“Look this way”: Using gaze maintenance to facilitate the detection of children’s false reports

Hannah Lawrence, Lucy Akehurst, Amy-May Leach, Julie Cherryman, Aldert Vrij, Megan Arathoon & Zara Vernham

Abstract

In two experiments, we investigated whether imposing a secondary task is an effective technique for detecting child deceit. First, 85 children aged 8 to 11 years old provided either a true or false report of a recent school event. At interview, some children were asked to gaze towards either the interviewer's face (IF) or a teddy bear's face (TF), whereas some children were given no gaze instruction. In both the IF and TF conditions, lie-tellers provided significantly fewer details than truth-tellers. A total of 192 adult evaluators then judged the credibility of 10 children's reports from one of the three ‘gaze’ conditions with and without guidance on level of detail. Evaluators discriminated truths from lies successfully when judging children instructed to look at IF, but not when children were asked to gaze towards TF. Evaluators who received guidance demonstrated better discrimination between true and false reports than evaluators who received no such information.

Child deception research has focused on both the developmental origins of children's lie-telling behaviours, and the forensic implications of deceptive child testimonies going undetected (see Talwar & Crossman, 2012 for a review). Past research has painted a bleak picture: Children not only have the potential to lie in forensic interviews (Tye, Amato, Honts, Devitt & Peters, 1999), but, when the video-recordings of their statements are presented to legal professionals (e.g. police officers, judges), they experience great difficulty in uncovering false testimonies (Bala, Ramakrishan, Lindsay & Lee, 2014; Leach, Talwar, Lee, Bala & Lindsay, 2004). Thus, if children do decide to provide deceptive reports, then they could easily slip through the net resulting in miscarriages of justice that are damaging to both the victims and defendants (O'Donohue, Benuto & Fanetti, 2010). Clearly, more effective deception detection strategies are needed.

Cognitive processing is an important factor in deception (Zuckerman, DePaulo & Rosenthal, 1981), particularly for children whose growing cognitive abilities are closely related to their ability to maintain false reports (Talwar & Crossman, 2011). Indeed, children's development of global executive functioning (Gordon, Lyon & Lee, 2014) as well as their development of specific executive functions, such as inhibitory control, working memory, executive planning and forward search planning, significantly contribute to their ability to conceal incriminating information when questioned (Alloway, McCallum, Alloway & Hoicka, 2015; Evans & Lee, 2011; Talwar & Lee, 2008; Williams, Leduc, Crossman & Talwar, 2016). Furthermore, lie-telling proficiency follows the developmental patterns of cognitive processes, such as inhibitory control (Debey, De Schryver, Logen, Suchotzki & Verschuere, 2015). This suggests that child lie-tellers, who are still developing certain

cognitive skills that might facilitate their lie-telling, might be affected by any interview technique that impacts upon these skills.

Growing research into adult deception has highlighted cognitive lie detection (CLD) as a promising strategic tool. Based on the well-established premise that lying is more cognitively demanding than truth-telling (e.g. Christ, Essen, Watson, Brubaker & McDermott, 2009; Hartwig, Granhag, Strömwall & Kronkvist, 2006; Mann & Vrij, 2006), CLD manipulates cognitive load, which refers to information-processing demands (associated with attentional and working memory) (Block, Hancock & Zakay, 2010), and transforms it into a system variable (Vrij, 2015). As a result, CLD techniques exaggerate behavioural differences between truth-tellers and lie-tellers, ultimately leading to impressive improvements in correct judgements of truths (57% for standard approach to 67% for CLD approach) and correct judgments of lies (47% for standard approach, 67% for CLD) (Vrij, Fisher & Blank, 2015). By taxing these cognitive load further, CLD decreases lie-telling performance. Children should be particularly susceptible to the negative effects of increased cognitive demand because their developing cognitive abilities, which already reveal their deceit, would be put under further strain.

Imposing cognitive load

Imposing cognitive load transforms the cognitive demand experienced by interviewees into a system variable through the addition of a secondary task (Vrij, 2015). Knowles (1963) proposed that each person has a limited pool of attentional resources that are differentially allocated to tasks according to difficulty. A difficult task, such as lie-telling, would draw more resources from this pool than a less difficult task, such as truth-telling. Lie-tellers would, therefore, have fewer resources (than truth-tellers) remaining if the pool were finite.

This asymmetry in the availability of cognitive resources for truth-tellers and lie-tellers has two consequences for lie-tellers when a secondary task is imposed. First, lie-tellers experience an overall increase in cognitive demand, working at or near to full attentional capacity. This means that lie-tellers exhibit more behavioural cues indicative of cognitive load compared to truth-tellers. Second, interference between the tasks may arise. When working at cognitive capacity, performance will depend on a person's ability to divide his or her attention in accordance with task demands. Attention can be flexibly allocated from moment to moment (Kahneman, 1973): As the secondary task becomes more difficult, additional resources can be allocated. If the tasks share a particular pool of resources, then diverting resources from the primary task to the secondary task should result in a trade-off (i.e. decreasing performance for the primary task and increasing performance for the secondary task).

Imposing cognitive load in order to detect deception could be particularly effective with a younger population whose ability to manage their attentional resources has not yet fully matured. Before the age of 11 years, children find it difficult to differentially allocate their attention in dual-task processing (Irwin-Chase & Burns, 2000). Furthermore, research has shown that the increase in cognitive load experienced, when moving from single tasks to dual-tasks, is greater for children than it is for adults (Karatekin, 2004). Although 10-year-olds can allocate their attention similarly to adults, their control over attention management in response to task difficulty is not yet fully developed. In the context of the current study, this suggests that child lie-tellers may overcompensate for the rising demands of a secondary task, diverting too many resources away from the primary task of lie-telling,, thus decreasing their performance on this task. It is also possible that children may prioritise the primary task, sacrificing their performance on the secondary task.

To date, two studies have examined the effects of cognitive lie detection techniques on children. Firstly, Liu et al. (2010) asked unanticipated questions of children aged 10 to 12 years old about a non-experienced life event. They found that, compared to truth-tellers, child lie-tellers were more likely to respond to unexpected questions. Secondly, Saykaly, Crossman, Morris and Talwar (2016) imposed cognitive load by asking children to falsely allege or deny play with a certain toy using the 'reverse order' interview instruction. Their results revealed that reverse order recall made it harder for child lie-tellers to maintain their reports compared to child truth-tellers, suggesting that telling a story backwards does increase cognitive demands. In summary, both these studies indicated that, when children have to perform a secondary task (i.e. answering a difficult question) at the same time as maintaining their false reports, their ability to maintain the lie is negatively affected. In the current experiment, the secondary task, introduced at interview, was an instruction to maintain gaze with either the interviewer's face or a teddy bear's face. A secondary task that has yet to be investigated with children.

Gaze maintenance

Using a systematic approach, Glenberg, Schröder and Roberston (1998) demonstrated that as the cognitive demands (i.e. cognitive difficulty) of a task increase, adults naturally avert their gaze. This cognitive strategy of gaze aversion is functional, as adults performed better on moderately difficult questions when they disengaged from (i.e. closed their eyes), rather than engaged with (i.e. looked at the interviewer's nose), disruptive visual components in their environment. Looking towards a visual/social stimulus, therefore, interfered with their task performance when the cognitive demands of the task were moderate. This behavioural response to avoid cognitive overload has also been investigated with children. Doherty-Sneddon, Bruce, Bonner, Longbotham and Doyle (2002) compared gaze aversion behaviour in children aged 5 and 8 years old in response to easy (low cognitive load) and

difficult (high cognitive load) questions. Results revealed that the older children averted their gaze away from the questioner's face more frequently in response to rising question difficulty (i.e. cognitive effort), but that this gaze pattern was only observed for younger children and for certain types of questions. This suggests that gaze aversion is used as an overt response to cognitive effort more consistently with increasing age. In addition, there is evidence to support that the primary function of gaze aversion is to manage cognitive demands rather than as a response to social difficulty. Doherty-Sneddon and Phelps (2005) measured gaze aversion in 8-year-old children who were questioned either face-to-face or via live video link. Results revealed that question difficulty strongly influenced gaze aversion in both interview conditions. In the current study, it was anticipated that, as children's ages ranged from 8 to 11 years old, they would attempt to use gaze aversion to reduce cognitive effort more so in the 'lie-telling' condition where cognitive load is higher than in the 'truth-telling' condition. Requiring interviewees to maintain gaze during questioning, as was the case for this study, would disable this coping mechanism for lie-tellers and maintain the increased cognitive demands of providing a false report. Furthermore, as maintaining gaze is not a natural behaviour, it would be necessary for interviewees in this experiment to intentionally remind themselves to comply with our gaze instruction, creating additional cognitive load.

In a previous study, maintaining eye contact was used to impose cognitive load on adult interviewees (Vrij, Mann, Leal, & Fisher, 2010). The researchers found that requiring eye contact elicited two cognitive cues (out of 14 cues) that discriminated lie-tellers from truth-tellers; namely, deceitful accounts contained fewer spatial details and were more chronological compared to truthful accounts. No significant differences were elicited between truth-tellers and lie-tellers when interviewees were given no 'eye contact' instruction. In terms of detection accuracy, the small difference in elicited cues only improved lie detection accuracy from 44% in the 'control' condition to 53% in the 'eye

contact' condition. If an improvement in accuracy rates is dependent on the exaggeration of behavioural differences between truth-tellers and lie-tellers, then eliciting two cognitive cues did not suffice. Vrij et al. (2010) suggest that these findings may be the product of anxiety for lie-tellers rather than increased cognitive load. Alternatively, previous research has shown that, even when adults find maintaining gaze with a person's face to be more difficult than either looking at the floor or closing their eyes, this does not result in them performing worse in the former condition compared to the latter two gaze conditions (Doherty-Sneddon, Bonner & Bruce, 2001).

On the contrary, the difficulty that children experience when instructed to direct their gaze does translate into poorer performance compared to a control condition involving no gaze instruction. In their first experiment, Doherty-Sneddon et al. (2001) compared the effect of gaze instruction (look at the speaker vs. look at the floor vs. close your eyes) on both adults' and 10-year-old children's task performance. Like adults, most children (83%) found looking at the floor or closing their eyes to be the easiest (least cognitively demanding) conditions. Results showed that, when children looked at the floor, this reported ease translated into them performing significantly better than when they looked at the speaker. This difference in task performance was also found across Doherty-Sneddon et al.'s subsequent experiments for different tasks and for a younger age group (6 years old). Children, therefore, experienced great difficulty in moderating the negative effects of gaze maintenance, with looking towards a face resulting not only in increased levels of cognitive demand for children, but also diminished task performance (Doherty-Sneddon et al., 2001). Thus, it was anticipated, for the current study, that children's interview performance would be affected by gaze maintenance.

Experiment 1

The aim of Experiment 1 was to investigate whether an instruction to maintain gaze would exaggerate differences between children's true and false reports. With a view to the future practical value of this research, it was important to consider how appropriate an instruction to maintain gaze would be with a child population. As maintaining gaze has already been linked to anxiety (Vrij et al., 2010), asking a child to look at an interviewer's face may intimidate some interviewees. In this study, we instructed some of the children to look at a face stimulus considered to be less intimidating; a teddy bear's face. It should be noted that toys can be useful in child witness interviews (Wilson & Powell, 2001), and a teddy bear was chosen because it has a face and is non-gender specific.

In this experiment, we predicted that lie-tellers would experience more dual-task interference than truth-tellers when instructed to maintain gaze. That is, lie-tellers' ability to provide a detailed account would be more negatively associated with their level of gaze compliance, compared to truth-tellers (Hypothesis 1). Secondly, we anticipated that this dual-task interference would exaggerate subtle differences in level of detail between true and false reports. Thus, it was expected that child lie-tellers would provide reports that were significantly less detailed than those provided by child truth-tellers, and this difference in detail would be greater for children instructed to maintain gaze compared to children given no gaze instruction (Hypothesis 2).

Method

Participants. Eighty-five children (37 boys, 48 girls) aged 8 to 11 years old (M = 10.46 years, SD = .81 years) were recruited from four primary schools in the United Kingdom. Participant information sheets were sent home to children's legal guardians who returned a signed written consent form. The general procedure was outlined to the children to obtain their verbal assent to participation, but they were naïve to the specific purpose of the

study and to the anticipated effect of maintaining gaze. All children, who were asked to lie, complied with the request to lie. Verification was sought from teachers that they had not taken part in the event that they were interviewed about. All children received a certificate and a stationery set in exchange for taking part.

Procedure. The experiment took place in two quiet areas of each school and involved the Principal Investigator (PI) who ran the study and a Research Assistant who conducted all interviews and was blind to the aims and hypotheses of the study. All children were tested individually.

The PI invited each child to take part in a short interview about a recent event at their school, thus events differed across schools. These events included a school sports day, a visit to the local cathedral, a school play, and a music concert. Children were randomly assigned to a Veracity condition within each year group in each school so that there were roughly equal numbers of truth-tellers and lie-tellers for each of the four events. Truth-tellers (n = 39, M_{age} = 10.28 years, SD_{age} = .83 years) were interviewed after they had experienced the event and were asked to provide a truthful recollection of what happened. Lie-tellers (n = 46, M_{age} = 10.43 years, SD_{age} = .81 years), on the other hand, were interviewed about an event that they had not experienced and were asked to convince the interviewer that they had already taken part in the event, when in fact they had not. This is similar to the veracity allocation carried out by other researchers interested in eliciting false allegations from children (e.g. Akehurst, Köhnken & Höfer (2001); Brunet et al., 2013; Lyon, Malloy, Quas & Talwar, 2008).

Within their veracity groups, children were also randomly assigned to a Gaze Instruction condition: Look at the interviewer's face (IF, n = 28, M_{age} = 10.25 years, SD_{age} = .80 years) or Look at the teddy bear's face (TF, n = 29, M_{age} = 10.52, SD_{age} =.74 years) or No

gaze instruction (Control, n = 28, M_{age} = 10.32 years, SD_{age} = .91 years). The teddy bear was seated on the interviewer's lap throughout all interviews (i.e. for all conditions). Prior to the interview, children in the IF and TF conditions were instructed by the PI to maintain gaze with the relevant face stimulus as much as they possibly could throughout the interview (i.e. to look at it as much as they could remember to do so). All children then received a sheet listing general themes that they could tell the interviewer about (e.g. talk about who was there, what happened, when it happened). This does not constitute coaching as neither truth-tellers nor lie-tellers were told exactly what they should say and they did not rehearse their story with the PI. Providing children with these themes was anticipated to elicit longer statements, allowing for more cues to deceit to occur (Leal, Vrij, Warmelink, Vernham & Fisher, 2015; Vrij, 2015). All participants were given approximately three minutes to prepare themselves before the PI escorted them to the interview room. Before entering the interview room, children in the IF and TF condition were given a final reminder by the PI to maintain gaze with the relevant face stimulus. Thsis was done out of earshot of the interviewer so that she remained blind to the aims and hypotheses of the study.

The interview protocol reflected the initial stages of a Cognitive Interview (Fisher & Geiselman, 1992): A rapport-building phase (that took place off-camera) was followed by two open-ended questions. First, an invitation to provide a free, uninterrupted narrative (e.g. tell me everything that happened when you took part in your school sports day), and then, secondly, a request, to all interviewees, to provide one additional piece of information about an aspect of the event that they had not already mentioned. No other questions were asked. All children were video-recorded, and their interviews later transcribed. All interviewees were asked the following question, which served as a manipulation check: Where were you instructed to look during the interview? The response options were 'interviewer's face', 'teddy bear's face' or 'no instruction given'.

Coding for detail. Two independent coders rated the children's interview transcripts for number of details included. To make the coding more precise, all transcripts were coded for five different types of details; visual details (e.g. "white clay head" contains three visual details), auditory details (e.g. "the teacher told us to take deep breaths" contains one auditory detail), spatial details (e.g. "he stood behind the curtain" contains one spatial detail), temporal details (e.g. "at the end we left" contains one temporal detail), and action details (e.g. "we played football" contains one action detail). One coder coded all of the transcripts for the current study, whilst the second coder rated a random sample of 20 transcripts. Considering that general level of detail is a reliable indicator of veracity (DePaulo et al., 2003), total number of details was calculated for each interviewee, by adding together the scores for all five detail types. Intra-class correlation coefficients (ICCs) were calculated for the two coders. Inter-rater reliability was high, with all ICCs demonstrating high levels of agreement between coders (visual details, ICC = .96; auditory details, ICC = .98; spatial details, ICC = .94; temporal details, ICC = .96; action details, ICC = .92; and total number of details, ICC = .98).

Coding for gaze maintenance. To provide an objective measure of gaze behaviour, two different independent judges, using INTERACT 14.0 software (Mangold, 2015), coded all interviews (from start to end) for the amount of time (in seconds) that the child interviewees gazed towards the interviewer's face (IF) and the teddy bear's face (TF). The duration of these gaze patterns for both face stimuli were then added together to give the total number of seconds spent gazing at the IF and the TF for each child. Percentage of time spent gazing towards both the IF and the TF were calculated by taking the total number of seconds spent gazing towards each face stimuli, dividing it by the total length of the interview in seconds and multiplying it by 100. Percentage of time spent gazing elsewhere was calculated by adding together the percentages for IF and TF and subtracting this total from 100. First,

both raters coded 17 interviews (20% of the total) to check for inter-rater reliability. Inter-rater reliability was high for time spent looking at the interviewer's face (ICC = .99) and at the teddy bear's face (ICC = .91). Rater 1 then coded the next 40% of the video recordings (n = 34) and Rater 2 coded the remaining 40% of the video recordings (n = 34). Percentage of time spent gazing at each face stimulus was calculated for each child by dividing the time spent gazing at the stimulus (in seconds) by the total duration of the interview (in seconds) and multiplying the result by 100.

Results

Manipulation checks. All 85 children correctly indicated where they had been asked to look during the interview. To test level of compliance more objectively, two-way ANOVAs were performed with Veracity and Gaze Instruction as the between-subjects factors. These were conducted to investigate differences in percentage of time spent gazing at (a) the interviewer's face, (b) the teddy bear's face, and (c) elsewhere (i.e. towards neither face stimulus). Figure 1 displays the distribution of gaze behaviour across 'veracity' conditions and Figure 2 across 'gaze instruction' conditions.

In terms of gazing towards the interviewer's face, there was a significant main effect of Veracity, $F(1, 79) = 5.78$, $p = .019$. Children providing a false report (M = 45.80%, SD = 22.41) spent a higher percentage of their interviews looking at the interviewer's face than children providing a true report (M = 35.24%, SD = 23.72), d = .46, 95% CI [.03, .89]. There was also a significant main effect of Gaze Instruction, $F(2, 79) = 10.50$, $p<.001$. Pairwise comparisons using Bonferroni adjustment showed that children instructed to look at the interviewer's face (M = 55.93%, SD = 24.97) spent a greater portion of the interview gazing at the interviewer's face than children instructed to look at the teddy bear's face (M = 31.61%, SD = 23.41, $p<.001$, d = .98, 95% CI [.42, 1.52], or given no gaze instruction (M =

28.71%, *SD* = 17.49, *p* = .001, *d* = 1.04, 95% CI [.47, 1.59]). There was no difference between these latter conditions, *p* = 1.00. There was no significant interaction effect, *F*(2, 79) = 1.10, *p* = .34.

In terms of gazing towards the teddy bear's face, there was no significant main effect of Veracity, *F*(1, 79) = .32, *p* = .57. There was, however, a significant main effect of Gaze Instruction, *F*(2, 79) = 9.50, *p*<.001. Pairwise comparisons using Bonferroni adjustment showed that instructing children to gaze at the teddy bear's face (*M* = 16.77%, *SD* = 18.77) resulted in a higher percentage of time looking at the teddy bear's face than instructing children to look at the interviewer's face (*M* = 5.22%, *SD* = 5.21, *p* = .001, *d* = .83, 95% CI [.29, 1.37]), or giving no gaze instruction (*M* = 4.66%, *SD* = 2.66, *p*<.001, *d* = .90, 95% CI [.35, 1.44]). There was no difference between these latter conditions, *p* = 1.00. There was no significant Veracity X Gaze Instruction interaction effect, *F*(2, 79) = .28, *p* = .76.

Finally, in terms of gazing elsewhere, there was a significant main effect of Veracity, *F*(1, 79) = 7.15, *p* = .009. Truth-tellers (*M* = 56.66%, *SD* = 24.22) spent a higher proportion of the interview looking elsewhere compared to lie-tellers (*M* = 44.48%, *SD* = 21.37), *d* = .54 (95% CI [.10, .97]). There was also a significant main effect of Gaze Instruction, *F*(1, 79) = 7.99, *p* = .001. Pairwise comparisons using Bonferroni adjustment showed that children given no gaze instruction (*M* = 61.73%, *SD* = 17.74) spent more time looking elsewhere compared to children instructed to look at the interviewer's face (*M* = 38.85%, *SD* = 24.21), *p*<.001, *d* = 1.08 (95% CI [.51, 1.64]). Percentage of time looking elsewhere did, however, not differ between children in the 'control' condition and those in the 'teddy bear's face' condition (*M* = 49.63%, *SD* = 22.61), *p* = .10. There was also no significant difference in percentage of time spent gazing elsewhere between children in the 'interviewer's face' condition and child in the 'teddy bear's face condition, *p* = .20. There was no significant interaction effect, *F*(2, 79) = .80, *p* = .45.

In sum, children were able to comply with the instruction to look at the interviewer's face or the teddy bear's face. That said, although our instructions did increase time spent gazing toward a specific face stimulus, overall compliance was relatively poor as the average participant complied with their gaze instruction for less than 50% of their interview. Furthermore, children in the 'teddy bear's face' condition only spent 16% of the time looking at their specified stimulus and just as much time looking at the interviewer's face and elsewhere as children in the 'control' condition. This lack of compliance may be because gazing at a static toy when responding to a person is an unnatural behaviour. It could also be because the location of the teddy bear was problematic; staring at the interviewer's lap may have seemed strange.

Hypotheses-testing. Preliminary analyses revealed no significant effects of child age, child gender, or specific activity reported (e.g. sports day, school trip) during the interview, on any of the dependent variables. The data for all participants were, therefore, combined for subsequent analyses.

Dual-task interference. We investigated whether lie-tellers experienced more dual-task interference than truth-tellers, when given the secondary task of maintaining gaze with either the interviewer's face or the teddy bear's face whilst being questioned. The 'performance operating characteristic' (POC, Norman & Bobrow, 1975) of truth-tellers and lie-tellers was calculated separately for children in both 'gaze instruction' conditions. By calculating Pearson's correlations between the total number of details included in the interviewee's account (i.e. level of detail) and the time they spent gazing towards either the interviewer's or the teddy bear's face (i.e. level of gaze compliance), we were able to examine to what extent the two tasks interfered with one another. High levels of interference would be characterised by a strong negative correlation between performances on both tasks

(i.e. increasing compliance with the gaze instruction resulting in decreasing level of detail in responses).

First, when the secondary task required interviewees to look at the interviewer's face, findings revealed a weak, negative correlation for truth-tellers, r = -.28, p = .40, and a small to moderate, positive correlation for lie-tellers, r = .39, p = .16. Although these correlations are not significant, this may be due to the effect of a limited sample size. Following the suggestion of Ferguson (2009), we therefore looked at the effect size of these correlations as "effect sizes are resistant to sample size influence, and thus provide a truer measure of the magnitude of effect between variables" (p. 532). Interpreting these r values as effect sizes (Field, 2013), the data showed that there was a small effect for truth-tellers and a medium effect for lie-tellers. This suggests that there was mild interference between truth-tellers' ability to provide detailed answers and their compliance with the gaze instruction. However, it also shows that there was no interference for lie-tellers, whose level of detail in fact increased with their level of compliance with the gaze instruction. Second, when interviewees were instructed to look at the teddy bear's face, there was no correlation between level of detail and compliance with the gaze instruction for truth-tellers, r = -.04, p = .91, nor for lie-tellers, r = .08, p = .78.

Level of detail. Preliminary analyses showed that true reports (M = 750.79, SD = 670.31) contained significantly more words than false reports (M = 508.33, SD = 560.88), $t(83) = 1.82$, p = .037, d = .40 (95% CI [-.37, .82]). As longer reports allow for more details to occur, length of statement would have an effect on our analysis of total detail. To take this effect into account, length of statement (in words) was entered as a covariate in our analyses. This is similar to previous work by Strömwall and Granhag (2005) when analysing reality monitoring scores.

First, a 2 (Veracity) x 3 (Gaze Instruction) ANCOVA was performed with total number of details as the dependent variable. There was a significant main effect of Veracity, $F(1, 78) = 8.44$, $p = .005$, a significant main effect of Gaze Instruction, $F(2, 78) = 3.16$, $p = .048$, and a significant Veracity X Gaze Instruction interaction effect, $F(2, 78) = 4.22$, $p = .018$. Descriptive statistics for each of the experimental cells are displayed in Table 1.

Of interest for the hypotheses is the Veracity X Gaze Instruction interaction effect. Separate ANCOVAs were conducted: first, for each of the Gaze Instruction conditions with Veracity as the independent variable, and second, for each of the Veracity conditions with Gaze Instruction as the independent variable. When children were instructed to gaze at the interviewer's face, truth-tellers provided significantly more details compared to lie-tellers, $F(1, 25) = 8.53$, $p = .007$, $d = .92$ (95% CI [.13, 1.70]). Similarly, when children were instructed to look at the teddy bear's face, truth-tellers provided more details in their statements than lie-tellers, $F(1, 26) = 5.88$, $p = .023$, $d = .83$ (95% CI [.058, 1.59]). Veracity did not have a significant effect on the number of details provided by children who were given no gaze instruction, $F(1, 25) = .24$, $p = .63$. Irrespective of whether they were providing a true report or a false report, children in the control condition included the same amount of detail.

For children who provided a truthful account, there was a significant effect of Gaze Instruction condition, $F(2, 35) = 4.04$, $p = .026$. Post-hoc testing using Bonferroni adjustment revealed that truth-tellers who looked at the interviewer's face provided more details than truth-tellers who were given no gaze instruction, $p = .03$, $d = .48$ (95% CI [-.31, 1.25]). There was no difference in quantity of detail between truth-tellers looking at the interviewer's face and those looking at the teddy bear's face, $p = 1.00$, and no difference between truth-tellers looking at the teddy bear's face and those in the control condition, $p =$

.14. For children who provided a fabricated account, there was no significant effect of Gaze Instruction, $F(2, 35) = .55$, $p = .58$.".

Discussion

The analysis of the association between providing a detailed account and complying with the gaze instruction revealed a small positive effect for lie-tellers in the 'interviewer's face' condition. That is, the more the lie-tellers looked at the interviewer's face the more details they gave. This was contrary to Hypothesis 1. Furthermore, the instruction to look at the teddy bear's face did not elicit dual task interference for the lie-tellers nor for the truth-tellers. Our theoretical assumption posited that lie-tellers, who have a more cognitively demanding primary task compared to truth-tellers, would reach the limit of their resources when a secondary task was imposed (Knowles, 1963), and, therefore, experience a high level of dual-task interference (Kahneman, 1973). However, our analysis of lie-tellers' dual-task interference does not support this theoretical assumption. Indeed, the positive relationship between level of detail and gaze compliance for lie-tellers instructed to look at the interviewer's face completely contradicts our hypothesis. This could be due to the cognitive resources required for each task originating from separate (limited) resources. Multiple resource theory (Wickens, 2002) posits that tasks that are structurally dissimilar, such as answering interview questions (verbal) and maintaining gaze (visual/social), will interfere less. This may explain why imposing cognitive load through constructing/maintaining a lie (verbal) and telling the lie in reverse order (verbal) had greater success in previous studies (Saykaly et al., 2016; Vrij et al., 2008) because the two tasks use similar cognitive processes.

An alternative explanation could be that the effect of gaze maintenance on task performance can vary dependent on the relevance of the visual stimulus to the primary task (Doherty-Sneddon et al., 2001). It could be the case, in the current study, that child lie-tellers

instructed to look towards the interviewer's face found the information communicated by her face more task-relevant than truth-tellers. Lie-tellers, who are more concerned with appearing honest than truth-tellers (Vrij, 2015), might have monitored the interviewer's face for feedback on how their deception was being received and used this to modify their responses (e.g. to say more to appear honest). However, this tactic works to their disadvantage, as longer statements are more likely to contain cues to deceit (Vrij et al., 2015). This would particularly be the case for child interviewees who tend to reveal their deceit verbally (Talwar & Lee, 2002). Furthermore, these unanticipated findings might be explained by differences in children's developing cognitive capabilities that are associated with lie-telling ability, such as executive functioning (Talwar & Crossman, 2011). Child lie-tellers in our study may have had good working memory skills that allowed them to look at the interviewer's face whilst telling their false report. Future research should investigate whether the effects of imposing cognitive load are moderated by children's growing cognitive development.

Interestingly, truth-tellers instructed to look at the interviewer's face did experience some dual-task interference. This unexpected finding requires further investigation. As memory can be data-limited (i.e. limited by a person's ability to recall a past experience), it could be that factors other than gaze compliance influenced our child truth-tellers' ability to provide a detailed account. Finally, the absence of dual-task interference for children instructed to look at the teddy bear's face could be explained by the teddy bear's face not being as cognitively effortful to look at as the interviewer's face. As the teddy bear's face did not provide any relevant feedback, it was not necessary for the interviewees to monitor it for suspicion. Nevertheless, we suggest caution in interpreting these correlations due to their non-significant nature.

Irrespective of the findings for dual-task interference, significant differences in level of detail between child truth-tellers and child lie-tellers were only elicited when a secondary task was imposed. For children instructed to look at the interviewer's face, these findings are in line with previous work with adults (Vrij et al., 2010), which has also found exaggerated behavioural differences between truths and lies when gaze was maintained. For children instructed to look at the teddy bear's face, these findings extend current knowledge and demonstrate that gazing towards a non-human stimulus could act as a less threatening, but still effective, substitute in practice. Although exaggerated differences occurred when a dual-task was imposed, it remains unclear from a theoretical standpoint why this was the case. The dual-task processes involved in providing a narrative and maintaining gaze require further examination to understand the theory behind this effect. Indeed, further probing of the significant interaction suggests that using different gaze instructions does not have an effect on false reports but rather has an effect on true reports. Thus, these exaggerated differences could be due to gaze maintenance facilitating longer truthful accounts rather than inhibiting false accounts. Our findings suggest that the request to look at the interviewer's face elicited true reports that were significantly more detailed than when no gaze instruction was provided. This may be due to the demeanour of our interviewer; supportive interviewers have been shown to elicit longer reports (Vrij, 2015). However, it is not within the scope of this research to draw any firm conclusions regarding these results. Furthermore, these findings should be interpreted with caution. Due to small experimental cell sizes, there is a risk of Type I error. This study, therefore, requires replication with a larger sample size to verify that the interaction effect remains significant.

In this study we were not able to examine the memory accuracy of the truth-tellers' detailed reports. Based on the information provided by the schools, we were only able to establish whether the children had taken part in the events or not, but, due to the scope of the

events, we were unable to capture all of the information regarding the events to code for correct and incorrect details. Future research is required to explore the relevance and accuracy of the reports provided by truth-tellers in the 'gaze instruction' conditions to understand the specific benefits of eliciting more details in true reports.

In the current study the interview protocol was short and non-elaborative. Using open-ended questions did allow us to go beyond the majority of past research, which has primarily focused on forced-choice questions using temptation resistance paradigms, to examine how gaze maintenance would affect children's longer narratives. However, this does not reflect interview protocols in real-life police investigations with child witnesses, where a variety of question types are used. We can, therefore, not generalize these findings to a whole police interview, but only to the beginning of the police interview where an uninterrupted free narrative is requested. Finally, our study represents a 'best case scenario' in which a child provides a long narrative. As we reduced our interview protocol to focus on two open-ended questions, it was important to facilitate long responses by providing all of the children with examples of the type of information they could provide and some time to prepare. Child witnesses typically provide shorter statements than both their adolescent and adult counterparts (Jack, Leov & Zajac, 2014); this may be due to them not knowing what level of detail is required at interview (Lamb, Orbach, Hershkowitz, Esplin & Horowitz , 2007). Future research should continue to test the generalizability of these findings by using a procedure where no examples are provided.

Despite the exaggerated difference in level of detail elicited between child truth-tellers and child lie-tellers in the dual-task gaze condition (compared to the single-task control condition), the major concern still remained whether evaluators would be able to discriminate between lie-tellers and truth-tellers more effectively when child interviewees

were instructed to maintain gaze compared to when no gaze instructions were given. We investigated this issue in Experiment 2.

Experiment 2

In Experiment 2, we tested the prediction that evaluators would discriminate better between truth-tellers and lie-tellers instructed to maintain gaze, than truth-tellers and lie-tellers who were given no gaze instruction (Hypothesis 3).

We also examined whether telling evaluators that truth-tellers provide more detail in their reports than lie-tellers would improve discrimination accuracy. Previous research into training to improve lie detection has shown that informing evaluators about empirically-supported verbal cues to deceit has the largest effect on their detection accuracy (Hauch, Sporer, Michael & Meissner, 2014). Overall, level of detail has been found to be a key indicator of veracity (DePaulo et al., 2003). It is also one of the general characteristics coded for in Criteria-Based Content Analysis (Steller & Köhnken, 1989) that has received the most support for distinguishing between child truth-tellers and child lie-tellers in the predicted direction (Vrij 2005). It was, therefore, anticipated that evaluators who received this guidance regarding detail would demonstrate better discrimination than evaluators who received no guidance (Hypothesis 4). It was further predicted that an improvement in discrimination, as a result of guidance, would be most pronounced when judging the credibility of children instructed to maintain gaze, due to a greater difference in detail being elicited in these conditions in Experiment 1 (Hypothesis 5).

Successful discrimination depends on whether evaluators can interpret behavioural cues correctly. It was, therefore, important to recognise that gaze aversion can be perceived as a strong indicator of deception (Global Deception Research Team, 2006), even though this cue is non-diagnostic (DePaulo et al., 2003). We could not rule out the possibility that gaze

behaviour perceived to be somewhat 'strange' might impact on evaluators' judgments of credibility. Half of the evaluators were, therefore, played visual-audio clips of the children's interviews, and the other half were played audio-only clips. We anticipated that evaluators who watched the visual-audio presentations displaying the gaze maintenance behaviour would demonstrate a truth bias because gaze maintenance might be interpreted as a sign of truthfulness (Vrij et al., 2010) (Hypothesis 6).

Method

Participants. A sample of 192 adult evaluators (89 males, 103 females) with an age range of 18 to 76 years (M = 27.14 years, SD = 11.71 years) was recruited. One hundred and ten participants (52% of the total sample) were undergraduate students who received 0.5 course credit for their participation. The further 82 participants were members of the general public recruited via convenience sampling. The non-student participants were not compensated for their participation.

Interview clips. A total of 30 interview clips were selected from the sample of 85 children in Experiment 1. There were ten clips per 'Gaze Instruction' condition; within each of those three sets of ten clips, there were five truth-tellers and five lie-tellers. In the first round of the interview clip selection process, all recordings that contained noise interference (e.g. school bell, road traffic) were excluded (n = 20). Second, clips in which the first free recall lasted longer than 300 seconds were removed (n = 7). This criterion was chosen to limit the total duration of the study (50 minutes maximum), reducing potential fatigue effects on evaluators' performance. The remaining 58 clips were divided by Gaze Instruction condition (IF, n = 22; TF, n = 16; CONTROL, n = 20), and five truth-tellers and five lie-tellers were randomly selected for each condition. The final thirty clips were edited down so that they only contained the child interviewee's first free recall. This selection process

resulted in an even distribution of gender (3 boys to 2 girls, or 2 boys to 3 girls) in each Veracity x Gaze Instruction cell, except for the false reports in the 'control' condition, which were all provided by boys. It was not anticipated that this would bias results as no response bias has been previously found for adults judging boys' credibility (Talwar, Crossman, Gulmi, Renaud & Williams, 2009). Interview clips lasted from 53 seconds to 239 seconds (M = 135.67 seconds, SD = 56.16 seconds). A 2 (Veracity) x 3 (Gaze Instruction) ANOVA was performed to ensure that there were no significant differences in length of clip across conditions. There was no significant main effect of Veracity, $F(1, 24) = .13$, $p = .72$, no significant main effect of Gaze Instruction, $F(2, 24) = .05$, $p = .96$, and there was no significant Veracity X Gaze Instruction interaction effect, $F(2, 24) = .62$, $p = .55$. For each 'gaze instruction' condition, four random rotations of the ten clips were created to reduce order effects.

Guidance on detail. Evaluators who received guidance were provided with a sheet stating that truth-tellers provided more detail overall in their accounts compared to lie-tellers, as this has been reported in previous deception research (DePaulo et al., 2003) and was also found in Experiment 1. To help evaluators understand what the experiment meant by the term 'detail', five different types of detail were presented in a table. For each type of detail, a description and an example of that detail were provided (i.e. 'visual detail refers to what the interviewee said that they saw. For example, a red hat contains two visual details'). Participants were advised to refer back to the guidance sheet as much as they found useful when watching/listening to the interview clips and were able to ask the experimenter for clarification on these types of detail before and during the experiment.

Procedure. The study took place in a quiet environment with few distractions. In order to prevent evaluators from working on the assumption that they would be presented with equal numbers of truth-tellers and lie-tellers, two steps were taken. First, participants

were informed that they would be asked to evaluate the veracity of twelve child interviews in turn (actually they only evaluated ten clips in total). Second, they were told that it was just as likely for a child to be telling the truth as it was for them to be telling a lie..

First, evaluators were randomly assigned to a Gaze Instruction condition. That is, they judged the credibility of ten interview clips (five truth-tellers and five lie-tellers) from only one of the Gaze Instruction conditions in Experiment 1 (IF vs. TF vs. Control). Evaluators who were provided with guidance on detail received this at the beginning of the experiment. Half of the evaluators watched all of the interview clips in visual-audio format, whilst the other half listened to all interview clips in audio-only format. Participants who watched visual-audio presentations of the interviewees in the 'interviewer's face' and the 'teddy bear's face' conditions were informed that the child interviewees had been asked by the experimenter to direct their gaze during the interviews. Evaluators then watched and/or listened to the clips, one at a time, via a computer. Headphones were provided. To record their credibility judgments, evaluators were given a hard copy answer booklet. Following each interview clip, evaluators were asked to decide if the child interviewee was lying or telling the truth.

Participants' dichotomous judgments (truth or lie) for each clip were used to measure hits (proportion of deceitful clips correctly identified as deceitful) and false alarms (proportion of truthful clips incorrectly identified as deceitful) for subsequent signal detection analysis.

Results

Accuracy. Overall accuracy (M = 51.72%, SD = 16.23) was not significantly different from chance, $t(191) = 1.47$, $p = .14$, but truth accuracy (M = 60.62%, SD = 20.56) was significantly above chance, $t(191) = 7.16$, $p<.001$, $d = .52$ (95% CI [.37, .67]), and lie accuracy (M = 42.81%, SD = 21.23) was significantly below chance, $t(191) = -4.69$, $p<.001$,

560 *d*= .34 (95% CI [.19, .48]). When evaluators judged the credibility of children instructed to
561 look at the interviewer's face (*M* = 58.91%, *SD* = 16.44), they performed significantly better
562 than chance, *t*(63) = 4.33, *p* <.001, *d* = .54 (95% CI [.28, .80]). When judging children
563 instructed to look at the teddy bear's face (*M* = 47.97%, *SD* = 15.45) or children given no
564 gaze instruction (*M* = 48.28% *SD* = 14.54), they were no better than chance (*p*s>.05).
565 Moreover, when evaluators were guided to look out for differences in detail (*M* = 53.96%, *SD*
566 = 17.07), they were better than chance, *t*(95) = 2.27, *p* = .025, *d* = .23 (95% CI [.03, .43]), but
567 not when no guidance was provided (*M* = 49.48%, *SD* = 15.11), *t*(95) = -.34, *p* = .74.

568 **Signal detection analysis.** The application of signal detection theory to deception
569 detection research has been largely recommended because it provides an opportunity to
570 measure two conceptually different parameters of accuracy (Meissner & Kassin, 2002);
571 *discrimination accuracy* - ability to discriminate lie-tellers from truth-tellers (in this
572 experiment, referred to as *d'*), and *response bias* – tendencies to favour a particular response
573 (truth or lie) (in this experiment, referred to as β). Means and standard deviations for
574 discrimination accuracy and response bias across all conditions are displayed in Table 2.

575 ***Discrimination accuracy.*** A 3 (Gaze Instruction) x 2 (Guidance Provision) ANOVA
576 was performed with participants' sensitivity scores (*d'*) as the dependent variable to examine
577 their ability to discriminate between truth- and lie-tellers.

578 First, there was a significant main effect of Gaze Instruction, *F*(2, 180) = 10.84,
579 *p*<.001. Post-hoc analyses using Bonferroni adjustment revealed that evaluators
580 discriminated better between children's truthful and deceptive accounts when the
581 interviewees were instructed to look at the interviewer's face compared to when the
582 interviewees were instructed to look at the teddy bear's face, *p*<.001, *d* = .66 (95% CI [.30,
583 1.02]), and when the interviewees were given no particular gaze instruction, *p*<.001, *d* = .67

(95% CI [.32, 1.03]). Evaluators' performance did not differ significantly between those instructed to look at the teddy bear's face and for those given no instruction (p = 1.00).

Second, there was a significant main effect of Guidance Provision, $F(1, 180) = 4.20$, p = .042. Pairwise comparisons using Bonferroni adjustment showed that evaluators who received guidance discriminated better between veracity groups than evaluators who received no guidance, d = .27 (95% CI [-.014, .55]).

Finally, there was a significant Gaze Instruction X Guidance Provision interaction effect, $F(2, 180) = 4.88$, p = .009. We performed univariate analyses to test the effect of providing guidance within each Gaze Instruction condition. There was a significant main effect of Guidance Provision for evaluators judging the credibility of child interviewees instructed to look at the teddy bear's face, $F(1, 62) = 12.10$, p = .001. For evaluators in the 'teddy bear's face' condition, those who received guidance (M = .22, SD = .76) were able to discriminate better than those who received no guidance (M = -.38, SD = .63), d = .87 (95% CI [.35, 1.38]). There was no significant main effect of Guidance Provision for evaluators assigned to the 'interviewer's face' condition, $F(1, 62) = 1.27$, p = .26, or the 'control' condition, $F(1, 62) = 1.15$, p = .29. There were no other significant interaction effects (p-values >.05).

In a second level of analysis, *d'* values were compared to 0 (no ability to differentiate between children's truths and lies) using one-sample t tests. With regard to Gaze Instruction, evaluators could reliably discriminate child truth-tellers from child lie-tellers in the 'interviewer's face' condition, $t(63) = 4.32$, p<.001, d = .54 (95% CI [.28, .80]), but not in the 'teddy bear's face' condition, $t(63) = -.87$, p = .39, nor the 'no gaze instruction' condition, $t(63) = -.87$, p = .38. For Guidance Provision, evaluators were able to discriminate reliably when provided with guidance, $t(95) = 2.30$, p = .024, d = .23 (95% CI [.03, .44]), but not when guidance was withheld, $t(95) = -.20$, p = .84.

Finally, we compared *d'* scores to 0 for the significant interaction between Gaze Instruction and Guidance Provision. When evaluators judged the credibility of children instructed to look at the interviewer's face, they were able to discriminate lie-tellers from truth-tellers whether guidance was provided (M = .55, SD = .85), $t(31) = 3.63$, p = .001, d = .64 (95% CI [.26, .1.02]), or not (M = .32, SD = .75), $t(31) = 2.43$, p = .021, d = .43 (95% CI [.063, .79]). For children instructed to look at the teddy bear's face, evaluators were not able to discriminate between children's truths and lies when provided with guidance (M = .22, SD = .76), $t(31) = 1.65$, p = .11, nor when there was no guidance provision (M = -.38, SD = .63), $t(31) = -3.46$, p = .002, d = .61 (95% CI [.23, .98]). That is, evaluators labelled the groups incorrectly (i.e. they tended to label lie-tellers as truthful and truth-tellers as deceitful). Finally, when children were given no gaze instructions, evaluators were not able to discriminate truthful from fabricated reports, with guidance provision, (M = -.17, SD = .78), $t(31) = -1.25$, p = .22, or without guidance provision, (M = .018, SD = .64), $t(31) = .16$, p = .88.

Response bias. Participants' response bias (β scores) was investigated to see whether they tended to identify children as lie-tellers or truth-tellers in any particular condition. A three-way ANOVA, with Gaze Instruction, Guidance Provision and Modality of Presentation of the clips as between-subjects factors, revealed significant main effects of Gaze Instruction, $F(2, 180) = 5.05$, p = .007, and Modality of Presentation, $F(1, 180) = 6.55$, p = .011. First, responses were more biased when judging the credibility of children instructed to look at the interviewer's face (M = 1.21, SD = .49) compared to children instructed to look at the teddy bear's face (M = 1.02, SD = .37), p = .020, d = .46 (95% CI [.10, .81]), and children given no particular gaze instruction (M = 1.01, SD = .38, 95% CI [.92, 1.11]), p = .019, d = .45 (95% CI [.10, .80]). Response bias did not significantly differ between evaluators judging child credibility in the latter two gaze conditions (p = 1.00). Second, evaluators demonstrated

more bias in the 'audio-only' condition (M = 1.16, SD = .48) than in the 'video-audio' condition (M = 1.01, SD = .35), d = .36 (95% CI [.07, .64]). There was no significant main effect of Guidance Provision and there were no significant interaction effects (*p-value*s >.10).

Using one-sample t tests, each β was compared to 1 (no bias). In signal detection theory, β values below 1 signify a tendency to respond *yes* (or *lie* in the current study), whereas values above 1 signify a tendency to respond *no* (or *truth* in the current study; Stanislaw & Todorov, 1999). Therefore, the subsequent analyses examined the existence and the nature of the bias. With regard to Gaze Instruction, evaluators who judged the credibility of children instructed to look at the interviewer's face were significantly biased to respond 'truth', $t(63) = 3.46$, $p = .001$, $d = .43$ (95% CI [.18, .69]), whereas no significant response bias was found for evaluators who judged children instructed to look at the teddy bear's face, $t(63) = .35$, $p = .73$, nor for evaluators who judged children in the 'no gaze instruction' condition, $t(63) = .30$, $p = .77$. In terms of Modality of Presentation, evaluators in the 'audio only' condition displayed a significant truth bias, $t(95) = 3.18$, $p = .002$, $d = .33$ (95% CI [.12, .53), whereas evaluators in the 'video-audio' condition showed no bias, $t(95) = .17$, $p = .87$.

Discussion

Instructing child interviewees to maintain gaze with the interviewer's face enabled evaluators to discriminate between true and false reports to a better degree than when no instruction was given, (in spite of a significant truth bias). However, discrimination accuracy was not affected when child interviewees were instructed to gaze towards the teddy's bear face. Thus, Hypothesis 3 was partially supported. The ability to accurately detect deception for evaluators rating children instructed to gaze at the interviewer's face may be due to differences in details provided by child truth-tellers and child lie-tellers. The cognitive lie detection approach posits that the ability to discriminate between truths and lies should increase with the activation and exaggeration of cognitive behavioural differences (Vrij,

2015). Considering that significant behavioural differences were elicited for both children instructed to look at the interviewer's face and children instructed to look at the teddy bear's face, it is possible that the exaggeration of these cues might need to reach a certain threshold, beyond which they become more apparent to an evaluator. It is possible that this threshold was only reached when child interviewees were instructed to look at the interviewer's face, in turn, facilitating evaluators' credibility judgments, but the threshold was not met when the children were asked to look at the teddy bear's face.

Informing evaluators that truth-tellers provide more detailed reports compared to lie-tellers did improve their ability to detect deception, thus supporting Hypothesis 4. However, it is difficult to conclude to what extent evaluators applied this guidance to the interview clips. Although training in verbal content cues is recommended because it leads to the highest training effects, it is also important to note that false information regarding cues to deceit can work as effectively as true information (Hauch et al., 2014). To encourage evaluators to engage more with the guidance and base their final credibility judgments on this specific information, it would be better to use methods such as the Psychologically Based Credibility Assessment Tool (Evans, Michael, Meissner & Brandon, 2013) that include the rating of diagnostic cues in the final credibility assessment.

Contrary to Hypothesis 5, the provision of guidance was not more beneficial when judging children who were instructed to maintain gaze compared to those in the 'control' condition. Indeed, the only benefit of providing guidance was that it protected evaluators in the 'teddy bear's face' condition from incorrectly labelling child veracity. As children in this condition were neither maintaining eye contact, nor free to look where they wished, their 'strange' gaze behaviour of looking at the interviewer's lap might have been interpreted incorrectly as suspicious. Directing evaluators' attention towards what the child was saying, through the use of our guidance, and encouraging them to base their credibility judgments on

684 the child's verbal behaviour, may have detracted from the misinterpretation of their 'strange'
685 gazing towards the teddy bear.

686 Finally, although we predicted in Hypothesis 6 that evaluators who watched the
687 visual-audio presentations displaying the gaze maintenance behaviour would demonstrate a
688 truth bias, this was not the case. This lack of truth bias might be due to evaluators
689 interpreting gaze maintenance behaviour differently from that suggested by the general
690 deception literature. On the one hand, gaze aversion is believed to be a cue to deceit (Global
691 Deception Research Team, 2006), but, on the other hand, nonverbal behaviour that deviates
692 from the expected norm, such as staring, can also be perceived to be 'fishy' (Bond et al.,
693 1992). It is not known to what extent gaze behaviour influenced evaluators' judgments, or
694 how much suspicion evaluators attached to this nonverbal cue; however, the lack of bias
695 might suggest that opposing interpretations may have cancelled each other out. Alternatively,
696 informing evaluators that children had been instructed to divert their gaze may have made
697 them more aware of their own bias.

698 For the current study evaluators were exposed to ten interview clips. This may have
699 led to evaluators comparing cues and information across interviews. In real police
700 investigations and court proceedings, it is likely that these comparisons will occur between
701 children's statements, adult's statements and physical evidence. Future research should try to
702 replicate this scenario to understand how a police officer or juror might judge the credibility
703 of a child both in isolation and in comparison to other sources.

704 **General Discussion**

705 We conducted the first empirical investigation exploring the use of gaze maintenance
706 to detect deception in child witnesses during investigative interviews. Similar to Vrij et al.
707 (2010), we predicted that the interview strategy would magnify differences in level of detail

between children's true and false reports. We also expected that the exaggeration of this cue would facilitate evaluators' ability to discriminate children's lies from truths.

The present findings show that gaze maintenance can be effective for determining the credibility of child witnesses. In Experiment 1, lie-tellers provided significantly fewer details in their reports compared to truth-tellers but only when they were instructed to look towards either the interviewer's face or a teddy bear's face. No significant difference was elicited when a secondary task was absent. In Experiment 2, we found that the exaggeration of this diagnostic cue facilitated evaluators' discrimination accuracy, but this was only when children were instructed to look at the interviewer's face.

Theoretically, the effect of imposing a secondary task on interviewee performance remains unclear. The findings of Experiment 2 make it difficult to discern whether the secondary task had any negative impact on truth-tellers' memory or whether lie-tellers experienced any additional cognitive load. The latter issue may be due to the nature of the secondary task in this study and the difficulty in pinning down the exact cognitive mechanisms involved. As previously mentioned, the development of certain cognitive skills is closely linked to children's proficiency to tell and maintain lies (Talwar & Crossman, 2011). It may therefore be wise, in future, to provide cognitive measures of the specific executive functions that the imposed secondary task aims to affect to be able to establish whether (a) there is a link between these cognitive skills and the performance on the tasks, and (b) whether children's ability to perform these cognitive skills predicts the effectiveness of imposing cognitive load. When testing dual-task methodologies, it would also be beneficial to obtain baseline measures of an individual's performance on single tasks (Task A only and Task B only) to which their performance on a dual-task (Tasks A and B simultaneously) could be compared.

Our findings provide further support for the practical value of manipulating cognitive load as a potential means for discriminating between children's true and false reports. In particular, the results demonstrate that the effects of imposing cognitive load are not limited to asking children to tell their stories backwards. This is beneficial because Saykaly and colleagues (2016) found that reverse order recall can adversely affect the accuracy of both truthful and deceptive statements, suggesting that it might not be helpful in real police investigations. In our study, requiring child interviewees to perform the secondary task of maintaining gaze had a positive effect on truth-tellers, eliciting more information from them than when no gaze instruction was given. This finding is in line with the primary goal of any investigative interview, which is to extract as much information as possible from the interviewee. This finding could be due the interviewer's supportive demeanour, which has been found with adults to elicit more details from truth-tellers than lie-tellers (Mann et al., 2013). Further investigation is required to determine whether it is the combined effect of a gaze maintenance instruction to witnesses and supportive interviewer behaviour that helps truth-tellers but not lie-tellers, rather than the technique on its own.

A practical limitation of using gaze maintenance with child interviewees may be its appropriateness in certain contexts. Maintaining gaze with an authoritative figure, such as a police officer, might be an intimidating task for children. Although none of the children instructed to look at the interviewer's face reported any discomfort, the average child did not maintain gaze for more than half of their interview. A recent school event is far less traumatic to talk about than incidents of physical and/or sexual abuse, which can be the main focus of police investigations involving child witnesses. Future research must examine the scope of the beneficial effects elicited in this study and balance them with potential discomfort in certain contexts. As such, the preliminary findings relating to an instruction to concentrate on the less intimidating teddy bear (or similar) should be extended.

Maintaining gaze, particularly with an interviewer's face, is an effective strategy for judging the credibility of children. Future research should continue to explore the application of dual-task processing to child interviews by examining strategies that target children's under-developed executive functioning, with a view to creating more appropriate secondary tasks for this potentially sensitive context.

References

Akehurst, L., Köhnken, G., & Höfer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6*(1), 65-83. doi: 10.1348/135532501168208

Alloway, T. P., McCallum, F., Alloway, R. G., & Hoicka, E. (2015). Liar, liar, working memory on fire: Investigating the role of working memory in childhood verbal deception. *Journal of Experimental Child Psychology, 137*, 30-38. doi: 10.1016/j.jep.2015.03.013

Bala, N., Ramakrishnan, K., Lindsay, R. C. L., & Lee, K. (2014). Judicial assessment of the credibility of child witnesses. *Alberta Law Review, 42*, 995-1017. Retrieved from http://heinonline.org/HOL/Page?handle=hein.journals/alblr42&div=43&g_sent=1&collection=journals

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*(3), 330-343. doi: 10.1016/j.actpasy.2010.03.006

Bond, C. F., Omar, A., Pitre, U., Lashley, B. R., Skaggs, L. M., & Kirk, C. T. (1992). Fishy-looking liars: Deception judgment from expectancy violation. *Journal of Personality and Social Psychology, 63(6)*, 969-977. doi: 10.1037/0022-3514.63.6.969

Brunet, M. K., Evans, A. D., Talwar, V., Bala, N., Lindsay, R. C. L., & Lee, K. (2013). How children report true and fabricated stressful and non-stressful events. *Psychiatry, Psychology & Law, 20*(6), 867-881. doi: 10.1080/13218719.2012.750896

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex, 19(7)*, 1557-1566. doi: 10.1093/cercor/bhn189

Debey, E., De Schryver, M., Logan, G. D., Suchotzki, K., & Verschuere, B. (2015). From junior to senior Pinocchio: A cross-sectional lifespan investigation of deception. *Acta Psychologica, 160*, 58-68. doi: 10.1016/j.actpsy.2015.06.007

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charleton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129(1)*, 74-118. doi: 10.1037/0033-2909.129.1.74

Doherty-Sneddon, G., Bonner, L., & Bruce, V. (2001). Cognitive demands of face monitoring: Evidence for visuospatial overload. *Memory & Cognition, 29(7)*, 909-919. doi: 10.3758/BF03195753

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental Psychology, 38(3),* 438-445. doi: 10.1037/0012-1649.38.3.438

Doherty-Sneddon, G., & Phelps, F. G. (2005). Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition, 33(4)*, 727-733. doi: 10.3758/BF03195338

Evans, A. D., & Lee, K. (2011). Verbal deception from late childhood to middle adolescence and its relation to executive functioning skills. *Developmental Psychology, 47*(4), 1108-1116. doi: 10.1037/a0023425

Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition, 2(1),* 33-41. doi: 10.1016/j.jarmac.2013.02.002

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532-538. doi: 10.1037/a0015808

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* London: Sage.

Fisher, R., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, IL: Charles C Thomas.

Glenberg, A. M., Schröder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition, 26(4),* 651-658. doi: 10.3758/BF03211385

Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology, 37(1),* 60-74. doi: 10.1177/0022022105282295

Gordon, H. M., Lyon, T. D., & Lee, K. (2014). Social and cognitive factors associated with children's secret-keeping for a parent. *Child Development, 85*(6), 2374-2388. doi: 10.1111/cdev.12301

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews when training to detect deception works. *Law and Human Behavior, 30(5),* 603-619. doi: 10.1007/s10979-006-9053-9

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2014). Does training improve the detection of deception? A meta-analysis. *Communication Research,* 1-61. doi: 10.1177/00993650214534974

Irwin-Chase, H., & Burns, B. (2000). Developmental changes in children's abilities to share and allocate attention in a dual task. *Journal of Experimental and Child Psychology, 77*, 61-85. doi: 10.1006/jecp.1999.2557

Jack, F., Leov, J., & Zajac, R. (2014). Age-related differences in the free-recall accounts of child, adolescent and adult witnesses. *Applied Cognitive Psychology, 28*(1), 30-38. doi: 10.1002/acp.2951

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *International Journal of Psychophysiology, 52*, 7-21. doi: 10.1016/j.ijpsycho.2003.12.002

Knowles, W. B. (1963). Operator loading tasks. *Human Factors: The Journal of the Human Factors and Ergonomic Society, 5*(2), 155-161. doi: 10.1177/001872086300500206

Leach, A-. M., Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). "Intuitive" lie detection of children's deception by law enforcement officials and university students. *Law and Human Behavior, 28(6),* 661-685. doi: 10.1007/s10979-004-0793-0

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

Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance

telephone calles. *Legal and Criminological Psychology*, *20*(1), 129-146. doi: 10.1111/lcrp.12017

Liu, M, Granhag, P. A., Lanström, S., Roos af Hjelmsäter, E., Strömwall, L., & Vrij, A. (2010). "Can you remember what was in your pocket when you were stung by a bee?": Eliciting cues to deception by asking the unanticipated. *Open Criminology Journal, 3*(1), 31-36. doi: 10.2174/1874917801003010031

Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development*, *79*(4), 914-929. doi: 10.1111/j.1467-8624.2008.01167.x

Mangold (2015). INTERACT 14 User Guide. Retrieved from the Mangold International GmbH (Ed.) website: http://www.mangold-international.com

Mann, S., & Vrij, A. (2006). Police officers' judgements of veracity, tenseness, cognitive load and attempted behavioural control in real life police interviews. *Psychology, Crime & Law, 12(3)*, 307-319. doi: 10.1080/10683160600558444

Mann, S., Vrij, A., Shaw, D. J., Leal, S., Ewens, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2013). Two head are better than one? How to effectively use two interviewers to elicit cues to deception. *Legal and Criminological Psychology, 18*(2), 324-340. doi: 10.1111/j.2044-8333.2012.02055.x

Meissner, C. A. & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26(5)*, 469-480. doi: 10.1023/A:1020278620751

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*(1), 44-64. doi:10.1016/0010-0285(75)90004-3

O'Donohue, W., Benuto, L., & Fanetti, M. (2010). Children's allegations of sexual abuse: A model for forensic assessment. *Psychological Injury and Law, 3*(2), 148-154. doi: 10.1007/s12207-010-9075-y

Saykaly, C., Crossman, A., Morris, M., & Talwar, V. (2016). Question type and its effect on children's maintenance and accuracy during courtroom testimony. *Journal of Forensic Practice, 18*(2), 104-117. doi: 10.1108/JFP-01-2015-0010

Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31(1)*, 137-149. doi: 10.3758/SF03207704

Steller, M., & Köhnken, G. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp.135-154). Deventer, the Netherlands: Kluwer.

Strömwall, L. A., & Granhag, P. A. (2005). Truths: Effects on adults' judgments and reality monitoring scores. *Psychiatry, Psychology and Law, 12*(2), 345-356. doi: 10.1375/pplt.12.2.345

Talwar, V., & Crossman, A. (2011). From little white lies to filthy liars: The evolution of honesty and deception in young children. *Advances in Child Development and Behaviour, 40*(140), 139-179. doi: 10.1016/B978-0-12-386491-8.00004-9

Talwar, V., & Crossman, A. (2012). Children's lies and their detection: Implications for child witness testimony. *Developmental Review, 32*(4), 337-359. doi: 10.1016/j.dr.2012.06.004

Talwar, V., Crossman, A. M., Gulmi, J., Renaud, S. J., & Williams, S. (2009). Pants on fire? Detecting children's lies. *Applied Developmental Science, 13*(3), 119-129. doi: 10.1080/10888690903041519

Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development, 26*(5), 436-444. doi: 10.1080/01650250143000373

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development, 79(4),* 866-881. doi: 10.1111/j.1467-8624.2008.01164.x

Tye, M. C., Amato, S. L., Honts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science, 3(2),* 92-109. doi: 10.1207/s1532480xads0302_4

Vrij, A. (2005). Criteria-based content analysis: A qualitative reviews of the first 37 studies. *Psychology, Public Policy, and Law, 11*(1), 3-41. doi: 10.1037/1076-8971.11.1.3

Vrij, A. (2015). A cognitive approach to lie detection. In P.A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception: Current challenges and cognitive approaches* (pp. 203-229). Chichester: John Wiley & Sons.

Vrij, A., Fisher, R. P., & Blank, H. (2015). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*. doi: 10.1111/lcrp.12088

Vrij, A., Mann, S., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior, 32*(3), 253-265. doi: 10.1007/s10979-007-9103-y

Vrij, A., Mann, S., Leal, S., & Fisher, R. (2010). 'Look into my eyes': Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law, 16(4),* 327-348. doi: 10.1080/10683160902740633

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177. doi: 10.1080/14639220210123806

Williams, S., Leduc, K., Crossman, A., & Talwar, V. (2016). Young deceivers: Executive functioning and antisocial lie-telling in preschool aged children. *Infant and Child Development*. doi: 10.1002/icd.1956

Wilson, J.C. & Powell, M. (2001). *A guide to interviewing children: Essential skills for counsellors, police, lawyers and social workers.* Crows Nest: Allen & Unwin.

Zuckerman, M., DePaulo, B.M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1-59). New York: Academic Press.

941 Table 1

Mean and Standard Deviations for Total Number of Details as a Function of Veracity and Gaze Instruction

	True Report		False Report		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
Interviewer's face	178.77	152.50	79.00	42.91	125.32	117.72
Teddy bear's face	152.46	86.93	92.94	56.81	119.62	76.65
Control	117.15	99.39	114.80	119.17	115.89	108.42
Total	95.52	79.27	149.46	116.21	120.27	101.03

Table 2

Discrimination Accuracy (d') and Response Bias (β) as a Function of Gaze Instruction, Guidance Provision and Modality of Presentation

	d'		β	
	M	*SD*	*M*	*SD*
Gaze Instruction				
Look at interviewer's face	.43***	.80	1.21**	.49
Look at teddy bear's face	-.08	.75	1.02	.37
No instruction (control)	-.08	.71	1.01	.38
Guidance Provision				
Yes	.20*	.84	1.12	.48
No	-.02	.73	1.05	.36
Modality of Presentation				
Video-audio	-.04	.77	1.01	.35
Audio only	.22**	.80	1.16**	.48

Note. Statistical tests compared *d'* to 0 and β to 1.

* $p<.05$ ** $p<.01$ *** $p<.001$

942

943