2020

Gender Differences in Moral Influences on Adolescents' Eyewitness Identification

Toni Spring
CUNY Queens College

Herbert D. Saltzstein
CUNY Graduate Center

Leeann Siegel
University of Pennsylvania

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

Gender Differences in Moral Influences on Adolescents' Eyewitness

Identification

KEYWORDS

Eyewitness identification;

gender; adolescent

development; decision-making;

moral development

Toni Spring, Herbert D. Saltzstein, and Leeann Siegel

Gender Differences in Moral Influences on Adolescents' Eyewitness

Identification

In this study, 232 (89 11- to-12-year-olds, 71 13- to-14-year-olds; 72 15- to-16-year-olds) students recruited from grades 6th–11th in an urban public high school participated in a study of eyewitness identification. The focus of this study was on the effects of age, gender and moral orientation on decisional bias and, as a secondary outcome, on accuracy (using signal detection analysis). The primary purpose of this and previous studies in this series is to uncover implicit moral decision-making in decisional bias. In this study the perpetrator, the bystanders and the foil were all females. Prior to completing the eyewitness identification task, participants were given instructions that emphasized either (a) fairness and crime prevention, or (b) neither. These instructions had no discernible effect on accuracy but, as in past studies, younger participants (below the age of 13) had lower decisional criteria, resulting in a higher rate of false alarms/positives. Further, those who judged the transgression as worse had a lower decisional criterion, indicating more false alarms. Females were more accurate than the males in identifying the female perpetrator and scored significantly higher on how bad they would feel if they were the victim than did the males.

There is a long history of psychological studies focused on children's and adolescents' face recognition and eyewitness identification (e.g., Brewer, Weber, & Semmler, 2005; Cultice, Somerville, & Wellman, 1983; Cutler & Penrod, 1995; Malpass & Lindsay, 1999; Meissner, Tredoux, Parker, & MacLin, 2005; Pozzulo, 2017; Shapiro & Penrod, 1986; Wixted & Mickles, 2014). These studies include findings of age differences. For example, young children have been shown not only to be less accurate when making forced choices but also more susceptible to interrogators' influence even when this influence is exercised subtly (e.g., Ceci & Bruck, 1993; Keast, Brewer & Wells, 2007; Muensterberg, 1908). In general, the relationship between actual eyewitness performance and confidence appears to be inconsistent (e.g., Keast, Brewer & Wells, 2007). Further, attempts have been made to improve children's eyewitness identification (i.e., using structured interview protocols and unbiased questioning) with limited success (e.g., Lamb et al., 2007, Roebers, Moga, & Schneider, 2001; Sporer et al., 1995).

There is also substantial evidence of gender differences in face recognition, much of it showing that females perform better than males on facial recognition tasks (e.g., Herlitz & Loven, 2013; Hills & Lewis, 2010; Loven, Herlitz, & Rehnman, 2011; Mason, 1986; Rehnman & Herlitz, 2006; Meissner et al., 2005). This suggests that there might also be a gender difference in eyewitness identification when the 'perpetrator' and 'foils' are females. Thus, testing some of our past findings (below) featuring a female perpetrator and foils is important for fully understanding eyewitness identification phenomena.

However, we should note that we bring a different perspective to these phenomena than has heretofore been the case. Virtually all of the studies cited above

examined eyewitness identification/facial recognition primarily as an exercise in

perceptual judgment, memory and general decision-making. However, a moment's

reflection makes clear that eyewitness identification involves not just (visual) perception,

memory and (general) decision-making, but also a kind of implicit moral decision-

making or moral judgment. Our reasoning is as follows: a false positive/alarm threatens

to send an innocent person to jail whereas a false negative/miss risks letting a guilty

person go free. Eyewitness identification therefore clearly involves, *inter alia,* a kind of

moral decision at an implicit level of functioning. Thus, our focus is not just to discover

predictors of the accuracy of eyewitness identification, as do most studies in this research

area, but rather to use the task as a type of measure of implicit moral judgment/decision-

making.

Yet, how do we to study eyewitness identification as moral decision-making?

Two complementary strategies come to mind: First, while the visual input remains

identical, the moral meaning may be varied by means of a voice-over that provides

variations in the meaning of the event (e.g., whether intended or not or creating small or

large damage). Second, one may examine the eyewitness's performance not just in terms

of general accuracy, but also in terms of the propensity to make false positive

identifications/false alarms relative to false negatives/misses. We use the eyewitness

identification task and framework as a vehicle for studying implicit moral decision-

making.

This is where signal detection theory (SDT) methodology becomes essential. The

SDT (Macmillan & Creelman, 2005; Swets, 1986) approach to measurement enables one

to differentiate between (a) overall accuracy (how well a person can discriminate between

'Signal Present' and 'Signal Absent'/ sensitivity (d') and (b) decisional bias, which is of

primary interest in our studies and which represents the minimum level of (internal)

certainty needed to decide that a signal was present (termed C). High bias scores reflect a

more stringent or conservative decisional bias, indicating fewer false alarm judgments

relative to misses, corrected for chance and lower bias scores reflect a looser or less

conservative decisional bias, indicating more false alarms relative to misses.

The specific measure of decisional criterion used (here, C2 and C3) indicates

where the decisional criterion is set: for C2 (between 'a little sure that it isn't the person',

'a little sure that it is the person'*)* and for C3, 'very sure it is the person' and the other

three alternatives). Both measures of C indicate whether the eyewitness has a propensity

to make one type of error over another, (e.g. false alarms over misses) independent of

accuracy, but use different criteria for a positive and negative judgment.

SDT methodology has a long history in the study of perceptual judgments (e.g.

Macmillan & Creelman, 2005; Meissner, et al., 2005; Swets et al., 1986) and in studies of

memory (e.g., Van Zandt, 2000; Vickers, 1979). In this study and in others in this

program of research, we have used this SDT methodology, along with (a) variations in

the framing of the visually identical event to have different moral meaning and (b)

measurements of not just overall accuracy (sensitivity) but also the propensity to make

false alarms (decisional bias) as tools to study implicit moral decision-making. It is the

two measures of C (decisional criteria) which is of special interest in our studies

although, of course, we also look at sensitivity (d').

In our past research (Spring, Saltzstein & Peach, 2012; Spring & Saltzstein, &

Vidal, 2015; Spring & Saltzstein, 2017) , involving six separate studies, we have found

that the way in which the event (film) is described or framed (e.g., the act was intended but caused little material damage vs. unintended but caused major material damage) interacts with the age of the eyewitness to affect the rate of false positives, so that younger children (10-12 years of age) had lower decisional bias scores (reflected in more false alarms) than did older children (13-15 years of age) when major damage was done even if it was unintended. Whereas the reverse was true when the act was intended but led to little material damage. This relates to Piaget's well-known finding in the *Moral Judgment of the Child* (1932) that young children (below the age of adolescence) often ignore the actor's intentions and instead judge the morality of an act in terms of the material outcomes of the act. We have found (Spring & Saltzstein, 2017) that this focus on material outcomes rather than intentions by younger children (roughly 10- to 12-year olds) is revealed in eyewitness identification in the form of false alarms, i.e., more false alarms when the outcome is bad and fewer false alarms when the outcome is minor even though the intentions may be malevolent.

Consequently, as a result of these previously published studies, we have come to think of these eyewitness judgments as a form of implicit moral decision, which may deviate from explicit moral decision in terms of its relationship to the age of the eyewitness and the features of the act (Spring & Saltzstein, 2017). We should note that the above applies to children between the ages of 10-15, and not to 7- to 9-year-old children whose eyewitness identification was *not* affected by the 'framing' of the filmed event. Initially, we were puzzled by this aberration from our general findings; however, systematic post-task interviewing solved the puzzle. Consistently, in the post-identification interviews, when the participants were asked in child friendly terms, why

false alarms/misses are bad to make as a witness, the 7- to 9 year olds indicated that they did not think of false alarms and misses as having any moral significance, a finding in and of itself whereas, the older children, ages 10 and older, did make their decisions in a way that frequently comports with moral considerations, e.g., intentions. For this reason, we no longer include 7- to 9-year olds as participants in our studies. These findings help confirm our underlying assumption about eyewitness identification, namely that eyewitness identification for older children (those above 9 years of age), adolescents and adults is a type of implicit moral decision-making.

Nevertheless, there has been an important, if unintended, limitation in this program of research: although all the studies included male and female child and adolescent participants/eyewitnesses, those filmed (perpetrator and foils) were all males. This error of omission was unintentional and unplanned. Perhaps, the omission was due to the general observation (e.g., Denno, 1994) that males commit significantly more crimes than females. However, these findings of lower incidence of crime by females primarily pertain to crimes of violence, which for the most part do not figure into our research. Nonetheless, a general gender difference of crime rate should not blind us to the goals of understanding the 'psychologies' of both genders. Therefore, we decided to conduct a comparable study featuring young women as the 'perpetrator'/target and 'foils'/bystanders to see whether our findings would be replicated.

Our major premise, which we believe that we have now established as valid and relevant, is that the eye-witnessing of suspects in crimes involve, *inter alia*, making an implicit moral decision. Therefore, when examined as a type of implicit moral judgment, there are two relevant strands of reasoning about morality with a long history in

philosophy, namely deontological (e.g., Kant, 1785/1964; Rawls, 1971) and consequentialist (e.g., Mills, 1861/1998). Put succinctly, these might be represented as (a) 'being fair or just' and (b) 'acting so as to maximize good and minimize bad outcomes' respectively. In this context, they might be represented as (1) 'you wouldn't want to convict an innocent person' and (2) 'don't let the perpetrator get away with it'. We believe that these two moral orientations may influence the propensity to make one or the other type of error, namely false alarms or misses differently for different age groups and genders.

Further, Gilligan (1977) has proposed and, on occasion, demonstrated that while there is much overlap, males favor a deontological or justice orientation towards moral situations, whereas females favor a consequentialist orientation where the goal is to do or produce 'good'. This claim has found support in some of her later studies (e.g., Gilligan, 1988) but has been challenged in others (e.g., Bucciarelli, 2015; Held, 1995: Jaffee & Hyde, 2000; Keller 1995). A related finding has also been reported by Hoffman (e.g., 1987) who in a number of studies has found that females, young girls and women have an earlier and stronger capacity for empathy, on average, than males have. While this is surely a complex claim and one that remains in dispute (e.g., Klein & Hodges, 2001), it deserves attention from researchers because of its potential relevance to eyewitness identification. This rationale may then serve as the theoretical and empirical context from which our study arises.

First, we wished to test whether age differences in decisional bias, which were found in earlier research involving all male actors, would also occur when all female actors were those to be judged. Second, we wished to see whether prompting a focus on

fairness and prevention 'being fair or just' and 'acting so as to maximize good and minimize bad outcomes' would have a differential effect on eyewitness performance, and third, whether any effects would be different by the age and/or, most especially, the gender of participants. Thus, one approach (set of instructions) focuses on finding and punishing the culprit whereas the other focuses on not convicting an innocent person. These two goals permeate eyewitness identification and make it a difficult task. Rather than contrast these two, we highlighted them in one (experimental) condition, but omitted them from the other (control) condition. Our expectation was that younger children would focus on the 'identifying and punishing the culprit' orientation whereas the older children would focus on a 'not identifying and punishing' orientation. For example, we hypothesized that younger children (11- to 12-year-olds) would commit more false alarms in the experimental condition whereas the older children and adolescents (13- to 14-year-olds and 15 – to 16-year-olds) would commit fewer false alarms. In other words, the younger children would be less cautious and the older children more cautious. Furthermore, females would not only be more accurate in their judgments but make fewer false positives or false alarms than males in an effort to avoid being unfair.

Method

Participants

The participants were 232 (89 11- to 12-year-olds, 71 13- to 14-year-olds, 72 15- to 16-year-olds) students recruited from grades 6th– 11th in a public high school in New York City, with the usual parental permission and student assent. There were no

consequences for nonparticipation. Students participated in small groups in their classroom. One hundred and twenty-one participants were randomly assigned to the experimental condition and 122 in the control condition.

The experimenters were the first author, a Ph.D. psychologist, and a high school student research assistant in his senior year at another NYC public high school.

Design and Stimulus Materials

The film: Simulating a typical high school event, four female students, with medium length brown hair pulled back with a rubber-band, are working on a group project at a table in a classroom. They are engaged in a conversation while working at the table for 30 seconds; while the film pans the faces of all four women. A fifth female student enters the classroom, approaches the group, and offers a distraction (sharing Girl Scout Cookies). Through the editing of the film, the profiles (left, center, right) of each face lasted 23 seconds. None of the students in the film had discerning physical features such as jewelry, tattoos, or head ornaments. One of the original four students takes the opportunity to steal an iPhone from one of the other students and puts it in her pocketbook while the other females are in conversation and preoccupied. All girls then proceed to sit down at the table. One girl visibly searches for her phone and appears to be quite upset when the she realizes it is gone. The video then fades to a black PPT slide. There is no audible sound of the actors. Immediately after the film viewing, in the experimental condition, on a black PPT slide, a voice-over describes the events and informs the participants (eyewitnesses) that the "stealing" of the iPhone will result in an expulsion from school, and then the voice-over emphasizes the need to preserve fairness by not convicting an innocent person, and the need for consequences, thereby,

discouraging the transgression from occurring again. In the control condition, on a black PPT slide a voice-over describes the events and informs the participants (eyewitnesses) that this theft will result in expulsion from school without any emphasis on preserving fairness or any of the negative consequences of choosing incorrectly. The film lasted approximately 138 seconds. In order to accommodate the school schedule, clusters of participants were randomly assigned to one of the conditions.

Following the film, the participants were asked to write down everything that they could recall and to complete 4 questions in the questionnaire. In the first two questions, participants explicitly judged "how bad" for each of the following questions: "how bad was what the student did, stealing another student's iPhone?"; "how bad would you feel if another student stole your iPhone?" The participants were asked to choose on a Likert scale with the following response options: not bad, a little bad, pretty bad, and very, very bad. The third question asked "what type of punishment would you suggest for the student who stole the iPhone?" Response options included: a) no punishment (but noted on the student's academic file for the stealing); b) in school detention (1-2 days); c) out of school detention (3-days); and d) expulsion from school. For the fourth question, participants were asked, "Do you recall the instructions that were given to you at the end of watching the film?" This question served as a manipulation check. After it was clear that the participants understood the film and completed their responses, they began the eyewitness facial recognition/identification task.

No opportunity was given for the participants to discuss the events among themselves. Judging from the participants' written recital of the events immediately after viewing the film, we found that the events in the film were very comprehendible even for

the youngest participants. The participants were asked to identify who of the six

individuals, presented seriatim, was the perpetrator and how certain they were that the

individual was the person who stole the iPhone using a four-point scale, from 'very sure

it was not the girl' to 'a little sure it was not the girl' to 'a little sure it was the girl' to

'very sure it was the girl'. There were a total of 54 successive judgments, which resulted

from six individuals (the four female students at the craft table, the female student who

brought cookies into the classroom, and a female student with similar facial features who

was not in the film at all) who were each presented from three perspectives (frontal, right

and left profiles) for a total of 18 photos. These 18 photos were presented in three

identical sequences for a total of 54 presentations. The sequences were random except

that the same person never appeared consecutively. In this study, the sequential lineup

presentation mode was used rather than a simultaneous lineup because sequential lineups

have sometimes been found to decrease lineup bias (Carlson & Carlson, 2014; Carlson,

Gronlund, & Clark, 2008; Gronlund, Carlson, Dailey, & Goodsell, 2009; Lindsay et al.,

1991; Pozzulo, 2017) although this claim has recently been called in question (Mickes,

Flowe & Wixted, 2012). Following the facial recognition presentation task, participants

answered three more questions in counter-balanced order regarding the seriousness

("badness") of the types of mistakes witnesses can make. These questions asked

participants to rate on a three-point scale how bad they thought it would be to do each of

the following things during an eyewitness identification task: "Saying it is the person

who stole the phone, when it really wasn't" and "Saying it was not the person who stole

the phone, when it really was." They were also asked why they selected their responses to

these questions in order to help assess their moral reasoning. The final question was to

check the extent to which participants recalled the instructions they were given prior to the eyewitness identification task, "Could you describe what you were thinking about when deciding that it was or was not the student who took the iPhone?"

Procedure

Prior to viewing the video participants were told that they are going to watch a video that was taken from hidden cameras in a high school and instructed that their job was to act as a witness and to watch the film closely. In the experimental condition (Fairness and Consequences) both were included in a counter-balanced order: "Please look at the individual profile of faces. We are interested to see if you can identify the thief. She may or not be in these photos. It is important that you catch the right person because you wouldn't want someone who didn't steal anything to be expelled; it wouldn't be fair. There's been a lot of theft at school and we want to stop it. So, it's important we catch the person who is doing it. We must make an example that there are consequences like being expelled for stealing." In the control condition only the stem was included, "Please look at the individual profile of faces. We are interested to see if you can identify the thief. She may or not be in these photos."

The 18 color profile photo slides were presented on a PPT format timed for 7 seconds for each profile of the students in the film and the one student who was not in the film at all. Judgments were made for each presentation of an individual profile. Participants circled one choice: that they were either "very sure it isn't", "a little sure it isn't", "a little sure it is" or "very sure it is". Groups of participants were arranged according to their availability at the school. Assignment to the Experimental and Control conditions was random for groups of participants. There were 111 participants in the

Control condition (54 males and 57 females) and 121 participants in the Experimental

condition (57 males and 64 females).

Analyses

The data, both the signal detection measures (decisional bias) and traditional

moral judgment measures, were analyzed using multivariate analysis of variance

(MANOVA) and structural equation models (SEM). Analyses were carried out using

SPSS version 25 and Stata version 15. The key independent variable of interest was

assignment to experimental condition; the experimental condition was exposed to

reminders of the need for fairness and consequences, while those in the control condition

received no such reminders. We were also interested in the effects of age and gender and

how condition might interact with age and gender.

Outcome variables of interest:

Using signal detection theory analysis, we assessed participants' tendencies to

make false alarm judgments over misses using decisional bias criterion C3, i.e.,

contrasting the alternative "very sure it is" the thief with the other three alternatives,

independent of sensitivity (d'). As noted, this criterion was chosen because past research

(Pozzulo, 2017) has found that young children are particularly prone to very confidently

identify the perpetrator and because in past studies we have consistently found clearest

results when using this criterion. Our other focus was on decisional bias C2: the split

between "a little sure it is" and "a little sure it isn't." In addition, the participants were

asked to rate how serious the act, actor and outcome for the victim was for the 'crime'

and by indicating what the punishment should be for stealing a student's phone.

Other variables we looked at included sensitivity (d'), the ability to discriminate between the target and the target + noise (Heeger, 2009). Higher d' scores indicate that participants are more capable of discriminating between the perpetrator (target) and the foils/others in the film and one who was not (target + noise). We also looked at an explicit moral judgment measure, Focus More On, which reflected whether participants self-reported focusing more on fairness or (crime) prevention. Additionally, we looked at participants' self-reports of 'how bad they would feel' if their iPhones were stolen (Ufeelbad), recommendations for punishment, opinions regarding the severity of the threat and judgments regarding which type of error it would be worse to make: a false alarm or miss. Each of these latter measures is described in greater detail below.

In addition to our main interest in children's eyewitness testimony itself, we were also interested to see whether the pattern of findings for the traditional, explicit measures and implicit measures (especially our measures of decisional bias, described above) showed similar patterns (effects of age, condition, and interactions between the two) or rather diverged as we have found in our past research, cited above.

Results

Implicit Measures/Signal Detection Analysis.

Summary statistics for our two measures of decisional bias, C3 and C2, as well as d' (d-prime) for participants in each experimental condition, gender and age group are included in Table 1 and Figure 1.

Decisional bias: As in the prior studies described above, C3 was related to age, (F(2, 226) =4.436 =, p<.001, eta^2 =.038); as was C2, (F(2,226) =8.279, p<.001, eta^2 =.068). This suggests that younger children, 10-12-year olds, used a laxer criterion than did older

(15- to-16-year-old) children, when making eyewitness identification, thus producing

more false alarms (see figure 1). This age difference has been found in all of our studies,

as reviewed in the introduction, and now extends to the eyewitness identification of

females. This relationship with age was not affected by whether instructions, as described

(emphasizing the need for fairness and prevention vs. neither), were introduced.

However, the effect of age on C2 did differ for participants depending on the measure of

Ufeelbad, their responses to the question "how bad would you feel if your iPhone was

stolen," which was asked on a four-point Likert scale (response options: not bad, a little

bad, bad, very bad) (p<0.02). For participants who responded that they would feel 'not

bad' or 'a little bad', C2 showed a negative relationship with age, while C2 showed a

positive relationship with age for participants who reported that they would feel 'pretty

bad' or 'very, very bad' (see Figure 2). Novel for studies in this series, there was a near

significant difference in sensitivity (d') for gender, $F(1, 226) = 3.810$, p < .052, eta^2

=.017 indicating that females, in the study, were more accurate than males (see Figure 1).

Explicit Moral Judgment Measures: *Focus on Prevention or Fairness*

A significant negative relationship was found between focus more on prevention

(*Focus More On*) and age ($F (2,226) = 12.42$; p< .001; eta^2= .099), such that the younger

children focused more on prevention whereas the older children focused more on

fairness, which is consistent with the actual signal detection measures described above.

Gender also showed a marginally significant relationship with Focus More On, ($F(1,226)$

$= 3.419$; p<.066; eta^2= .015) with females tending to focus more on fairness compared

with males. Additionally, Focus More On was negatively correlated with C3 (decisional

criteria) in the expected direction. That is, those participants who focused more on prevention than on fairness had lower C3 scores, indicating more false alarms.

Using structural equation modelling, we tested whether a model that included the direct effects of Focus More On, gender, and age on C3, along with the indirect effects of gender and age through Focus More On C3 would have higher incremental validity compared to a model that included only the main effects of each variable on C3. Surprisingly, we did not find evidence to support this hypothesis.

Confidence Judgments

How bad would you feel? (Ufeelbad) A significant difference for gender was found for the question "how bad would you feel if your iPhone was stolen," which was assessed by a Likert scale (response options: not bad; a little bad; pretty bad; very, very bad) ($F(1,226) = 5.029$, p $<.026$, eta^2 .022) with girls reporting feeling worse. In addition, the interaction between gender and age group had an effect on *Ufeelbad* that approached statistical significance ($F(2,226) = 2.704$ p$<.069$, eta^2 .023), such that *Ufeelbad* tended to decrease with age for boys, but not for girls (see Figure 3).

Judgments regarding punishment, severity of the theft, and which error is worse to make. A significant effect was obtained for age group on the four-point Punishment scale (no punishment to expulsion), $F(2,26) = 7.246$ p $< .001$, partial eta squared = .06). Not surprisingly, the younger (11- to-12-year-old) children advocated a harsher punishment. A significant effect was also obtained for age group and on the scale "how bad was what the student did," which was measured using a Likert scale (response options: not bad; a little bad; pretty bad; very, very bad) ($F(2,226) = 6.196$ p $< .002$, partial eta squared = .052). Younger children (11- to-12-year olds) rated the theft as a

worse crime than did older children. A significant effect was obtained for age group on 'which error is worse', the decisional criteria for "false alarms, misses or both are errors are equally bad" (F (2,226) = 4.755 p < .009, partial eta^2 = .04). Eleven to 12-year-old children were also more likely than the older children to judge misses as worse than false alarms (see Table 2).

Discussion

The findings indicate that the effect of age on decisional bias is robust. While prior research (Spring, Saltzstein & Peach, 2012; Spring & Saltzstein, & Vidal, 2015; Spring & Saltzstein, 2017) using an eyewitness investigation task involving a male perpetrator and male foils has demonstrated the effect of age on decisional bias, this study shows that the same effect occurs when a female perpetrator and foils are involved. A new finding was that female adolescents and pre-adolescents used a more stringent criterion when identifying a (female) thief than do adolescents and pre-adolescents males. The differences involving gender are particularly interesting since past research (briefly summarized in the introduction) has shown that females have also been found to be *more accurate* when identifying female faces than male faces. However, including the indirect effects of gender and age through one attitudinal measure, Focus More On 'fairness or prevention' on C3, the split between "a little sure it is' and 'very sure it is' the perpetrator, surprisingly did not improve the predictive validity of our model. This might be interpreted as supporting the idea that this kind of decision-making is implicit, i.e., outside of conscious awareness.

Further, the prompting of the more justice-based and compassion-based orientations had no effect whatsoever, possibly indicating that these 'promptings' by the

researchers may not have been necessary possibly because the participants, pre-adolescents

and adolescents, already bring these orientations to the task and therefore any instructions

are ineffective, This interpretation is consistent with the idea that, in general, females tend

to be more 'compassionate' than males, as mentioned in the introduction, but here

demonstrated on an eyewitness identification task, which as we have emphasized earlier,

may be treated as involving implicit moral decision-making.

Thus, this study has both confirmed some previous findings, for example, the

relationship between age and decisional criteria, and found some new ones, especially the

tendency for girls to use a more conservative decisional criterion than boys when

identifying a perpetrator. What has been confirmed is that both males and female

perpetrators are more likely to be positively misidentified (indicated by the low

decisional criterion, C scores) by younger eyewitnesses.

Finally, and more generally, the data re-affirm the usefulness of treating

eyewitness identification as involving *inter alia,* a kind of implicit moral decision-

making. The findings with regards to gender are particularly interesting in that they

expand our understanding of gender and social cognition, in general, and in particular

gender and eyewitness decision-making.

ACKNOWLEDGEMENTS

Table 1. C2, C3 and d-prime by Gender, Age Group & Experimental Condition

	C2 mean (SD)	C3 mean (SD)	d-prime mean (SD)
Gender			
Male (*n*=111)	.49 (.56)	1.38 (.55)	.65 (1.11)
Female (*n*=121)	.43 (.56)	1.34 (.57)	.96 (1.23)
Age Group			
11- to-12 (*n*=89)	.30 (.47)	1.25 (.55)	.68 (1.10)
13- to-14 (*n*=72)	.45 (.60)	1.33 (.56)	.84 (1.13)
15- to-16 (*n*=71)	.66 (.57)	1.52 (.53)	.95 (1.32)
Experimental Condition			
Experimental (*n*=121)	.43 (.58)	1.35 (.58)	.80 (1.24)
Control (*n*=111)	.49 (.55)	1.37 (.53)	.82 (1.12)

Note: Age was a significant predictor of both C3 ($p<.001$) and C2 ($p<.001$).

Additionally, gender predicted dprime, although the effect was only marginally

significant ($p<.052$).

Table 2. Age Group Differences in Judgments Regarding Punishment, Which Error is Worse to Make, and Severity of Theft

	11- to-12 (*n*=89) Proportion of respondents	13- to-14 (*n*=72) Proportion of respondents	15- to-16 (*n*=71) proportion
Punishment			
No punishment/academic note (1)	.08	.13	.15
In school detention (2)	.17	.38	.39
Out-of-school detention (3)	.53	.38	.35
Expulsion (4)	.22	.11	.11
Which error is worse to make			
Miss	.43	.31	.22
Equally bad	.06	.04	.01
False alarm	.52	.65	.76
Severity of theft			
Not bad	.00	.01	.01
A little bad	.01	.01	.04
Pretty bad	.42	.68	.51
Very, very bad	.57	.30	.43

Note: The values in each cell of the table represent the proportion of respondents in each age group who gave the indicated response. Age group was a significant predictor of punishment (p<.001), which error is worse to make (p<.009), and severity of theft (p<.002). Compared to older children, younger children tended to think stealing the iPhone was a worse crime and advocated for harsher punishments for the perpetrator of the crime. Younger children were also more likely than older children to judge misses as worse than false alarms, although the majority of participants in each age group judged false alarms to be the worse error.
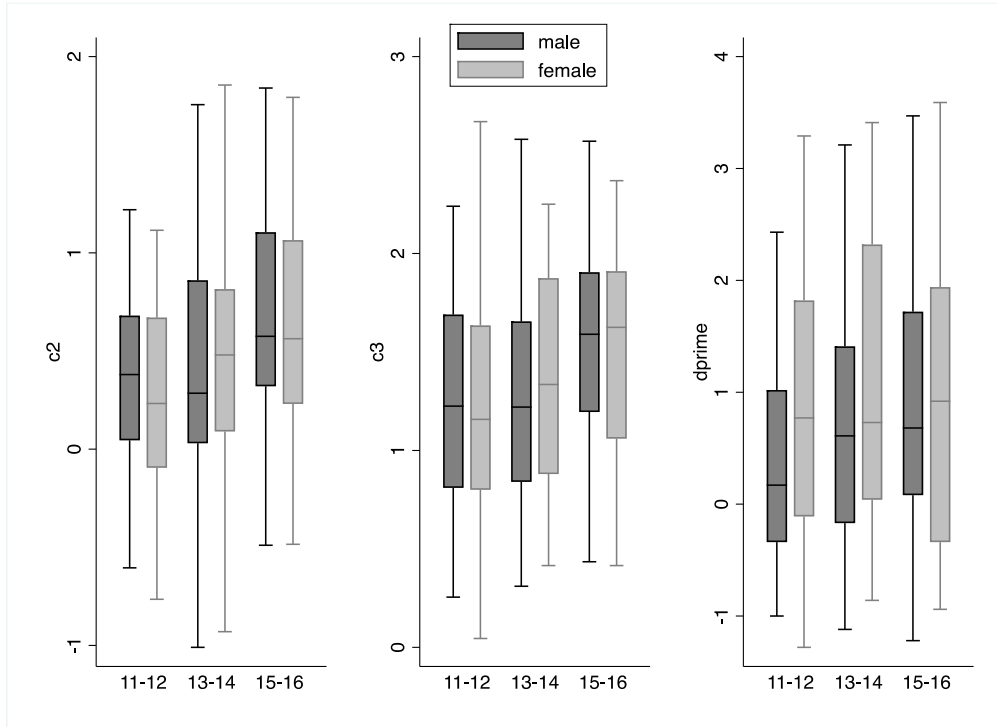
Figure 1. Boxplots of C2, C3 and d-prime by gender, age group and experimental

condition Note: The boxplots display the median and interquartile ranges for C2 (left

panel), C3 (middle panel) and d-prime (right panel) for groups of participants broken

down by age and gender. In these plots, different age groups are indicated along the x-

axes, and different genders are distinguished by the colors of the bars; the bars for males

are dark gray while the bars for females are light gray. C2 and C3 are both measures of

the decisional criterion participants used when deciding whether each person displayed

in the photo lineup was the guilty party with C2 representing a split between 'a little sure

that it isn't the person' and 'a little sure that it is the person' and C3 representing a split

between 'very sure it is the person' and the other three alternatives. d-prime is a measure

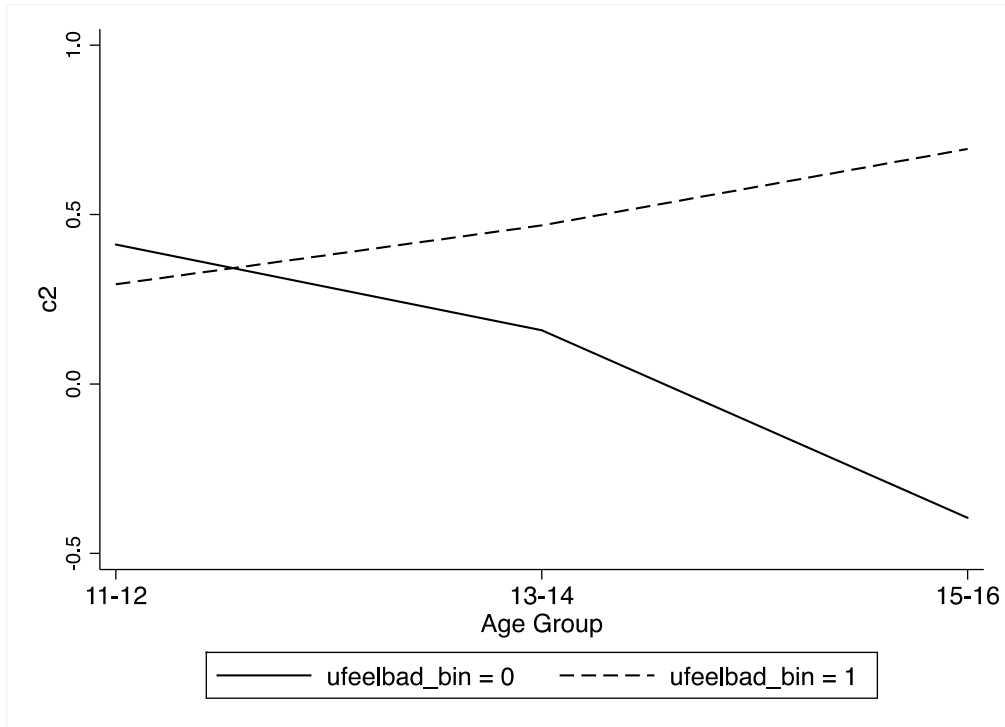*of participants' overall accuracy.*

Figure 2.

Effect of age on C2 differs on Ufeelbad

Note: Ufeelbad_bin = 0 indicates participants who responded 'not bad' or 'a little bad' in response to our question about how they would feel if their iPhone was taken, whereas Ufeelbad_bin = 1 indicates participants who responded 'pretty bad' or 'very, very bad.' There was a significant interaction between Ufeelbad and age (p<0.02). For participants who responded 'not bad' or 'a little bad', C2 showed a negative relationship with age, while C2 showed a positive relationship with age for participants who reported they would feel 'pretty bad' or 'very, very bad'.
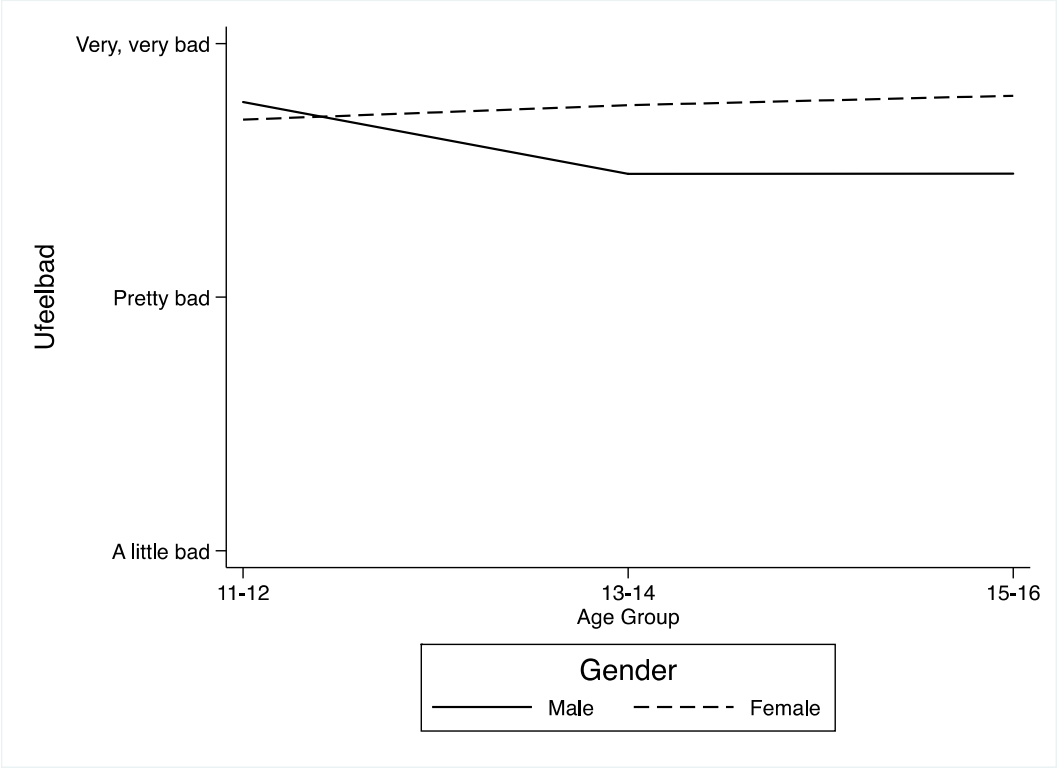
Figure 3. Effect of age on Ufeelbad differs for males and females

Note: The interaction between gender and age group had an effect on Ufeelbad that

approached statistical significance (p<.069). Ufeelbad tended to decrease with age for

boys, but not for girls.

References

Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness identification. In *Psychology and law: An empirical perspective* (pp. 177–221). New York, NY: Guilford Press.

Bucciarelli, M. (2015). Moral dilemmas in females: Children are more utilitarian than adults. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01345

Carlson, C. A., & Carlson, M. A. (2014). An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *3*(2), 45–53. https://doi.org/10.1016/j.jarmac.2014.03.004

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *14*(2), 118–128. https://doi.org/10.1037/1076-898X.14.2.118

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

Cultice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' memory monitoring: Feeling-of-knowing judgments. *Child Development*, *54*(6), 1480. https://doi.org/10.2307/1129810

Cutler, B. L., & Penrod, S. D. (1995). Mistaken identification: The eyewitness, psychology, and the law. *Mistaken Identification: The Eyewitness, Psychology, and the Law.*, x, 290–x, 290.

Denno. D. (1994). Gender, crime and the criminal law defenses. *Journal of Criminal Law and Criminology*, *85*(1), 80–180.

Gilligan, C., & Attanucci, J. (1988). Two Moral Orientations: Gender Differences and

Similarities. Merrill-Palmer Quarterly, 34(3), 223-237. Retrieved July 16, 2020, from

www.jstor.org/stable/23086381

Gilligan, C. (1977). In a different voice: Women's conceptions of self and of morality.

Harvard Educational Review, *47*(4), 481–517.

https://doi.org/10.17763/haer.47.4.g6167429416hg5l0

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the

sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *15*(2), 140–

152. https://doi.org/10.1037/a0015082

Heeger, D. (2009, October 12). Signal Detection Theory. Retrieved September 9, 2019, from

https://www.cns.nyu.edu/~david/handouts/sdt/sdt.html

Held, V. (1995). The Meshing of care and justice. *Hypatia*, *10*(2), 128–132.

https://doi.org/10.1111/j.1527-2001.1995.tb01374.x

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition:

A meta-analytic review. *Visual Cognition*, *21*(9–10), 1306–1336.

https://doi.org/10.1080/13506285.2013.823140

Hills, P. J., & Lewis, M. B. (2011). The own-age face recognition bias in children and adults.

Quarterly Journal of Experimental Psychology, *64*(1), 17–23.

https://doi.org/10.1080/17470218.2010.537926

Hoffman, M. L. (1987). *The contribution of empathy to justice and moral judgment.* In N.

Eisenberg & J. Strayer (Eds.), *Cambridge studies in social and emotional development.*

Empathy and its development (p. 47–80). Cambridge University Press

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin*, *126*(5), 703–726.

Kant, I. (1964). *Groundwork of the metaphysics of morals* (H. J. Paton, Ed. & Trans.). New York, NY: Harper & Row.

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. https://doi.org/10.1016/j.jecp.2007.01.007

Keller, J. (1997). Autonomy, relationality, and feminist ethics. *Hypatia*, *12*(2), 152–164. https://doi.org/10.1111/j.1527-2001.1997.tb00024.x

Klein KJK, Hodges SD. Gender differences, motivation, and empathic accuracy: When it pays to understand. Personality and Social Psychology Bulletin. 2001;27:720–730.

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: a review of research using the NICHD Investigative Interview Protocol. *Child abuse & neglect*, *31*(11- to-12), 1201–1231. doi:10.1016/j.chiabu.2007.03.021

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, *76*(6), 796–802. https://doi.org/10.1037/0021-9010.76.6.796

Lovén, J., Herlitz, A., & Rehnman, J. (2011). Women's own-gender bias in face recognition memory: The role of attention at encoding. *Experimental Psychology*, *58*(4), 333–340. https://doi.org/10.1027/1618-3169/a000100

Macmillan, N. A., & Creelman, C. D. (2005). Detection Theory. New York: Psychology

    Press, https://doi.org/10.4324/9781410611147.

Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring lineup fairness. *Applied Cognitive*

    Psychology, *13*(S1), S1–S7. https://doi.org/10.1002/(SICI)1099-

    0720(199911)13:1+<S1::AID-ACP678>3.0.CO;2-9

Mason, S. E. (1986). Age and gender as factors in facial recognition and identification.

    Experimental Aging Research, *12*(3), 151–154.

    https://doi.org/10.1080/03610738608259453

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. https://doi.org/10.3758/BF03193074

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of

    eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus

    sequential lineups. *Journal of Experimental Psychology: Applied, 18,* 361–376.

    http://dx.doi.org/10.1037/a0030609

Mill, J. S., & Crisp, R. (1998). *Utilitarianism*. Oxford; New York: Oxford University Press.

Münsterberg, H. (1908). *On the witness stand; essays on psychology and crime*. Retrieved

    from //catalog.hathitrust.org/Record/001744300

Pozzulo, J. (2017) *The Young Eyewitness.* Washington, APA.

Rawls, J. (1971). *A theory of justice* (Rev. ed). Cambridge, Mass: Belknap Press of Harvard

    University Press.

Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory*, *14*(3), 289–296.

https://doi.org/10.1080/09658210500233581

Roebers, C. M., Moga, N., & Schneider, W. (2001). The role of accuracy motivation on children's and adults' event recall. *Journal of Experimental Child Psychology*, *78*(4), 313–329. https://doi.org/10.1006/jecp.2000.2577

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

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence^accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315–327. https://doi.org/10.1037/0033-2909.118.3.315

Spring, T., & Saltzstein, H. D. (2017). Decisional bias as implicit moral judgment. *International Journal of Developmental Science*, *11*(3–4), 81–86. https://doi.org/10.3233/DEV-160192

Spring, T., Saltzstein, H. D., & Peach, R. (2013). Children's eyewitness identification as implicit moral decision-making: Children's eyewitness identification. *Applied Cognitive Psychology*, *27*(2), 139–149. https://doi.org/10.1002/acp.2871

Spring, T., Saltzstein, H. D., & Vidal, B. (2015). A moral developmental perspective on children's eyewitness identification: Does intent matter? *Archives of Scientific Psychology*, *3*(1), 1–7. https://doi.org/10.1037/arc0000011

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.

Van Zandt, T. (2000). ROC curves and confidence judgements in recognition memory.

Journal of Experimental Psychology. Learning, Memory, and Cognition, *26*(3), 582–600.

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

Press.

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection

model of eyewitness identification. *Psychological Review*, *121*(2), 262–276.

https://doi.org/10.1037/a0035940