Children’s eyewitness identification performance: Effects of a Not Sure response option and accuracy motivation

Neil Brewer*, Amber Keast and James D. Sauer
Flinders University, Adelaide, South Australia, Australia

Purpose. Children who witness crimes are sometimes asked to view a photospread lineup to see if they can identify the culprit. Here, we examined the effectiveness of two manipulations designed to assist in overcoming the tendency of child eyewitnesses to choose from lineups, a tendency that manifests in false identifications from target-absent lineups.

Methods. Children aged around 12 years (N = 432) were randomly assigned to control, Not Sure, or accuracy motivation conditions in order to examine both target-absent and -present identification test performance across multiple sets of stimulus materials.

Results. The Not Sure option did not affect false alarms or hit rates, but the combination of accuracy motivating conditions and the Not Sure option reduced guessing, with overall rates of false identifications falling by 12.2%. The latter effect was much stronger under some stimulus conditions but not detectable under others, indicating that the manipulations could improve but not impair performance. While there were beneficial effects at the group level, the manipulation did not produce a general change in individual children’s decision strategies.

Conclusions. Eyewitness identification test manipulations that reliably reduce false alarms, without lowering hits, by child witnesses have not yet been identified. Here, we showed how a simple-to-implement accuracy motivation manipulation reduces false alarms under some conditions, while also highlighting the importance of evaluating the efficacy of interventions across varied encoding and test conditions. However, developing a procedure that can produce a generalized improvement in decision making at the individual level remains a major challenge for eyewitness researchers.

Some children who witness crimes are asked by police to view a photoarray or lineup to see if they can identify the offender, as illustrated by UK field data (Pike, Brace, & Kynan, 2002) and US case studies (Innocence Project, 2008) on the involvement of young people in police identification tests. A positive identification of a suspect by a child...
witness has the potential to influence police investigations and the outcomes of criminal proceedings. The fallibility of adult eyewitnesses exposed to identification tests is well documented (Brewer, Weber, & Semmler, 2005; Cutler & Penrod, 1995; Wells, Memon, & Penrod, 2006). There is also an extensive body of evidence highlighting problems with children’s identifications. In particular, children and young adolescents (e.g., ages spanning 9–13 years) are more likely than adults to make a positive identification (i.e., to choose) from a lineup, particularly when presented with a culprit- or target-absent lineup (e.g., Keast, Brewer, & Wells, 2007; Parker & Carranza, 1989; Parker & Ryan, 1993; Pozzulo & Lindsay, 1998). Positive identifications of innocent suspects are, in turn, known to be major factors underlying wrongful convictions (Innocence Project, 2008). Given the negative consequences that may arise from a witness’s identification of a suspect, it is not surprising that one focus of eyewitness identification research has been the examination of procedures designed to reduce the tendency of children to make positive identifications from culprit-absent lineups.

Researchers hoping to demonstrate the efficacy of interventions for improving children’s identification performance have to overcome a number of significant obstacles. First, the identification of interventions that reliably overcome the tendency by children to make positive identifications has thus far proven to be extremely elusive, suggesting the possibility that the only effective means of reducing children’s positive identifications from culprit-absent lineups might be via manipulations that encourage a stricter decision criterion and, in turn, produce significant reductions in hit rates for target-present lineups. Second, there are several methodological difficulties that are closely associated with working in the area of children’s identification decisions. Obtaining sufficiently large samples of child participants is often difficult for researchers in any area of psychological inquiry, but this difficulty is exacerbated when reliant on the categorical data and limited number of observations per participant (typically one) provided by the eyewitness identification paradigm. While it may be tempting to conclude that any significant results that emerge with small samples are indicative of a relatively powerful and reliable effect, Brewer et al. (2005, p. 182) provided a compelling demonstration of how the chi-squared statistics for identification decisions derived from quite small samples are frequently misleading with respect to the data patterns in the much larger sample from which they were drawn. A further difficulty that usually accompanies the combination of small samples and restricted number of data points per participant is that, in any individual study, there is usually minimal variability in the encoding and test stimuli and conditions, thereby severely limiting the extent to which conclusions can be generalized to other conditions (cf. Lindsay, Read, & Sharma, 1998; Sauerland & Sporer, 2009; Wells & Windschitl, 1999).

Given these methodological difficulties, any research involving child witnesses that demonstrates reliable effects of interventions across diverse stimulus materials and conditions is particularly valuable. Likewise, research that overcomes the significant methodological difficulties associated with children’s identification research, but fails to find convincing support for the efficacy of theoretically motivated interventions, still can provide a valuable guide for other researchers and practitioners about the likelihood of success when considering certain intervention practices with child witnesses. In this context, the present study examined two different approaches to modifying children’s identification test responding: (1) providing a specific option for withholding a response, namely a Not Sure response option. (2) Combining this withholding option with a battery of conditions designed to enhance children’s motivation to be accurate.
Previous research has explored a number of different approaches for dealing with children’s strong propensity to make a positive identification from a lineup. Some studies have included an additional response option (e.g., a Not Here card, a Mr Nobody card, or a silhouette lineup figure) that still allows children to point to a stimulus when rejecting a lineup (Beal, Schmitt, & Dekle, 1995; Davies, Tarrant, & Flin, 1989; Zajac & Karageorge, 2009), with the performance outcomes mixed. Beal et al. found no effect with 5- to 7-year-old children, whereas Davies et al. did with children up to 11 years. Zajac and Karageorge found a strong effect on target-absent performance, though it remains unclear whether children were responding to the silhouette because it resembled the target. Others have allowed trials or practice on mock lineups to prepare witnesses for possible lineup outcomes (Davies, Stevenson-Robb, & Flin, 1988; Parker & Ryan, 1993), also with mixed results. Davies et al. found no effect with 7 to 12-year-olds but Parker and Ryan did with children of similar age. Taking quite a different approach with children aged 10–14 years, Pozzulo and Lindsay (1999) used an elimination lineup which required the witness first to pick out the lineup member that looked most like the target and then to decide if it actually was the target. They reported fewer false identifications from target-absent lineups, without any deterioration in target-present performance.

Despite the occasional positive outcomes with some procedures, there are a number of features of previous studies that suggest that more research is needed to clarify the effectiveness of such interventions. For example, some studies have only used target-absent lineups so it is unclear if the manipulation induced a general criterion shift that would also have undermined target-present performance. Most studies have used only one set of stimulus materials, thereby raising queries about the generality of the findings across encoding and test conditions. Further, given the great difficulties experienced by most researchers in recruiting child participants, sample sizes in many studies have been small, making it difficult to assess the reliability of the findings.

In the present study, we accommodated these methodological concerns within an examination of the effectiveness of interventions motivated by theory and empirical work on the strategic control of memory reporting. In a programme of research focusing on memory reporting, not eyewitness identification, Koriat and Goldsmith (1994, 1996) have emphasized how the accuracy of a person’s memory reports is under strategic control. People reporting on the contents of their memory are potentially able to screen out (i.e., not report) incorrect answers, providing they have the option of free report. Consistent with this proposition, Koriat and Goldsmith found that giving participants the option of withholding answers led to improved accuracy, without affecting output quantity. Moreover, participants were able to further increase accuracy when this dimension of performance was emphasized.

These findings in the free report context underpin Koriat and Goldsmith’s (1996) theorizing on the role of metamemorial monitoring and control processes in the strategic regulation of memory reporting and their effects on memory outputs and accuracy. Specifically, Koriat and Goldsmith propose a monitoring mechanism, which is used to assess the likely accuracy of potential memory reports and a control mechanism that determines whether to volunteer the best candidate answer. The decision as to whether or not to volunteer a candidate answer depends on the assessed probability that the answer is correct, and situational demands that influence the setting of the response criterion. Effective control of reporting is then reflected in the extent to which the volunteering or withholding of answers is sensitive to the monitoring output.
Applying this theoretical perspective to the identification test context suggests the possibility of being able to enhance children’s identification accuracy. For example, the traditional lineup or identification test could be considered to be a forced selection test as the Not Present or Not There response option may still be interpreted as a response that is either correct or incorrect. In contrast, providing an additional and clear, no choice response option such as Don’t Know or Not Sure could be considered conceptually equivalent to the withholding of responses permitted in the free report paradigm. Translating the lineup procedure into something that more closely parallels a free report format by the inclusion of a Don’t Know or Not Sure response option may enhance children’s identification accuracy – assuming, of course, that they are able to monitor the likely (in)accuracy of their best candidate identification response. Similarly, the addition of accuracy emphasis via a combination of instructions and incentives emphasizing accuracy, in conjunction with a Don’t Know or Not Sure response option, may further enhance accuracy (again assuming effective monitoring of likely accuracy).

There are some empirical grounds for thinking that children, at least from 8 years up, may be able to monitor the accuracy of their memory reports. For example, when subjected to open-ended and unbiased questioning about a witnessed event, children aged 8 years and older (a) make confidence judgments that reliably discriminate correct from incorrect memory reports (Howie & Roebers, 2007; Roebers, 2002; Roebers & Howie, 2003) and (b) display a positive linear relationship between confidence judgments and the accuracy of memory reports (Howie & Roebers, 2007).

There are also, however, grounds for querying the sensitivity of children’s monitoring. Some studies have reported overconfidence in children’s memory reports (Allwood, Granhag, & Jonsson, 2006; Schneider & Bjorklund, 1998), especially when questioning is (mis)leading (Howie & Roebers, 2007), thereby indicating a monitoring problem. Keast et al.’s (2007) failure to detect meaningful confidence–accuracy relationships for 12-year-old children’s eyewitness identification performance suggests poor monitoring of identification response accuracy though they noted the possibility that children may actually monitor effectively but simply fail to withhold positive identification responses despite having assessed their memory as fragile. Such problems with the regulation of memory reporting are highlighted elsewhere by the finding that, while children may accurately judge which to-be-learned items are more difficult and, hence, require more study time, their allocations of study time do not align with their metacognitive assessments (Koriat, 2002; Schneider & Lockl, 2002). Should it be the regulation of reporting, rather than the monitoring of accuracy, that is the main problem for children in the identification context, manipulations that emphasize the desirability of withholding a response might well prove effective.

What then is known about children’s responsiveness to instructions or manipulations that emphasize withholding a response? A number of studies have shown that children as young as 4–5 years will withhold reporting items, using a Don’t Know response option, especially if the instructions emphasize why it is important to withhold responses that may be inaccurate (e.g., Howie & Dowd, 1996; Mulder & Vrij, 1996; Nesbitt & Markham, 1999; Roebers, Moga, & Schneider, 2001). Moreover, Roebers et al. (2001) found that overall accuracy for children aged 6–8 years was lowest when forced to answer questions, and highest when a free report option existed and there was an incentive for accuracy.

Taken together, these findings suggest that a combination of a strong accuracy motivation manipulation with an option to withhold responding such as a Don’t Know or Not Sure response option might be effective for improving children’s identification...
performance. This could occur by encouraging a stricter reporting criterion that should reduce positive identifications from target-absent lineups and, yet, perhaps not undermine target-present accuracy given that the target will, on average, be a relatively strong match to the child’s memory of the offender. Three previous studies (Dekle, Beal, Elliott, & Huneycutt, 1996; Memon & Rose, 2002; Pozzulo & Lindsay, 1997) have reported that child witnesses (aged 5–6, 8–9, and 10–14 years, respectively) seldom used a Don’t Know option. Pozzulo and Lindsay (1997) also found that extended instructions which highlighted the negative consequences of identifying the wrong person had no impact on target-absent performance of children aged 10–14 years, although a Don’t Know option was not provided in this condition. In sum, while there may be doubts about whether children’s identification performance will be modified by merely providing a response option that permits no choice, the combination of such an option and a strong accuracy motivation manipulation has not been examined. We elected not to include the accuracy motivation manipulation by itself (i.e., without a no choice option) as the opportunity to withhold a response appears to be crucial to screening out incorrect answers.

Accordingly, we contrasted children’s identification performance on both target-absent and -present lineups across three conditions: a control condition, a condition which allowed the witness to make no choice, and an accuracy motivation manipulation (combined with the no choice option). We elected to use Not Sure rather than Don’t Know as the no choice option because it caters for children who do not know the answer and those who might perceive pressure to pick but are clearly not sure. The accuracy motivation manipulation had two components in an attempt to boost the impact of the manipulation. First, children received instructions informing them of the importance of making a correct decision and the possible negative consequences of a false identification. Second, children received feedback for correct answers through a points game in which children received 10 points for a correct rejection/identification, 5 points for a Not Sure response, and 0 points for a false identification or incorrect rejection. The feedback was provided after each identification response. While in actual forensic contexts (a) the accuracy of a child’s identification would generally not be known and (b) attempting multiple identifications would not be common, training of children on practice lineups with feedback on performance certainly could be implemented if it was known to be effective.

We deliberately targeted only older children ($M = 12$ years 1 month) for several complementary reasons. Our access to child participants was such that we had to make a strategic decision either to have a limited number of participants at each of several age groups or a sufficient number at one age group to provide a decisive evaluation of the manipulations. We chose the latter strategy because (a) the limited evidence of successful interventions in previous identification studies mainly involves older children and (b) if the manipulations had no impact on older children’s performance, it seems unlikely that they would be effective with younger children. In other words, we focused on first testing the interventions with children who were most likely to benefit.

We also had children view four different stimulus events and make four lineup decisions to examine the generality of the findings across encoding and test conditions. Previous use of these same stimulus materials in our laboratory (and elsewhere) has shown that they produce quite different patterns of identification performance (i.e., choosing and identification decision accuracy), though these patterns are not obviously linked to stimulus characteristics such as exposure duration. The stimulus materials vary in a number of ways, including target features and/or distinctiveness, stimulus exposure...
duration, the closeness of the encoding-test stimulus match, the target-foil relationships, etc. In other words, the variations likely mirror some of the types of variations seen in the real world. As a number of researchers have argued recently (e.g., Lindsay et al., 1998; Sauerland & Sporer, 2009), assessing identification performance given the variability in encoding and test situations that is likely to pertain in real-world identification contexts is crucial for establishing reliable and, hence, generalizable phenomena and effects.

Importantly, this methodological approach permits an examination of identification performance at two different levels. By examining the impact of an intervention on identification decision accuracy of children across multiple stimulus conditions, it is possible to determine whether any beneficial effect is due to (a) the production of an improvement in decision making under specific stimulus conditions only, which will be reflected in variations in accuracy across stimulus materials or (b) the promotion of an improved decision-making strategy at the level of the individual child, reflected in a higher proportion of accurate decisions per individual across the four stimulus sets. Outcome (b) is clearly the preferred outcome in that it signals that the intervention is capable of effecting generalized improvements in children’s decision making in the identification context. But, while outcome (a) is not the preferred outcome, it is informative in that it indicates that the intervention is at least likely to be effective under some, but not all, conditions. Thus, if the intervention is not particularly demanding to implement, its routine use may be warranted.

Method

Participants and design
Four hundred and thirty-two children (205 males, 227 females) recruited from upper primary and lower middle school grades across the city of Adelaide, South Australia, participated. Ages ranged from 9 years 6 months to 14 years 1 month, but most children were within a tight age band ($M = 12$ years 1 month, $SD = 0.26$ years). Children were randomly assigned to one of three instruction conditions: control, Not Sure response option, or accuracy motivation (combined with the Not Sure response option). Neither the proportion of males and females nor the children’s ages differed significantly between conditions. All participants viewed four simulated crimes and attempted identifications from two target-present and two target-absent lineups.

Materials
The four video clips of simulated crimes varied in terms of nature of the event, exposure duration, just as happens in real crimes. The events were:

**Stimulus A**
A thief stealing a video player from a house (duration = 30 s). A young man approached the front door of a house and tried to open it. Upon finding the door locked, he moved to the large window next to the front door, removed the screen from the window, opened the window and entered the house through the window, exiting from view. After several seconds, he emerged from the window, carrying a video player. He left the premises through the front gate, still carrying the video player.
**Stimulus B**

A thief trying to break into a car (duration = 16 s). A young man jiggled with the front driver's side door of a car parked in front of a house while looking around suspiciously. After a few seconds, he managed to open the door. Then, when someone from the house yelled out, he looked up and then ran off down the street.

**Stimulus C**

A thief stealing a credit card from a restaurant (duration = 115 s). A young man entered a restaurant and waited while a customer placed a credit card on a counter for a waiter to process. After the customer departed, leaving the card with the waiter, the thief engaged the waiter in conversation about a reservation and, subsequently, stole the credit card when the waiter was distracted by the telephone.

**Stimulus D**

Thief breaking open the door of a warehouse (duration = 105 s). A car drove up to a warehouse car park and reversed into a park. Three men got out of the car and looked around the outside of the warehouse. One of the men (the focus of the camera) then used a screwdriver to open the warehouse door, looking around suspiciously as he did so. After opening the door, the target and one accomplice entered the warehouse and removed a number of boxes which they loaded into the car. After doing this twice more, a police car arrived.

**The lineups**

The lineups for all four targets consisted of eight colour photos (4 × 5.75 cm), arranged in two rows of four and shown simultaneously on a 17-in. computer screen (resolution of 1024 × 768 pixels). All foils, and each target's replacement in the relevant target-absent lineup, were selected from photo pools using the match-description strategy recommended for lineup composition (Technical Working Group for Eyewitness Evidence, 1999; Wells, 1993). All photos provided a front view from the chest up. Below each photo was a number (1–8), and at the lower centre of the screen was a button marked *Not There*. For the *Not Sure* and the accuracy motivation conditions, there was also a *Not Sure* button located at the bottom of the screen to the left of the *Not There* button. The original photos were scanned with a resolution of 100 × 144 pixels. All targets wore different clothing to that seen in the video. The target and the target's replacement (in the target-absent lineup) were randomly assigned among positions 1–8 for each participant. An identification response required a mouse-click on one of the photos, on the *Not Present* button, or (depending on the condition) on the *Not Sure* button.

**Experimental manipulations**

In all conditions, children were instructed, prior to each lineup, that the thief may or may not be present in the lineup. Children in the *Not Sure* condition were told ‘If you DON'T KNOW or are NOT SURE if the thief is in the lineup (row of faces) that is OK. You can click on the “Not Sure” button. It is better to click on the “Not Sure” button than to guess’.
Children in the *Accuracy Motivation* condition received the *Not Sure* response option and instructions, and additional manipulations designed to increase their motivation to be accurate. First, they were told ‘Making a correct decision is really important because if the wrong person is picked out he might get into trouble and even go to jail for something that he didn’t do’. They were also encouraged to be accurate through the use of a points game. The points game was explained to the children as follows:

Let’s see how good you are as an eyewitness. If the thief is there and you correctly pick him out you will get 10 points. If the thief is NOT there and you click ‘Not there’ you will get 10 points. If you click ‘Not Sure’ you will get 5 points. BUT, if you make a mistake and pick the wrong face or click ‘Not There’ when the thief is there, you will get 0 points.

Children in this condition received feedback on each lineup decision via the computer after making their decision. The feedback included information on both the accuracy and number of points obtained for the most recent identification decision, as well as a tally of their total points so far. Children in the control condition did not have a *Not Sure* response option or any accuracy motivation manipulations. As extra instructions were given to those in the accuracy motivation condition, we included filler instructions (e.g., about working quietly), matched for the number of words, in both the control and *Not Sure* option conditions so that children in all conditions were required to read the same number of words between seeing the video clip and being presented with the lineup.

**Procedure**

Children took part in the experiment in groups of 6–25 (depending on the number of children per class who returned the parental consent form and on the available computer resources), but each child was randomly assigned to the experimental conditions. Children were supervised throughout by the experimenter, a research assistant, and the children’s class teacher or school principal. They ensured that children worked individually, did not talk, or engage in other distracting behaviours.

The experiment was conducted on individual PCs. The initial screen asked the children to ‘Please sit quietly and wait for your instructions’. Once all children were seated the experimenter provided brief verbal instructions about the nature of the task and the need for the children to work independently. Children were then directed to click on the ‘Next’ button to begin the experiment. After viewing the first video clip, children received the appropriate instructions and information for their condition and were then presented with the corresponding lineup. This procedure was repeated for the remaining three video clips. Each child viewed two target-present and two target-absent lineups, beginning with a target-present lineup and alternating between target-present and -absent. The order in which the videos and corresponding lineups were presented was counterbalanced, resulting in 24 different video combinations. Given that half of the lineups for each video were target absent and the other half target present, there were 48 different video/lineup combinations for each condition with three children per condition run in each of these 48 combinations.

**Results**

We first examined whether the manipulations improved accuracy under specific stimulus conditions, thereby improving performance at the group level. As indicated
previously, this is not the most sought after outcome, but it at least indicates whether an intervention is likely to be effective under some, even if not all, conditions.

Table 1 shows the breakdown of identification responses combined across all stimulus sets for target-absent and present lineups. This overall pattern is of interest from the perspective of the likely generality of the findings, given that real-world encoding and test conditions are likely to vary widely. Given the lack of independence of observations for the four stimulus sets, a hierarchical log-linear analysis is not appropriate. Instead, we present a descriptive analysis of the combined data, followed by inferential statistics on data for the individual stimulus sets.

Table 1. Overall frequencies and percentages of identification responses across all stimulus sets for each instruction type

<table>
<thead>
<tr>
<th>Instruction type</th>
<th>Control</th>
<th>Not Sure</th>
<th>Accuracy motivation</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Target absent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False ID</td>
<td>146</td>
<td>50.7</td>
<td>151</td>
<td>52.4</td>
</tr>
<tr>
<td>Not Present</td>
<td>142</td>
<td>49.3</td>
<td>91</td>
<td>31.6</td>
</tr>
<tr>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>46</td>
<td>16.0</td>
</tr>
<tr>
<td>Target present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct ID</td>
<td>111</td>
<td>38.5</td>
<td>115</td>
<td>39.9</td>
</tr>
<tr>
<td>Foil ID</td>
<td>89</td>
<td>30.9</td>
<td>77</td>
<td>26.7</td>
</tr>
<tr>
<td>Not Present</td>
<td>88</td>
<td>30.6</td>
<td>61</td>
<td>21.2</td>
</tr>
<tr>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>35</td>
<td>12.2</td>
</tr>
</tbody>
</table>

For target-absent lineups, the accuracy motivation condition produced an overall reduction in both false identifications (12.2%) and correct rejections (9.4%). In the Not Sure condition the effect was constrained to a reduction in correct rejections (17.7%). Use of the Not Sure option when presented with target-present lineups primarily reflected a decline in Not Present responses (9.4%) in the Not Sure condition. In contrast, in the accuracy motivation condition, the use of Not Sure responses meant fewer foil (9.7%) and, to a lesser extent, correct identifications (5.5%). In sum, the Not Sure condition appeared merely to substitute Not Sure responses for lineup rejections. The accuracy motivation manipulation, however, appeared to reduce the tendency to guess at a positive identification. Given that this latter trend prevailed across multiple sets of stimuli, and is firmly based on data from a large sample, it is encouraging even if not a particularly strong pattern. In other words, when examined across stimuli, the accuracy motivation manipulation – a relatively simple procedure – reduced guessing.

While the overall performance changes provide the most relevant information about the to-be-expected overall, or average, impact of the manipulations, an examination of

---

1 Overall, there was a slight trend for choosing from target-absent lineups to increase as children proceeded through the lineups, but this was not affected by instruction type.
performance for each of the various sets of stimulus materials provides a useful
guide as to the possible upper and lower boundaries for performance improvement.
Tables 2 (target absent) and 3 (target present) show the identification response patterns
for each of the four sets of stimulus materials.

Table 2. Frequencies and percentages of identification responses from target-absent lineups for each stimulus set and instruction type

<table>
<thead>
<tr>
<th>Instruction type</th>
<th>Control</th>
<th>Not Sure</th>
<th>Accuracy motivation</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Stimulus A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False ID</td>
<td>28</td>
<td>38.9</td>
<td>38</td>
<td>52.8</td>
</tr>
<tr>
<td>Not Present</td>
<td>44</td>
<td>61.1</td>
<td>27</td>
<td>37.5</td>
</tr>
<tr>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>7</td>
<td>9.7</td>
</tr>
<tr>
<td>Stimulus B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False ID</td>
<td>52</td>
<td>72.2</td>
<td>47</td>
<td>65.3</td>
</tr>
<tr>
<td>Not Present</td>
<td>20</td>
<td>27.8</td>
<td>12</td>
<td>16.7</td>
</tr>
<tr>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>13</td>
<td>18.1</td>
</tr>
<tr>
<td>Stimulus C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False ID</td>
<td>16</td>
<td>22.2</td>
<td>29</td>
<td>40.3</td>
</tr>
<tr>
<td>Not Present</td>
<td>56</td>
<td>77.8</td>
<td>35</td>
<td>48.6</td>
</tr>
<tr>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>8</td>
<td>11.1</td>
</tr>
<tr>
<td>Stimulus D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False ID</td>
<td>50</td>
<td>69.4</td>
<td>37</td>
<td>51.4</td>
</tr>
<tr>
<td>Not Present</td>
<td>22</td>
<td>30.6</td>
<td>17</td>
<td>23.6</td>
</tr>
<tr>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>18</td>
<td>25.0</td>
</tr>
</tbody>
</table>

For target-absent lineups, the most clear-cut observations to be made from
inspection of the data patterns are (1) unlike previous research demonstrating a
reluctance of children to use a Don’t Know option, children used the Not Sure option
for all stimuli. (2) A fall in false identifications (compared with the control condition)
produced by the Not Sure manipulation for two stimuli (B and D) was offset by an
increase for the other two stimuli (A and C). (3) A substantial fall in false identifications
under the accuracy motivation manipulation for two (B and D) of the four stimuli. For
each stimulus set, a 3 (instruction type) × 2 (response: false ID vs. no ID) chi-squared
probed the impact of instruction condition. Significant $\chi^2(2, N = 216) = 11.34,
p < .01, w = 0.23,$ and $9.69, p < .01, w = 0.21,$ for stimuli B and D, respectively,
reflected meaningful reductions (26.4 and 25.0%) in false identifications produced by
the accuracy motivation manipulation and, to a lesser extent, by the Not Sure
manipulation. The average decrease in false identifications across stimuli was 12.2% for
the accuracy motivation condition, whereas false identifications rose by an average
of 2.0% for the Not Sure condition.

For target-present lineups, clear patterns are more difficult to detect, with the
following the more apparent: (1) children again used the Not Sure option for all stimuli.
(2) Reduced guessing produced by the Not Sure manipulation is suggested by fewer
correct identifications for one stimulus (D) and reduced foil identifications for another (B),
although more correct identifications were evident for another target (C). (3) Reduced guessing produced by the accuracy motivation manipulation is suggested by fewer correct identifications for two stimuli (C and D) and fewer foil identifications for two stimuli (B and D). For target-present lineups, a 3 (instruction type) × 3 (response: correct ID vs. foil ID vs. no ID) chi-squared detected only two significant effects for (a) stimulus B, \( \chi^2(4, N = 216) = 10.16, p < .05, \omega = 0.22 \), reflecting fewer foil identifications in the accuracy motivation condition especially and (b) stimulus C, \( \chi^2(4, N = 216) = 12.06, p < .05, \omega = 0.24 \), reflecting the discrepancy between correct identifications for the Not Sure and accuracy motivation conditions. The average decrease in correct identifications across stimuli was 5.6% for the accuracy motivation condition, whereas correct identifications rose by an average of 1.2% for the Not Sure condition.

Table 3. Frequencies and percentages of identification responses from target-present lineups for each stimulus set and instruction type

<table>
<thead>
<tr>
<th>Instruction type</th>
<th>Response</th>
<th>Control N</th>
<th>Control %</th>
<th>Not Sure N</th>
<th>Not Sure %</th>
<th>Accuracy motivation N</th>
<th>Accuracy motivation %</th>
<th>Overall N</th>
<th>Overall %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus A</td>
<td>Correct ID</td>
<td>41</td>
<td>56.9</td>
<td>39</td>
<td>54.2</td>
<td>39</td>
<td>54.2</td>
<td>119</td>
<td>55.1</td>
</tr>
<tr>
<td></td>
<td>Foil ID</td>
<td>12</td>
<td>16.7</td>
<td>13</td>
<td>18.1</td>
<td>10</td>
<td>13.9</td>
<td>35</td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td>Not Present</td>
<td>19</td>
<td>26.4</td>
<td>12</td>
<td>16.7</td>
<td>13</td>
<td>18.1</td>
<td>44</td>
<td>20.4</td>
</tr>
<tr>
<td></td>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>8</td>
<td>11.1</td>
<td>10</td>
<td>13.9</td>
<td>18</td>
<td>8.3</td>
</tr>
<tr>
<td>Stimulus B</td>
<td>Correct ID</td>
<td>38</td>
<td>52.8</td>
<td>39</td>
<td>54.2</td>
<td>36</td>
<td>50.0</td>
<td>113</td>
<td>52.3</td>
</tr>
<tr>
<td></td>
<td>Foil ID</td>
<td>24</td>
<td>33.3</td>
<td>17</td>
<td>23.6</td>
<td>12</td>
<td>16.7</td>
<td>53</td>
<td>24.5</td>
</tr>
<tr>
<td></td>
<td>Not Present</td>
<td>10</td>
<td>13.9</td>
<td>8</td>
<td>11.1</td>
<td>11</td>
<td>15.3</td>
<td>29</td>
<td>13.4</td>
</tr>
<tr>
<td></td>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>8</td>
<td>11.1</td>
<td>13</td>
<td>18.1</td>
<td>21</td>
<td>9.7</td>
</tr>
<tr>
<td>Stimulus C</td>
<td>Correct ID</td>
<td>21</td>
<td>29.2</td>
<td>32</td>
<td>44.4</td>
<td>14</td>
<td>19.4</td>
<td>67</td>
<td>31.0</td>
</tr>
<tr>
<td></td>
<td>Foil ID</td>
<td>16</td>
<td>22.2</td>
<td>12</td>
<td>16.7</td>
<td>13</td>
<td>18.1</td>
<td>41</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>Not Present</td>
<td>35</td>
<td>48.6</td>
<td>24</td>
<td>33.3</td>
<td>43</td>
<td>59.7</td>
<td>102</td>
<td>47.2</td>
</tr>
<tr>
<td></td>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>4</td>
<td>5.6</td>
<td>2</td>
<td>2.8</td>
<td>6</td>
<td>2.8</td>
</tr>
<tr>
<td>Stimulus D</td>
<td>Correct ID</td>
<td>11</td>
<td>15.3</td>
<td>5</td>
<td>6.9</td>
<td>6</td>
<td>8.3</td>
<td>22</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>Foil ID</td>
<td>37</td>
<td>51.4</td>
<td>35</td>
<td>48.6</td>
<td>26</td>
<td>36.1</td>
<td>98</td>
<td>45.4</td>
</tr>
<tr>
<td></td>
<td>Not Present</td>
<td>24</td>
<td>33.3</td>
<td>17</td>
<td>23.6</td>
<td>22</td>
<td>30.6</td>
<td>63</td>
<td>29.2</td>
</tr>
<tr>
<td></td>
<td>Not Sure</td>
<td>–</td>
<td>–</td>
<td>15</td>
<td>20.8</td>
<td>18</td>
<td>25.0</td>
<td>33</td>
<td>15.3</td>
</tr>
</tbody>
</table>

Clearly, the most compelling observation from the analysis of data for the individual stimulus sets is that, under some (as yet unspecified) conditions, accuracy motivation conditions can have a meaningful impact on what has been generally considered to be the most problematic aspect of children's identification test performance (i.e., choosing from target-absent lineups). Under other conditions, there is neither positive impact; nor is there a negative impact.

Next, we examined the impact of the manipulations on identification decision accuracy across all stimulus conditions to determine if either manipulation promoted an
improved decision-making strategy at the level of the individual child that would translate into an overall improvement in accuracy at the individual level. As noted earlier, this would be a more impressive finding than the group-level findings described above because it would signal an intervention capable of producing reliable improvement across conditions, something that (to our knowledge) has not been reported in the child identification literature to date.

Correct decisions comprised correct positive identifications and lineup rejections; incorrect decisions were positive identifications of innocent foils and incorrect lineup rejections (see Table 4 for data summary). A 3 (instruction type) × 2 (target presence) analysis of variance (ANOVA) on number of correct decisions made by each participant did not detect a significant effect of instruction type, $F(2, 411) = 0.24, ns, f = 0.02$. On average, individual children made few accurate decisions (out of 4 decisions made, the mean number correct was less than 0.5) and neither intervention improved overall decision accuracy at the level of the individual. In other words, while the accuracy motivation manipulation improved performance at the group level under some conditions, neither manipulation produced an improvement in decision-making performance at the individual level.

### Table 4. Number correct identification decisions by instruction type and target presence

<table>
<thead>
<tr>
<th>Instruction type</th>
<th>Target presence</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present Memory</td>
<td>Absent Memory</td>
<td>Overall</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Control</td>
<td>0.39</td>
<td>0.34</td>
<td>0.49</td>
<td>0.34</td>
<td>0.44</td>
</tr>
<tr>
<td>Not Sure</td>
<td>0.45</td>
<td>0.41</td>
<td>0.39</td>
<td>0.41</td>
<td>0.42</td>
</tr>
<tr>
<td>Accuracy motivation</td>
<td>0.40</td>
<td>0.39</td>
<td>0.49</td>
<td>0.43</td>
<td>0.44</td>
</tr>
</tbody>
</table>

**Discussion**

to date, attempts to improve children’s eyewitness identification performance and, specifically, to restrict their tendency to choose from target-absent lineups, have met with limited success. We are not aware of any studies that have demonstrated generalized improvements in children’s decision making in the identification context. We are also not aware of studies that have tested the efficacy of interventions across multiple stimuli and hence are able to assert with any confidence that an intervention is effective (and not detrimental) across a number of stimulus conditions.

Here, we examined the effectiveness of manipulations motivated by theorizing on the strategic control of memory reporting, theorizing which suggests that the availability of a clear option of withholding a response, coupled with conditions designed to motivate accuracy, may assist in overcoming children’s tendency to choose from lineups. In so doing, we also addressed several methodological issues that have hampered many previous research efforts in this area. Specifically, we examined both

\[ \text{The target presence main effect was not significant. A weak interaction effect was detected, } F(2, 411) = 4.02, p < .05, f = 0.10, \text{ but the only difference detected by simple effects analyses (Bonferroni correction) was between target-present and -absent accuracy for the control condition.} \]
target-absent and -present performance across multiple sets of stimulus materials, with a relatively large sample size. Where we have been unable to demonstrate an efficacious intervention, our sample size does at least allow future researchers to make a reasonable assessment of the likelihood that a similar intervention will be successful.

The key findings are as follows. First, the absence of any discernible positive trends with our large sample provides a strong pointer that use of the Not Sure format, by itself, for encouraging children to withhold a response (when in doubt) is likely to be ineffective in enhancing identification accuracy. In contrast to previous research which found that children seldom used a Don't Know option (Dekle et al., 1996; Memon & Rose, 2002; Pozzulo & Lindsay, 1997), children in both the Not Sure and accuracy motivation conditions used the Not Sure response option for approximately 16–22% of target-absent lineups, and 12–15% of target-present lineups. This suggests that Not Sure may be a more attractive response option for children than a Don't Know option, although it is of course possible that the computerized presentation of the lineup and associated instructions ensures that such response options, regardless of their exact wording, are more salient and, hence, used. However, while children used the Not Sure option, this option (by itself) produced no overall drop in false identifications from target-absent lineups. Rather, it reduced (correct) lineup rejections.

Second, the examination of individual children’s decision accuracy across all four sets of stimuli revealed no significant differences between experimental conditions. In other words, neither the Not Sure nor the accuracy motivation manipulation promoted any change in decision strategy at the individual level that allowed children to perform consistently more effectively across the different stimulus conditions. Thus, a key issue for future research will be to continue the search for identification procedures that produce a general change in children’s decision-making strategies in the identification task. As foreshadowed earlier, it may well turn out to be the case that the only way of reducing false alarms may be to effect a conservative criterion shift that also reduces hits.

Although discovering that either of these manipulations could effect a general change in individual children’s decision making would have represented an impressive breakthrough in dealing with a hitherto intractable problem, the absence of such an effect does not mean that the manipulations tested here should be dismissed as possible interventions for child witnesses. Across stimuli, the combination of the accuracy motivation conditions and the Not Sure response option appeared to reduce guessing, with overall rates of false identifications falling by 12.2%. Importantly, for two of the four stimuli, the drop in false identifications was much larger (> 25%), with only minor falls in hit rates.

One possible interpretation of these patterns is that their unevenness – presumably reflecting the complex variations in encoding and test conditions across stimuli – make them difficult to interpret. But, it is exactly these same variations that will characterize real-world encoding and test conditions, variations that only the use of multiple stimuli will allow researchers to detect. Thus, our data suggest that an accuracy motivation manipulation such as that used here (a) is unlikely to impair children’s identification performance and (b) will, under some stimulus conditions, meaningfully reduce the likelihood of false alarms. Given the minimalist nature of the intervention, it would be relatively easy to incorporate in identification procedures with child witnesses.

Of course, important questions remain for future research. Given that we cannot control real-world encoding conditions or even (given current knowledge) many lineup conditions (e.g., the most appropriate collection of foils), and that these conditions are
likely to vary widely, the impact of an intervention measured over a range of conditions is clearly of paramount interest. Nevertheless, knowing when the intervention is likely to work best would still prove useful in refining our understanding of the underlying mechanism(s). As we discussed earlier, there are many variables (e.g., encoding conditions, target features/distinctiveness, encoding-test stimulus match, lineup foil characteristics) that are likely to interact in shaping identification performance. Unfortunately, our current state of knowledge simply does not allow us to pinpoint any individual variable as the locus of any performance differences between different stimulus sets. Clearly, length of exposure to the target was not the critical factor as the accuracy motivation manipulation produced major improvements at both the long and short exposures (stimuli B and D) as well as no change at similar exposure durations (A and C). At present, we know that the encoding and test conditions associated with the largest improvement produced by the accuracy motivation condition were those (for whatever reason) associated with relatively high rates of false alarms, although care needs to be taken when drawing inferences from only four stimulus sets. Indeed, perhaps the false alarm rates for the two stimulus sets that showed little or no improvement were just too close to the floor to provide any realistic chance of improvement. Thus, it is possible that the overall impact detected here underestimates likely real-world effects where encoding conditions may be worse, retention intervals longer, and so on.

Further research could also illuminate the key components of our accuracy motivation intervention. Given the difficulty researchers have had improving children’s performance, we designed our intervention hoping to produce maximum impact. Koriat and Goldsmith’s (1994, 1996) theory suggested that a combination of accuracy motivation conditions (i.e., instructions and feedback/incentives) and the Not Sure would be desirable, and that it is unlikely that the accuracy motivating conditions would be effective in the absence of a clear option of withholding a response.

Future work may, however, reveal that the feedback/incentive component of the manipulation is unnecessary, as is tentatively suggested by a post hoc examination of our data. We examined whether decision accuracy improved across lineups, reflecting the feedback/incentive taking hold. As (i) all participants encountered lineups in the order of target-present, -absent, -present and -absent and (ii) accuracy differences between target-present and -absent conditions reflecting things like target distinctiveness, idiosyncrasies of lineup composition (including the similarity of the target and the target’s replacement to the foils) and so on are confounded with lineup order, our analysis of this issue used only two levels of order: order 1 included the first target-present and -absent lineups, order 2 included the second. A 3 (instruction type) × 2 (target presence) × 2 (order) ANOVA on number of correct decisions revealed a significant effect of order only, but the effect size was below the cutoff for small, \(F(1, 287) = 5.57, p < .05, f = 0.05\). Further, decision accuracy was lower for the second two lineups than for the first two \((M = 0.46, SD = 0.38) vs. M = 0.42, SD = 0.37\), with the accuracy motivation data patterns virtually identical to the overall. Thus, the feedback/incentive component of the accuracy motivation manipulation clearly did not improve decision accuracy at the individual level, suggesting that the accuracy motivation instructions alone may be all that is necessary. Of course, it remains possible that some form of pre-identification test training using feedback and a tangible reinforcer could improve performance.

To ensure that our contrasts were not underpowered, as is often the case in children’s identification studies, we only sampled children at the upper level of the
children’s age range where previous research has suggested at least some grounds for optimism about the possibility of effective intervention. Whether these manipulations would be more or less effective with younger children remains an empirical question. For example, older children may be more likely to benefit if they tend to be overly confident in their ability to make a correct decision (cf. Keast et al., 2007). Conversely, younger children may benefit more if the manipulation reduces their vulnerability to influence by authority figures. Answering additional questions such as these will require a huge research investment to secure sufficient samples, exposure to multiple stimuli, and so on. One of the other practical constraints associated with ensuring we had a sufficiently large sample was that the retention interval between the crime and the lineup was very short, unlike what might normally apply in actual cases. Whether our manipulations would exert a greater (or lesser) influence as memory strength faded over a long retention interval is difficult to predict.

Finally, an important part of our protocol involved obtaining multiple identifications from each participant. This is an unusual procedure in this area of research and we are unsure as to exactly how it may affect children’s responding. At first glance, the varied data patterns obtained for the different stimuli might seem to call into question this approach. But in fact, given that conditions outside the laboratory will vary in all sorts of (often unpredictable) ways, this diversity of materials and data provides a very important message for eyewitness researchers. Data from any one or two of these stimulus sets would have suggested quite different conclusions about the effectiveness of the manipulations used in this research. For example, using just stimulus set B or D, or indeed both of them, would have produced what would appear to be powerful (and perhaps eminently publishable) evidence for the conclusion that accuracy motivation will produce generalized reductions in false identifications. Only later would other researchers discover that generalized improvements are not inevitable.

Despite the limitations of this research, it is clear that, at least under some encoding and test conditions, meaningful improvements in children’s identification performance can sometimes be produced by accuracy motivation conditions that are easy to implement. Clearly, it would represent a major advance in this area if researchers could also find ways of effecting a global change in children’s decision-making strategies on the identification task.

Acknowledgements
This research was supported by grants from the Australian Research Council.

References


Received 18 June 2009; revised version received 14 August 2009