City University of New York (CUNY)

CUNY Academic Works

Publications and Research

Queens College

2012

Children's Eyewitness Identification as Implicit Moral Decision-Making

Toni Spring
CUNY Queens College

Herbert D. Saltzstein
CUNY Graduate Center

Roger Peach
CUNY Lehman College

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/qc_pubs/431

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

Children's Eyewitness Identification as Implicit Moral Decision-Making

Toni Spring, Herbert D. Saltzstein, & Roger Peach

Correspondence to: Toni Spring, Department of Elementary and Early

Childhood Education, Queens College, CUNY, Flushing, USA.

E-mail: toni.spring@qc.cuny.edu

This is an Accepted Manuscript of an article published by Wiley Online Library in Applied Cognitive Psychology, Appl. Cognit. Psychol. (2012) and is available online: in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/acp.2871

Abstract

Why are young children particularly prone to make false positive errors or false alarms when identifying a wrongdoer? In three studies the problem was approached using a signal detection analysis, focusing on the moral costs of false alarms, as understood at different points in development. The findings are: (1) decisional criteria became more conservative, indicating fewer false alarms, with age in three studies, (2) children's beliefs about the seriousness of false alarms and misses changed from (a) a non-moral concern to (b) a moral concern for false negatives or misses to (c) a moral concern for false alarms. (3) These findings were replicated in two demographically different communities. More critically, (4) how the filmed event is framed e.g., as a moral transgression (stealing) or a pro-social (helping) act (Study 1); and as intentional with little damage or unintentional with major damage (Study 3), interacts with age in influencing decisional criteria.

Key Words: Moral Decision-Making, Eyewitness Identification, Signal Detection, Moral

Intentionality

Psychologists have had a long-term interest in children's eyewitness identification for both theoretical and practical reasons. The policy implications are obvious (e.g., Ceci & Bruck, 1995; NY Times editorial, 8/26/11; Poole & Lamb, 1998) given the controversies surrounding the veracity of children's testimony and about the processes involved in obtaining testimony, especially as to whether young children's testimony is reliable and especially likely to lead to false convictions (e.g., Ceci & Bruck, 1993; Cutler, & Penrod, 1995; Pezdek & Roe, 1994). Thus, interest in children's eyewitness identification has spawned a huge research literature. It has been studied from the point of view of memory, suggestibility, cognitive development, decision-making, etc., and combinations of these perspectives. This intersection of traditional sub-fields of study, e.g., memory and suggestibility, (e.g., Ceci & Bruck, 1993) has also lent particular theoretical interest to the study of children's eyewitness identification. Here we focus on eyewitness identification as moral decision-making.

Several studies (e.g., Pozzulo & Lindsay, 1999; Roebbers, Moga, & Schneider, 2001; Wells & Loftus, 2003; see Keast, Brewer & Wells, 2007 for mixed results) have shown that younger children are not only inclined to make errors in general when trying to identify the guilty person, but in particular, to make more *false positive errors* or *false alarms* (i.e., declare an innocent person guilty) than *false negative errors* or *misses* (failing to identify the true culprit).¹ Further, confidence is no guarantee of accuracy (e.g., Sporer et.al, 1995; Keast et.al, 2007).

We begin with the question, 'What is different about eyewitness identification and other forms of person identification?' The difference between choosing someone as (i) guilty or as innocent of a transgression and (ii) remembering whether someone was present or as absent (at an event) is that choosing someone as guilty or as innocent is to make a moral decision, that is, a

decision with obvious implications for fairness to others and for the welfare of society, whereas simply identifying someone as being present at an event has no such implications. In a sense, we are specifying an answer to the suggestion raised by Brewer, Weber, and Semmler (2005): ‘They [young children] may have a less sophisticated understanding of the purposes of an identification test and the consequences of choosing or not choosing, making an incorrect choice, or even making a correct choice’ (p. 187).

Thus, our central assumption is that eyewitness testimony has moral consequences, which may be understood differently at different points in moral and cognitive development. What might these implications be? Our overarching hypothesis is based on the seminal research on children’s moral judgments by Jean Piaget (1965/1932). According to Piaget, children at different points in development view the purposes and fairness of punishment differently. He found that younger children believe in the superiority of what he termed *expiative justice*, harsh and arbitrary punishments, in contrast to older children’s preference for *restitutive justice*, punishment which is proportional, moderate, and related to the offense. Younger children also approve of all forms of *collective punishment*, the punishment of the group for the misbehavior of one of its members; whereas older children only do so under certain special conditions. He also posited that young children believe that wrongdoing, even with unintended consequences, demands punishment, and the harsher, the better. He found that younger children tend to follow consequences rather than intentions in judging culpability. The importance of moral intentionality and outcome will be addressed when we describe Study 3.

Despite this potential link between children’s moral development, especially their attitudes towards punishment, and their eyewitness identification, there are, to our knowledge, no published empirical studies that have examined children’s identification errors within a moral

decision-making frame- work. Thus, what we believe is needed to understand the development of children's eyewitness identification fully is to place it in the framework of the development of moral decision-making.

Rather than simply ask about the seriousness of *false alarms* and *misses* in a de-contextualized and non-moral context, we decided to set the inquiry in the context of a facial identification or eyewitness task.² Three studies are reported here, which have the following characteristics: (i) Each participant is individually shown a specially prepared brief film and then performs an identification task in which the participant is asked to identify whether each of a series of face photos, presented sequentially, is the guilty person or not and how confident the 'witness' is of his or her judgments; (ii) in two of the studies, the visually identical film was 'framed' (verbally described by means of a voice-over) so as to have different moral implications. That is, each participant saw the 54 photos *sequentially*. Thus, there are nine signal plus noise (perp) trials and 45 noise only (foils) trials per participant. The first study framed the film as either (a) a transgression, stealing, or (b) a pro-social act, helping. The second study used only the stealing framing but introduced instructions to discourage either false alarms or misses. In the third study, the act was framed as either (c) an intended transgression resulting in little material damage or (d) an unintended transgression resulting in major material damage. (iii) At the end of the identification task, each participant was then asked about the seriousness of false alarms or misses in a child-friendly way, described in the succeeding discussions.

Our initial prediction was that younger children would choose the misses as worse with justifying reasons, which stressed the importance of punishing the guilty party, but that with age this would shift to a choice of false alarms as worse because of the desire to avoid punishing an innocent person, in accordance with Western philosophical and legal concerns. Of course, we did

not expect well-formed philosophical or legal justifications in support of these judgments. The exact point of transition could not realistically be predicted because as prior research in the area of moral development has shown, these points of transition depend on socio-cultural factors, such as socio-economic status (SES), as well as how the questions are posed (e.g., Kohlberg, 1969). With much developmental research, it is the direction of change rather than the exact points of transition that is deemed critical.

Our research strategy utilizes the signal detection theory (Green & Swets, 1966) as a useful tool in analyzing such decisions in that it separates two aspects of performance: (i) sensitivity or discriminability between signal (here, the perpetrator) and noise (the foils) and (ii) bias or the criterion set for making a positive identification. The latter is an index of the relative balance of false alarms to misses corrected for chance. Unlike many studies of eyewitness identification, it is this measure of decisional bias, termed bias, where the decision-maker implicitly sets his or her criterion for making a positive identification, rather than the overall accuracy (sensitivity), which is the focus of interest.

The studies have these purposes: the first was to demonstrate the feasibility of studying children's, in particular, younger children's (6 to 9 years old) judgments using a signal detection method and demonstrating the two predicted phenomena: (i) that bias or the index of the decisional criterion implicit in the identification judgments increases (indicating fewer false alarms) with age if the actor being judged committed a moral transgression; (ii) that the children's explicit judgments of which kind of decisional error is worse also changes with age; and (iii) that these two kinds of indices, one involving actual performance on the eyewitness/facial recognition task and the other the beliefs about the seriousness of false alarms/false negatives, are related.

Although detection-theory procedures have been applied extensively to the investigation of many sensory modalities (e.g., Gescheider, 1985) and memory (e.g., Miller & Wolford, 1999), there have been fewer studies of children's memory that have used signal detection-theory procedures (Cultice, Somerville, & Wellman, 1983; Gee & Pipe, 1995; Locksley, Stangor, Hepburn, Grosovsky, & Hochstrasser, 1984; Meissner, Tredoux, Parker, & MacLin, 2005; Pezdek & Banks, 1996), and, to our knowledge, even fewer have used the full signal detection procedures to examine children's decision-making in an eyewitness identification context, for example, sometimes too few trials were used in the studies.

Our *first hypothesis* is that younger children's judgments use a less stringent decisional criterion (indicated by a lower bias score) than would older children and adolescents, over and above any differences in sensitivity or accuracy. Hence, younger children would make more false alarms than older children and adolescents.

Study 1

The purpose of the first study was to explore (1) whether children as young as 6 or 7 years could understand and perform the task, (2) whether decisional bias or the tendency to use a more stringent decisional criterion, committing *fewer* false alarms, increases with age, and (3) whether beliefs about decisional errors, whether false alarms or misses are worse, also change with age.

Method

Participants

Forty-eight children were recruited from a Roman Catholic parochial school in one of the outer-boroughs of New York City. This community and school itself are characterized by great

ethnic diversity and include many children whose families were from Latin America and Asia.

All children were fluent in English. Twenty-five participants were male and 23 female.

Twenty-four 6-9-year-olds (14 male and 10 female, $M = 7.5$, $SD = 1.2$ mos.) and 24 11-to 14-year-olds (11 male, 13 female, $M = 12.6$, $SD = .84$ mos.) participated with parental permission.

Design and Procedure

The design was in a 2 age groups by 2 conditions (moral transgression /stealing vs. pro-social/helping) factorial design. Twenty-four participants were randomly assigned to the *stealing* and 24 to the *helping* conditions. All participants were individually shown a short video depicting four men, all dressed in identical red t-shirts, practicing soccer in a locker-room. They enter the locker room together, put their gear away and then at the end of the video one of the men leaves carrying two sports bags, one with a New York Yankees logo on it, which had been brought into the locker room by one of the other men.

Each child was given instructions both before and after seeing the film, which were designed to set the meaning of or frame the act, as a moral transgression (stealing) or pro-social moral act (helping). The participants in the *Stealing* condition were asked to pretend that they were witnesses to a crime and that their task was to identify the man who *stole* the bag. Those in the *Helping* condition were asked to pretend that the man took the bags to the owner in order to help him, and the owner wanted to find the man to give him a reward.

After seeing the film, they were asked to tell what they had seen, i.e., what had happened in the film, and if the participant did not mention the nature of the act (stealing or helping) the experimenter did. These instructions were given just before the child began the task to ensure that the framing of the film was remembered and understood. All children were then told that they were going to see three series of photographs, presented sequentially, and for each

photograph they were going to be asked whether or not it was the man who took the bags, and how sure they were whether it was or was not the man. Their judgments were made using a 4-point scale, consisting of two blue and two red smiling faces (see Appendix) to clearly distinguish the choices. The scale was carefully explained to each participant, and each child was familiarized with it and its use until he/she was proficient. Specifically, for each photograph, they were then asked to indicate whether they were *very sure that it was the man* who took the bag, or *a little bit sure that it was the man* who took the bag, or *a little bit sure that it was not the man* who took the bag, or *very sure it was not the man* who took the bag. This procedure is known in the signal detection literature as the confidence rating procedure and reduces the number of trials need to obtain reliable measures (Gescheider, 1985). A set of 18 photographs consisted of the man who had taken the bags away at the end of the video (the perpetrator or ‘perp’) and 5 foils (3 men who had not taken the bags away at the end of the video and 2 foils who were not in the video at all). The photos were ‘head-and-shoulders’ shots, with each individual wearing the same red t-shirt, as had been worn in the video. The perpetrator or ‘perp’ and the foils were all photographed against an identical, neutral background. Each individual was presented one-third of the time frontally, one-third in left profile and one-third in right profile. Within each of the three presentation sets, order of photo presentation was determined randomly, except that no individual’s photographs appeared consecutively in any one set. In all, children were exposed to 54 photographs (trials), with the ‘perp’ being presented nine times, and the foils being presented 45 times.

We should note that we understand that this lineup procedure is unusual and may not elicit the most accurate judgment. For example, the elimination line-up (e.g., Pozzulo & Lindsay, 1999) produces more accurate identification under certain conditions. However, maximizing

accuracy was not our immediate aim. Rather, this procedure was used in order to help uncover the underlying processes of eyewitness judgments, in particular *bias* or the relative preponderance of false alarms-to-misses errors.

At the conclusion of this phase of the study, children were asked a series of questions about the seriousness of false alarms and misses. This was done by individually showing them the last photos in which they asserted that the photo was the ‘man who did it’ and the last one in which the child asserted that the photo was not of the ‘man...’ (in counter-balanced order). We then asked; *let’s make believe that you made a mistake...after all everyone makes a mistake once in a while....how bad a mistake would that be: very bad, a little bad, or not so bad, and why?* The child was then asked if he or she had any questions, was praised for their cooperation, allowed to choose a colored pen as a reward and given a certificate of appreciation.

Data Analysis

There are two independent parameters that signal detection theory provides. The first parameter is *sensitivity* or discriminability, indicated by d' . The higher the d' score the more the participant is able to discriminate the perpetrator from the foils (in signal detection terms, the signal from the signal + noise). The second parameter, in which we are especially interested, is termed *bias*, which is an index of the decisional criterion used. Higher scores indicate *fewer* false alarms in relation to misses over-and-above what one would expect by chance, i.e., a more conservative or cautious decisional bias. There are three possible measures of bias or decisional criteria for each participant (C1, C2, & C3). The three measures were computed by dividing the responses at the three possible dividing points among the four alternatives: (1) *very sure that it is not the man*, (2) *a little bit sure that it is not the man*, (3) *a little bit sure that it is the man*, and (4) *very sure that it is the man*. The measure used was based on the division between (4) “very sure

that it is the man” and the other three alternatives (1, 2 & 3): i.e., “a little sure that it is the man”, “a little sure that is not the man” and “very sure that it is not the man” all counting as a “no” response. This decision to focus on this measure was based on the consistency of the results when it was used and also on the finding that young children are not only more prone to make more false alarms, but to identify positively *with certainty* (e.g., Brewer et al., 2005). The verbal answers to the questions at the end of the session about the seriousness of false alarms and misses were coded by independent assessors, who were graduate students and were blind to the participant’s age and experimental conditions.

Results

Eyewitness Performance

The measure of sensitivity, d' , increased significantly with age, $F(1, 44) = 9.84, p < .003, \eta_p^2 = .18$, and there was no interaction between age and condition, $F(1, 44) = 1.95, p > .10, \eta_p^2 = .03$ (See Figure 1). More critical to our hypotheses were the data for *bias* (C3) (See Figure 2). As noted above, here we report results for responses which indicated a *very sure* positive identification and the other three alternatives. Significantly lower bias scores indicate that the younger children used a looser criterion, i.e., committed relatively more false alarms in the *stealing* than *helping* condition. The overall ANOVA revealed a significant age effect, $F(1, 44) = 10.08, p = .003, \eta_p^2 = .19$, with the bias index increasing with grade (age) and a significant interaction effect between age and condition, $F(1, 44) = 4.76, p = .04, \eta_p^2 = .10$, as expected. When we tested for age effect by condition, we found: For the helping condition the age comparisons were not significant; $F(1, 22) = 0.54, p > .40$. However, the age effect was significant in the stealing condition was: $F(1, 20) = 5.72, p < .001, \eta_p^2 = .37$. The bias scores are

higher for the older children than the younger children indicating, as expected, that as children become older, they are particularly less likely to make false alarms errors.

Eyewitness Beliefs

We now turn to the children's beliefs about false alarms and misses as gleaned from the interviews. Inter-rater reliability for coding the results of the interviews was .85 for a sample of 45 responses from 15 interviews. Responses to the question as to whether it was worse to make a *false alarm* or a *miss* type of error were coded into four categories: (a) a false alarm is worse, (b) a miss is worse, (c) there is no difference between the two errors, and (d) a confused or 'don't know' response.

It became apparent that we needed to revise our original hypothesis, namely, that younger children would choose misses and older children would choose false alarms as worse. The younger children, who gave specific responses, chose all three possibilities, that is, false alarms are worse, misses are worse and both equally bad. This surprising lack of preference for misses became clearer when the reasons justifying their choices were examined. Many of these reasons did not refer to the moral costs of false alarms or misses, but focused instead on non-morally relevant considerations, especially task difficulty. Thus, the major results for the children's beliefs were: (1) many of the younger children did not justify their judgments of the seriousness of *false alarm* and *misses* by reference to the (moral) outcomes for the accused whereas the older children did; and (2) the older children judged false alarms as worse than misses, but primarily in the stealing condition. Test of association by chi-square between age and 'which is worse' response reached significance, $\chi^2(2) = 8.08, p < .02$.

Discussion

The data, in general, supported the performance hypothesis, that younger children would employ a more lax criterion when making eyewitness judgments specifically of *moral* transgressions. That is, in the moral transgression (stealing) condition, children in the younger age groups had lower bias scores, indicating that they were more prone to make false alarms, than children in the 11-14 years age group. Children in the youngest age group were more likely to be very sure that a person was guilty of stealing when he was not than those in the older age groups. However, this age effect was not found in the pro-social (helping) condition. This indicates that the moral context in which identification judgments are made influences the decision-making process, as expected.

Our second hypothesis concerning age differences in beliefs had to be modified. Some of the youngest (6-9 years old) children gave essentially non-moral reasons for explaining which kind of error was worse. These reasons usually focused on how difficult or easy the task was, and *made no mention of the importance of convicting the guilty or protecting the innocent*. We thus concluded that the framing of an event of their moral meaning affected the child's eyewitness performance in ways that paralleled their beliefs about the meaning of *false alarms*. However, as the study had been conducted at a special, parochial school, before further investigating these findings, we decided to test their *external validity* or generalizability, by replicating the main hypotheses with samples of children from two different populations.

Study 2

The first aim of the second study was to replicate the preliminary findings of the first study that younger children bring a decisional bias favoring false alarms to the eyewitness identification tasks (when the task is framed as stealing) with more narrowly defined age groups in two communities, which differed markedly in SES, thus testing for external validity. The

second aim was to investigate whether children's eyewitness performance and beliefs can be influenced by instructions, which either emphasized the dangers of false alarms or of misses. Such a test would indicate whether a 'mind-set' which young children might bring to the task was easily modified. Studies of suggestibility, of course, have a long history in developmental psychology (e.g., Ceci, S. J., & Bruck, M., 1993; Munsterberg, 1908; Roebbers, 2002; Yuille, & Wells, 1991) and suggestions have, of course, been used to study the malleability of adult and occasionally, child witness, but never, to our knowledge, specifically designed to affect the decisional criteria brought to an eyewitness task, as distinguished from affecting a particular identification judgment.

We should note that although there is a huge literature on socio-economic and cultural differences in moral judgment and reasoning (e.g., Turiel, 2002); we know of no studies which contrast children's eyewitness identification by socio-economic background. Indeed, our focus was not on these differences, as interesting as they might be, but on testing the external validity or generalizability of the developmental differences by studying them in two communities.

Method

Participants

Sixty-nine participants (36 males and 33 females) were drawn from three age groups; younger children 7-9 years old; intermediate children aged 10-12 years old; and early adolescents 13-15 years old recruited from two public schools in two communities differing SES (see succeeding discussions). The participants in the three age groups from first (lower SES) school community were: $M = 7.3$, $SD = 1.6$ months.; $M = 11.5$, $SD = 2.4$ months.; $M = 13.5$, $SD = 1.8$ months. Those from the second (upper-middle SES) school community were $M = 7.8$, $SD = 1.8$ months.; $M = 11.1$, $SD = 2.2$ months.; and, $M = 13.4$, $SD = 1.2$ months.

These school communities were in Nassau County and Suffolk County on Long Island, NY, and differed in general SES. The poverty rate (indicated by the percentage of students eligible for free or reduced cost lunches for the 2003-04 school year) was 66.9% and in the other was 7.9%. Most of the participants in the first school community were of Latino background whereas most of those in the second school community were of non-Latino European background. However, it should be noted that the communities were chosen primarily to contrast SES rather than ethnicity. The children from the first school community were recruited from two public schools; those in the second school community were recruited through the Parent Teacher Association and through a local buying cooperative. A few of the latter were seen privately at home. All others were seen in classrooms or similar rooms.

Design

The study had a factorial design with three between-subject factors: two school communities, three age groups and two experimental biasing conditions. Three age groups were included to better hone in on the developmental trends.

Procedure

The procedure was very similar to that used in the first study, but was administrated by computer to allow for one experimenter (the first author) to conduct the study. Also, only the *stealing* framing condition, and not the helping condition, was used since the developmental trends were significant and substantial only in that condition in the first study. Participants were randomly assigned to the two biasing instructions.

The questions about which kind of errors were worse and why were essentially the same as in Study 1. Two different biasing instructions were then given randomly to the children prior to viewing the sequential line-up. In trying to bias participants towards making false alarms

errors and away from making misses the children were told, “*You don’t want to let someone who really stole the backpack to get away with it, do you?* (Pause) *Don’t you think that the person who stole the backpack should get punished?*” In trying to bias children towards making misses and away from false alarms errors they were told, “*You don’t want someone who didn’t steal anything to get punished, do you?* (Pause) *Don’t you think that it would be unfair to punish someone who didn’t steal anything?*” These instructions were designed to alter the decisional criteria of the participants and were introduced before each of the three sequences of judgments.

Data Analysis

The method of calculating the different parameters of performance were the same as in Study 1, and again the bias measure presented is based on the difference between the two positive judgments, *a little sure it was the man* and *very sure it was the man*.

Results

The results for eyewitness identification (see Figure 3) were as follows: (1) The sensitivity measure (d') was not significant for age, $F(2, 68) = 6.62, p = .13, \eta_p^2 = .87$. (2) The results for C3, presented in Figure 4, were consistent with the first study: Overall, younger participants were significantly more likely to say that they were very sure it was the perp when it wasn’t than older participants: $F(2, 66) = 7.66, p < .001, \eta_p^2 = .20$. Post-hoc analyses showed that early adolescents (13-15 years) in both communities were less likely to say they were very sure it was the perp when it was not than both the younger (ages 7-9 years) and the older (ages 10-12 years) children. In addition, children in the upper-middle SES community generally were significantly less likely to say they were very sure it was the perp when it wasn’t than those in the lower SES community $F(1, 68) = 10.87, p < .002, \eta_p^2 = .13$. (3) Surprisingly, biasing

instructions had absolutely no effect on children's performance in either community neither as a main effect $F(1, 68) = .228$ $p > .60$, $\eta_p^2 = .004$, nor in interaction with age, $F(2, 68) = .08$, $p > .9$.

The findings show that the younger eyewitnesses and those from the lower SES community both are more willing to say that they are definitely sure that someone is the guilty individual than did the older children. This pattern both confirms the earlier developmental finding just reported, and also shows that community background (a combination of social class and ethnicity) also affects performance bias.

In addition to which kind of error was worse (false alarms or misses), reasons to justify these choices were coded for concern for fairness to the accused or concern for punishing the guilty. The number of reasons focusing on (a) punishing the guilty was subtracted from (b) the number focusing on protecting the innocent. Non-moral reasons, for example, focusing on task difficulty, were scored as zero. The results of ANOVA were as follows: (1) There was a significant interaction of age group by school, $F(2, 63) = 3.79$, $p < .028$, $\eta_p^2 = .108$. As Figure 5 shows, the age trend was linear for the lower SES community, but not for the upper SES community. It should be noted that the fact that these differences in the relationship between age and judgment differ for the two communities is *not* due to ceiling effects since *not* all of the adolescents in the higher SES community choose false alarms as worse or were more concerned with protecting the innocent. The unexpected finding for the adolescents in the upper-middle class sample, is discussed later.

Discussion

The second study has confirmed the two findings relating (a) decisional bias and age and (b) beliefs about false alarms or misses and age, in two samples from very different communities, thus helping to establish external validity. The exception to the aforementioned

was the upper-middle class adolescents' expressed beliefs. Comments made by some of these adolescents help to elucidate their beliefs. The teenagers from upper-SES community simply did not consider taking a gym bag as 'such a big deal'. This led us to use a film depicting a more serious 'crime' for Study 3.

Our hunch is that the community differences may be explained by different experiences with crime, but this remains purely conjectural. Although intriguing and of considerable interest for policy, we decided not to focus on these demographic differences, which we leave for another day, and rather continue studying the basic developmental phenomena. There was a complete absence of any suggestibility effect. Perhaps, the child's moral orientation towards punishment is too robust and strongly entrenched to be influenced by such transitory interventions.

Study 3

Although we were encouraged by (a) the feasibility of studying young children's eyewitness identification using a signal detection method and (b) by the apparent robustness of the developmental course of performance bias, we were not as satisfied with the framing in Study 1 because the *pro-social* condition did not involve a transgression at all and the reason for searching for the pro-social helper may not have been as motivating as was the search for the moral transgressor (thief). That is, the participants may have questioned why it was necessary to identify someone who had performed a morally good act. Thus, we needed two moral rationales that were equally plausible, but had different moral significance, which depended on the age or developmental status of the child.

Therefore, for the third study we decided to contrast two kinds of *moral* rationales against one another, which according to past research as well as the concept of *mens rea*, have different

significance as to moral responsibility (see succeeding discussions). Like in Study 1, the rationales were established by framing the *visually same* event as having two different moral meanings by means of verbal descriptions. Further, these two kinds of framings or meanings were informed by what is known about how children judge moral encounters. Since Piaget's (1932/1965) groundbreaking research on children's moral judgments, it has been known that there is a gradual shift from using material outcome as a criterion for judging culpability to judging according to the actor's intentions. Although there have been numerous challenges to the phenomenon on methodological and empirical grounds, (e.g., Karniol, 1978; Keasey, 1978) and especially to Piaget's explanation for it, the demonstrations of the shift to what in the law is termed, the *mens rea*, is reported to be a robust one *when intentions and outcome are confounded* (e.g., Keasey, 1978). Thus, the general conclusion we draw from the moral judgment literature is not that children are unaware of the intentions behind an act, but rather that they ignore or give less weight to intentions when the outcome is bad. Thus, although young children may literally be aware of the bad intentions, they don't consider them relevant to a judgment of culpability. Therefore, we framed or narrated the same filmed event (described later) to be either innocently intended with disastrous outcome, a devastating fire (thus featuring the bad outcome) or malevolently intended with minor outcome (featuring the bad intentions). It should be noted that the confounding of the kind of intention with the kind and level of outcome was by design because it is primarily when intentions and outcome are confounded that developmental differences emerge. This fits with Piaget's claim that young children are aware of the intentions behind an act, but simply consider them largely irrelevant for judging moral culpability and focus (in his terms, center) on the outcome instead.

Method

Participants

One hundred participants recruited from a public elementary and junior high school, in New York City, participated in a study of eyewitness identification, with the usual parent consent. After removing 14 subjects' data due to incompleteness, the final sample size was 86 participants, consisting of twenty-eight 7- to 9-year-olds (14 boys, 14 girls, $M = 7.9$, $SD = 8.1$), thirty 10- to 12-year-olds (16 boys, 14 girls, $M = 10.9$, $SD = 9.7$), and twenty-eight 13- to 15-year-olds (10 boys, 18 girls, $M = 13.4$, $SD = 3.6$).

Stimulus Materials

The same filmed act of setting a fire (described later) was framed in two different ways: (a) as involving "neutral" intentions, which resulted in a fire that ruined the restaurant (bad outcome) and (b) as a "bad" intentional act of trying to set a fire that did not result in any damage to the restaurant (bad intention) by different voice-overs. One stressed the neutral intentions but bad outcome and the other the bad intentions but neutral outcome.

The film was professionally shot in a restaurant in lower Manhattan. All the actors were male, Caucasian, 5ft 11 in. to 6ft 1 in. in height, ages 22-24, dressed in jeans and crew neck sweaters. 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. The two 'framed' versions employed in the study are presented below.

Both versions begin with: "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.” Then the text proceeds differently in the two conditions:

A. Voice over for Film Version 1 (*Neutral Intentions with Bad Outcome*): "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 on fire and the restaurant gets ruined!" (Show smoke in the garbage can. Fade to black. Sirens heard.)

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

Procedure

To check on moral judgments of the event, after viewing the film, the participants were first asked three questions in counter-balanced order; "How bad was the man in the film?", "...what the man did in the film?", and "... what happened to the restaurant?" Following this, they were given the eyewitness identification task. The task was to choose the perpetrator, who either intentionally or unintentionally started a fire, either one that fizzled out or one that created major damage. Each participant received the following instructions prior to viewing the lineup. "Now we would like you to find the man who started the fire [or tried to start a fire]. He may or may not be in the photos." Each participant was then sequentially shown a series of photos (by computer), which included (a) the perpetrator, (b) each of the three other men at the restaurant, (c) the waiter (bystander), and (d) a man, who was not in the film at all.

As in the previous studies, a signal detection procedure was used by presenting the six individuals each presented frontally and in right and left profiles. The series again consisted of

three exposures of the 18 photos (six individuals by three angles), yielding 54 judgments in all, randomly ordered as in the previous studies. All photos were head and shoulder shots with left/right and frontal profiles. Lineup targets were dressed in dark blue t-shirts and were filmed against the same background. All participants were familiarized with the procedure so that they were proficient in its use before the actual judgments began. As in the previous studies this required a short training procedure for the youngest participants in which they were asked about something that they were certain was true and something they were certain was untrue and shown how to represent each option. At the end of photo-identifications, each participant was asked how bad it would be to say someone did it when he did not (a false alarm) and how bad would it be to say someone did not do it when he did (a miss), and why it would be bad?

Data Analysis

As in the previous studies and for reasons already noted, the C3 measure of *bias* was the focus of our analysis. The method of calculating beliefs about “which is worse” is the same as reported earlier, in describing study 2. Reliability of coding based on a sample of 38 was 0.95.

Results

Analysis of the sensitivity (d') measure showed (Figure 6) a significant age effect, $F(2, 80) = 3.32, p < .04, \eta_p^2 = .08$, but no trace of a condition or interaction effect, $ps > .20$. The measure of bias (C3), the relative balance of false alarms versus misses (shown in Figure 7), again shows a significant condition effect, $F(1, 80) = 8.25, p < .005, \eta_p^2 = .09$, with the youngest (7- to 9-year-olds) age group committing more false alarm errors than the older children (10- to 12- year-olds) and adolescents. However, this main effect is modified by an interaction age and condition $F(2, 80) = 4.42, p < .015, \eta_p^2 = .10$. Post-hoc tests revealed significant differences between conditions only for the 10- to 12- year olds ($p < .05$).

Beliefs and Reasons Concerning False alarms and Misses

The hypothesis that children's beliefs about the seriousness of false alarms and misses change with age was first tested by generating a *Which Is Worse Reason Composite Score* for each participant, on the basis of the reasons given to justify "Why is it bad to say the man did it when he didn't" or "...didn't do it when he did?" as already described earlier. ANOVA shows that the effect of age was significant for the *Which is Worse Reason Composite*, $F(2, 80) = 6.30$, $p = .003$. $\eta_p^2 = .136$, but not for condition, $F(1, 80) = 2.23$, $p > .10$. $\eta_p^2 = .027$. 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. Thus, the *Which is Worse Reason Composite* measure increased with age regardless of condition, indicating an increasing concern for the innocent accused, as expected (See Figure 8).

Relationship between Performance Bias and Beliefs about Which is Worse

The third question posed was whether the measure of performance bias was related to their stated beliefs about the seriousness of false alarms/misses. Figure 9 shows that those who declared misses to be worse had marginally significantly lower bias scores, indicating a higher rate of false alarms, than those with higher bias scores, reflecting fewer false alarms but only in the no-fire condition. ANOVA revealed a marginally significant relationship within the latter condition, $F(1, 51) = 2.98$, $p < .09$, $\eta_p^2 = .06$.

Moral Judgments

To examine explicit moral judgment beliefs about the two events (fire and no fire), a Moral Judgment Composite Score was generated for each participant. 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 first two items were

correlated ($r = .35$), as expected, and the third was not, thus indicating that the participants, in general, 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?). Surprisingly, the expected interaction effect between age and condition (Neutral Intent Bad Outcome Fire /Bad Intent Neutral Outcome No Fire) was not obtained, $F(2, 80) = .57, p > .50, \eta_p^2 = .01$. However, condition did have a significant effect, $F(1, 80) = 8.74, p < .004, \eta_p^2 = .10$. All age groups judged the no fire (bad intention) act as *worse* than the fire (bad outcome) act. Also, there was a non-significant trend for moral judgments to get less severe with age regardless of the type of transgression (film). The discrepancy between these moral judgment scores and the decisional criteria (bias scores) reflecting the balance of false alarms and misses is discussed later.

Discussion

The major findings of this study are (1) younger (7-9 years old) children appear to use a general strategy, 'when in doubt, positively identify' regardless of the framing of the act whereas the next oldest group (10-12 years old) used a laxer criterion, indicated by more frequent false alarms, when identifying the perpetrator in the bad outcome (fire) condition than in the bad intention (no fire) condition. The adolescents (13-15 years old) did not significantly respond differently to the two framing conditions. (2) When choosing which was worse, false alarms or misses, (a) many of the youngest children reasoned about false alarms and misses in non-moral terms (i.e., without regard to the consequences of their identification); (b) the next oldest group reasoned in moral terms, but thought misses were worse, whereas (c) the adolescents thought false alarms *were* worse mostly for moral reasons. These results did not appear to depend on condition (fire/no fire). (3) When asked explicitly about how bad the act and actor were, all age groups thought the bad intention act and actor were worse.

It was surprising that the explicit moral judgment questions (the moral judgment composite index) did not reveal the expected interaction between age and condition, with younger children judging the bad outcome act as worse and the adolescents judging the bad intention act worse. This absence of support for the assumed developmental shift from an outcome to an intention orientation may be because of the vivid mode of presentation (film rather than the standard verbal format). In any case, generally younger and older children and adolescents tended to judge the bad intended act as worse despite the severity of the outcome! However, eyewitness identification judgments, in particular, the false alarms rate, as indicated by the bias index, *did* show a developmental difference. Perhaps, the false alarms acted as a sort of implicit measure reflecting the influence of outcome on the 10- to 12- year- olds, who differentiated between the two conditions (framings of the act) although the explicit (moral judgment composite) showed the opposite condition difference (i.e., bad intention was worse than bad outcome). One might ask why this was not also true for the youngest group of participants (7- to 9- year- olds). We believe that the key to this puzzle lies in our finding that the youngest children do not approach the identification task as having moral consequences and, therefore, any moral significance. Thus, these youngest children appear to be insensitive to the moral framing of the act.

General Discussion

We began by posing the question as to why young children are prone specifically to make more false alarms and our results suggest a partial answer. The answer lies in the following findings: First, we have confirmed in all three studies, that indeed the decisional criterion used by younger children is laxer, allowing for more false alarms, over and above any differences in overall accuracy, in comparison to older children and especially with adolescents. Further, as

evidenced in Studies 1 and 3 this decisional difference is constrained by how the event is framed. Rather than an overall, perhaps impulsive, tendency to guess or affirm, it appears to be limited to events that involve moral transgressions (Study 1) and further depends on the nature of the moral transgression, for example, whether malevolently intended with little damage or unintended but with major damage, although this interacts with the age of the participant (Study 3), as described and discussed earlier. One may ask ‘What about the age of the child makes him/her more cognizant of the moral implications of their eyewitness decision?’. A possible critical fact may be the child’s increasing ability and propensity to consider the implications of their acts from the others’ point of view, usually termed, perspective taking (e.g., Piaget, 1929; Selman, 1971)

Third, children’s stated beliefs about the seriousness of false alarms and misses, as distinguished from their actual performance, systematically change with age. Namely, many of the younger children either do not appear to treat these errors as having any moral significance or, if they do, they believe misses are worse; older children do generally recognize the moral significance of the two kinds of errors, but generally believe misses are worse. It is only in adolescence that the majority of participants think false alarms are worse *for moral reasons*.

Fourth, we have obtained some evidence of a relationship between (a) the shift in actual *performance* from committing more false alarms than misses and (b) the shift in *beliefs*. However, the nature of the relationship is somewhat different from that initially expected: Those who do not view the task as moral are more prone to employ a laxer decisional criterion than those who view the task as a moral one, but it does not appear to make as much of a difference whether the latter group think false alarms or misses are worse! It should be noted that the number of participants in the non-moral beliefs group was relatively small (only 11 in the no fire condition). Clearly, this finding needs further exploration.

Finally, despite the fact that participants at all ages in the third study explicitly recognize the greater culpability of an act based on malevolence despite minor outcome, compared to the unintended act resulting in major damage in *their explicit moral judgments* (the *composite moral judgment measure*), the middle age group (10-to 12- year- olds) use a laxer criterion, that is, committing more false alarms, when the act involves unintended but major damage.

In a sense, one might view the identification task as a type of implicit moral judgment recently documented by other researchers (e.g., Sinnott-Armstrong, 2002) at least for the 10- to 12- year- olds. For this age group, despite the fact that they *explicitly* recognize that a bad intended act is to be condemned despite the minor outcome, an unintended act with major outcome (e.g., a destructive fire) elicits a laxer decisional criterion when actually identifying the culprit. Perhaps, in the emotionally-arousing fire condition, the identification task may operate as an implicit moral judgment or, to put it more starkly, *damage was done; someone must pay*. Thus, those who are accused of accidentally causing great damage may be at greater risk of being falsely identified by children, as reflected in looser decisional criteria, that is, more false alarms, regardless of the explicitly recognized good intentions, philosophical and legal principles to the contrary. These findings may have important implications for actual legal practice, and demonstrate—we believe—the usefulness of bringing a moral-decisional lens to bear on the phenomena.

References

- Brewer, N., Weber, N., & Semmler, C. (2005) "Eyewitness Identification" pp. 177-221 in
Brewer, N. and Williams, K. D. (Eds.) *Psychology and Law: An empirical perspective*.
New York: Guilford Press.
- 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., Bruck, M. (1995). Jeopardy in the courtroom: A scientific analysis of children's
testimony. Washington, DC: American Psychological Association.
- Cultice, J. C., Somerville, S. C., and Wellman, H. M. (1983). Preschoolers' Memory.
Monitoring: Feeling-of-Knowing Judgments. *Child Development*, 54, 1480-1486.
- Cutler, B. L. & Penrod, S. D. (1995). Mistaken Identification: The eyewitness, psychology, and
the law. New York: Cambridge University Press.
- Gee, S., & Pipe, M-E. (1995). Helping children to remember: The influence of object cues on
children's accounts of a real event. *Developmental Psychology*, 31, 746- 758.
- Gescheider, G. A., *Psychophysics: Methods, Theory, and Application*, 2d ed., Lawrence Erlbaum
and Associates, Hillsdale, N. J., 1985, pp 174 – 191.
- Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York:
John Wiley.
- Karniol, R. (1978) Children's use of intention cues in evaluating behavior. *Psychological
Bulletin*, 85, 76-86.
- Keasey, C. B. (1978). Children's developing awareness and usage of intentionality and motive.
In C. B. Keasey (Ed.), *Nebraska Symposium on Motivation* (Vol. 25, pp. 19-260).
Lincoln: University of Nebraska Press.

- Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology*, 97(4), 286-314.
- Kohlberg, L. (1969) Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of Socialization Theory and Research*. Chicago: Rand McNally.
- Locksley, A., Stangor, C., Hepburn, C., Grosovsky, E., & Hochstrasser, M. (1984). The ambiguity of recognition memory tests of schema theories. *Cognitive Psychology*, 16, 421-448.
- Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, 106, 398-405.
- Munsterberg, H. (1908) *On the witness stand: Essays on psychology and crime*. New York: Clark, Boardman, Doubleday.
- Pezdek, K., & Banks, W.P. (Eds.) (1996). *The Recovered Memory/False Memory Debate*. San Diego: Academic Press.
- Pezdek, K., & Roe, C. (1994). *Memory for childhood events: How suggestible is it?* *Consciousness and Cognition*, 3, 374-387.
- Piaget, J. (1932/1965). *The moral judgment of the child*. New York: New York Free Press.
- Poole, D.A., & Lamb, M.E. (1998/2003). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association.
- 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.

- 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.
- Selman, R. (1971). Taking another's perspective: Role-taking development in early childhood. *Child Development*, 42, 1721-1734.
- Sinnott-Armstrong, W. (2002) Moral relativity and intuitionism. *Philosophical Issues*, 12, 305-328.
- Sporer, S.L., Penrod, S.D., Read, J. D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychology Bulletin*, 118, 315-327.
- Turiel, E. (2002) *The Culture of Morality: social development, context and conflict* (Cambridge, England, Cambridge University Press.
- Wells, G. L. & Loftus, E. F. (2003). Eyewitness memory for people and events. In A. Goldstein (Ed.), *Comprehensive handbook of psychology*, Volume 11, *Forensic psychology*. New York: John Wiley and Sons.
- What did eyewitnesses really see? (2011, August 26). *The New York Times*, p. A18.
- Yuille, J. C., & Wells, G. I. (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, D.C.: American Psychological Association.

Appendix: 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'.*

Author Note

This paper is based, in part, on pre-dissertation research (study 2) and dissertation research (study 3) by the first author and supervised by the second author. The first and second authors co-wrote the paper, with assistance from the third author. Spring is an Adjunct Assistant Professor at Queens College of CUNY; Saltzstein is Professor of Developmental and Social Psychology at the CUNY Graduate School; and Peach is a Research Associate at Lehman College of CUNY.

Acknowledgements

The authors also wish to thank Carla Martinez, Vienna Messina, David Barnard, and Yoko Takagi for aid in collecting data, coding responses, and analyzing data. A special thanks to Teri Cao for the professional filming. Thanks, are also extended to the school administrators, teachers and students of the participating schools. Some of this research was supported by a grant from PSC-CUNY to the second author. We also wish to thank Brian Cox and Jay Verkuilen for expert advice on signal detection methodology, and the two anonymous reviewers for helping to clarify this research report.

Footnotes

It is customary in the Signal Detection literature to use the terms: *false alarms* and *misses* instead of false positives and false negatives. Of course, none of these terms were used with the children.

² We use the term eyewitness identification rather than facial identification even though this research is not conducted in a forensic setting because (a) the whole person was visible in the task and (b) more importantly, because the purpose was to identify a wrongdoer not just any person.

Figure Captions

Figure 1. Relationship between sensitivity (d') and age by condition (Study 1).

Figure 2. Relationship between bias (C3) and age by condition (Study1).

Figure 3. Relationship between sensitivity (d') and age by school (Study 2).

Figure 4. Relationship between bias (C3) and age by school.

Figure 5. Relationship between total moral reasons and age by school (Study 2).

Figure 6. Relationship between sensitivity (d') and age by condition (Study 3).

Figure 7. Relationship between bias (C3) and age by condition. (*Note.* Bad outcome [Fire]. Bad intent [No fire].)

Figure 8. Relationship between 'Which is worse reason' composite scores and age group by condition. (*Note.* Conditions: Bad intent neutral outcome [No fire] and Neutral intent bad outcome [Fire].)

Figure 9. Relationship between bias (C3) and moral vs. non-moral reasons.