bad a mistake would that be: very bad, a medium bad or a little bad, and why? Then the two photos were held up and the child asked to choose, which kind of mistake is worse and why? These reasons were coded according to whether they showed concern for convicting the wrong person, letting the guilty party go free [both moral concerns] and other, non-moral concerns (e.g., task

difficulty).

The results of the follow-up study were as follows: (1) age differences in accuracy were consistent with the Saltzstein and Peach's study and previous research; older children and young adolescents were better able to differentiate the perpetrator from foils than younger children in both communities; (2) differences in bias require some explanation. Using the confidence rating system method, there are three possible measures of bias based on how the four confidence ratings are dichotomized, for example, between *very sure* it wasn't the man and *a little sure* it wasn't the man. The only dichotomy which showed consistent differences in bias was based on dichotomizing between a *little sure* that it was the man to *very sure* it was the man. Here, the results were significant and consistent, as indicated by the confidence intervals, namely, the early adolescents in both communities showed a significantly stricter criterion than the younger children. There was also a main effect for community such that those from the disadvantaged community were more willing to say that they're definitely sure that someone is the guilty individual than did the advantaged community; i.e., their bias scores were lower, than the children from the advantaged community. It should be noted that the age differences were obtained in both communities. This pattern both confirms the earlier findings just reported (Saltzstein & Peach, 2005) and also shows that socio-cultural community background, defined as a combination of social class and ethnicity, also affected performance bias; (3) biasing instructions had no effect on children's performance in either community.

When we turn to the interview data however, the results are more complicated and

not altogether consistent with performance. The reasons were coded as to whether they focused on concern for fairness to the accused or concern for punishing the guilty or a non-moral or “I don’t know” response. These two moral reason responses were combined and contrasted with other (non-moral reasons e.g., "it (the task) wasn't that difficult") and also contrasted with one another. We also examined which kind of error they chose as worse, regardless of the qualifying reason they gave.

The results of the follow-up study were: (1) a tendency for the older children and those from the advantaged community to give more moral reasons (justifications), (2) a greater tendency for older children, but only in the disadvantaged community, to express more concern for protecting the innocent than for punishing the guilty; (3) a tendency for older children to choose false positives as a worse outcome, but again only in the disadvantaged community. These interactions between age and community were not due to ceiling effects since not all of the older children in the advantaged community chose false positives as worse. Confirming previous age differences in sensitivity, the predicted age differences in the stringency (or looseness) of the judgmental criterion used were obtained in both higher and lower SES samples; (4) biasing instructions had no effect on eyewitness performance; and, (5) younger children did choose false negatives as worse than false positives whereas older children and adolescents chose the reverse.

The primary purpose of this dissertation is to build upon the previous research by examining whether developmental differences in the child's understanding of the moral meaning of the act affect their eyewitness performance in identifying the perpetrator of the act. Data from Saltzstein and Peach (2005) suggest that the framing of an event in moral and neutral terms affects the eyewitness’s performance in identifying the

perpetrator, especially as to the witness's decisional criteria, depending on the age of the child.

The Present Study and Research Questions

In this dissertation study the focus is on both these aspects of participants' performance (accuracy and bias), but especially on bias (decisional criteria). There is reason to believe that, in addition to overall age changes in accuracy, the balance between false positives and false negatives changes with age, specifically, that younger children typically use a more lenient criterion and thus are more likely to misidentify an innocent person as the culprit, committing a false positive whereas older children and adolescents are more likely to use a stricter criterion and therefore are more likely to commit a false negative or let a guilty person go free. Research with younger children (6-8), older children (10-12), and young adolescents (13-15) support this conclusion (Spring, 2006). The focus of the current study is to extend the research to look at the effect of the intentionality of a transgression and how this might make eyewitnesses more or less stringent in the eyewitness identification task and to see whether this can affect the balance between false positives and false negatives, depending on the age of the witness. In other words, with regard to eyewitness performance, changes in the way the individual frames or understands the act should affect their eyewitness identification, in particular the decisional criterion used (C) that is different for the different age groups.

In addition, this study will investigate children's beliefs about making a false positive or a false negative error. According to Western jurisprudence, false positives are worse than false negatives, as in, "better to let ten guilty men go free than to convict one

innocent man". But, do (Western) children also think that? Developmental theory and the previous research cited suggest that they do not; that younger children may believe it is worse to let someone who is actually guilty go free than to convict an innocent person. This is suggested by the developmental course of children's moral judgments, as described above.

This study will explore the effect of differential framing (construal) on eyewitness identification of moral transgressions with age by focusing on the outcome of the act versus the intentions of the actor. The identical act of setting a fire was framed differently, by changing the intentions of the actor and the severity of the outcome. This difference in framing should affect the child witness's performance in interaction with age. Thus, interaction effect between the child's age and framing on decisional bias (C) was expected, rather than a simple main effect for age and for framing. In this dissertation study the question is how children of different ages use information about intentions and outcome in identifying the perpetrator, and in particular, in setting a decisional criterion. We expect to find that a developmental difference in the child's understanding of the moral meaning of the act will affect their eyewitness performance in particular the decisional criterion used (C). Specifically, we expect that younger children (7-9 and 10-12 year-olds) would use a looser criterion and commit more false positive errors in the unintentional/bad outcome framing than in the bad intention/neutral outcome framing, whereas the adolescents and adults will use a stricter criterion. In other words, younger children and older children will view the bad outcome/neutral intent (fire) setting as more morally important and therefore commit more false positive errors and use a lower decisional criterion in that framing condition. In contrast, the adolescents and

adults will view the bad intent/neutral outcome (no fire) act as more morally significant and will use a stricter criterion (be more cautious) regardless of condition framing. Regarding beliefs, when participants are asked which type of error is worse (their beliefs of *how bad* it is to make a false positive or false negative error) we expect that younger children will believe that the false negative errors are worse than false positive errors independent of their performance and of experimental (framing) condition as was the result in previous research (Saltzstein & Peach, 2005; Spring, 2006).

The research design to test this hypothesis is a 2X4 factorial design; two conditions (neutral intent/ bad outcome (fire) and bad intent/neutral outcome (no fire)) by four age groups (7-9, 10-12, 13-15, and 19-25). Thus, the primary purpose of this investigation is to follow up the previous findings to demonstrate that developmental differences in the child's understanding of the moral meaning of the act affect their eyewitness performance in identifying the perpetrator of the act and, in particular, the leniency or stringency in the decisional criterion that children use. Our focus is on both these aspects (accuracy and decisional criterion) of the participants' performance, but particularly on bias (here the decisional criterion set). There is reason to believe that, in addition to overall age changes in (general) accuracy, the balance between false positives and false negatives changes with age, specifically, that younger children typically use a more lenient criterion and thus are more likely to misidentify an innocent person as the culprit, committing a false positive error whereas older children and adolescents are more likely to commit a false negative error (or let a guilty person go free). The focus of this study is to extend the research to look at the effect of framing the act to emphasize the intentions behind the act or the bad outcome resulting from the act (intentionality and

outcome of the transgression) and how these two factors might affect the stringency of the eyewitness' decision to identify and thus might affect the balance between false positives and false negatives. In other words, if the framing of the act affects eyewitness identification, changes in the way the individual frames the act, including developmental influences on the framing, should affect their eyewitness identification, in particular the decisional criteria used.

Method

Participants

One hundred and thirty-six participants recruited from Hunter College Community School, Forest Hills Junior High, and Queens College, CUNY New York City, participated with the study. Fourteen adult participants were removed from the study due to incomplete data (leaving some face recognition choices blank) on the eyewitness identification task making the final sample size, 122 participants, ages 7-9 (14 male, 14 female, M = 7.9), 10-12 (16 male, 14 female, M = 10.9), 13-15 (10 male, 18 female, M = 13.4), and 19-25 (8 male, 28 female, M = 23.8). School administrators distributed consent forms to all students in the appropriate grades who attend each school. Parent permission for the three younger groups and participant permission for the adolescent and adult group was obtained.

Design

The design of this study is a 2X4 design; film versions (intended to cause a fire/no fire neutral outcome; did not intend to cause a fire a neutral intent/fire bad outcome) by four age groups (7-9, 10-12, 13-15, and 19-25). The main data of sensitivity, bias, false positive and false negative errors was analyzed using signal detection. The moral judgment data (how bad was the act, the actor and the outcome) and the moral reason data (which error is worse and why) were coded by two independent raters and reliability was on (average) 95%.

Materials

Stimulus events. A professionally-produced film was shot in a small restaurant in SoHo, New York City. All dubbing and editing was done at the NYU School of Film. Actors in the film were hired through the NY Casting Agency. Each participant in the study was shown a brief film depicting a group of five young men in a restaurant celebrating one of their birthdays. After two minutes, the waiter tells them that they must leave the restaurant immediately and one of the men throws a lit birthday cake into the garbage can. There are two versions of the film, which frame it differently so as to give each one a distinct moral meaning by means of a voice over (See below). There was no audible dialogue among the actors in either film version. The film lasted 127 seconds, with the perpetrator visible for the entire time. He was filmed from various angles, with his face available from right profile for 90 seconds, left profile for 10 seconds and full profile for 27 seconds.

A. Voice over for Film Version 1 (*Neutral Intentions with Bad Outcome*):

"Look, there are four men at that table with a lit birthday cake! It must be someone's birthday. Uh oh, there is an announcement that the restaurant is closing. They must leave quickly. They have no time to enjoy their cake!" [They blow out the candles and one man throws the cake in the garbage. The candles are smoking.] “That guy is throwing the cake away so that they can get out of the restaurant quickly.” "Uh, oh! He doesn’t see that the candles are still burning and smoking when he throws out the cake.” “Oh, look, the garbage can catches fire and the restaurant gets ruined!" [Show smoke in the garbage can. Fade to black. Sirens from a fire truck are heard.] The film lasts for 127

seconds.

B. Voice over for Film Version 2 (*Bad Intentions with Neutral Outcome*):

"Look, there are four men at that table with a lit birthday cake! It must be someone's birthday. Uh oh, there is an announcement that the restaurant is closing. They must leave quickly. They have no time to enjoy their cake!" [They blow out the candles and one man throws the cake in the garbage. The candles are smoking] "Hey, that one guy is really angry that they can't eat the cake. So, he throws the cake with the candles still burning and smoking into the garbage can. He really hopes that the restaurant catches on fire, oh, but look nothing has happened!" [Show non-smoking garbage can. Fade to black.] The film lasts for 127 seconds.

Procedure

The face recognition task

The participant's task is to choose the perpetrator of the fire or the perpetrator who had unsuccessfully tried to start a fire, in one case the neutral intentional perpetrator (Version 1) and in the other the intentional perpetrator (Version 2). Each participant received instructions prior to viewing the lineup. The instructions were: "Now we would like you to identify the man who started the fire [or tried to start a fire]. He may or may not be in the photos." (See Appendix for child assent/script). Each participant was then individually shown a series of photos (by computer), which included (a) the man who set the fire or perpetrator, (b) each of the three other men at the table or "foils," (c) one man who is the waiter (bystander), and (d) one man, who was not in the film. The procedure

for the adults was the same except that they participated in groups of four (four individual computers in one room). The foils matched the general physical description of the perpetrator, following the recommended practice for line-up composition (e.g., Brewer, Weber, & Semmler, 2005: Wells, 1993).

The experiment was conducted in the schools' computer laboratory classrooms. The researcher and one research assistant were present to ensure that participants attended to the computer monitor. Child participants were seated individually in front of a computer and shown the film on their computer monitors. The adult participants were interviewed with individual computer monitors, but in groups of four. To check for understanding participants were then asked to describe what the event that took place in the film. To check on moral judgments of the event the participants were asked three questions in counter balanced order; "How bad was the *man* in the film (who started the fire/ who tried to start a fire)?", "How bad was what the *man did* in the film?", and "How bad was what happened to the *restaurant*?"

The photos were shown one at a time, and the child was asked whether this was the man who set the fire (or tried to set the fire) and how sure he/she was. For each photograph the participants were asked whether or not it depicted the man who started a fire, and how sure they were whether it was or was not the man: whether they were *very sure* that it was the man who started a fire, or *a little bit sure* that it was the man who started a fire, or *a little bit sure that it was not* the man who started a fire, or *very sure it was not* the man who started a fire. The confidence rating scale consisted of a row of smiley faces: two red smiley faces over *very sure* that it is the man, one red smiley face

over *a little bit sure* that it is the man, one blue smiley face over *a little bit sure it isn't* the man, and two blue smiley faces over *very sure that it isn't* the man. Participants chose their confidence level by clicking the mouse on the appropriate button. The experimenter explained how the scale worked and confirmed that the children had understood the procedure by verbally asking them to respond to the four following instructions:

1. '*Click the button that would tell me that you were very sure about something - like your name';*
2. '*Click the button that would tell me that you are just a little bit less sure than the last time';*
3. '*Click the button that would tell me how sure you are that today is Sunday';*
4. '*Click the button that would tell me how sure you are just a little bit less sure that it isn't Sunday'*.

In order to perform a Signal Detection analysis, it was necessary that the witnesses judged a relatively large number of judgments. This was formed by presenting the six individuals (1 perp, 3 foils in the group in the film, 1 bystander (waiter) in the film, and 1 foil not seen in the film) each presented frontally and in right and left profile. The series consisted of 3 trials of the 18 photos [6 individuals X 3 angles].The series of photos were randomly ordered except that no photos of the same individual appeared consecutively. For each photo participants indicated whether or not that photo depicted the perpetrator and rated their confidence (*very sure that it is not the man, little bit sure that it is not the man, a little bit sure that it is the man, very sure that it is the man*). All photos were head and shoulder shots with left/right and frontal profiles. Lineup targets

were dressed in the same clothing (dark blue t-shirt) and were filmed against the same background. They were familiarized with the procedure so that they were proficient in its use before the actual judgments began. At the end of photo-identifications, each participant was asked how bad it would be to say someone did it when he didn't (a false positive) and how bad would it be to say someone did not do it when he did (a false negative), and why it would be bad.

Accuracy and Decisional Criterion (Bias response) Analysis

As mentioned in the procedure section, each participant saw three trials of 18 photos each for a total of 54 photos. For each of the photos participants chose among four responses: *very sure that it is not the man, a little bit sure that it is not the man, a little bit sure it is the man, very sure it is the man* in the film. These four responses are needed, in signal detection parameters, to obtain a ROC curve with fewer trials, whereby having three criterion points better fit a (d') parameter. (See previous section on SDA in the introduction.)

Participants' responses were entered into SPSS and calculated in terms of Signal + Noise (SN) what can be called positive and negative hits; and Noise (N) what are the false positive and false negative errors. Each participant had a sum of hits and errors that were entered into a signal detection program using an equal-variance normal distribution model. The parametric measures are based on Macmillian and Creelman's (1991) review that the signal detection measures of sensitivity and criterion bias (decisional criterion) can be separately calculated. The sensitivity (d') measure corresponds to: $\boldsymbol{d' = Z\ (HR) - Z\ (FAR)}$ where: the proportion of *false alarm rates* (FAR) is subtracted from the

proportion of *hit rates* (HR). The more able a participant is to discriminate between the perpetrator and foil faces the larger d′ will become.

A response criterion is the tendency to respond more frequently with either *"yes"* or *"no"* on the basis of a looser or stricter response criterion (Macmillan & Creelman, 1991). The response bias equation is: **C= ½ (Z (HR) – Z (FAR)** where "C" is defined as the distance between the criterion and the neutral point, where neither response of *"yes/no"* is favored. Decision criteria can be influenced by three main factors: (a) The instructions given to the participant; (b) The frequency of signal trial and no-signal trial; and (c) The relative cost of making errors (false alarms and misses) and the relative benefit of making correct responses (hits and correct rejections).

The measure of decisional criterion (bias) in this study was obtained as follows: There is one measure of sensitivity (*d'*) across the face recognition task with three measures of decisional criteria to each participant (C1, C2, C3). The three measures of decisional criteria are the bias points on the participant's responses to the four response scale: *very sure that it is <u>not</u> the man, a little bit sure that it is <u>not</u> the man, a little bit sure that is the man, and very sure that it is the man.* I have analyzed the three decisional criteria corresponded to each participant, I am only reporting the C3 scores. Reporting the C3 makes sense in terms of the literature that cites the differences that are especially found between those who are very confident and those who are not confident in their judgments. Individual differences of sensitivity (*d'*) and decisional criterion (C3) were obtained. For addressing the problem of encountering when either a hit or false alarm rate equaled zero, I added a "1" constant to the "0" choices in any of the four judgment

categories, i.e., "*very sure it is the man"*. The individual scores of d' and C3 enabled me to correlate the performance and individual measures from the belief measures [e.g., each participant was asked how bad it would be to say someone did it when he didn't (a false positive) and how bad would it be to say someone did not do it when he did (a false negative), and why it would be bad]. When using aggregate data ANOVA's were calculated to see the significant relationship between groups.

Results

Eyewitness Performance

Sensitivity (d'). The first goal of the present study was to examine whether the age groups differed in their ability to discriminate the perpetrator from the foils. The prediction was that the younger children would have lower sensitivity (*d'*) than the older children, adolescents, and adults in both conditions (Fire: neutral intent/bad outcome and No Fire: bad intent/neutral outcome). A signal detection analysis with aggregated data using confidence intervals was conducted to assess whether the children's sensitivity differed according to the type of condition under examination. These analyses did not reveal a significant difference across the two conditions [Neutral intent bad outcome (fire) vs. Bad intent neutral outcome (no fire)] within each age group (See Table 1). The one-way ANOVA for each age group was not significant (See Tables 1 and 2 and Figure 5). However, adolescents' appeared better at recognizing the "perp" from "foils" than the younger two age groups. The data for the adult sample were omitted since the interview procedure differed in the administration the adults were grouped by fours (four computers in one classroom) and the diversity of the adult population may have had an effect on their sensitivity scores. When removing their data, a clear and significant linear age trend emerges from the youngest to the adolescent samples (Table 3). A trend analysis was performed and sensitivity (*d'*) was found to increase linearly with age. Table (3) 'Trend Analysis' and Figure 6 shows this significant age trend in accuracy (*d'*).

Table 1
Means and Confidence Intervals for Sensitivity (d') by Age Across and by Condition (N=122)

	Age Across			By Condition					
				No Fire			Fire		
Age Group	*n*	*M*	95% CI	*n*	*M*	95% CI	*n*	*M*	95% CI
7-9	28	.578	.303, .852	14	.506	.118, .894	14	.649	.261, 1.037
10-12	30	.692	.427, .957	15	.839	.465, 1.214	15	.544	.169, .919
13-15	28	1.068	.793, 1.342	14	.979	.591, 1.367	14	1.156	.768, 1.544
19-25	36	.828	.586, 1.070	18	.917	.575, 1.259	18	.738	.396, 1.080

Table 2
Analysis of Variance for Sensitivity (d') by Age

Source	*SS*	*df*	*MS*	*F*	*p*
Between	3.762	3	1.254	2.367	.074
Within	52.510	118	.530		
Total	66.271	121			

Note. Age was collapsed across conditions.

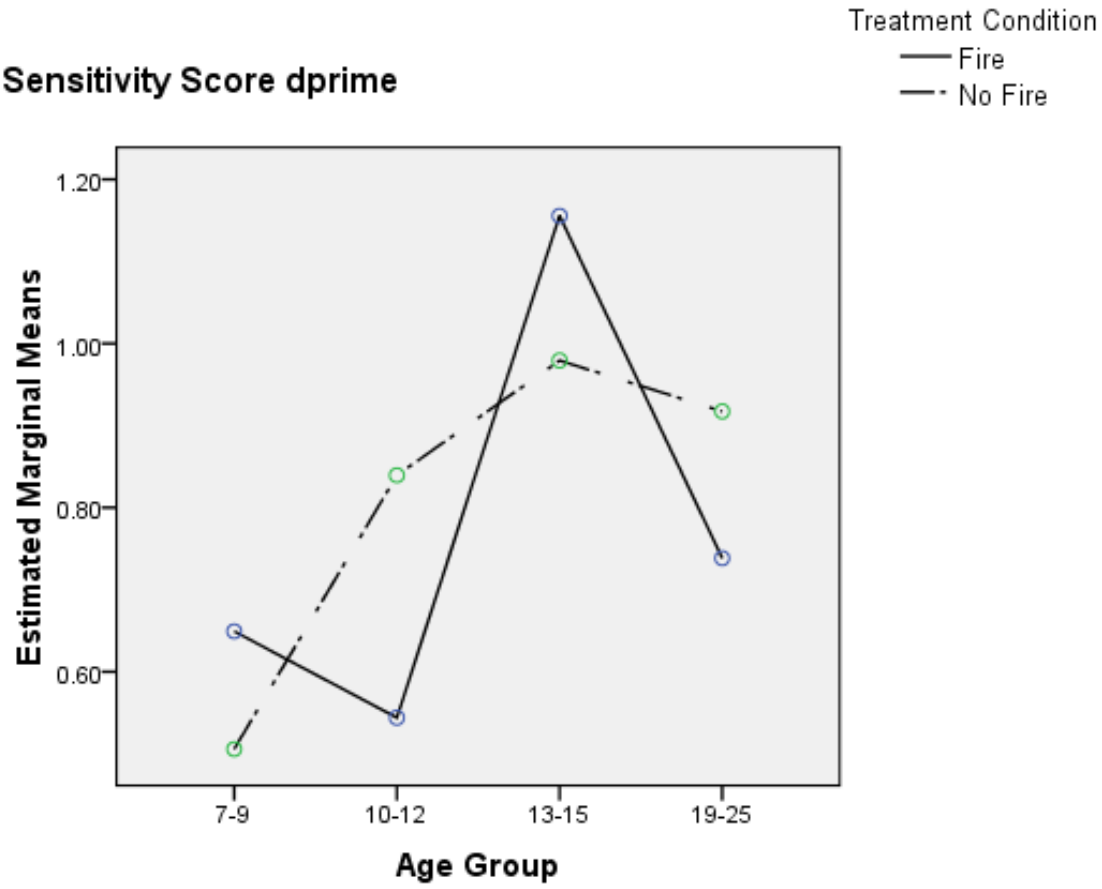


Figure 5.
ANOVA sensitivity (*d'*) by condition[1] and age.
[1]Neutral intent/bad outcome (fire); Bad intent/neutral outcome (no fire).

Table 3
Trend Analysis Sensitivity (d') by Age and Collapsing Across Condition

Source	***SS***	***df***	***MS***	***F***	***p***
Age Group	3.696	2	1.848	3.322	.041
Condition	2.907E-5	1	2.907E-5	.000	.994
Total	49.306	85			

Note. Excluding adult sample.

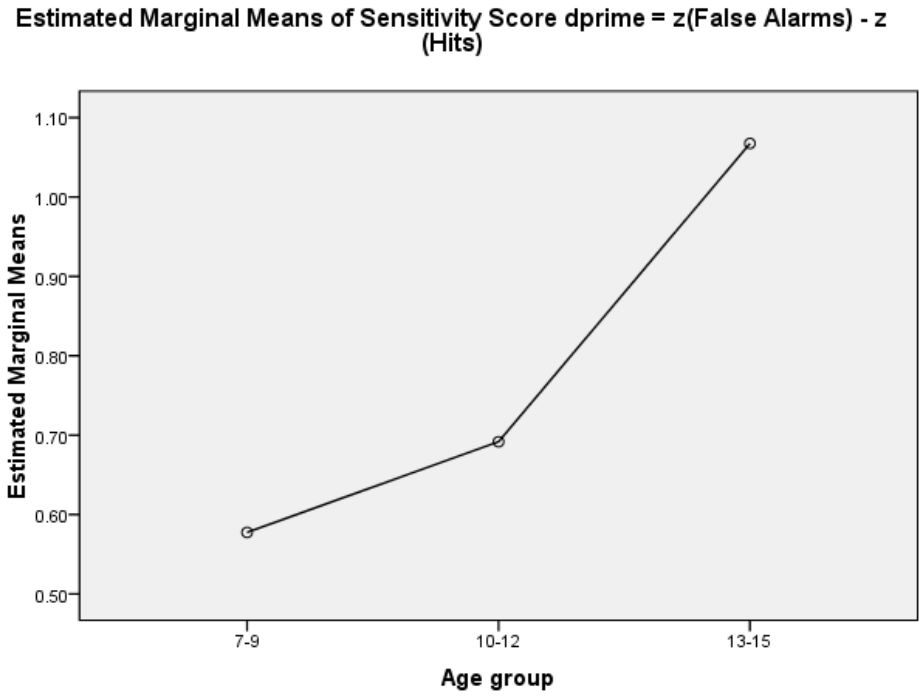


Figure 6.
Trend analysis sensitivity (*d'*) by age[1].
[1]excluding adult sample.

Decisional Criteria (C3)

An ANOVA was performed to examine whether younger participants are more likely to have lower bias scores in the confidence and accuracy measures. These analyses revealed two significant interaction effects: 1) the dependent variable C1 Bias Level for Confidence 1 (*'little sure it is not the man'* versus the other three choices) showed an interaction of 'mean centered age' X Condition Fire. 2) The dependent variable C3 Bias

Level for Confidence 3 (*'very sure that it is the man'* versus the other three choices) also basically showed the same interaction of 'mean centered age' X Condition (Table 4). The results for C1 parallel those for C3. Table 5 shows the results of bias level C3 by condition and age group.

There are two ways to analyze the interaction: 1) controlling for condition and comparing ages; and 2) controlling for age and comparing conditions. It was decided to compare age, controlling for condition to focus on age comparisons.

Table 4
Interaction of bias level and condition

Bias	Predictors in model	R^2 (*model p*)	Beta (*p*)
C1	Age X Fire	.085 (.015)	-.230 (.010)
C3	Age X Fire	.068 (.040)	-.207 (.021)

Table 5
Criterion (C3) by Condition and Age Group

Age Group	Fire	N	Decisional Criteria (C3)	DC-95% CI
7-9	No	14	1.348	(1.136-1.560)
	Yes	14	1.368	(1.156-1.579)
10-12	No	15	1.549	(1.345-1.753)
	Yes	15	.963	(.758-1.167)
13-15	No	14	1.436	(1.224-1.648)
	Yes	14	1.264	(1.052-1.475)
19-25	No	18	1.278	(1.091-1.465)
	Yes	18	1.301	(1.115-1.488)

In the Bad Intention Neutral Outcome (No Fire) Condition, $F = 3.84$, $p = .012$, the 10-12 year-olds use a *stricter* criterion than in the other three age groups (See Table 6 and Figure 7). However, in the Neutral Intention Bad Outcome (Fire) Condition, $F = 2.85$, $p = .045$, the 10-12 year-olds use a significantly *looser* criterion (Table 7 and Figure 8). Separate Independent t-tests were performed to see the age comparisons between the 10-12 yr. olds and the other three age groups [Bad Intent Neutral Outcome (No Fire) Condition 7-9 yr. vs. 10-12 yr., $F = .001$, $p = .016$; 10 yr. vs. 13-15 yr., $F = .024$, $p = .066$; 10-12 yr. vs. 19-25 yr., $F = .049$, $p = .033$; Neutral Intent Bad Outcome (Fire) Condition 7-9 yr. vs. 10-12 yr., $F = 1.43$, $p = .193$; 10-12 yr. vs. 13-15 yr., $F = .006$, $p = .396$; 10-12 yr. vs. 19-25 yr., $F = .367$, $p = .051$] (See Table 9). Alternatively, when controlling for age, the 10-12 year-olds were the only group that showed a significant difference between conditions (See Table 8 and Figure 9).

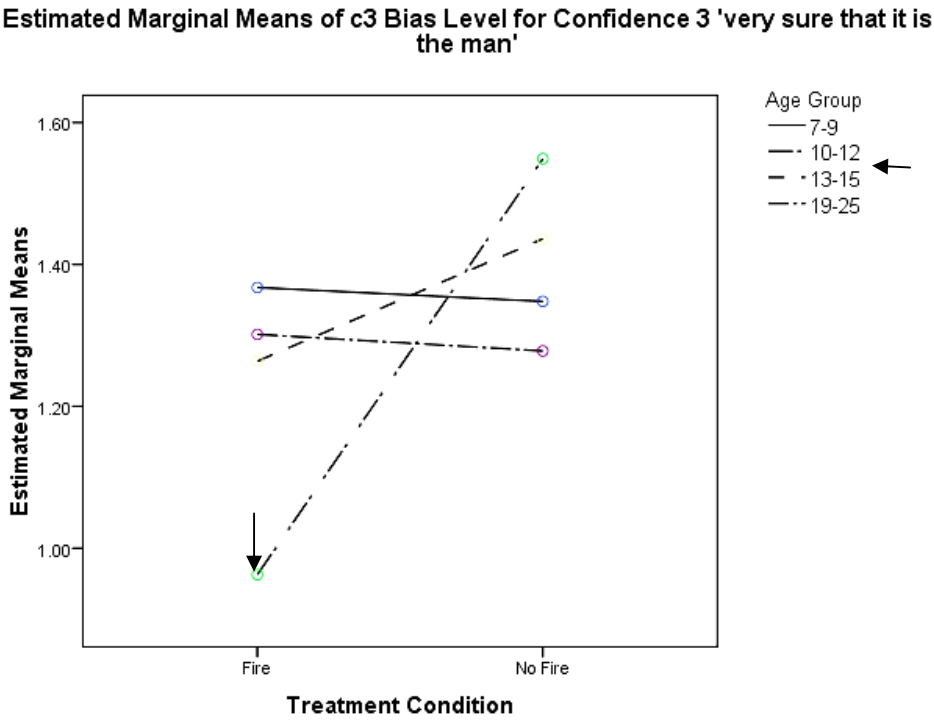


Figure 7
Bias C3 (Very sure that it is the man) by condition

Table 6
Comparing Age Groups for Bad Intent No Fire Neutral Outcome Condition and C3

Source	*SS*	*df*	*MS*	*F*	*p*
Between	1.859	3	.620	3.846	.012
Within	6.944	57	.066		
Total	8.803	60			

Table 7
Comparing Age Groups for Neutral Intent Bad Outcome Fire Condition and C3

Source	*SS*	*df*	*MS*	*F*	*p*
Between	1.442	3	.620	2.855	.045
Within	9.600	57	.168		
Total	11.042	60			

Table 8
Tests of Between-Subjects Dependent Variable: C3 for 10-12 year-olds only and condition

Source	*SS*	*df*	*MS*	*F*	*p*
Condition	2.578	1	2.578	15.116	.001
Within	4.776	28	.171		
Total	7.355	29			

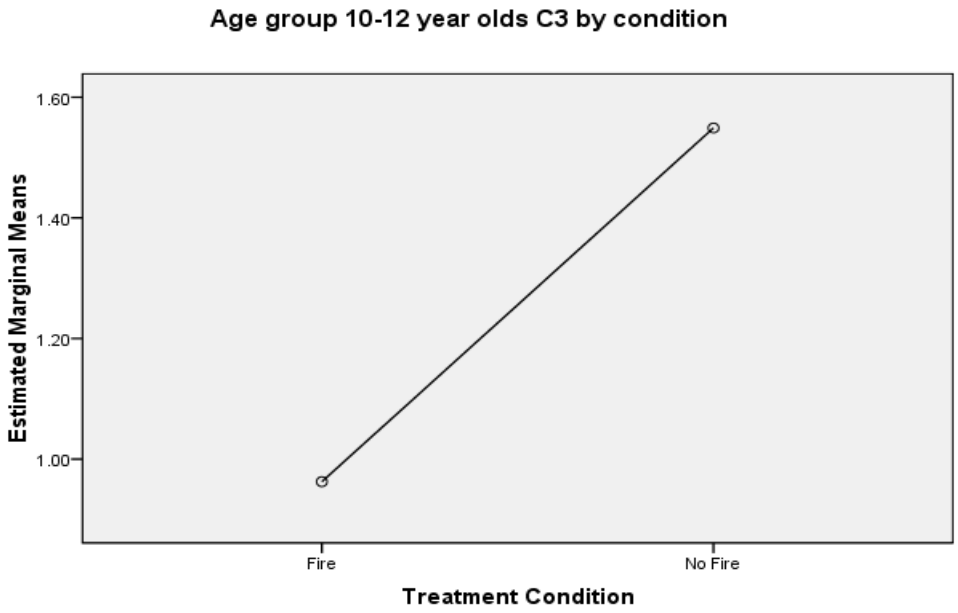


Figure 8.
10-12 year-olds C3 (*Very sure that it is the man*) by condition
Note. Lower the score looser the criterion.

Table 9

Comparison of Age Groups for C3 in Two Conditions: Bad Intent Neutral Outcome (No Fire) and Neutral Intent Bad Outcome (Fire)

Compared Age Groups	*t*	*df*	*p*	95% CI
Bad Intent Neutral Outcome (No Fire)				
7-9 yr. vs. 10-12 yr.	2.5*	27	.016	.0829, .7267
10-12 yr. vs. 13-15 yr.	-1.9	27	.066	-.623, .0212
10-12 yr. vs. 19-25 yr.	-2.23*	31	.033	-.647, -.029
Neutral Intent Bad Outcome (Fire)				
7-9 yr. vs. 10-12 yr.	-1.3	27	.193	-.510, .1081
10-12 yr. vs. 13-15 yr.	.86	27	.396	-.1557, .3816
10-12 yr. vs. 19-25 yr.	2.03	31	.051	-.001, .543

Note. C3: Very sure that it is the man.

*$p < .05$.

False Positive Errors

Analyses of false positive errors were performed to confirm the basic hypothesis that younger children are more inclined to commit more false positive errors than the older children and adult participants. The ANOVA analysis shows that across both conditions the youngest (7-9 year-olds) age group committed more false positive errors than the adolescents and adult participants (Figure 9). However, this main effect is superseded by an interaction age and condition. The results of a 2x4 ANOVA are presented in Table 10, F (3, 114) = 3.49, p = .018. Separate Independent t-tests were performed to see the age comparisons (See Table 11).

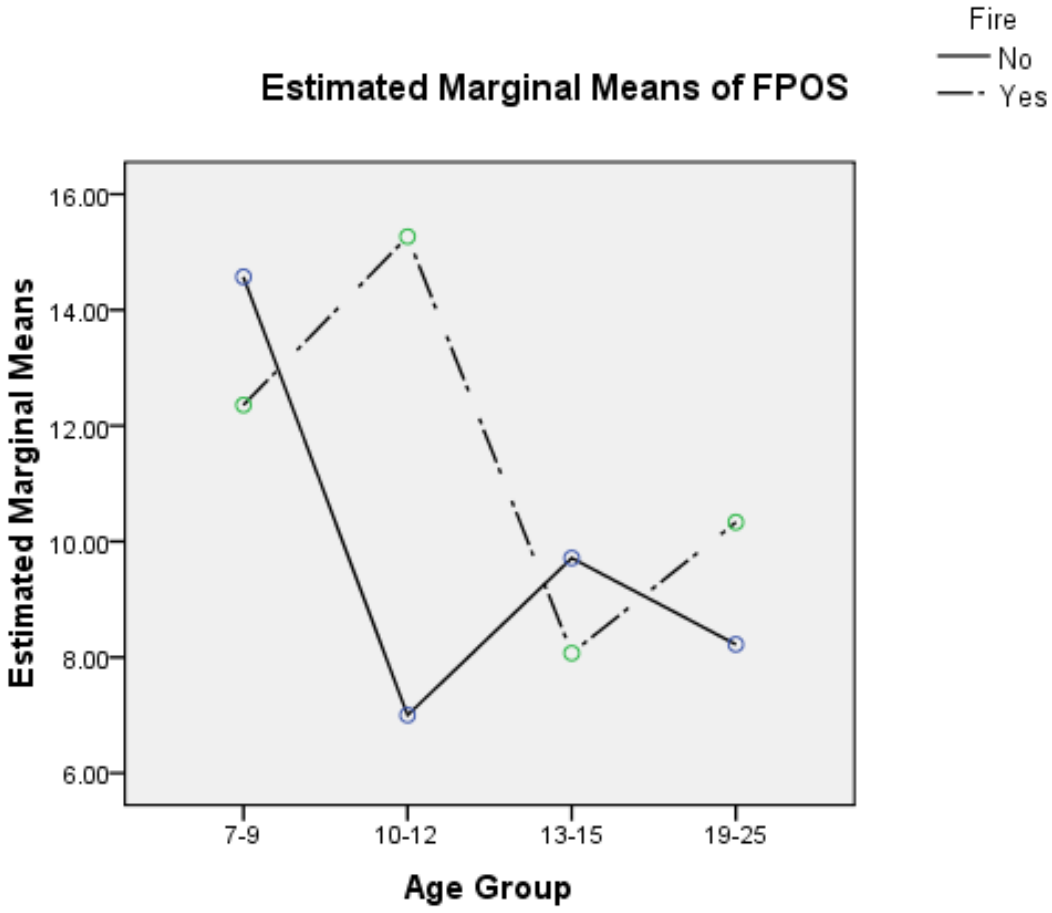


Figure 9.
ANOVA false positive errors by condition[1] and age.
[1] Neutral intent bad outcome (Fire), Bad intent neutral outcome (No fire).

Table 10

Analysis of Variance for False Positive Errors as a Function of Condition and Age Group

Source	*SS*	*df*	*MS*	*F*	*p*
		Between subjects			
Condition	80.20	1	80.20	2.20	.141
Age group	382.78	3	127.59	3.50**	.018
Age group x Condition	510.25	3	170.08	4.66**	.004
Within-group error	4158.473	114	36.48		

Note. Conditions: Bad intent neutral outcome (No fire) and Neutral intent bad outcome (Fire). Age groups: 7-9 yr., 10-12 yr., 13-15 yr., and 19-25 yr.

*$p < .05$. **$p < .01$.

Table 11

Comparison of Age Groups for False Positive Errors in Two Conditions: Bad Intent Neutral Outcome (No Fire) and Neutral Intent Bad Outcome (Fire)

Compared Age Groups	*t*	*df*	*p*	95% CI
Bad Intent Neutral Outcome (No Fire)				
7-9 yr. vs. 10-12 yr.	3.24**	27	.003	2.786, 12.35
7-9 yr. vs. 13-15 yr.	1.93	26	.065	-.314, 10.02
7-9 yr. vs. 19-25 yr.	3.35**	16	.004	2.333, 10.36
10-12 yr. vs. 13-15 yr.	-1.18	27	.248	-7.43, 2.001
13-15 yr. vs. 19-25 yr.	.80	15	.431	-2.42, 5.409
Neutral Intent Bad Outcome (Fire)				
7-9 yr. vs. 10-12 yr.	-1.08	27	.289	-8.43, 2.61
7-9 yr. vs. 13-15 yr.	1.68	26	.105	-.952, 9.524
7-9 yr. vs. 19-25 yr.	.789	30	.436	-3.21, 7.26
10-12 yr. vs. 13-15 yr.	-1.18	27	.248	-7.43, 2.001
13-15 yr. vs. 19-25 yr.	.809	15	.431	-2.42, 5.409

*$p < .05$. **$p < .01$.

Moral Judgment Composite as Mediator

To examine whether moral judgment beliefs mediate the effect of condition and age on the witness's C3 bias score, a Moral Judgment Composite Score was generated for each participant to determine if the combined moral judgments of the act, actor, and outcome would mediate the relationship between age, condition, and decisional bias. Participants were asked in counter balanced order "*How bad the man was in the film*", "*How bad was what the man did in the film*", and "*How bad was what happened to the restaurant*." The correlations of the moral judgments are presented in Table 12. It is important to note that the participants do not confuse outcome (*How bad was the restaurant?*) with the actor (*How bad was the man?*) or the act (*How bad was what the man did?*). A structured equation model was created and the model was supported only for the adults (See Figure 10).

Table 12
Pearson Product-Moment Correlations for the Moral Judgments

Moral Judgments	*M*	*SD*	1	2	3
1. How bad was what the man did?			---	.35**	-.03
2. How bad was the man?			.35**	---	-.31**
3. How bad was restaurant?			-.03	-.31**	---

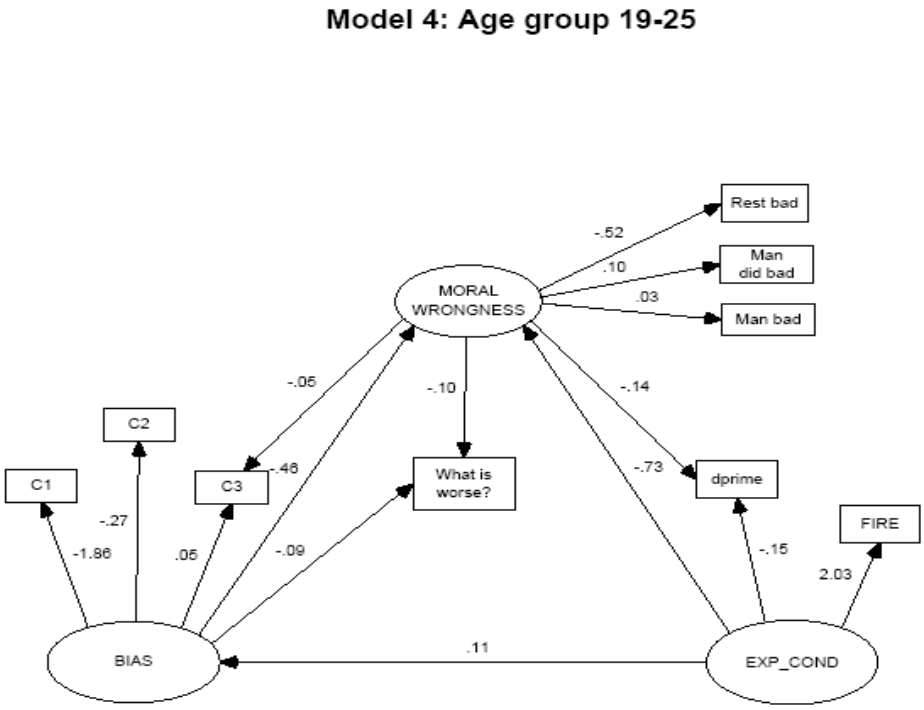


Figure 10.
Structured equation model for the adult sample.

To better understand why the structured equation model was not generally supported, the effect of age and condition on the Moral Judgment Composite scores (*how bad was the actor, how bad was the outcome of the restaurant, and how bad was the act*)

were analyzed. The expected interaction effect between age and condition [Neutral Intent Bad Outcome Fire /Bad Intent Neutral Outcome No Fire] was not found. Instead, there was a non significant tendency for moral judgments to get less severe with age regardless of type of transgression [film]. All age groups judged the No Fire [bad intention] act as worse than the Fire [bad outcome] act (See Table 13 and Figure 11). For the children and adolescents their moral judgments of the act did not operate as a mediating factor in their eyewitness identification.

Table 13
Analysis of variance composite score of Moral judgment

Source	*SS*	*df*	*MS*	*F*	*p*
Condition	21.825	1	21.825	29.323	.011
Age group	12.288	3	4.096	5.569	.096
Condition X Age group	2.207	3	.736	.360	.782

Note. Conditions: Bad intent neutral outcome (No fire) and Neutral intent bad outcome (Fire). Age groups: 7-9 yr., 10-12 yr., and 19-25 yr.

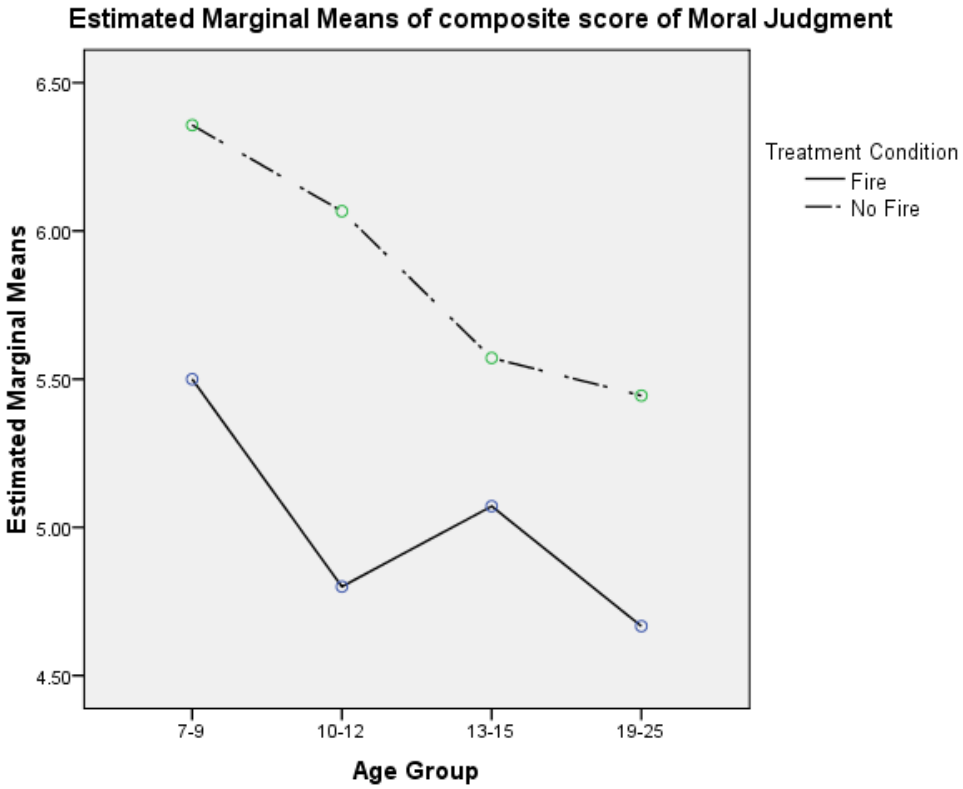


Figure 11.
Composite score of moral judgments (*How bad was the man? How bad was what the man did? How bad was what happened to the restaurant?)* by condition and age group. *Note.* Conditions: Bad intent neutral outcome (No fire) and Neutral intent bad outcome (Fire). Age groups: 7-9 yr., 10-12 yr., and 19-25 yr.

In summary, the results for eyewitness performance were as follows: (1) the sensitivity (*d'*) parameters were not significantly different across the two conditions, but did significantly increase with age up through adolescence. (2) The difference between the bias parameter distributions for the Neutral Intent Bad Outcome (Fire) Condition and Bad Intent Neutral Outcome (No Fire) Condition were significantly different at the p =.05 level, with bias being more stringent in the Bad Intent Neutral Outcome (No Fire) Condition, but only for the 10-12 year-olds (See Figures 12 and 13). (3) The 7-9 year-olds commit more false positive errors regardless of condition and adolescents and adults commit fewer false positive errors regardless of condition.

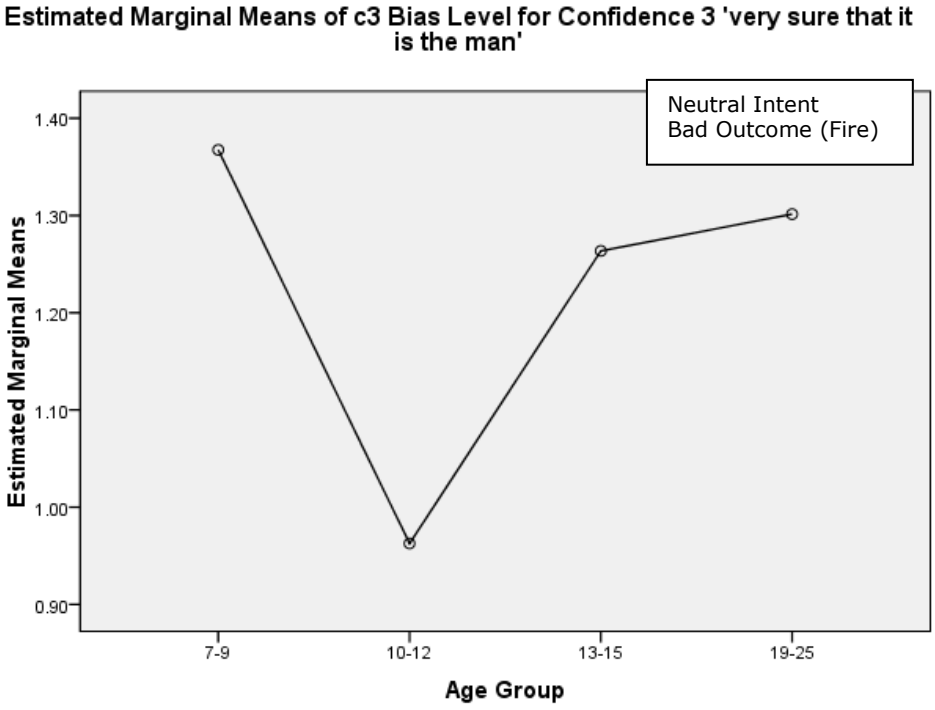


Figure 12.
Neutral intent bad outcome (fire) condition C3 (*very sure that it is the man*) by age group.

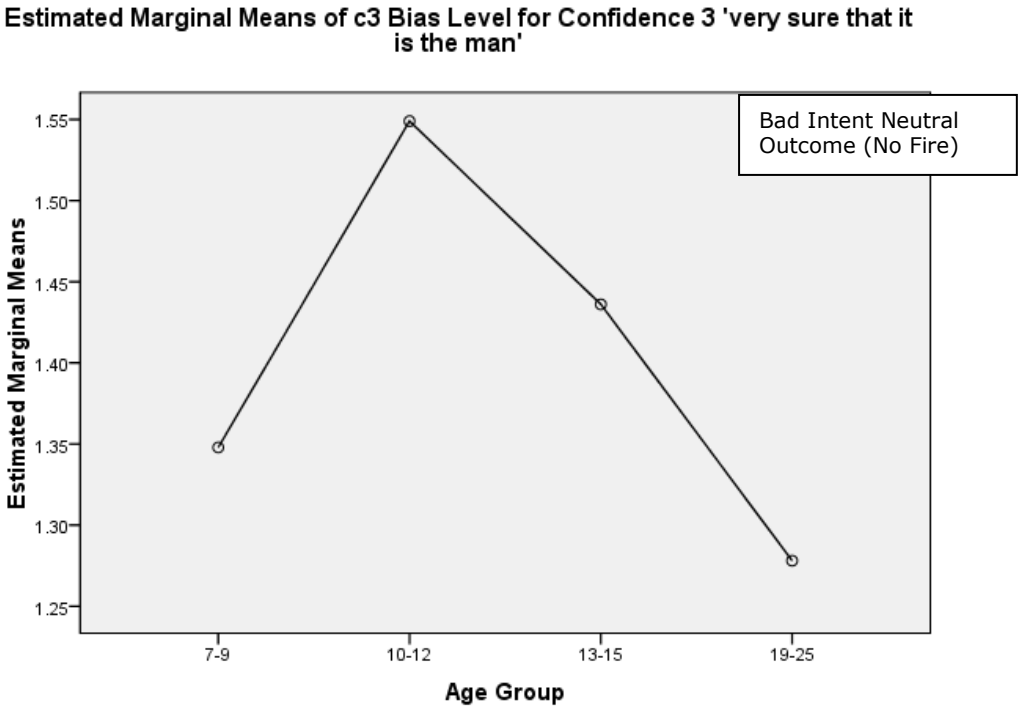


Figure 13.
Bad intent neutral outcome (no fire) condition C3 (*very sure it is the man*) by age group.

Beliefs and Reasons Concerning False Positive and False Negative Errors

To examine whether young children differ from older children and adults in their moral beliefs regarding which error is worse to make as a witness (F+ or F-) participants' responses to the follow-up beliefs questions were analyzed. We tested the hypothesis regarding beliefs by a) examining the simple choices as to which error is worse; and b) by analyzing the reasons why they chose either a false positive or a false negative error as worse (described below).

Participants' choices of whether it was the worse to make a false positive error or a false negative error were coded into four categories: 1. a false positive error is worse; 2. a false negative error is worse; 3. there is no difference between the two errors; and, 4. a confused or a 'don't know' response. Responses that asserted that both errors are equally

bad, or were confused received a 0 score. Results for the "Which Error is Worse" are presented in Table 14 and Figure 14. There is a significant age trend for the dependent variable "Which Error is worse?" Adults (age 19-25) and adolescents (age 13-15) asserted that false positive errors are worse in both conditions, whereas the youngest (age 7-9) asserted that false negative errors are worse. However, participants' beliefs about false positives and false negative errors were not related to sensitivity (*d'*) or decisional bias (C).

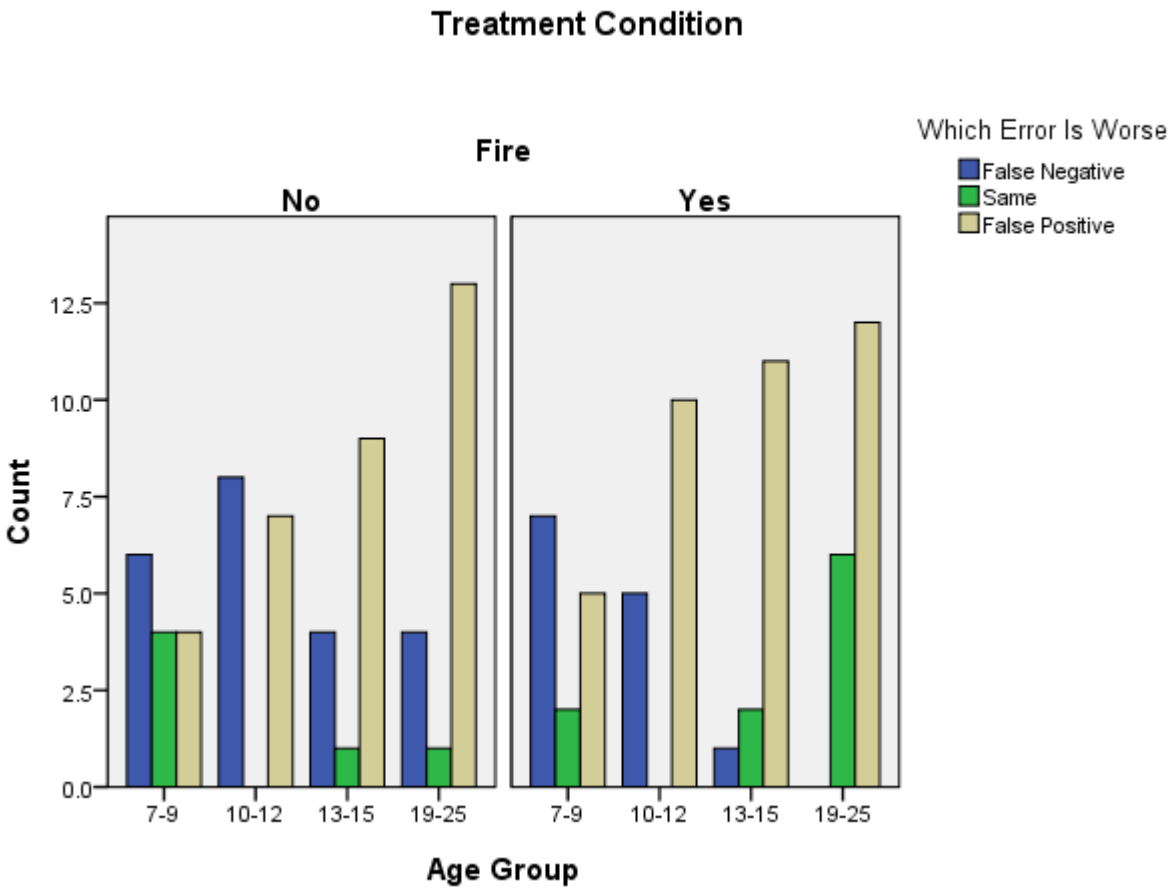


Figure 14.

Which type of error is worse to make as a witness? *Note.* Conditions: Bad intent neutral outcome (No fire) and Neutral intent bad outcome (Fire).

Table 14
Chi-Square Tests on Which Error is Worse to Make

	Value	*df*	*p*
Pearson Chi-Square	35.589	18	.008
Likelihood Ratio	41.566	18	.001
Linear-by-Linear	4.340	1	.037
Total	122		

Participants' reasoning as to why false positive errors and/or false negative errors are worse supports the "Which is Worse" finding. The responses were coded for whether the reasons focused on concern for fairness to the accused or concern for punishing the guilty. These two were combined and contrasted with other (non-moral reasons) and also contrasted with one another. A *Which Is Worse Reason Composite Score* was generated for each participant, based on the reasons given to justify "Why is it bad to make a false positive or a false negative error?" Each time a participant responded with a moral reason that showed concern for the guilty man going free they score a -1, and a +1 for each time they responded with concern for the innocent man accused. Participants who responded in non-moral terms or said "I don't know" scored a 0. ANOVA shows that age was significant for the *Which is Worse Reason Composite*, $F = 5.64$, $p = .001$, but not for condition, $F = 1.81$, $p = .181$ (See Table 15 and Figure 15).

A higher score indicates a greater concern for the innocent man being accused unjustly; whereas lower scores indicate concern for the guilty man going free (See Figure 16). Cross-tabulation with age group and the *Which is Worse Reason Composite* were constructed and the focus on the concern for the innocent man increases with age regardless of condition (See Table 16). A Tukey HSD was performed and significant differences between the youngest (7-9 yr) and the adolescent (13-15 yr) and between the

adult (19-25 yr) groups were found (See Table 17).

Table 15
Analysis of Variance for Which is the Worse Type of Error Reasons

Source	*SS*	*df*	*MS*	*F*	*p*
		Between subjects			
Condition	4.022	1	4.022	1.811	.181
Age group	37.627	3	12.542	5.648	.001
Age group x Condition	6.739	3	2.246	1.012	.390
Within-group error	253.175	114	2.221		
Total	301.508	121			

Note. Conditions: Bad intent neutral outcome (No fire) and Neutral intent bad outcome (Fire). Age groups: 7-9 yr., 10-12 yr., and 19-25 yr.

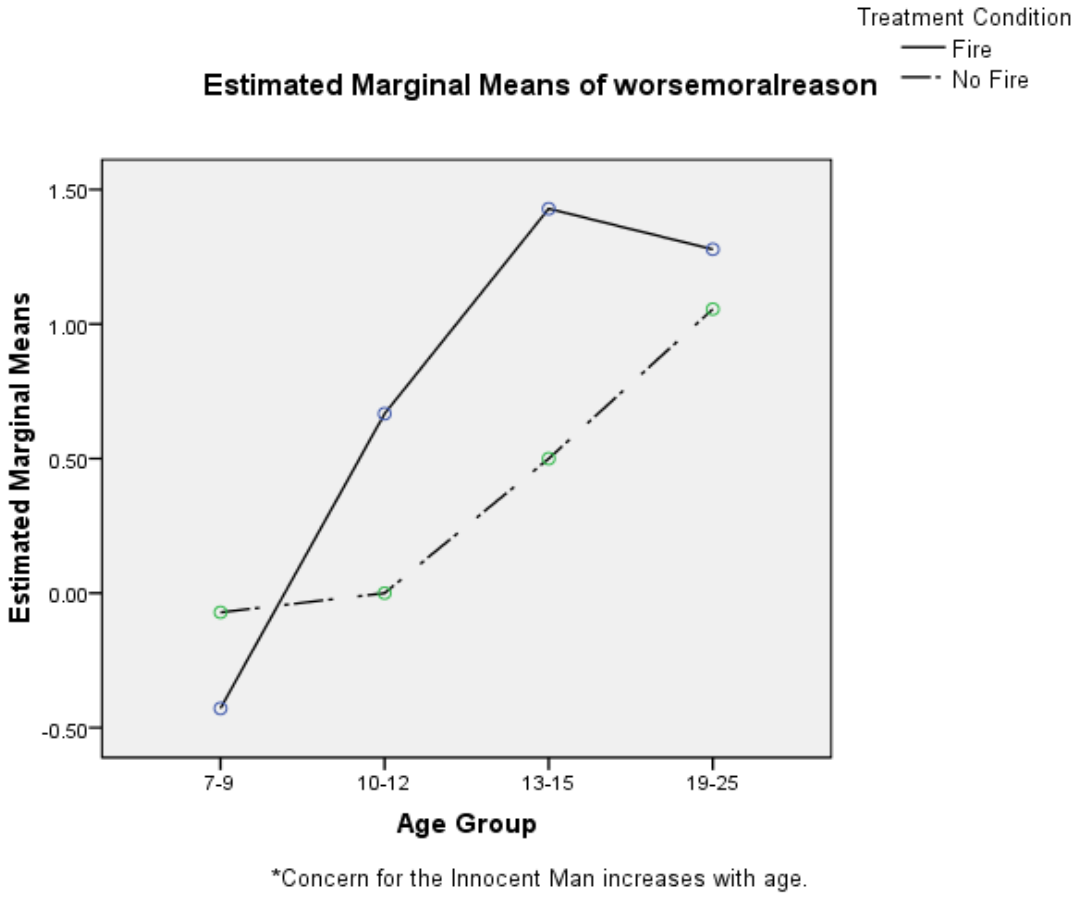


Figure 15.
Which is worse reason composite by condition and age group.
Note. Conditions: Bad intent neutral outcome (No fire) and Neutral intent bad outcome (Fire).

Table 16

Cross Tabulation Age Group by Which is Worse Reason Composite Score

			F-			non-moral		F+		
			-3	-2	-1	0	1	2	3	Total
Age Group	7-9	Count	0	2	5	12	4	1	4	28
		% within Age Group	.0	7.1	17.9	42.9	14.3	3.6	14.3	100.0
	10-12	Count	1	0	9	3	11	6	0	30
		% within Age Group	3.3	.0	30.0	10.0	36.7	20.0	.0	100.0
	13-15	Count	0	3	2	4	9	4	6	28
		% within Age Group	.0	10.7	7.1	14.3	32.1	14.3	21.4	100.0
	19-25	Count	1	0	3	8	13	7	4	36
		% within Age Group	2.8	.0	8.3	22.2	36.1	19.4	11.1	100.0
	Total Count		2	5	19	27	37	18	14	122

Table 17

TUKEY HSD Multiple Comparisons for Which is Worse Reason

Compared Age Groups	*MD*	*SE*	*p*	*95% CI*
7-9 yr. vs. 10-12 yr.	-.583	.391	.447	-1.604, .437
7-9 yr. vs. 13-15 yr.	-1.214*	.398	.015	-2.252, -.175
7-9 yr. vs. 19-25 yr.	-1.416*	.375	.001	-2.395, -.437

Based on observed means.

The error term is Mean Square (Error) = 2.221.

*. The mean difference is significant at the .05 level.

Discussion

The first goal of this study was to explore whether ability to discriminate the perpetrator from the foils differed by the age of the eyewitnesses. The eyewitnesses' results for sensitivity (*d'*) show a linear trend so that sensitivity increases with age for children and adolescents, but appears to level off for adults. The adult sample in this study came from both ethnically and socioeconomically diverse backgrounds whereas the children and adolescents came from mostly white middle-class socioeconomic backgrounds. The difference in ethnic and socioeconomic backgrounds may have affected the findings. In addition, the adult participants had individual computer monitors but were placed in groups of four in the computer laboratory classroom. However, the findings do confirm previous research by Chance and Goldstein (1984), which found that face recognition improves by age 12.

Age group effects were also found for decisional criterion C3 and C1, but these were dependent on condition. Another prediction of this study was based on the hypothesis that decisional criteria would reflect both age and the nature of the transgression. There was a significant interaction with age, but in a somewhat unexpected form. The 10-12 year-olds used a significantly looser criterion in the Neutral Intent Bad Outcome Fire condition than in the Bad Intent Neutral Outcome No Fire condition, as expected. What was unexpected was that the decisional criteria for the youngest age group (7-9 year olds) did not differ between the two film conditions (Fire/neutral intent vs. No Fire/bad intent). The

7-9 year old children apparently did not take the intentions of the actor or the outcome of the act into consideration in their eyewitness decision-making.

That is, was also hypothesized for the 7-9 and 10-12 year-old age groups (that their bias would be lower in the Fire/neutral intent condition) was actually found only for the 10-12 year olds. That the 10-12 year-old children performed as predicted and not the 7-9 year-olds seems puzzling at first. Yet, the previous studies, discussed (Saltzstein & Peach, 2005; Spring, 2006) found that young children do not think in morally relevant terms when making identification choices. It is possible that these younger children disconnected the act of choosing the person from the moral nature of the act and relied on the general strategy of over guessing, thus maximizing false positive errors. In contrast, the 10-12 year-olds had a lower bias score in the Neutral Intent Bad Outcome (Fire) Condition and a higher bias score in the Bad Intent Neutral Outcome (No Fire) Condition indicating that they did distinguish between the two moral framings of the film (conditions). In comparison, the adolescents and adults were much more cautious in their eyewitness identifications but, also didn't seem to distinguish between the two conditions. Adolescents and adults appear to have employed generally conservative decisional approach in both (Fire/Neutral Intention and No Fire/Bad Intention) conditions.

Calculating the frequency of when the 7-9 yr. olds responded to reasons why it is bad to commit false positive or false negative errors as "no difference" or "I don't know" and comparing the frequencies to the 10-12 yr. olds a difference

in scores is apparent (12 vs. 3). There is a significant difference found in the Chi-square between the 7-9 yr. olds and the 10-12 yr. olds [χ^2 (1, N = 58), = 8.155, p <.004] reasons for which is the worse error (false positive or false negative) to commit as a witness (See Table 18).

Table 18
Chi-Square Tests Using Frequency of Moral Responses vs. Non-moral Responses Between 7-9 yr. olds and 10-12 yr. olds (Which is Worse Moral Reason Composite scores)

	Non-moral answer	Moral answer	Total
7-9 yr. olds	12 (7.241)	16 (20.759	28
10-12 yr. olds	3 (7.759)	27 (22.241)	30
Total	15	43	58

df = 1

Pearson Chi-square statistics = 8.155

P. Value = 0.004

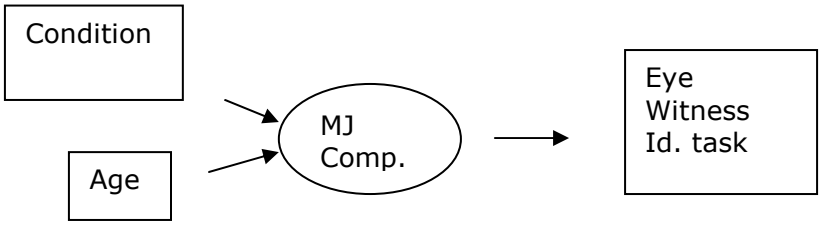
The 7-9 yr. old children may only have been concerned about identifying the 'perp' regardless of his act, and to have used a strategy of over-guessing, as if following the strategy 'if in doubt, identify', perhaps out of over confidence in their selections. Examining the absolute number of false positive errors confirms that younger children (ages 7-9 yrs.) do positively identify regardless of condition. These results confirm findings of past research (e.g., Cutler & Penrod, 1995; Pozzulo & Lindsay, 1998) on children in eyewitness identification studies. The adolescents and adults made fewer false positive errors in both conditions, and the

10-12 year-olds were again the only age group that differentiated between the two conditions, making more false positive errors in the Fire Condition/Neutral Intent, paralleling the results from the C measure of decisional bias. False positives are generally found to decrease with the advancement of age (e.g. Chance & Goldstein, 1984; Cutler & Penrod, 1995). Thus, the 7-9 yr. old children may have perceived the task differently than the older children and adolescents. It has been suggested that children's increased false positive error rates are a result of feeling pressure to make identification (Pozzulo & Lindsay, 1998). However, the mere presentation of a lineup may exert an implicit demand to select someone positively (Ceci, Ross, & Toglia, 1987). Young children may provide responses that they think experimenters want – in this case, as many positive hits to ensure their chances of picking the culprit.

Participants' beliefs about false positives and false negative errors are considered next. The youngest participants (7-9 yr. olds) believe that it is worse to make a false negative error in both conditions and the oldest participants believe that it is worse to make false positive errors in both conditions. That is moral beliefs as to which type of error is the worse type of error to make shifts from false negative to false positive with age regardless of condition. The adults (age 19-25) and adolescents (age 13-15) assert that false positive errors are worse in both conditions, conversely the youngest (age 7-9) assert that false negative errors are worse. This lends support to my previous study (Spring, 2006) as well as results by Saltzstein & Peach (2005). Young children beliefs' about moral transgressions and the severity of the outcome does not seem to make a difference

in their beliefs as to which error is worse to make. They do have more concern for letting the guilty man go free (false negative error) in both conditions.

The moral judgment composite measure (*how bad was the actor, how bad was the outcome of the restaurant, and how bad was the act*) did not mediate the age x condition effect on bias (C) for the children and adolescents although it did for the adults.



Thus, children's and adolescents' explicit moral judgments were not a mediating factor in their eyewitness identification tasks. What is not clear is whether this finding is due to a lack of understanding in how the younger participants interpreted the films, particularly as to the theoretically important differences between the actors' intent and the outcome of the act, or to something more general, the difference between making an explicit moral judgment and making an implicit judgment, as in setting decisional criteria.

Moral Judgment Composite Results: Examining the results of the Moral Judgment Composite [i.e., how wrong was the act and how bad was the person committing it?] helps clarify why the structured equation model did not hold since there was not the expected interaction effect between the type of transgression (film) and age for the Moral

Judgment Composite scores, which Piaget and many others had reported using the standard vignette methodology. Instead, in both conditions [films] the Moral Judgment Composite measure gets less severe (non-significantly) with age and the bad intention act is judged worse than the bad outcome act. Thus the premise of the study, positing a shift from outcome to intention as moral criterion, was not supported.

In this study, older children (10-12 year-olds) do appear to approach the eyewitness face recognition task in moral terms and do differentiate between the two conditions, but the pattern of their decisional criterion and stated beliefs differ markedly. In particular, the 10-12 year-olds' bias scores indicate a looser criterion in the Fire/Neutral Intent condition; however, they (non-significantly) tended to judge false positive errors as worse in that condition than in the No Fire/Bad Intent condition (See Figure 14). Why the 10-12 year-olds stated beliefs about which type of error is worse to commit appear to contradict their eyewitness identification performance is not clear. However, this conflict between the moral beliefs and the conservative and/or lax criterion of the 10-12 year-olds in this study may actually shed light on how children develop simultaneously between understanding the actor's guilt and culpability at different levels of awareness (implicit/explicit). Most of the recent literature on moral judgments has involved verbally presented vignettes, sometimes accompanied by graphics, and not more realistic film depictions of live events. In this study, films were used that simulated a real-life event and the intention and outcome were explicitly depicted. However, the participants' may have intuitively felt differently about the intent of the actor based on the outcome information that was provided in the two film versions.

What may be related is a phenomenon termed, *The Knobe Effect* at an implicit level. An experimental philosopher, Joshua Knobe (2003) has argued that an individual's intuition about whether an act was intentional depends, in general, on whether the outcome was good or bad. This has been called *The Knobe Effect*. In the current study, it is possible that the participants' *intuitively* felt that the actor in the Fire/Neutral Condition had intended to cause harm solely on the basis of the fire ruining the restaurant. Thus, identifying the perpetrator may have operated like an implicit judgment reflecting the eyewitness's intuitive feeling that the perpetrator should be judged as if he caused the fire intentionally even in the neutral intentions condition.

Thus, implicit as well as explicit judgments may also be influenced by the outcome information that was made available to the child. In other words, an intentional criterion is not really of the effect of extreme outcomes; it may really depend on the outcome information, despite philosophical and legal principles to the contrary. Additional research has shown that an eyewitness may become emotionally aroused by the bad outcome regardless of the actor's intention. For example, a study by Remjin and Crombag's (2007) used a videotaped event, where a policeman pushed another man who fell to the ground. When the participants in the study were told that the man had died as a result of the event, Remjin and Crombag (2007) found that effect knowledge of an event's outcome, "the more blame witnesses tend to attribute to the perpetrator" and tended to overestimate the amount of violence that the event involved.

Perhaps, the eyewitnesses in this dissertation study that viewed the neutral intent/ bad outcome condition felt an immediate need to attribute blame to the perpetrator and thus

judged this act to be more morally wrong and this inclined them to be sure to identify a 'perp' during the identification task all at an implicit level. If this is the case, it may mean that eyewitnesses who testify with full knowledge of the unintended bad outcome of an event as serious (Fire) may feel a need to identify the accused as the perpetrator, thus exacerbating their false positive error rate and weakening the decisional criterion. Furthermore, while the 10-12 year-old children's *explicit* judgments may depend on the actor's *intentions* and not the outcome, their *implicit* judgments (reflected in their decisional bias on the eyewitness task) may have been influenced by the *outcome.* If this is true, it is indeed interesting both theoretically and practically.

Conclusion

This study demonstrates the importance of the moral significance of the transgression for face recognition in the eyewitness context, specifically how different kinds of framing of the act affects decisional criteria. This approach may be particularly important when studying eyewitness identification by children and adolescence, whose moral framing of events is dramatically changing. The research strategy used here has rested on the combination of two very disparate approaches: a) decisional theory, in particular, signal detection theory, and b) moral development theory and research. I believe this synergy of two very different theoretical approaches may add to our understanding of children's eyewitness identification. Understanding children's abilities to differentiate the moral issues of an event and what decisional criterion that they use as an eyewitness is paramount in the criminal justice system.

A limitation to this study is in its lack of ecological validity. This study used a film to create a more realistic life event. However, it is a laboratory study and not an actual court or court-simulated event. The participants were all recruited from Queens and Manhattan, two boroughs in NY City and thus may not be representative of the general population. In the future, it would be informative to have more participants in the 7-9 year-old and 10-12 year-old age groups to further explore the findings reported here. In addition, if the intentions of the actor were held constant and the damage varied it might help clarify whether variations in outcome alone led to the pattern of results or to disentangle the moral framings of outcome and intent (i.e., a good intention with a bad outcome vs. a bad intention vs. a good outcome).

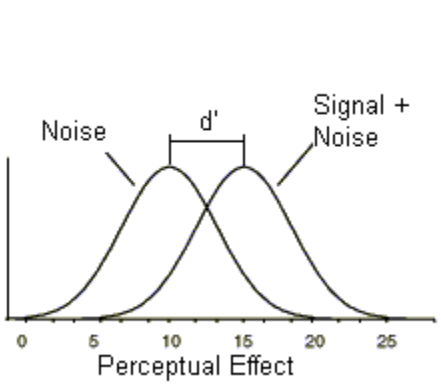
Future Research

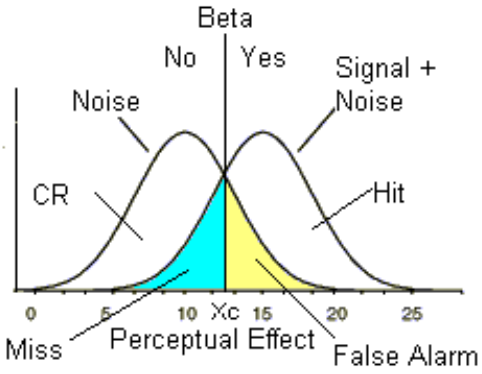
I believe that research that investigates how the moral framing and understanding of a crime influences eyewitness identification and decisional processes is sorely needed, especially with children and adolescents. Research findings in this area may provide guidance for forensic procedures in an effort to reduce wrongful convictions that are based on eyewitness identifications.

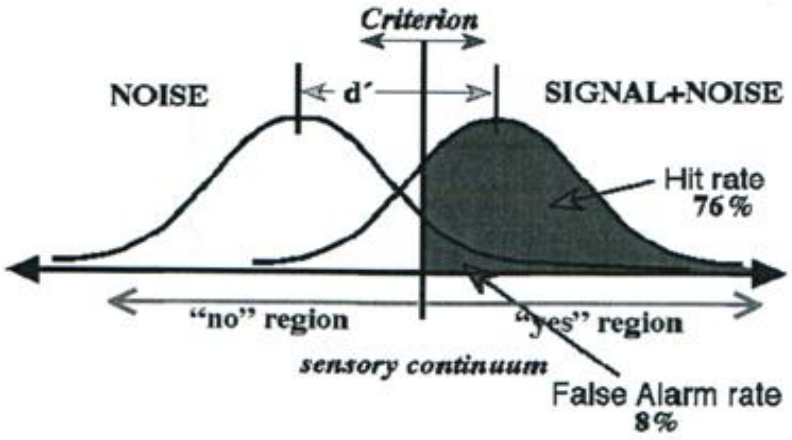
APPENDIX

Appendix A

Theory of Signal Detection





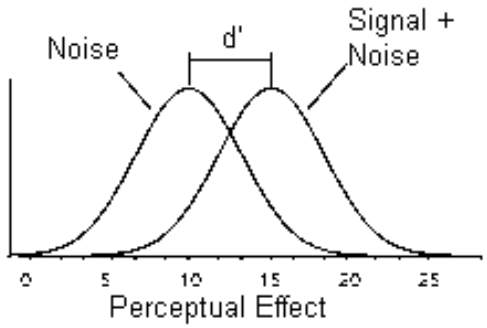


1. A) Dprime (*d'*) corresponds to the proportion of *false alarm rates* (FAR) subtracted from the proportion of *hit rates* (HR). The more accurate a participant is in discriminating between new and old items [in this study perp and foil faces] the greater the *d'*. In other words, if an individual is unable to discriminate any foils from perp, his/her hit rate will be equal to his/her false alarm rate, and *d'* will be zero.

B) Decisional Criterion (C) corresponds to the distance between the criterion and the neutral point, where neither response (yes or no) is favored. For instance, a negative C value will signify a bias towards responding "*yes*" and a positive C value will signify towards responding "*no*".

Appendix B

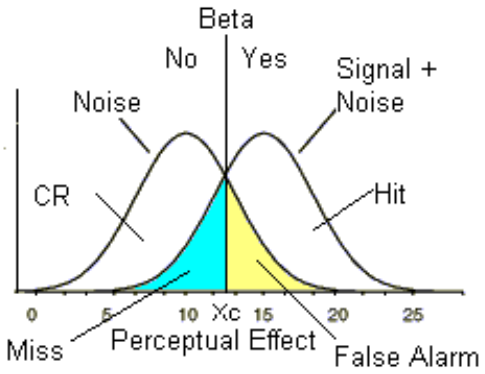
2. A) Response Criterion Analysis refers to a relationship of probability, possible responses, and stimuli in the environment.



a. The curve on the left represents the probability that the detection of the signal is due to noise alone.
b. The curve to the right represents the probability that the detection of the signal is to the presence of the signal.
c. The point in the center of the intersection represents the point where the detection of the signal is equally probable both to the presence of noise or signal plus noise.
d. The area of overlap is where an error occurs, either a miss or a false alarm.

There are three types of response criterion (Beta or C) in which an operator may adopt with their own implications.

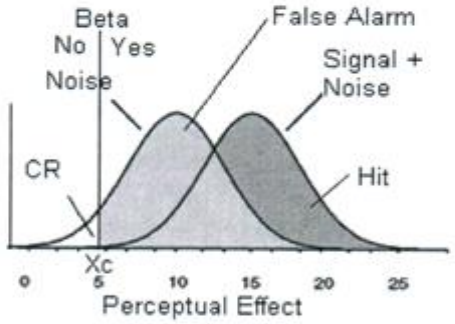
Neutral Criterion



Beta corresponds to Decisional Criterion (C)

There is an equal probability of getting a miss or a false alarm, C = 0

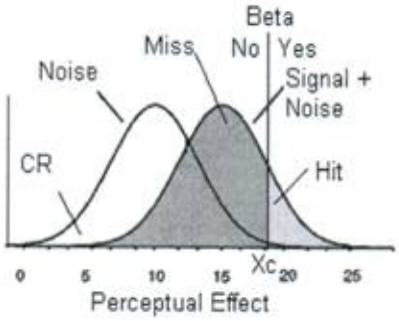
Liberal Criterion



Beta corresponds to Decisional Criterion (C)

The participant responds "*yes, signal present*" all of the time – leading to a high rate of false alarms and a high rate of hits, C< 0.

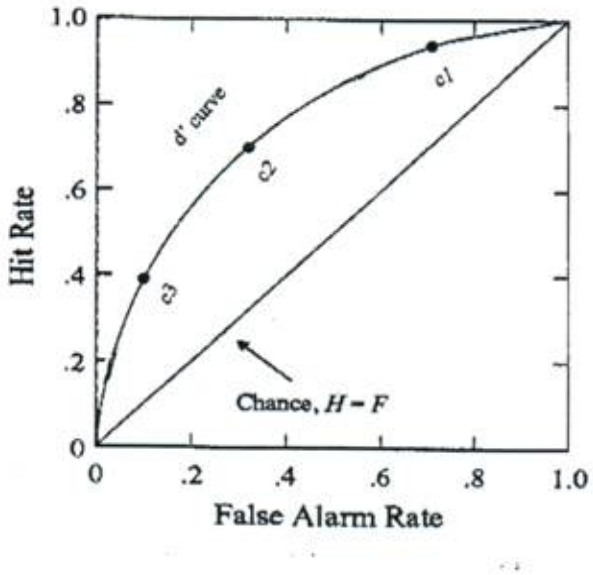
Conservative Criterion



Beta corresponds to Decisional Criterion (C)

The participant responds "*no, there is no signal present*" most of the time which leads to a low number of hits and a high number of misses, but few false alarms, C > 0.

2. B) Response Criterion Analysis with three points (C1, C2, and C3)



The graph above shows the decisional criterion (C3), where Z is equal to where

the participant puts the criterion in relation to the mean of the noise distribution. Note that the value of *d′* is constant while the decisional criteria change, because different criteria have different Z values. Therefore, d' measure is theoretically independent from C measure. Statistically "C" values are only related to "*d*'" values in a way that both measures are derived from Z value.

Appendix C

Table 19

Analysis of Variance for C1 ('little sure it is not the man') by Condition and Age.

Source	*SS*	*df*	*MS*	*F*	*p*
		Between subjects			
Condition	1.86	1	1.86	6.28**	.014
Age group	1.77	3	.59	1.98	.120
Condition X Age group	2.80	3	.93	3.14**	.028
Within-group error	33.89	114	.29		

Note. Conditions: Bad intent neutral outcome (No fire) and Neutral intent bad outcome (Fire). Age groups: 7-9 yr., 10-12 yr., 13-15., and 19-25 yr.
*p <.05. **p < .01

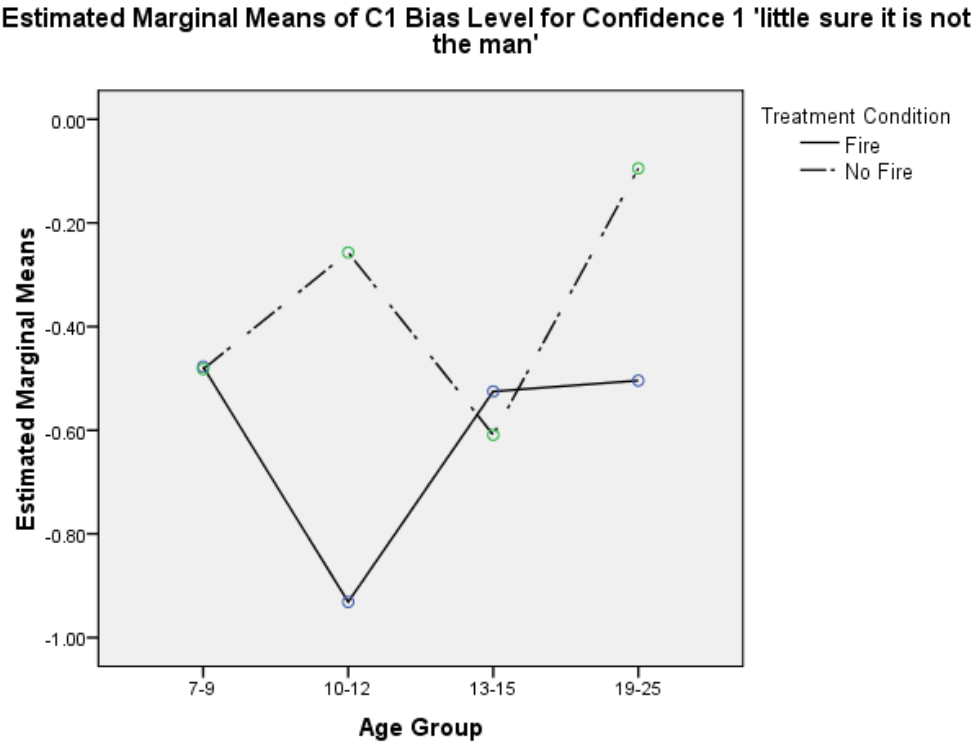


Figure16.

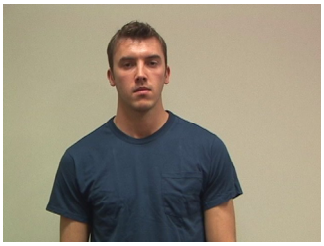
ANOVA C1 (‘*little sure that it is not the man*’) by condition[1] and age.
[1]Neutral intent/bad outcome (fire); Bad intent/neutral outcome (no fire).

Appendix D

The perpetrator/fire starter (first), bystander/waiter (last), and four “foils” for sequential lineup.





Appendix E

Assent for children under age 12 – script

"Hello, my name is _____. I go to school at a college in New York. Let me tell you what I do there. I study how kids and grown-ups think about things and make decisions. That's why I like to talk with kids about different things.

So, thanks very much for come here and helping me today. If there is something I ask you to do or say to you that you don't understand, let me know and I'll explain. If there's anything that I ask you to do that you don't want to do, let me know and we can stop right away and you can go back to your class. No one will be angry with you. Okay? But I think that you'll find it fun."

Before showing video: "I would like to tape record this interview so that I can listen to you and not have to worry about writing things down. Is that okay?" (Start recording if child agrees.)

"You are going to see a video. The video is not real, it is make believe or pretend like on TV, but I want you to watch very closely. Later I am going to ask you to pretend that you were really there and ask you some questions about what happened and who was there. Okay? Are you ready to start?

(video is shown)

"Now can you tell me what you saw on the video?" (If the child misses any critical elements of the action, the interviewer reminds them of what happened) (In counter balance order ask three questions) "How bad was the man?" not bad at all, a little bad, medium bad, very bad

"How bad was what the man did?" not bad, a little bad, medium bad, very bad

"How bad was what happened to the restaurant?" not bad, a little bad, medium bad, very bad

"So now, I'm going to show you some photos of people on the computer, one photo at a time, just like a TV show. You're going to be the witness, who sees a crime, and is trying to help the police by telling them who started the fire. So, I'm going to show you each person's photo, once from the front (gesture) and once from the left profile and once from the right profile (gesture). [Then the confidence scale is carefully explained to the child. See Appendix F.] "Do you understand what I would like you to do?"

(Show photo to child and say "Is this the man who started the fire? How sure are you?" Make sure that the child understand and uses the confidence scale. See Appendix G)

Randomize: After a series of the identifications, the interviewer says, "Thank you very much – you did just great. Now, I would like to ask you some more questions about what you saw and did, okay? You know that people can make mistakes and witnesses can make mistakes, right? (Showing relevant photos. See Appendix H for example.). You just said that this is the man who started the fire. What if you made a mistake and he wasn't the one who started the fire? How bad is it to make that kind of mistake: very bad, medium bad or not so bad? Why?" [False positive].

Now (picking up a different photo, that the child said was not the perp. See Appendix 9 for example.) "You told me that this man didn't set the fire. What if you made a mistake and that guy was really the one who set the fire, but you said he didn't? How bad is it to make that kind of mistake: very bad, medium bad or not so bad? Why?" [False negative].

"Which kind of mistake – saying someone did start a fire when he really didn't or saying

someone didn't start a fire when he really did – is worse? Why? "(As usual in this kind of research, answers are gently probed to disambiguate any unclear responses.)

Finally, the child is told: "Thank you very much. You've really helped me to better understand what kids think. Now, I've asked you a lot of questions -do you want to ask me anything about what we talked about? Were the answers hard? How? (The interviewer emphasizes that a lot of questions had to do with people's opinions and that there are no right or wrong answers about opinions. In addition, the interviewer cautions the children about fires and fire starting).

The children are then thanked again and given a pen and a certificate for participating. "You can go back to your class now."

Appendix F - Computer Practice

1. *'Click the button that would tell me that you were very sure about something - like your name';*

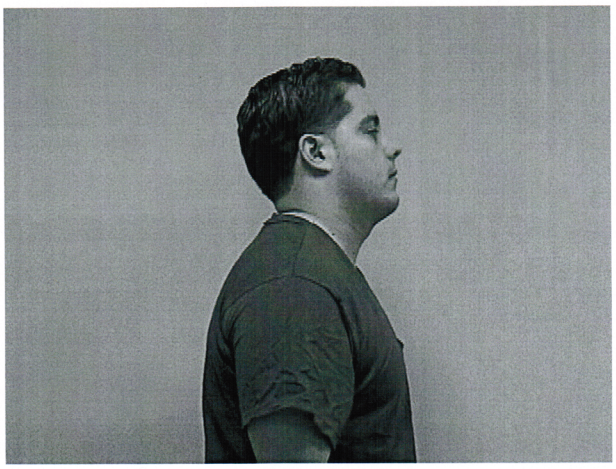
2. *'Click the button that would tell me that you are just a little bit less sure than the last time';*

3. *'Click the button that would tell me how sure you are that today is Sunday';*

4. *'Click the button that would tell me how sure you are just a little bit less sure that it isn't Sunday'.*

Appendix G: Computer Photo

eye witness testimony

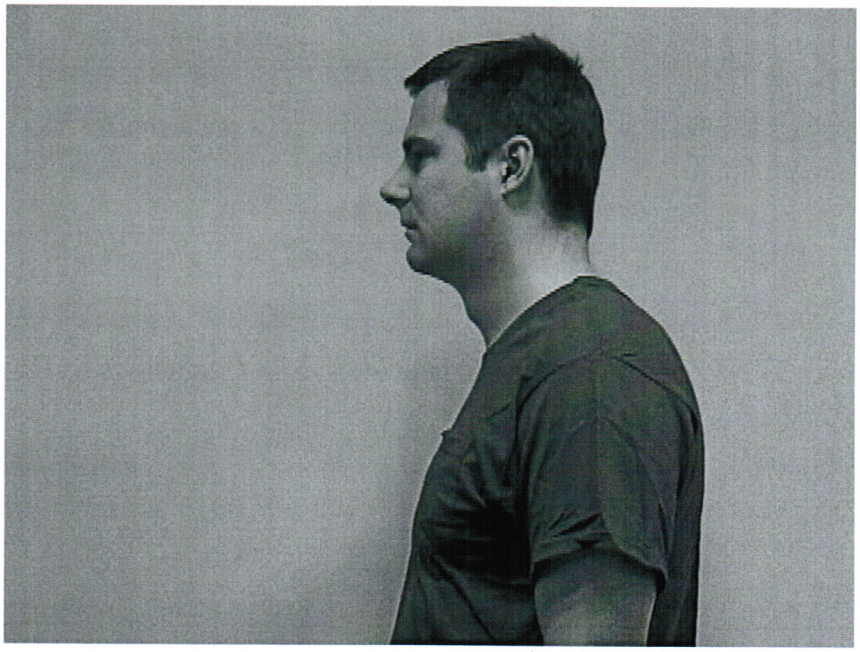


Very Sure It **Isn't** ☺☺ | A Bit Sure It **Isn't** ☺ | A Bit Sure It **Is** ☺ | Very Sure It **Is** ☺☺

Appendix H : Example of last photo of *very sure it is the man*

Very Sure It Is

Appendix I: Example of last photo of *very sure it is not the man*



Very Sure It Isn't

Appendix J

Face recognition trial sheet that recorded responses from the computer program print out.

ID_____ Date: Trial 1

Trial #	Photograph Details	Very Sure It Isn't	A Bit Sure It Isn't	A Bit Sure It Is	Very Sure It Is
1	Bernardo Right	☺☺	☺	☺	☺☺
2	Stan Center	☺☺	☺	☺	☺☺
3	Bob Center	☺☺	☺	☺	☺☺
4	Chris Left	☺☺	☺	☺	☺☺
5	Dean Right	☺☺	☺	☺	☺☺
6	David Right	☺☺	☺	☺	☺☺
7	Bernardo Center	☺☺	☺	☺	☺☺
8	Bob Right	☺☺	☺	☺	☺☺
9	Bernardo Left	☺☺	☺	☺	☺☺
10	Stan Right	☺☺	☺	☺	☺☺
11	David Center	☺☺	☺	☺	☺☺
12	Bob Left	☺☺	☺	☺	☺☺
13	Chris Center	☺☺	☺	☺	☺☺
14	Dean Left	☺☺	☺	☺	☺☺
15	David Left	☺☺	☺	☺	☺☺
16	Stan Left	☺☺	☺	☺	☺☺
17	Chris Right	☺☺	☺	☺	☺☺
18	Dean Center	☺☺	☺	☺	☺☺

References

Armsby, R. E. (1971). A reexamination of the development of moral judgment in children. *Child Development, 42*(4), 1241-1248.

Austin, V. D., Ruble, D. N., & Trabasso, T. (1977). Recall and order effects as factors in children's moral judgments. *Developmental Psychology, 48*, 470-474.

Baron, J., Granato, L., & Spranca, M., & Teubal. (1993). Decision-making biases in children and early adolescents: Exploratory studies. *Merrill-Palmer Quarterly,* 39(1), 22-46.

Berndt, T. J. & Berndt, E. G. (1975). Children's use of motives and intentionality in person perception and moral judgment. *Child Development, 46*(4) 904-912.

Brewer, N. & Day, K. (2005). The confidence-accuracy and decision latency-accuracy relationships in children's eyewitness identification. *Psychiatry, Psychology and Law, 12*, 119-128.

Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness identification. In N. Brewer, & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp.177-221). New York: Guilford.

Butler, S., Gross, J., & Hayne, H. (1995). The effect of drawing on memory performance in young children. *Developmental Psychology, 31*(4) 597-608.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology, 16*(4), 257-269.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*(3) 403-439.

Ceci, S.J., Ross, D. F. , & Toglia, M. P. (1987). Suggestibility of children's memory: Psychological Implications. *Journal of Experimental Psychology. 116*(1), 38-49.

Chance , J., & Goldstein A.G. (1984) Face-recognition memory: Implications for children's eyewitness testimony. *Journal of Social Issues, 40*, 69-85.

Cole, C.B., & Loftus, E.F. (1987). The memory of children. In S.J. Ceci, M.P., Toglia, & D.F. Ross (Eds.), *Children's eyewitness memory* (pp.178-208). New York: Springer-Verlag.

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.

Cutler, B.L. & Penrod, S.D. (1995). Mistaken Identification: The eyewitness, psychology, and the law. New York: Cambridge University Press.

da Silva, J. (2007). Accuracy and decision-making criteria in cross-race eyewitness identification. Unpublished doctoral dissertation, Graduate Center of the City University of New York.

Feinfield, K.A., Lee, P.P., Flavell, E. R., Green, F.L., & Flavell, J.H. (1999). Young children's understanding of intention. *Cognitive Development, 14*(3), 463-486.

Feldman, M., Frankel, F. & Sharp, D. (1971). Development of free recall learning in children. *Developmental Psychology, 4,* 109-123.

Fivush, R., & Shukat, J.R. (1995). Content, consistency, and coherence in early autobiographical recall. In M.S. Zaragoza, J.R. Graham, G.C.N. Hall, R. Hirschman, & Y.S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp.5-23). Thousand Oaks, CA: Sage Publications.

Flavell, J.H. (1999) Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50,* 21-45.

Flin, R.H., (1980). Age effects in children's memory for unfamiliar faces. *Developmental Psychology, 16*(4), 373-374.

Goodman, G.S., & Bottoms, B.L. (1993). *Child victims ,child witnesses: Understanding and improving testimony.* New York: Guilford Press.

Goodman, G. S., Rudy, L., Bottoms, B., & Aman, C. (1990). Children's concerns and memory: Issues of ecological validity in the study of children's eyewitness testimony. In R. Fivush & J. A. Hudson (Eds.), *Knowing and remembering in young children,* (pp. 249-284). New York: Cambridge University Press.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: John Wiley.

Heeger, D. (1998). Signal Detection Theory. http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology, 1* (6), 659-671.

Knobe, J (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16(2),* 309-324.

Lindsay, R.C., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84*(3), 315-321.

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

Loftus, E. F. (1979) *Eyewitness Testimony.* Cambridge MA: Harvard University Press. Seventh Edition.

Loftus, E, F., Loftus, G.R., & Messo, J. (1987). Some facts about "weapon focus." *Law and Human Behavior, 11,* 55-62.

Macmillan, N. A. & Creelman, C. D. (1991). Detection theory: A user's guide. Mahwah, NJ:Erlbaum. Second Edition.

Meissner, C. A. (2002). Applied aspects of the instructional bias effect in verbal overshadowing. *Applied Cognitive Psychology, 16*(8), 911-928.

Meisner, J. & Brigham, J.C. (2001). Thirty years of investigating the own-race bias in memory for faces. *Psychology, Public Policy and Law.7*(1), 3-35.

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.

Narby, D. J., Cutler, B.L., & Penrod, S. D. (1996).The effects of witness, target, and situational factors on eyewitness identifications. In S.L. Sporer, R. S., Malpass, & G. Kohnken (Eds.), *Psychological issues in eyewitness identification* (pp. 23-52). Mahwah, NJ: Lawrence Erlbaum.

Nummendal, S. G., & Bass, S. C. (1976). Effects of the salience of intention and consequence on children's moral judgments. *Developmental Psychology, 12*, 475-476.

Piaget, J., (1932/1965). *The moral judgment of the child.* New York: New York Free Press.

Pipe, M.-E. (1993). Children's testimony. In P. R. H. Webb, J. G. Adams, W. R. Atkin, L. L. Caldwell, & R.M. Henaghan (Eds.), *Family law guide* (6th ed., pp. 522-533). Wellington: Butterworths.

Pozzulo, J. D., & Lindsay, R. C. L. (1998) Identification accuracy of children versus adults: A meta analysis. *Law and Human Behavior*, *22* (5), 549-570.

Pozzulo, J. D. & Lindsay, R.C.L., (1999) Elimination line-ups: An improved identification for child eyewitnesses. *Journal of Applied Psychology, 84*(2), 167-176.

Read, J.D., Yuille, J.C., & Tollestrup, P. (1992). Recollections of a robbery: Effects of arousal and alcohol upon recall and person identification. *Law and Human Behavior, 16*(4), 425-446.

Remjin, C. A., & Crombag, H. F. (2007). Heuristics in causal reasoning and their influence on eyewitness testimony. *Psychology, Crime & Law, 13*(2), 201-211.

Roebers, C.M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology, 38*(6),1052-1067.

Roebers , C.M. & Howie, P. (2003) Confidence judgments in event recall: Developmental progression in the impact of question format. *Journal of Experimental Child Psychology, 85*(4), 352-371.

Roebers, C. M., Moga, N., & Schneider, W. (2001). The role of accuracy and motivation on children's and adults' event recall. *Journal of Experimental Child Psychology, 78*(4), 313-329.

Salmon, K., Bidrose, S., & Pipe, M.-E. (1995). Providing props to facilitate children's event reports: A comparison of toys and real items. *Journal of Experimental Child Psychology, 60*(1), 174-194.

Saltzstein, H., & Peach, R. (2005). *Children's eyewitness testimony as moral decision-making*. Unpublished manuscript, The Graduate Center of the City University of New York.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual Innocence: Five days to execution and other dispatches from the wrongly convicted.* New York: Doubleday.

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial recognition studies. *Psychological Bulletin, 100*(2),139-156.

Shultz, T. R. (1980). Development of the concept of intention. In W. A. Collins (Ed.), *The Minnesota Symposia on Child Psychology* (Vol.13). Hillsdale, NJ: Erlbaum.

Sophian, C., & Stigler, J. W. (1981). Does recognition memory improve with age? *Journal of Experimental Psychology, 32*(2), 343-353.

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

Spring, T. (2006, May). *A moral decision-making approach to children's eyewitness testimony: A replication and extension.* Paper presented at the meeting of the Association of Psychological Science, New York, NY.

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.

Swets, J.A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99*(2), 181-198.

Tsukomoto, N. & Saltzstein, H. (2009). *Collective punishment*. Unpublished manuscript, The Graduate Center of the City University of New York.

Van Zandt, T. *(*2000*).* ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 582-600.

Vickers, D. (1979) *Decision processes in visual perception.* New York: Academic Press.

Wells, G.L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, *36*(12), 1546-1557.

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

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*(3), 360-376.

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8*(3), 155-167.

Wells, G.L., Small, M. Penrod, S.D., Malpass, R.S., Fulero, S.M., & Brimacombe, C.A.E. (1998). Eyewitness identification procedures: Recommendations for line-ups and photo spreads. *Law and human Behavior, 22*(6), 603-647.

Yuill, N. (1984). Young children's coordination of motive and outcome judgments of satisfaction and morality. *British Journal of Developmental Psychology, 2*(1), 73-81.

Yuille, J.C., & Wells, G.L. (1991). Concerns about the application of research findings: The issue of ecological validity. In J. Doris (Ed.), The suggestibility of children's recollections: Implications for eyewitness testimony (pp. 118-128). Washington, DC: American Psychological Association.