Legal and Criminological Psychology (2010), 15, 195–207

The British Psychological Society

www.bpsjournals.co.uk

Perceptions about memory reliability and honesty for children of 3 to 18 years old

Daniel B. Wright[1]*, Florence Hanoteau[2], Camilla Parkinson[2] and Anna Tatham[2]

[1]Florida International University, Miami FL, USA
[2]University of Sussex, Brighton, UK

Purpose. To examine the relationship between perceived memory characteristics and age.

Method. Participants rated the reliability and honesty of children's memory for one of two events. The children's ages varied from 3- to 18-years-old.

Results. Participants ($N = 612$) believed that memory reliability increased with age, but the observed effect was non-linear. Perceived reliability increased rapidly for children from 3 to 6 years. After this, male participants believed memory reliability increased, but less than in early childhood. Female participants did not think memory reliability increased in middle childhood and adolescence. Further effects involving type of event, age of participant, and the gender of the eyewitness were observed for honesty and the relationship between these attributes and beliefs in guilt.

Conclusions. These findings stress the need for more research on development trends of memory in middle childhood and adolescence.

Understanding the development of episodic memory is a massive ongoing endeavour. It is clear that from birth through adolescence into adulthood memory develops, but not in a simple way (Brainerd, Reyna, & Ceci, 2008). It is important to understand people's folk beliefs about the development of memory because the judicial system often relies on the testimony of children. Over the past 30 years there has been a growing awareness that children, when questioned properly, can accurately report much information (e.g., Goodman & Melinder, 2007; Lamb *et al.*, 2009; Orbach *et al.*, 2000; Sternberg, Lamb, Esplin, Orbach, & Mitchell, 2001). The courts now allow much younger children to testify, but there remains much debate about the accuracy of children's memory (Goodman, 2006).

Adolescents often testify in court. The lack of research places jurors in a difficult situation. The courts ask them to judge the reliability and honesty of the testimony of

**Correspondence should be addressed to Professor Daniel B. Wright, Psychology Department, Florida International University, Miami, FL 33199, USA (e-mail: dwright@fiu.edu).*

DOI:10.1348/135532508X400347

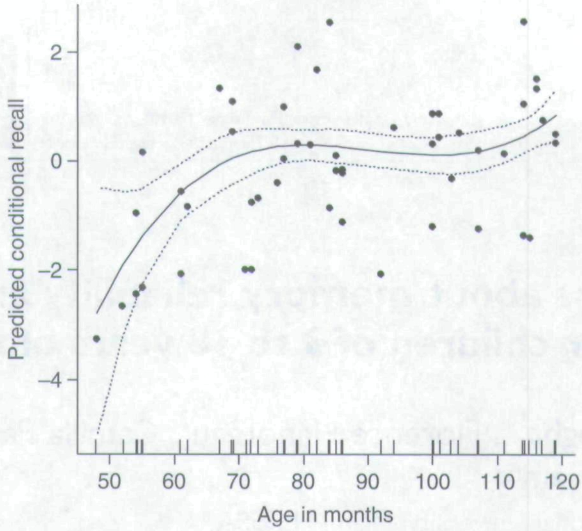

Figure 1. The relationship between remembering and age found in Study 2 of London et *al.* (2009). It shows memory increasing up to about 6 years old and then levelling off. The figure is reproduced from Wright and London (2009, Figure 7.11). The dashed lines show $\pm$ *SE*.

people of different ages and to use these to help them to reach a verdict. More research on adolescents' memory is needed and some studies have examined this (Melynk & London, 2009). One study that looked at memory over a broad age range was London, Bruck and Melnyk's (2009) study of children's memory over 1 year. Figure 1 from their data shows children's memory abilities increased up until about the age of 6-years-old (72 months) and then levelled off. While there are a few studies examining episodic memory across a broad range of children's ages, more research is necessary to establish a pattern of memory development and how it varies across situations. The purpose of our research is to see how people rate the reliability and trustworthiness of children's memory across an age range spanning younger children through adolescents. Given the complex relationship between age and actual memory abilities, we are particularly interested in the shape of the relationship between age and perceived memory qualities. For example, is the relationship linear or does the relationship follow a more curved pattern.

While we believe our research is applicable to courtroom settings, the design of our study was guided more by the basic scientific questions, listed above, about people's beliefs of children's memories than by making sure the materials match a specific legal example. There has been much research on people's beliefs about children's memory. There are several methods that have been used to ascertain how good people think children's memory is.

The simplest method to find out how people judge of the memory of children is to ask people directly. For example, Magnussen and colleagues (2006) asked a sample of Norwegians: 'When small children tell about events they have experienced, do you think they remember better, as well as, or worse than adults?'. They found only 18% thought the children's memories would be worse. Most thought small children were the same or better than adults. This is at odds with the beliefs of eyewitness testimony experts. Kassin, Tubb, Hosch, and Memon (2001) used the direct approach with eyewitness testimony experts. They asked the experts whether they agreed or disagreed with the following statement: 'young children are less accurate as witnesses than are adults'. Sixty-three percent of experts agreed.

There are limitations to this simple direct approach. One problem is that people may have difficulty interpreting the statements without further contextual information. For example, if we assume that memory development follows a complex pattern, the precise ages to which these statements refer may be critical. Further, people's beliefs about children's memory may depend on situational characteristics not specified in the short statements used by Magnussen *et al.* (2006) and Kassin *et al.* (2001). Leippe and Romanczyk (1987, 1989) argue that the impact of any testimony should relate to both the perceived memory capability of the witness and the witness' trustworthiness. Because there may be more motivation to lie in certain cases and children's capabilities for lying show a developmental trend (Bussey & Grimbeek, 2000), the relationship between accuracy and age could be quite complex. This means that the 'correct' answer to Kassin *et al.*'s survey question arguably is 'it depends'.

Most research under the heading 'juror decision-making' uses a different method to ascertain people's beliefs about children's memories. Participants are presented with an event description which includes testimony from a witness and are asked to make various ratings. The researchers usually have multiple versions of the event description, with slight differences in the testimony, and give a different version to each participant (i.e., a between-subjects design). This allows researchers to test if these differences affect the participants' ratings.

One of the first of these studies was by Loftus (1974). She had three versions of a brief description about a criminal event. The difference among the versions was whether an eyewitness identified the defendant. When the event was described without the eyewitness identification only 18% of participants rendered a guilty verdict, but this went up to 72% when an identification was made. Importantly, the percentage only dropped to 68% when the witness was described as having poor vision. Using this method she showed the large impact that even unreliable eyewitness testimony can have on juror decision-making. Since then many researchers have presented event descriptions to participants to show how people weigh different types of information when making attributions of guilt (for a review, see Winter & Greene, 2007; for a recent example, see Wright, 2007).

There are several studies that have explored how people judge children's memory using this method. Consider McCauley and Parker's (2001) study. Like us they were interested in people's perception of children's memory. They gave a nine page event description to participants. The event was a robbery, a sexual assault by a stranger, or a sexual assault by an acquaintance. McCauley and Parker varied whether the victim, who also provided the critical eyewitness testimony, was 6 or 13 years old. Participants rendered a verdict (guilty or not guilty) and made ratings of the victim's memory and honesty. They found the victim's age was not significantly related to verdict, but that the perceived memory reliability was slightly higher for the 13 year old victim than for the 6 year old victim, $M_{13-\text{year-old}} = 2.85$ versus $M_{6-\text{year-old}} = 2.71$, on a 0–4 scale with high scores corresponding to better perceived memory. This difference, however, was only significant for male participants. There were no age differences for honesty ratings.

While we have similar aims to McCauley and Parker (2001), our design has some important differences. First, because they had only two ages it is not possible to estimate the form of the relationship between age and participants' attributions. Therefore, we had 11 different ages: 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 12-, 15-, and 18 years old. Nightingale (1993, Expt 2) took a similar approach by varying age in 1-year increments between 6 and 14 years olds. We have opted for finer discrimination at younger ages (3 through 10 years old) than above 10 years old because we expect greater differences at these younger ages. We also start at a younger age than McCauley and Parker, and Nightingale. Second, particularly for measuring trustworthiness and the verdict, the fact that the key witnesses in McCauley and Parker and in Nightingale were also the victims confounds trying to assess judged memory effects. In our scenarios the key witness is a bystander. The results on perceived memory accuracy and age have been mixed: sometimes children are believed to be less accurate than adults, sometimes more accurate than adults, and sometimes no differences are detected. Part of the reason for this mixture of results is the complexity of the task faced in many cases (see discussion in Bottoms, Golding, Stevenson, Wiley, & Yozwiak, 2007). The ratings for a bystander eyewitness are a more pure measure of perceived memory qualities than for a victim eyewitness. While in many cases the eyewitness is also the victim, it is important to disentangle these roles for the current research. Once the relationship between perceived memory and age is better understood, research can be done to examine if the relationship holds for victim–witnesses.

Our primary interest is with the age of witness variable, but we manipulated other aspects for the event description both to test for the generality of any age effects and for exploratory purposes. Like McCauley and Parker we varied the event-type. We had the witness either view a physical assault or a sexual assault. The witness was either female or male. Participants were from a community sample (rather than a university sample) and indicated their age and gender. These are examined in an exploratory fashion.

We use a shorter scenario than McCauley and Parker for two reasons. The first is because we wanted to make sure that the procedure would be short enough not to adversely affect participant recruitment. The second is to ensure that participants focus on the age manipulation. Our interest is whether participants are able to conceive of children of these different ages and how they perceive their memories. Therefore, embedding the age manipulation in a long transcript could produce null results because people fail to remember the witness' age after reading a long transcript.

Method

Participants and design

A power analysis was conducted to help to determine the sample size. Several different analyses were planned. The most complex we envisaged involved simultaneously testing a four degree of freedom effect (the age variable split into two cubic curves joined smoothly at the median: the default curve for many statistical packages). We entered this into the software G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) with $\alpha = .05$, power $= .80$, and a 'small' effect size to detect. This suggests a sample size of 602, so our aim was to get approximately 600 participants. Most of the individual tests that we examined were much simpler. When just a single degree of freedom test is evaluated with this sample size and $\alpha = .05$, there is a power of approximately .93 to detect a 'small' effect.

People were approached in commercial areas in the South-East of England and asked if they would take part in a brief study. Six hundred and twelve took part. Five were excluded from all analyses for not responding either to the memory reliability or to the honesty questions. Six did not answer the verdict question so are excluded from analyses which included the verdict questions. Twelve people did not indicate their gender and 22 did not indicate their age. These participants are excluded for tests including these variables. Of those who indicated, 56% were female and the average age was 33 years old (*SD* = 14). We did not ask for any further demographic information.

Participants were randomly allocated to 1 of 44 conditions of an 11 × 2 × 2 between-subjects design. The factors were the eyewitness' age, eyewitness' gender, and the type of event. The ages were: 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 12-, 15-, and 18 years old. The eyewitness was either a girl or a boy. To avoid confusion we use girl and boy to refer to the gender of the eyewitness and female and male to refer to the gender of the participant. The event was either physical abuse in the home or sexual abuse in the school.

Materials and procedure

Two event descriptions were created (see Appendix A). One involves a child seeing his/her parents argue and alleged physical abuse of the child's mother in the family home. The other involves a child seeing a teacher and another student and alleged sexual abuse against the other student in the classroom. The descriptions are between 250 and 275 words. We felt this was enough to place the testimony within a meaningful context, but not so long that people would drop-out of the study. Participants were not compensated for their time so it was important that the task only took about five minutes. The description was printed on a single sheet of paper and was given to the participant along with, if needed, a pencil. The 44 (11 × 2 × 2) versions were randomly distributed within each of the three interviewers' workload.

After reading the description, participants were asked to render a verdict (guilty or not guilty), rate their belief that they believed the defendant was guilty (0–100% scale in 10% increments), rate how reliable they thought the eyewitness' memory was of the event (0–10 scale), and rate how honest they thought the eyewitness was (0–10 scale). They were also asked to indicate their age and gender. After this, participants were thanked and, if they wished, debriefed. The study received ethical approval from the School of Life Science's Research Governance Committee of the University of Sussex.

Results

It is unlikely that there is a simple relationship between age with perceived memory reliability and honesty. Most of the past research in this area has either asked participants simply to compare children with adults, without specifying the age (e.g., Kassin *et al.*, 2001) or listed only a few discrete ages (e.g., McCauley & Parker, 2001; a notable exception being Nightingale, 1993, Expt 2). This means that age has usually been treated as a categorical variable and the analysis fitted into the well-known ANOVA framework. Using 11 age categories allowed us to explore the relationship in more detail. Rather than just fitting a straight line through the data, we tested whether more flexible curves could improve the fit of the model. There are several choices of flexible curves. The method we used involved different numbers of specifically designed contrasts. For example, a common set of contrasts allows the relationship to be two cubic polynomials joined together at the median in a smooth way. Here 'smooth way' means that the second derivative of these curves is the same at this point (Hastie & Tibshirani, 1990; Wright & London, 2009). This requires four contrasts, or four degrees of freedom. We used the `bs` function from the freeware R (R Development Core Team, 2008). The syntax for running all the analyses and constructing all graphs in R is available from the first author.

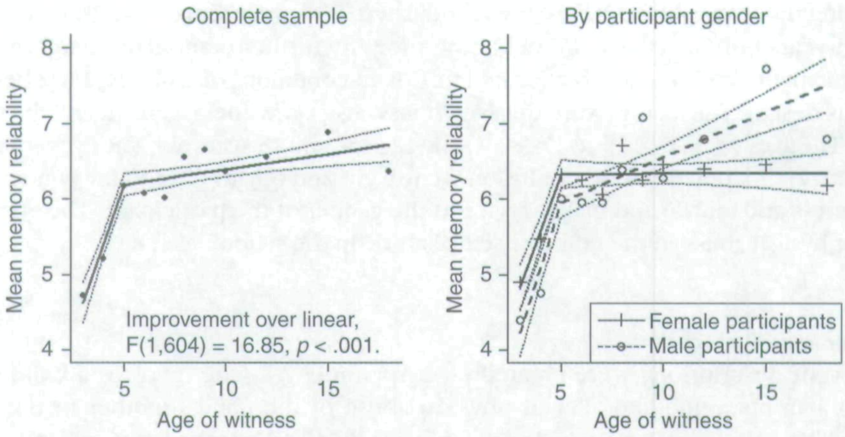

Figure 2. The relationship between perceived memory reliability and the witness' age. The left panel shows the data for all participants and the non-linear relationship is clear. The right panel showed the interaction by participant's gender. The lighter (dashed) lines are $\pm SE$.

Perceived reliability

We began by predicting perceived reliability with all the experimental variables (witness age, gender, and crime type) and their interactions. In a backward stepwise fashion, the highest order and least significant interaction was removed providing its *p*-value was greater than .05. The process stopped when no effect had a *p*-value greater than .05. All variables were removed except for age as a linear predictor, which was statistically significant, $F(1, 605) = 22.04$, $p < .001$, $\eta^2 = .04$. The left panel of Figure 2 shows the observed data with black dots. While this significant effect is consistent with the tendency for reliability to increase with age, the data points clearly show a non-linear relationship. Several models with an extra degree of freedom increased the fit significantly. For example, two straight lines connected at the median (8-years-old) produced a better fit, new $\eta^2 = .06$. The improvement was statistically significant, $F(1, 604) = 16.85$, $p < .001$, $\eta_p^2 = .03$. Locating this connection, which is called a *knot* in the statistics literature, at the median is the default for the software used. The model fit increased slightly with the knot at 5-years-old. This is the model depicted in the left panel of Figure 2. Other models with two degrees of freedom, like a single quadratic curve, also fit the data significantly better than the linear model. More complex curves (in terms of degrees of freedom) were examined, but none significantly improved the model.

Most participants indicated their age and gender. We observed no effects with participants' age on estimated reliability (*min.* $p > .30$). There was an interaction between participant gender and witness age, $F(1, 590) = 8.07$, $p = .005$, $\eta_p^2 = .01$, with male participants believing older witnesses more than female participants. This is depicted in the right panel of Figure 2.

Perceived honesty

Similar analyses were conducted for honesty. All the effects which included crime type were non-significant and removed. There was an interaction between the age and gender of the witness, $F(1, 603) = 6.02$, $p = .01$, $\eta_p^2 = .01$, which corresponds to an increase with age in believed honesty for girl witnesses but not for boy witnesses. The observed means are shown in Figure 3 with +s for the females and Os for the males.

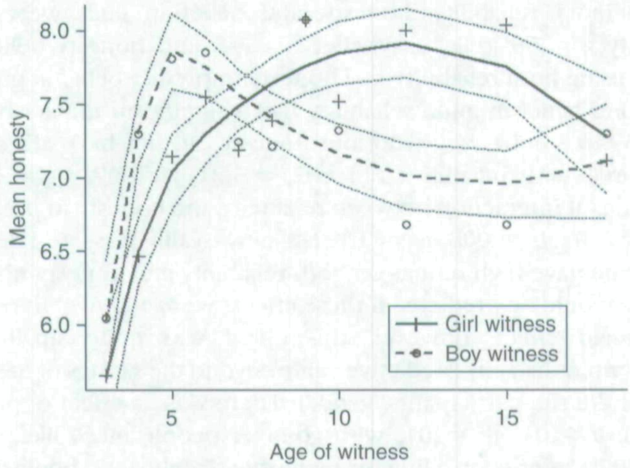

Figure 3. The relationship between honesty and the age of the witness, split by whether the witness was a girl or a boy. The lighter (dashed) lines show $\pm SE$.

We explored more complex (non-linear) relationships between age and perceived honesty. Because age was used both as a main effect and as part of an interaction the modelling was more complex than with reliability. We first included a model, which allowed the age effect to be composed of two straight lines connected at the median, and allowed these lines to be different for the girl and boy witnesses. This was a significant improvement, $F(2, 601) = 8.12$, $p < .001$, $\eta_p^2 = .03$. However, this still did not fit the data as well as more complex models. A model which allowed two quadratic equations connected at the median provided a further significant increase, $F(2, 599) = 5.16$, $p = .004$, $\eta_p^2 = .02$. The shape of the curves were significantly different for the girl and boy witnesses, $F(3, 599) = 3.25$, $p = .02$, $\eta_p^2 = .02$. We tried different locations for connecting these curves, and the optimal location was between 5 and 6 years old. The curves in Figure 3 show this model, with quadratic curves connected at 5 years and 5-months-old. Participants thought that honesty increased for boys and girls up to about 5 to 6 years old. After this point, participants thought boys became less likely to tell the truth while honesty ratings for girls continued to increase. Once the children were 18 years old, participants gave boys and girls similar ratings for honesty.

We looked to see if any of these effects were moderated by the participants' age or gender. No significant effects involving these variables were observed.

Predicting guilt

Six people did not render a verdict so were removed from these analyses. Two related measures of guilt were recorded: the binary score for verdict and the numeric variable for belief in guilt on a 0–100 probability scale. As expected, the two measures of guilt (belief in guilt and verdict) were closely related. Using a logistic regression with belief used to predict verdict, the association was statistically significant, $z = 11.24$, $p < .001$, Nagelkerke's $R^2 = .50$ (abbreviated as NR^2, see Appendix B). The point where there is a 50% chance of a guilty verdict, which is arguably the point of reasonable doubt (see Wright & Hall, 2007, for details), for the sample is 74.39% (SE = 1.54%, SE found with the `dose.p` function of Venables & Ripley, 2002). We will now report analyses on the numeric belief variable and then discuss differences for the dichotomous verdict variable.

Perceived memory reliability, honesty, and belief in guilt were all correlated: reliability/honesty $r = .59$, reliability/belief $r = .34$, and honesty/belief $r = .25$. The multiple r-value using both reliability and honesty to predict belief in guilt was $r = .34$. Thus, in predicting belief in guilt, reliability had a significant impact beyond honesty, $F(1, 602) = 37.39$, $p < .001$, $\eta_p^2 = .06$; but honesty did not have a significant impact beyond that shared with reliability, $F(1, 602) = 3.21$, $p = .07$, $\eta_p^2 = .00$. There was, however, a significant interaction between reliability and honesty in predicting belief in guilt, $F(1, 601) = 7.03$, $p = .008$, $\eta_p^2 = .01$. The new r-value was .36. This effect was due to participants who gave high ratings on both reliability and honesty not giving as high ratings of guilt as would be predicted if these effects were additive. In comparison with the reliability main effect, however, this effect was quite small. None of the experimental variables had any predictive value beyond the ratings of memory reliability and honesty. Nor did the participants' gender. There was an effect of participants' age, $F(1, 578) = 5.98$, $p = .01$, $\eta_p^2 = .01$, with younger people more likely to believe the defendant was guilty after controlling for memory reliability and honesty. The decrease was non-linear, only participants less than 30 years old had a higher likelihood of believing in guilt.

The analyses were repeated for the dichotomous verdict variable using a series of logistic regressions. Both perceived memory reliability, $\chi^2(1) = 57.25$, $p < .001$, $NR^2 = .14$, and perceived honesty, $\chi^2(1) = 26.15$, $p < .001$, $NR^2 = .06$, were related to verdict. As with the belief variable, the honesty contribution did not add predictive value beyond reliability, $\chi^2(1) = 0.88$, $p = .35$, $NR_p^2 = .002$, but the reliability contribution did add predictive value beyond honesty, $\chi^2(1) = 31.98$, $p < .001$, $NR_p^2 = .08$. The interaction was non-significant, $\chi^2(1) = 0.78$, $p = .38$, $NR_p^2 = .002$.

To allow comparison with the previous results we added the experimental and demographic variables, individually, to the model including reliability, honesty, and their interaction. Some effects were significant:

(a) there were more guilty verdicts for the sexual abuse, $z = 2.98$, $p = .003$;
(b) older participants gave fewer guilty verdicts, $z = 2.49$, $p = 0.01$;
(c) females gave fewer guilty verdicts, $z = 2.91$, $p = 0.004$.

The effects for crime type and participants' gender are different from those when predicting belief in guilt. This suggests that each may have direct effects in predicting verdict beyond belief in guilt. Controlling for belief in guilt there were more guilty verdicts for sexual assault than physical assault, $z = 3.89$, $p < .001$, and by female participants, $z = 3.04$, $p < .001$. These effects are shown in Figure 4. For those given the physical abuse case, the required level of belief in guilt to surpass 'reasonable doubt' was about 80%, but only 70% belief was necessary for the sexual abuse case. Similarly, for these data female participants were willing to render a guilty verdict with a belief in guilt of 70% while male participants required the more stringent level of 80%. In addition, participant age was significant, $z = 2.67$, $p = .01$. The necessary belief in guilt for a 50% probability of rendering a guilty verdict rose from 71% for 20-year-olds, to 80% for 50-year-olds, and to 89% for 80-year-olds.

Discussion

The developmental trajectory of episodic memory is an active research area within psychology. Most research shows that cognitive skills improve greatly during early

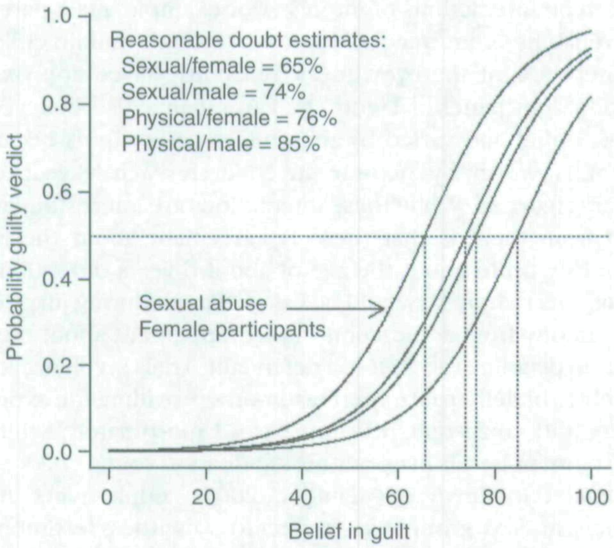

Figure 4. The probability of rendering a guilty verdict with the belief in guilt, broken down by the gender of the witness and the type of crime.

childhood and that some continue developing through adolescence (Blakemore & Choudhury, 2006). However, even as some cognitive abilities increase this may mean lower memory accuracy (Brainerd *et al.*, 2008). More research is necessary to understand the complex relationship between age and memory.

In a courtroom setting, testimony from child witnesses is often critical. In many cases, the child's testimony may be the only evidence against the alleged perpetrator (London, Bruck, Wright, & Ceci, 2008). Given the complexity of the relationship between age and memory, the judicial system puts jurors in a difficult situation. Sometimes with and sometimes without the help of expert witnesses, jurors have to judge the reliability and credibility of a child's testimony. The focus of this paper was how these judgments relate to the child's age. In order to examine this relationship, we used a between-subjects design where participants read an event description where the eyewitness was between 3 and 18 years old. We used 11 different age categories, which allowed us to look at the shape of the relationship.

The clearest finding was that participants believed memory reliability increased with age. Finding that people believed young children's memory was less reliable than adults is at odds with some research (e.g., Magnussen *et al.*, 2006), but in line with other research (e.g., Newcombe & Bransgrove, 2007) and with expert opinion (Kassin *et al.*, 2001). The mixed results within the literature may be because most previous research did not consider age across as large a range as we have. The increase after the age of 6 years old appears small and variable. We found it depended on characteristics of the eyewitness and the participant.

The increase in perceived reliability was not linear with age. Several statistical models that allow the perceived qualities to improve rapidly at first, but then for any improvement either to stop or to become less rapid, fit our data for memory reliability and honesty. The fact that, for example, two straight lines connected at a single point fit the data well (see Figure 2) does not mean that this model is correct. It just means that the simpler linear model must be rejected. Different research designs and meta-analytic techniques are necessary to fine-tune this relationship further.

There were several interactions of interest. For example, male participants believed the eyewitness' reliability continued to increase through middle childhood while the female participants thought the eyewitness' reliability stayed approximately constant during this period (right panel of Figure 2). Participants also believed that children's honesty in middle childhood varied by the child's gender: girls became more honest while boys did not. However, by the time the children reach 18 years old their honesty levels were similar (Figure 3). While these interactions are interesting in themselves, our main conclusion from them is that there is agreement about the rapid increase in reliability and honesty up to about the age of about 6 years old, but after that there is variation. This suggests that there would be disagreement during jury deliberation about the validity of testimony from adolescents and disagreement about the weight to place on such testimony in deciding the guilt of a defendant. Trials involving adolescents would be those most likely to benefit from expert testimony, providing the expert testimony was based on a strong and consistent research base. Unfortunately, while there is much research on development trends in eyewitness abilities in young children, there has been less on adolescents (Goodman & Melinder, 2007). Adolescents are a particularly important, yet understudied, group with respect to eyewitness testimony research. More episodic memory research on adolescence is necessary for researchers to be able to address the wealth of questions that can occur about adolescents' memory in court.

Finally, our interest was if simply saying that a witness was a certain age would affect people's responses. This is a causal hypothesis (Wright, 2006) regarding how people are affected by their beliefs about this age group. Thus, we do not provide any other information, like showing that the younger witnesses have worse mental skills, that might further increase the effects.

References

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*, 296–312.

Bottoms, B. L., Golding, J. M., Stevenson, M. C., Wiley, T. R. A., & Yozwiak, J. A. (2007). A review of factors affecting jurors' decisions in child sexual abuse cases. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness testimony: Volume 1, memory for events* (pp. 509–543). Mahwah, NJ: Lawrence Erlbaum Associates.

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*, 343–382.

Bussey, K., & Grimbeek, E. J. (2000). Children's conceptions of lying and truth-telling: Implications for child witnesses. *Legal and Criminological Psychology*, *5*, 187–199. doi:10.1348/135532500168083

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Goodman, G. S. (2006). Children's eyewitness memory: A modern history and contemporary commentary. *Journal of Social Issues*, *62*, 811–832.

Goodman, G. S., & Melinder, A. (2007). Child witness research and forensic interviews of young children: A review. *Legal and Criminological Psychology*, *12*, 1–19. doi:10.1348/135532506X156620

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman and Hall.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the 'general acceptance' of eyewitness testimony research. *American Psychologist*, *56*, 405–416.

Lamb, M. E., Orbach, Y., Sternberg, K. J., Aldridge, J., Pearson, S., Stewart, H. L., Esplin, P. W., & Bowler, L. (2009). Use of a structured investigative protocol enhances the quality of investigative interviews with alleged victims of child sexual abuse in Britain. *Applied Cognitive Psychology, 23*, 449–467.

Leippe, M. R., & Romanczyk, A. (1987). Children on the witness stand: A communication/persuasion analysis of jurors' reactions to child witnesses. In S. J. Ceci, M. P. Toglia, & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 155–177). New York: Springer.

Leippe, M. R., & Romanczyk, A. (1989). Reactions to child (versus adult) eyewitnesses: The influence of jurors preconceptions and witness behaviour. *Law and Human Behavior, 13*, 103–132.

Loftus, E. F. (1974). Reconstructing memory: The incredible eyewitness. *Psychology Today, 8*, 116–119.

London, K., Bruck, M., & Melnyk, L. (2009). Post-event information affects children's autobiographical memory after one year. *Law and Human Behavior, 33*, 344–355. doi:10.1007/s10979-008-9147-7

London, K., Bruck, M., Wright, D. B., & Ceci, S. J. (2008). How children report sexual abuse to others: Findings and methodological issues. *Memory, 16*, 29–47.

Magnussen, S., Andersson, J., Cornoldi, C., De Beni, R., Endestad, T., Goodman, G. S., Helstrup, T., Koriat, A., Larsson, M., Melinder, A., Nilsson, L. G., Ronnberg, J., & Zimmer, H. (2006). What people believe about memory. *Memory, 14*, 595–613.

McCauley, M. R., & Parker, J. F. (2001). When will a child be believed? The impact of the victim's age and juror's gender on children's credibility and verdict in a sexual abuse case. *Child Abuse & Neglect, 25*, 523–539.

Melnyk, L., & London, K. (2009). *Adolescents: The neglected eyewitnesses*. Symposium at Society Research in Child Developments, Denver, CO.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691–692.

Newcombe, P. A., & Bransgrove, J. (2007). Perceptions of witness credibility: Variations across age. *Journal of Applied Developmental Psychology, 28*, 318–331.

Nightingale, N. N. (1993). Juror reactions to child victim witnesses: Factors affecting trial outcome. *Law and Human Behavior, 17*, 679–694.

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect, 24*, 733–752.

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, see http://www.R-project.org.

Sternberg, K. J., Lamb, M. E., Esplin, P. W., Orbach, Y., & Mitchell, S. (2001). Use of a structured investigative protocol enhances young children's responses to free-recall prompts in the course of forensic interviews. *Journal of Applied Psychology, 86*, 997–1005.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S/S-Plus* (4th ed.). New York: Springer.

Winter, R. J., & Greene, E. (2007). Juror decision making. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandosky, & T. J. Perfect (Eds.), *Handbook of applied cognition* (2nd ed.). Chichester: Wiley.

Wright, D. B. (2006). Causal and associative hypotheses in psychology: Examples from eyewitness testimony research. *Psychology, Public Policy, and Law, 12*, 190–213.

Wright, D. B. (2007). The impact of eyewitness identifications from simultaneous and sequential lineups. *Memory, 15*, 746–754.

Wright, D. B., & Hall, M. (2007). How a 'Reasonable Doubt' instruction affects decisions of guilt. *Basic and Applied Social Psychology, 29*, 85–92.

Wright, D. B., & London, K. (2009). *Modern regression techniques using R*. London: Sage.

Received 31 October 2008; revised version received 2 December 2008

Appendix A

XXXX = age. Matthew was substituted for Mary for half of the participants (and son for daughter, his for her, and he for she).

Home/physical abuse description

John Rosenbury is on trial for physical abuse towards his wife Mary Rosenbury. Mary has decided to press charges against her husband although he denies the assault.

The prosecution is relying on the testimony of the couple's XXXX year old daughter, Mary, who is the only witness to the alleged assault. Mary says that her parents often shout at each other and that she is told to go to her room when this happens. However, one evening she says that she came out of her room because the screaming and shouting was scaring her. She says that when she went down the stairs she saw her father standing over her mother who was laying on the floor in the kitchen holding her arm. Mary's mother called to her to call the police when she saw her standing in the doorway which she went to do, but Mary says that her father took the phone off of her and dialled the ambulance himself. Mary said her father seemed angry and upset and had told her that her mother had fallen over during their discussion. Mary also stated that she noticed that there was water on the kitchen floor at the time of the incident.

Mr. Rosenbury denies any physical abuse towards his wife, claiming that her injuries were sustained by slipping on water that was on the floor and that she is blaming the accident on him in order to make him appear violent to gain custody of their daughter, Mary. The medical examination following the incident did not indicate whether the injury was acquired through physical abuse or was the result of a fall.

School/sexual abuse description

Mr. Perry is a school teacher on trial for sexually abusing 8-year-old Rebecca Jessops, a student who was in his care. The alleged incident occurred on 17 May, 1999 at 4 p.m. when the child was waiting for her mother to collect her from school. Laura Jessops, the child's mother filed a complaint against Mr. Perry for inappropriately touching her daughter after Rebecca told her that he touched her between her legs.

The prosecution is relying on the testimony of XXXX year old, Mary Rosenbury who is the only witness to the alleged assault. Mary says that on the 17 May, she had forgotten her homework when she left school and so returned to retrieve it. She returned to the school at 3.55 p.m. and says that the school was empty. She says that as she was walking through the corridor towards her classroom she heard a girl crying and looked into the classroom to see what was going on. She claims that she saw Rebecca pushing Mr. Perry away as he was touching her leg. Mary said that she heard Mr. Perry say to Rebecca that he was trying to help her and when asked by the prosecution if she saw Mr. Perry touch her between her legs she said no.

Mr. Perry denies any sexual abuse towards Rebecca Jessops, claiming that the child was distressed as her mother was late picking her up and so he was making efforts to comfort her. A medical examination of Rebecca after the incident was inconclusive of sexual contact.

Appendix B

Nagelkerke's R^2 is a measure of effect size for logistic regression (Nagelkerke, 1991). The equation for Nagelkerke's R^2 (1991) is:

$$\frac{1 - e^{-(L0-LM)/n}}{1 - e^{-L0/n}}$$

where *L0* is the deviance of the null model and *LM* is the deviance of the model. This is analogous to η^2 for ANOVAs. A partial Nagelkerke's R^2 is computed to show the effect of one variable after controlling for others, like η_p^2 for ANOVAs. For partial Nagelkerke's R^2 the deviance for the null model is replaced with the deviance for the model without the effect. A function for the package R was written that calculates both of these:

```
nagel <- function(model,base = 0){
  n <- length(model$residuals)
  if(is.numeric(base)) L0 <- model$null.deviance
  else L0 <- base$deviance
  LM <- model$deviance
  nagel <- (1 - exp((LM - L0)/n))/(1 - exp(-L0/n))
  print(paste("NR^2 = ", format(nagel,digits = 4)))
  return(nagel)}
```

When a generalized linear model (glm) is run in R a glm object is created. The `nagel` function calculates Nagelkerke's R^2 if a single glm object is entered (i.e., `nagel(glm1)`) and calculates partial Nagelkerke's R^2 if two glm objects are entered (i.e., `nagel(glm1,glm0)`). In this paper Nagelkerke's R^2 is abbreviated NR^2 and partial Nagelkerke's R^2 is abbreviated NR_p^2.