The Cost of Detecting Deception: Judging Veracity Makes Eyewitnesses Remember a Suspect Less Accurately but With More Certainty

KERRI L. PICKEL*, BRITTNEY M. KLAUSER and HEATHER M. BAUER
Department of Psychological Science, Ball State University, Muncie, IN, USA

Summary: The current study extends previous research demonstrating the detrimental effects of divided attention during encoding on eyewitness memory. Previous data indicate that judging the veracity of a suspect causes witnesses to scrutinize him or her carefully and requires relatively high cognitive effort. We therefore hypothesized that performing this task while simultaneously observing the suspect should impair witnesses’ memory for his or her appearance and message while ironically inflating their certainty and other testimony-relevant judgments. Our results supported these predictions. Moreover, inducing witnesses to be suspicious about the suspect’s truthfulness (Experiment 1) and motivating them to judge veracity as accurately as possible (Experiment 2) amplified the memory impairment effect and further increased several testimony-relevant ratings. Additionally, compared with witnesses who incorrectly identified the suspect in a line-up, those who made a correct decision expressed greater certainty about their line-up accuracy and also provided higher ratings on some other testimony-relevant measures. Copyright © 2013 John Wiley & Sons, Ltd.

Previous eyewitness memory research has established the harmful effects of dividing attention during encoding. Specifically, witnesses’ reports become less accurate when some of their attentional resources are diverted away from the perpetrator and toward other stimuli or tasks. For example, a visible weapon can capture witnesses’ attention, which leads them to allocate fewer resources to encoding the perpetrator’s physical appearance and to provide poorer descriptions of him or her (Fawcett, Russell, Peace, & Christie, 2013; Hope & Wright, 2007; Pickel, Ross, & Truelove, 2006). Additionally, listening to a foreign-accented message as opposed to an unaccented message requires a greater level of cognitive processing, making witnesses likely to describe the speaker less accurately and to misidentify him or her voice (Pickel & Staller, 2012). These examples suggest that dividing attention deters witnesses from elaboratively encoding details related to the perpetrator. As a result, subsequent memory performance suffers.

Another situation in which a witness might have to divide his or her attention is when the perpetrator converses with someone, such as a victim or a bystander, and the witness wishes to determine whether the perpetrator is lying. A bank manager, for example, might evaluate the truthfulness of a robber who claims to have a hidden weapon and an accomplice when deciding whether to cooperate or to activate an alarm, all while trying to encode the robber’s physical appearance for a police report to be submitted later. Pickel, Kulig, and Bauer (2013) hypothesized that, because judging the veracity of a perpetrator is cognitively demanding (Porter, McCabe, Woodworth, & Peace, 2007; Reinhard & Sporer, 2008, 2010; Vrij, Granhag, & Porter, 2010), performing this task should lead to less accurate memory reports. Past research does indicate that detecting deception is effortful.

People believe (incorrectly) that certain cues can distinguish true statements from lies, and they try to monitor and evaluate them while observing the target. The cues include nonverbal and paralinguistic behaviors (e.g., gaze aversion, posture shifts, fidgeting, and speech pauses), as well as details within the message content (e.g., logical consistency and negation; The Global Deception Team, 2006; Mann, Vrij, & Bull, 2004; Taylor & Hick, 2007; Vrij, Akehurst, & Knight, 2006; Vrij et al., 2010). However, there are no cues that uniquely reveal deception across communicators and contexts, and people generally cannot detect deception well (Bond & DePaulo, 2006), but they persist in examining cues like those listed earlier to judge veracity.

Pickel et al. (2013) appealed to Wickens’ (2002, 2008) multiple resource theory to support their predictions about the effects of judging veracity on memory (see Pickel et al. for a more complete discussion). The theory proposes that cognitive resources are limited so that multiple tasks performed simultaneously must compete for those resources. Therefore, it may be difficult or impossible to complete two very demanding tasks at the same time. Drawing upon this model, Pickel et al. anticipated that witnesses would find it hard to encode information about a perpetrator elaboratively while judging that person’s veracity because of the high level of effort required by each of the two tasks. Furthermore, if the judgment task is prioritized, witnesses’ subsequent descriptions of the perpetrator should be impaired compared with those provided by witnesses who simply observed the perpetrator without judging veracity.

The researchers tested this hypothesis in successive steps. First, they verified that the veracity judgment task is cognitively demanding by showing that participants’ performance on a concurrent secondary task suffered when they judged the veracity of a target person rather than merely observing him. Next, they conducted two methodologically similar experiments. In the first, the participants watched a video of an actor who portrayed a robber ordering a bank manager to withdraw cash from the bank’s vault and hand it over. Claiming to have a gun hidden in his pocket and two accomplices nearby, the robber threatens to harm the manager and others in the bank if the manager does not follow his detailed instructions precisely or tries to activate the alarm. The participants were asked to assume the role of the manager and,
in keeping with that role, to pay close attention. Compared with controls, witnesses who were instructed before watching the video to decide whether the robber was telling the truth or lying about the accomplices and weapon remembered his appearance and message less accurately.

Moreover, inducing witnesses to be suspicious about the robber’s claims (by pointing out that he might be bluffing to gain cooperation) amplified this memory impairment. Pickel et al. (2013) hypothesized that the latter result occurred because suspicion encourages observers to scrutinize a target particularly closely for cues they believe will reveal deception (e.g., Buller, Strzyzewski, & Comstock, 1991; Burgoon, Buller, Ebesu, & Rockwell, 1994; Forrest, Feldman, & Tyler, 2004; Millar & Millar, 1998); therefore, suspicious observers allocate even more cognitive resources to the already demanding veracity judgment task. Consequently, even fewer resources are left over for encoding details about the target’s appearance and message. In support of this explanation, suspicious witnesses reported trying to use a greater number of cues to judge veracity compared with witnesses in the other two conditions. Regarding suspicious witnesses’ memory performance, it is important to note that the deception cues they choose to scrutinize, such as gaze aversion and speech pauses, are not details that police investigators would probably ask witnesses about, like hair color and height. Accordingly, because suspicious witnesses expend more resources on the veracity judgment task than nonsuspicious witnesses and because they attend to the wrong details, they less accurately remember forensically relevant information about the target’s appearance and message. Pickel et al. (2013) replicated these findings in their second experiment using a different video and crime scenario.

Pickel et al.’s (2013) study was the first to demonstrate that the act of judging veracity worsens subsequent memory for the target. The present research was designed in part to replicate this basic memory impairment effect using a new stimulus video (with a new target person, crime scenario, and message) as well as the finding that inducing witnesses to be suspicious about the target’s truthfulness exaggerates the effect. In addition, we aimed to extend the previous study in three ways. First, although Pickel et al. measured witnesses’ ability to describe the targets and recall their messages, they did not include a line-up task. In the current study, we attempted to show that judging veracity impairs witnesses’ line-up identifications as it does performance on the other memory measures.

Second, we hypothesized that, like inducing suspicion, motivating witnesses to try their best to judge veracity intensifies the memory impairment effect because high motivation leads witnesses to scrutinize the target even more closely for deception cues than they otherwise would and to allocate even more resources to the judgment task. As Porter et al. (2007) noted, few studies have investigated the impact of motivating observers to judge veracity accurately. However, the research that does exist suggests that high motivation increases the cognitive demands of the veracity judgment task. Reinhard and Sporer (2008) manipulated participants’ motivation (which the authors called ‘task involvement’) to evaluate a target’s credibility by emphasizing to those in the high motivation condition the importance of their performance for the success of the study and for future psychological research. Participants in the low motivation condition were told their data would be used in an introductory methods course. The researchers predicted that, compared with less motivated judges, highly motivated ones would experience a greater level of cognitive load as they used systematic (or central route) processing rather than less effortful heuristic (or peripheral route) processing (Chaiken & Trope, 1999) to evaluate the target’s truthfulness. The results supported the prediction; highly motivated participants utilized more complex judgment-relevant information, whereas those in the low motivation condition relied exclusively on easier-to-use nonverbal cues. Reinhard and Sporer (2010) later replicated their findings, and in the only other published study that addresses this issue, Forrest and Feldman (2000) similarly argued that motivation to judge veracity causes observers to work harder as they closely examine the target for cues associated with deception.

Our third goal related to extending Pickel et al.’s (2013) research was to measure certain dependent variables that might be affected by the veracity judgment task but were previously unexplored. Specifically, we were interested in witnesses’ expressions of their subjective feelings about the accuracy of their reports. For example, witnesses are commonly asked by police detectives how certain they are that they have accurately remembered the perpetrator and/or identified him or her in a line-up. Certainty (also termed ‘confidence’) is critically important during a criminal investigation. Police investigators and prosecutors are more willing to pursue charges against suspects identified by more certain witnesses (Bradfield, Wells, & Olson, 2002). Moreover, jurors’ verdicts are heavily influenced by the amount of certainty verbalized by witnesses (Cutler, Penrod, & Dexter, 1990; Luus & Wells, 1994), and in fact, certainty is the ‘primary factor’ they consider (Bradfield et al., 2002, p. 112).

The problem with this situation is that certainty is not perfectly correlated with accuracy, although jurors logically assume there is a strong positive relationship (Sporer, Penrod, Read, & Cutler, 1995). Instead, certainty is malleable, as a number of studies have shown. For example, one important variable that alters expressed certainty independently of accuracy is whether the witness received confirming feedback after identifying the suspect in a line-up. In the original research on this issue, Wells and Bradfield (1998) discovered that confirming feedback (‘Good, you identified the actual suspect’) given by the line-up administrator after witnesses made a false identification increased their subsequent certainty ratings compared with disconfirming or no feedback. The authors noted that this result is especially alarming given that post-identification feedback is legally permissible in the USA, as the courts have typically been more concerned about prohibiting line-up administrators from influencing witnesses’ choices. Furthermore, in addition to certainty, confirming feedback inflated several other ‘testimony-relevant judgments’, which Wells and Bradfield grouped into three general categories: (i) retrospective judgments about the witnessed event (e.g., ‘How good of a view did you get of the gunman?’); (ii) judgments related to the witnesses’ identification experience (e.g., ‘How easy or difficult was it for you to figure
out which person in the photos was the gunman?’); and (iii) summative judgments (e.g., ‘How willing would you be to testify in court that the person you identified was the [gunman]?’). The finding that confirming feedback increases ratings of certainty and other testimony-relevant judgments has been well replicated (e.g., Bradfield et al., 2002; Wells & Bradfield, 1999; Wells, Olson, & Charman, 2003; see also a meta-analysis by Douglass & Steblay, 2006), and it is particularly worrisome when one considers that the U.S. Supreme Court has explicitly advised that jurors consider some of the specific judgments that are distorted by feedback, including witnesses’ certainty, opportunity to view the perpetrator, and degree of attention paid to the perpetrator during the crime event (Manson v. Braithwaite, 1977). Courts in some other nations have made similar recommendations (Douglass & Steblay, 2006).

We hypothesized that the act of judging the veracity of a target would inflate testimony-relevant judgments. We expected this result to occur because judging veracity would cause witnesses to study the target for deception cues, giving them the false sense that they paid close attention to that individual in general, encoding various forensically related details connected to him or her, including his or her physical appearance and message. In turn, witnesses’ belief that they encoded the target elaboratively should lead them to express high levels of confidence in their memory of the target as well as inflated subjective ratings of their opportunity to observe and identify him or her and their willingness to stand by their reports. Moreover, we thought both suspicion and motivation would intensify these effects because they lead to still greater scrutiny of the target.

In sum, we expected judging veracity to produce two main results, both of which should be strengthened by inducing suspicion or motivating witnesses to try their best. First, memory for the target should be impaired, whether measured by description accuracy or line-up performance. Second, despite this reduction in memory, witnesses should ironically express greater certainty about their recollections, as reflected in their testimony-relevant judgments.

**EXPERIMENT 1**

In this experiment, we hoped to replicate, using new stimuli, Pickel et al.’s (2013) result that judging the veracity of a target individual impairs witnesses’ memory for his or her appearance and message and that inducing witnesses to be suspicious about the target’s truthfulness intensifies this effect. We created a new stimulus video that included a different crime scenario, target, and message than those used previously. In contrast with the earlier stimulus, the target’s message concerns his intentions rather than his past behavior; as a side note, investigating true and false statements about future purpose of his visit to the school. They were asked to suppose

We additionally predicted that judging veracity would inflate witnesses’ certainty in the accuracy of their memory performance (along with other testimony-relevant judgments) by leading them to conclude (mistakenly) that they had studied and elaboratively encoded many kinds of details about the suspect, not just deception cues. Furthermore, we thought suspicion would increase this effect by encouraging witnesses to allocate even more resources to the veracity judgment task.

Finally, we expected witnesses who correctly identified the suspect in the line-up to give higher ratings for the testimony-relevant judgments than witnesses who made incorrect identifications. Bradfield et al. (2002) obtained this result. They proposed that witnesses do not make testimony-relevant judgments online (i.e., during the witnessed event or the identification task) and instead construct their responses later, when the judgments are requested. As part of this process, witnesses can assess the degree of similarity between the line-up member they identified and their memory of the individual they saw during the witnessed event to draw inferences that allow them to make the testimony-relevant judgments. For example, witnesses who correctly identified the target would notice a strong resemblance between that person and their memory of the target, and so, they would conclude that they must have had a good view during the witnessed event.

**Method**

**Participants**

The participants (N = 136) were introductory psychology students at a Midwestern U.S. University who were fulfilling a course requirement. They ranged in age from 18 to 49 years (M = 19.07, SD = 2.78). Most participants were female (61%) and White (87%).

**Materials and procedure**

In groups of 2–10, the participants watched a video (running time 35 seconds) of a male actor portraying a visitor to a high school checking in with the vice principal and explaining the purpose of his visit to the school. They were asked to suppose that there have recently been multiple thefts of electronic equipment from the school, as well as vandalism. Because of lack of funds for security cameras, it has become the vice principal’s job to screen visitors in an effort to keep unauthorized individuals out and to protect students, teachers, and school property. The participants were instructed to imagine themselves as the vice principal and that, in keeping with this role, they should attend to the visitor carefully. In the video, the
visitor’s body is visible from the waist up, and he faces the camera. He is carrying a backpack over his shoulder, which would make it possible for him to conceal and carry away items stolen from the school. He claims that he intends to go to the art teacher’s room to talk with her about a project she wants to assign to her students.

Before playing the video, the experimenter instructed the witnesses who were randomly assigned to the control group simply to watch the video carefully. Those in the Judge Veracity condition were told they would need to decide whether the suspect was being truthful or lying about his reason for wanting to enter the school. Witnesses in the Judge Veracity/Suspicion condition were given the same task, but in addition, the experimenter pointed out to them that the suspect might not have legitimate business at the school, and therefore, he might be lying. Suspicion has been manipulated similarly in previous research (Buller, Strzyzewski, & Comstock, 1991; Forrest et al., 2004; McCormack & Levine, 1990; Millar & Millar, 1998; Toris & DePaulo, 1985).

After viewing the video, the witnesses completed a written questionnaire with five sections. The first section requested a binary veracity judgment regarding whether the suspect was truthful or deceptive about his reason for wanting to gain entrance to the school. The witnesses also used 11-point scales to rate their certainty in their judgment and, as a manipulation check, the amount of effort they put into making it. Higher numbers on the scales represented greater certainty and effort. We expected witnesses in the Judge Veracity/Suspicion condition to report that the task required more effort than those in the Judge Veracity condition. Because the control witnesses did not decide whether the suspect was lying while watching the video, their effort ratings were not directly comparable with the other two groups, and we made no specific prediction about their ratings.

The questionnaire’s second section inquired about the appearance of the suspect, including his clothing and physical features. Most of these questions offered alternatives and then asked witnesses to elaborate. For example, one item required witnesses to specify whether the suspect wore glasses and, if they said yes, to write down the type (e.g., sunglasses) and a description (e.g., black plastic frames). Witnesses could choose to indicate that they did not know the answer. Two coders who were blind to condition independently used a scoring key to identify the number of correct and incorrect details the witnesses reported on the questionnaire about the suspect’s appearance. As an example, the man wore khaki pants, so ‘khaki’ was one correct detail a witness could have reported when describing the pants. Any detail not listed on the key was considered incorrect. For example, some witnesses erroneously stated that the suspect wore jeans, and others reported the wrong shirt color. Interrater reliability was calculated using a sample of 50 questionnaires and was found to be high, \( r = .97 \).

The third section of the questionnaire was an open-ended item asking witnesses to recall the suspect’s message in as much detail as possible. As with memory for the suspect’s appearance, the coders counted the number of correct and incorrect details witnesses identified regarding the content of his message. As an example, the man stated that he wanted to see the art teacher, so ‘teacher’ was one correct detail a witness could have reported. Interrater reliability was calculated based on a sample of 50 questionnaires and was found to be high, \( r = .97 \).

After finishing the third section, the witnesses individually attempted to identify the suspect in a six-person, target-present, simultaneous photo line-up presented as two rows of three photos. We constructed the line-up according to expert recommendations (Malpass, Tredoux, & McQuiston-Surrett, 2007). The fillers were selected by the authors so that they generally resembled the suspect without matching him so closely that it would be difficult to discern a difference between the fillers and the suspect. The suspect’s position was randomly distributed in the line-up from one participant to the next, and no line-up member stood out from the others on the basis of one or more attributes. We opted for a simultaneous line-up because, although sequential line-ups are increasingly used by U.S. law enforcement agencies, the former type is still more common (National Institute of Justice, 2009). Although choosing a filler and rejecting a target-present line-up are both errors, the two actions often involve different cognitive processes (Sporer et al., 1995). For example, witnesses may identify a filler when they retrieve details that place that line-up member in the wrong context. In contrast, witnesses may fail to choose anyone because they realize they have no relevant memory of the target individual. We wanted to avoid the complication of having some witnesses choose and others reject the line-up. Therefore, following a commonly used strategy (e.g., Bradfield et al., 2002), we did not explicitly advise our witnesses that the suspect might not appear in the line-up. Consequently, all witnesses selected someone.

When the witnesses had completed the line-up task, they returned to the questionnaire. The fourth section asked for the 12 testimony-relevant judgments (Table 1) developed by Bradfield et al. (2002). The final section of the questionnaire required witnesses to report demographic information. Upon finishing the questionnaire, the witnesses were debriefed, thanked, and dismissed.

Results

Manipulation check

We predicted that, among witnesses who knew they would have to judge the suspect’s veracity, suspicious witnesses would report expending more effort on this task compared with those who were not made suspicious. A simple analysis of variance revealed a difference across the three conditions, \( F(2, 133) = 5.56, \ p = .005, \ \eta^2 = .08 \). A follow-up Student Newman–Keuls procedure with alpha set at .05 clarified that witnesses in the Judge Veracity/Suspicion condition put significantly more effort into the task (\( M = 6.78, \ SD = 1.65 \)) than those in the Judge Veracity (\( M = 5.84, \ SD = 2.21 \)) and control (\( M = 5.31, \ SD = 2.48 \)) conditions. Judge Veracity and control ratings did not differ significantly.

Memory for the suspect’s appearance

The correct and incorrect details witnesses reported were analyzed using a simple analysis of variance. When the main analysis was significant, a Student Newman–Keuls procedure
Table 1. Testimony-relevant judgments

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appearance certainty:</td>
<td>[When you described the suspect], how certain were you that the information you gave was accurate?</td>
</tr>
<tr>
<td>Message certainty:</td>
<td>[When you recalled the suspect’s message], how certain were you that the information you gave was accurate?</td>
</tr>
<tr>
<td>Line-up certainty:</td>
<td>[When you made your line-up decision], how certain were you that your decision was accurate?</td>
</tr>
<tr>
<td>Time to identify:</td>
<td>After you were first shown the line-up, how long do you estimate it took you to make a decision?</td>
</tr>
<tr>
<td>View:</td>
<td>How good of a view did you get of the man in the video?</td>
</tr>
<tr>
<td>Face:</td>
<td>How well were you able to make out the specific features of the man’s face?</td>
</tr>
<tr>
<td>Attention:</td>
<td>How much attention were you paying to the man’s face while watching the video?</td>
</tr>
<tr>
<td>Basis:</td>
<td>To what extent do you feel that you had a good basis to provide information about what the man looked like and said?</td>
</tr>
<tr>
<td>Easy to remember:</td>
<td>How easy or difficult was it for you to remember what the man looked like and said?</td>
</tr>
<tr>
<td>Willingness to testify:</td>
<td>On the basis of your memory of the man, how willing are you to testify in court that the information you provided about him is accurate?</td>
</tr>
<tr>
<td>Memory for strangers:</td>
<td>Generally, how good is your memory for the faces of strangers you have encountered on only one prior occasion?</td>
</tr>
<tr>
<td>Clear image:</td>
<td>How clear is the image you have in your memory of the man you saw in the video?</td>
</tr>
</tbody>
</table>

with alpha set at .05 was used to determine which conditions differed significantly.

In regard to the number of correct details, we found a significant effect (Table 2), $F(2, 133) = 28.36, p < .001, \eta^2 = .43$.

The Judge Veracity/Suspicion witnesses remembered fewer correct details than those in the other conditions, and Judge Veracity witnesses remembered fewer details than controls. We also discovered a significant effect for incorrect details, $F(2, 133) = 11.89, p < .001, \eta^2 = .18$. Witnesses who were made suspicious reported significantly more incorrect details than those in the other groups, and the Judge Veracity group reported significantly more incorrect details than controls.

Memory for the suspect’s message

The message details were analyzed using a simple analysis of variance, followed by a Student Newman–Keuls test with alpha set at .05. The number of correct details differed across groups (Table 2), $F(2, 133) = 10.64, p < .001, \eta^2 = .16$. Witnesses who were made suspicious about the suspect’s intent reported significantly fewer correct details than those in the other conditions. Also, witnesses in the Judge Veracity condition reported fewer correct details than controls.

A significant effect was also found for the number of incorrect details, $F(2, 133) = 15.24, p < .001, \eta^2 = .23$. Individuals in the Judge Veracity/Suspicion condition made significantly more errors in remembering the suspect’s message than those in the Judge Veracity and control conditions, and witnesses in the Judge Veracity group reported more incorrect details than controls.

Line-up accuracy

To look for differences in line-up accuracy across groups, we used a binary logistic regression with veracity judgment condition as the predictor variable. The omnibus test revealed that the predictor significantly affected line-up accuracy, $\chi^2(2, N = 136) = 8.04, p = .02$. Cramer’s $V = .24$ (Table 2). We next ran contrasts to determine which pairs of conditions differed significantly. The results indicated that a greater proportion of control than Judge Veracity/Suspicion witnesses identified the suspect, Wald $\chi^2(1, N = 91) = 7.47, p = .006$. Moreover, the proportion of Judge Veracity witnesses who made a correct choice did not differ from the proportion in the control condition, Wald $\chi^2(1, N = 90) = 2.01, p = .16$, or the Judge Veracity/Suspicion condition, Wald $\chi^2(1, N = 91) = 1.96, p = .16$. In sum, suspicious witnesses were less likely than controls to identify the suspect, and the performance of the Judge Veracity witnesses did not differ from either of the other groups.

Testimony-relevant judgments

We analyzed the 12 testimony-relevant judgments using a multivariate analysis of variance with two factors: veracity...
judgment condition (with three levels) and line-up accuracy (with two levels). When a main effect of veracity judgment condition was found to be significant, a Student Newman–Keuls procedure with alpha set at .05 was used to determine which conditions differed significantly. Results and statistics are displayed in Tables 3 and 4.

**Main effects of veracity judgment condition.** In comparison with witnesses in the control condition, those who were told they would have to judge veracity prior to watching the video had significantly higher ratings on five of the 12 testimony-relevant judgments, as shown in Table 3. Specifically, the Judge Veracity group reported greater certainty in their memory for the suspect’s message and their line-up decision, and they indicated that they had a more favorable view of the suspect, they paid more attention to his face, and they had a better image of him in their memory. When witnesses were made suspicious in addition to being told in advance that they would have to judge the suspect’s veracity, these effects were exaggerated, and the ratings rose even higher, as shown by significant differences between the Judge Veracity/Suspicion condition and the other two conditions.

Two testimony-relevant judgments were not significantly increased by judging veracity alone, but if the witnesses were also made suspicious, their ratings were higher than controls’. Specifically, compared with controls, Judge Veracity/Suspicion witnesses reported greater certainty in the accuracy of their description of the suspect and more willingness to testify in court.

**Main effects of line-up accuracy.** Main effects of line-up accuracy were discovered for several measures (Table 4). Compared with witnesses who made incorrect decisions, those who identified the suspect reported being more certain about their line-up performance, taking less time to make a line-up choice, paying more attention to the suspect’s face, and having a clearer image of the suspect in memory. There was also a marginally significant main effect of line-up accuracy such that individuals who correctly identified the suspect rated their view of him as slightly better than individuals who did not.

**Interactions.** For one testimony-relevant judgment (attention to the suspect’s face), we obtained a significant interaction, $F(2, 130) = 3.09, p = .05$. Simple effects analyses

Table 3. Experiment 1: Witnesses’ testimony-relevant judgments as a function of veracity judgment condition

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Control</th>
<th>Judge Veracity</th>
<th>Judge Veracity/Suspicion</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appearance certainty</td>
<td>6.36 (1.40)</td>
<td>6.89 (1.39)</td>
<td>7.30 (1.21)</td>
<td>3.11</td>
<td>.048</td>
<td>.05</td>
</tr>
<tr>
<td>Message certainty</td>
<td>7.11 (1.79)</td>
<td>7.78 (1.40)</td>
<td>8.65 (1.22)</td>
<td>7.96</td>
<td>.001</td>
<td>.11</td>
</tr>
<tr>
<td>Line-up certainty</td>
<td>5.60 (1.75)</td>
<td>6.36 (1.67)</td>
<td>7.09 (1.79)</td>
<td>9.13</td>
<td>&lt;.001</td>
<td>.12</td>
</tr>
<tr>
<td>Time to identify</td>
<td>3.49 (1.95)</td>
<td>3.69 (2.00)</td>
<td>3.78 (2.31)</td>
<td>0.08</td>
<td>.92</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>View</td>
<td>6.60 (1.50)</td>
<td>7.29 (1.62)</td>
<td>8.07 (1.41)</td>
<td>9.84</td>
<td>&lt;.001</td>
<td>.13</td>
</tr>
<tr>
<td>Face</td>
<td>6.51 (1.25)</td>
<td>6.00 (1.99)</td>
<td>6.04 (1.81)</td>
<td>1.77</td>
<td>.18</td>
<td>.03</td>
</tr>
<tr>
<td>Attention</td>
<td>5.13 (1.25)</td>
<td>5.87 (2.17)</td>
<td>6.80 (1.67)</td>
<td>11.09</td>
<td>&lt;.001</td>
<td>.15</td>
</tr>
<tr>
<td>Basis</td>
<td>6.60 (1.48)</td>
<td>6.18 (1.60)</td>
<td>6.43 (1.70)</td>
<td>0.53</td>
<td>.59</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Easy to remember</td>
<td>6.22 (1.77)</td>
<td>5.89 (1.85)</td>
<td>6.04 (1.90)</td>
<td>0.19</td>
<td>.82</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Willingness to testify</td>
<td>5.87 (1.62)</td>
<td>6.33 (1.57)</td>
<td>7.15 (1.80)</td>
<td>6.79</td>
<td>.002</td>
<td>.10</td>
</tr>
<tr>
<td>Memory for strangers</td>
<td>6.11 (2.18)</td>
<td>6.13 (2.00)</td>
<td>6.54 (2.27)</td>
<td>0.58</td>
<td>.56</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Clear image</td>
<td>6.04 (1.60)</td>
<td>6.69 (1.41)</td>
<td>7.43 (1.57)</td>
<td>10.10</td>
<td>&lt;.001</td>
<td>.13</td>
</tr>
</tbody>
</table>

Note: For the judgments, means are reported with standard deviations in parentheses. Ratings were made on an 11-point scale; higher numbers indicate greater quantities of each variable. Values in the same row that do not share the same alphabetical subscript differ significantly, $p < .05$. For the MANOVA, $df=2$ and 130.

Table 4. Experiment 1: Witnesses’ testimony-relevant judgments as a function of line-up accuracy

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Correct</th>
<th>Incorrect</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appearance certainty</td>
<td>6.82 (1.46)</td>
<td>6.91 (1.23)</td>
<td>0.01</td>
<td>.91</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Message certainty</td>
<td>7.67 (1.70)</td>
<td>8.22 (1.33)</td>
<td>1.02</td>
<td>.32</td>
<td>.01</td>
</tr>
<tr>
<td>Line-up certainty</td>
<td>6.60 (1.85)</td>
<td>5.87 (1.70)</td>
<td>11.15</td>
<td>.001</td>
<td>.08</td>
</tr>
<tr>
<td>Time to identify</td>
<td>3.32 (2.13)</td>
<td>4.30 (1.84)</td>
<td>6.11</td>
<td>.02</td>
<td>.05</td>
</tr>
<tr>
<td>View</td>
<td>7.40 (1.59)</td>
<td>7.17 (1.66)</td>
<td>3.58</td>
<td>.06</td>
<td>.03</td>
</tr>
<tr>
<td>Face</td>
<td>6.38 (1.65)</td>
<td>5.80 (1.80)</td>
<td>1.68</td>
<td>.20</td>
<td>.01</td>
</tr>
<tr>
<td>Attention</td>
<td>6.13 (1.80)</td>
<td>5.57 (1.93)</td>
<td>8.44</td>
<td>.004</td>
<td>.06</td>
</tr>
<tr>
<td>Basis</td>
<td>6.57 (1.52)</td>
<td>6.09 (1.70)</td>
<td>2.62</td>
<td>.11</td>
<td>.02</td>
</tr>
<tr>
<td>Easy to remember</td>
<td>6.26 (1.76)</td>
<td>5.65 (1.92)</td>
<td>3.05</td>
<td>.08</td>
<td>.02</td>
</tr>
<tr>
<td>Willingness to testify</td>
<td>6.46 (1.71)</td>
<td>6.46 (1.81)</td>
<td>0.92</td>
<td>.34</td>
<td>.01</td>
</tr>
<tr>
<td>Memory for strangers</td>
<td>6.29 (2.29)</td>
<td>6.22 (1.86)</td>
<td>0.18</td>
<td>.67</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Clear image</td>
<td>6.83 (1.65)</td>
<td>6.52 (1.56)</td>
<td>5.03</td>
<td>.03</td>
<td>.04</td>
</tr>
</tbody>
</table>

Note: For the judgments, means are reported with standard deviations in parentheses. Ratings were made on an 11-point scale; higher numbers indicate greater quantities of each variable. For the MANOVA, $df=2$ and 130.
clarified that there were significant differences across veracity judgment conditions for witnesses who correctly identified the suspect, \( F(2, 130) = 10.93, p < .001, \eta^2_p = .17 \), as well as for those who selected a filler, \( F(2, 130) = 6.92, p < .01, \eta^2_p = .11 \). However, among correct witnesses, the Judge Veracity/ Suspicion and the Judge Veracity ratings were higher than control ratings and did not differ from each other. In contrast, among incorrect witnesses, the ratings of the Judge Veracity/ Suspicion witnesses were higher than those of the other two groups, which did not vary significantly. No other interactions were significant (\( ps > .20 \)).

Veracity judgment
We used a binary logistic regression with veracity judgment condition as the predictor variable in order to compare the proportion of witnesses in each group who judged the suspect as deceptive. The omnibus test was significant, \( \chi^2(2, N = 136) = 7.29, p = .03 \), Cramer’s \( \chi = .23 \). The Judge Veracity/Suspicion group contained the highest proportion of witnesses who decided the suspect was lying (.80), and in the other two groups, an equal proportion thought he was lying (.58). Because the proportions in the latter two conditions were the same, there was no need to run contrasts, and we concluded that suspicious witnesses were more likely than witnesses in the other two conditions to judge the suspect as deceptive.

The witnesses also rated their confidence in the veracity judgment. A simple analysis of variance revealed no significant differences across veracity judgment conditions for this variable (\( p = .52 \)).

Discussion
As predicted, the control witnesses remembered the suspect’s message and appearance most accurately, followed by those who judged veracity, and the Judge Veracity/Suspicion group performed the worst. These findings suggest that judging veracity divides attention during encoding and leads to a decline in memory accuracy. Making witnesses suspicious exaggerated this memory impairment effect.

Although they remembered information less accurately than controls, witnesses who judged the suspect’s veracity reported higher ratings on five of their testimony-relevant judgments (message certainty, line-up certainty, attention, view, and clear image). As with the memory impairment effects, suspicion further increased these ratings. For two testimony-relevant judgments (appearance certainty and willingness to testify), witnesses who judged the suspect’s veracity provided ratings that were only slightly higher than controls’. However, if the witnesses were also made suspicious, judging veracity led to significantly increased ratings on these two measures. Further evaluation of the results can be found in the General Discussion.

EXPERIMENT 2
The first experiment replicated Pickel et al.’s (2013) finding that judging veracity impairs witnesses’ memory for a suspect’s appearance and message, and it produced the new result that judging veracity inflates testimony-relevant judgments. The main purpose of Experiment 2 was to examine whether, like inducing suspicion, motivating witnesses to judge veracity as accurately as possible exaggerates these effects. As explained earlier, we hypothesized that high motivation would encourage witnesses to expend more effort than they otherwise would as they studied the suspect, searching for deception cues. Our expectations were based on previous data indicating that motivation does in fact increase the amount of cognitive resources observers allocate to the judgment task (Forrest & Feldman, 2000; Reinhard & Sporer, 2008, 2010).

Method
Participants
Introductory psychology students (\( N = 125 \)) attending the same university as in Experiment 1 participated to fulfill a course requirement. They ranged in age from 18 to 33 years (\( M = 19.26, SD = 2.00 \)). The majority reported being female (65%) and White (81%).

Materials and procedure
In groups of 2–10, the participants watched the video used in the first experiment. Beforehand, they were randomly assigned to one of three conditions. The control witnesses were asked simply to watch the video carefully. Witnesses assigned to the Judge Veracity condition were told they would need to decide whether the suspect was being truthful or lying. Those assigned to the Judge Veracity/Motivation condition were likewise given the veracity judgment task, but they were also motivated to try their best to make an accurate decision. Specifically, these witnesses were told that some people are especially skilled at judging veracity if they try to do well, that they might be in this group, and that the three top-performing participants would receive a $50 award. Previous researchers (e.g., Porter et al., 2007) have similarly used monetary incentives to motivate participants to try to determine a target’s truthfulness as accurately as possible.

After viewing the video, the witnesses completed the same questionnaire as in Experiment 1. The procedures used for coding the correct and incorrect details witnesses reported about the suspect’s appearance were as before. Interrater reliability based on a sample of 50 questionnaires was high, \( r = .98 \). Regarding details witnesses reported about the content of the suspect’s message, interrater reliability based on a sample of 50 questionnaires was high, \( r = .96 \).

The witnesses completed the same line-up task as before, and everyone made a line-up identification. When they were finished, the witnesses were debriefed, thanked, and dismissed.

\(^1\) In reality, witnesses’ veracity judgments could not be classified as correct or incorrect because the target was an actor playing a role rather than an individual who was genuinely lying or telling the truth. Also, even if this were not the case, another problem is the strong probability that more than three witnesses would have made accurate judgments. However, the witnesses were not informed during the testing session that the target was only acting, and they were not told they would make only one veracity judgment, so the motivation information probably seemed plausible to them (none raised any concerns). The true situation was explained during debriefing, and three randomly selected witnesses received monetary awards after data collection was completed.
Results

Manipulation check
As in Experiment 1, witnesses were asked to rate how much effort they put into the veracity judgment as a manipulation check. We hypothesized that, among the witnesses who expected to judge veracity, motivated witnesses would scrutinize the suspect more carefully and thus would put more effort into the judgment. A simple analysis of variance, followed by a Student Newman–Keuls test, revealed a significant difference, $F(2, 122) = 12.98$, $p < .001$, $\eta^2 = .18$, such that witnesses in the Judge Veracity/Motivation condition provided higher ratings than witnesses in the other two conditions. In addition, participants in the Judge Veracity condition reported significantly fewer details than those in the other two conditions. Also, those in the Judge Veracity condition remembered fewer details than controls.

Regarding the number of incorrect details, a significant effect was also found, $F(2, 122) = 20.07$, $p < .001$, $\eta^2 = .25$, such that witnesses who judged veracity while motivated reported more incorrect details about the suspect’s appearance than those in the other conditions. In addition, participants in the Judge Veracity condition reported more incorrect details than controls.

Memory for the suspect’s appearance
The correct and incorrect details were analyzed using a simple analysis of variance, followed by a Student Newman–Keuls test, $\eta^2 = .14$, such that motivated witnesses who judged veracity reported significantly fewer details than those in the other two conditions. Also, those in the Judge Veracity condition remembered fewer details than controls.

Memory for the suspect’s message
The message details were analyzed using a simple analysis of variance, followed by a Student Newman–Keuls test, $\eta^2 = .39$, such that motivated witnesses who judged veracity reported significantly fewer details than those in the other two conditions. Also, those in the Judge Veracity condition reported significantly fewer details than controls.

Regarding the number of incorrect details, a significant effect was also found, $F(2, 122) = 38.38$, $p < .001$, $\eta^2 = .25$, such that witnesses who judged veracity while motivated reported more incorrect details about the suspect’s appearance than those in the other conditions. In addition, participants in the Judge Veracity condition reported significantly fewer details than controls.

We also found a significant effect for incorrect details, $F(2, 122) = 12.99$, $p < .001$, $\eta^2 = .18$, such that witnesses in the Judge Veracity/Motivation group made more errors than those in the other two conditions, which did not differ significantly.

Line-up accuracy
To compare line-up accuracy across groups, we used a binary logistic regression with veracity judgment condition as the predictor variable. The omnibus test was significant, indicating that the veracity judgment condition influenced line-up accuracy, $\chi^2(2, N = 125) = 7.12$, $p = .03$, Cramer’s $V = .24$ (Table 5). We then ran contrasts to determine which pairs of conditions differed significantly. We found that a greater proportion of control than Judge Veracity/Motivation witnesses identified the suspect, Wald $\chi^2(1, N = 84) = 6.50$, $p = .01$. In addition, the proportion of Judge Veracity witnesses who decided correctly did not differ from the proportion in the control condition, Wald $\chi^2(1, N = 83) = 2.50$, $p = .11$, or the Judge Veracity/Motivation condition, Wald $\chi^2(1, N = 83) = 1.10$, $p = .30$. Thus, the findings were that a greater proportion of control than Judge Veracity/Motivation witnesses correctly identified the suspect. The performance of the Judge Veracity condition fell between the other two conditions and did not differ from either.

Testimony-relevant judgments
As in Experiment 1, we analyzed the testimony-relevant judgments using a multivariate analysis of variance with veracity judgment condition and line-up accuracy as factors. When a significant main effect of veracity judgment condition was obtained, we used a Student Newman–Keuls procedure with alpha set at .05 to determine which conditions differed significantly. Results and statistics are presented in Tables 6 and 7.

Main effects of veracity judgment condition. In comparison with controls, witnesses in the Judge Veracity group reported higher ratings on five of the 12 testimony-relevant judgments (Table 6). Specifically, Judge Veracity witnesses reported greater certainty that they accurately described the suspect’s appearance and message and that they correctly chose the suspect in the line-up, and they claimed that they had a better view of the suspect and paid more attention to his face.

Increasing witnesses’ motivation in addition to telling them they would have to judge veracity amplified three of these effects. Witnesses in the Judge Veracity/Motivation condition provided higher ratings than witnesses in the other two conditions in terms of their certainty about their description of the suspect’s appearance and their line-up performance and the amount of attention they paid to the suspect’s face.

Table 5. Experiment 2: Witnesses’ memory for the suspect’s appearance and message and line-up identification accuracy as a function of veracity judgment condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Appearance details</th>
<th>Message details</th>
<th>Line-up</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Control</td>
<td>23.62 (4.36)</td>
<td>3.71 (1.64)</td>
<td>37.24 (10.24)</td>
<td>1.19 (1.04)</td>
</tr>
<tr>
<td>Judge Veracity</td>
<td>19.93 (2.10)</td>
<td>5.20 (1.78)</td>
<td>32.24 (9.14)</td>
<td>1.51 (1.52)</td>
</tr>
<tr>
<td>Judge Veracity/Motivation</td>
<td>17.50 (2.76)</td>
<td>6.57 (2.63)</td>
<td>27.38 (11.25)</td>
<td>2.95 (2.26)</td>
</tr>
</tbody>
</table>

**Note:** For appearance and message details, means are reported with standard deviations in parentheses. For line-up accuracy, the table shows the proportion of witnesses who correctly identified the suspect. Values in the same column that do not share the same alphabetical subscript differ significantly, $p < .05$. Cost of detecting deception 321

Two testimony-relevant judgments were not significantly inflated by simply telling witnesses to judge veracity, but when they were also motivated to try their best, their ratings were higher than control witnesses. In particular, motivated witnesses reported having a clearer image of the suspect in memory and a greater willingness to testify in court.

**Main effects of line-up accuracy.** We obtained significant main effects of line-up accuracy for six testimony-relevant judgments (Table 7). In comparison with witnesses who made incorrect decisions, those who identified the suspect reported being more certain about their description of the suspect’s appearance and their line-up selection. They also estimated taking less time to complete the line-up decision, reported having a better view of the suspect, said they paid more attention to the suspect’s face, and claimed to have a clearer image of the suspect in their memory. No other main effects of line-up accuracy were significant.

**Interactions.** None of the interactions between veracity judgment condition and line-up accuracy were significant (ps > .10).

**Veracity judgment.** We used a binary logistic regression with veracity judgment condition as the predictor to compare the proportion of witnesses in each group who thought the suspect was deceptive. The omnibus test was significant, \( \chi^2(2, N=125)=6.48, p=.04 \), Cramer’s V = .23. Therefore, we ran contrasts to determine which groups differed significantly. We discovered that the proportion of witnesses who decided the suspect was lying was higher in the Judge Veracity/Motivation condition (.61) than in the Judge Veracity condition (.57), \( \chi^2(1, N=84)=5.34, p=.02 \). Additionally, the proportion of Judge Veracity witnesses who judged the suspect as deceptive (.61) did not differ from the proportion in the control condition, \( \chi^2(1, N=83)=1.3, p=.23 \). In sum, the Judge Veracity/Motivation witnesses were significantly more likely than witnesses in the other conditions to conclude that the suspect was lying.

A simple analysis of variance that compared witnesses’ certainty regarding their veracity judgment across veracity judgment conditions revealed a significant effect, \( F(2, 122)=12.17, p<.001, \eta^2=.17 \). A Student Newman–Keuls test clarified...
that witnesses who judged veracity while motivated ($M = 7.31$, $SD = 1.79$) reported significantly more certainty than those in the control ($M = 5.55$, $SD = 1.64$) and Judge Veracity ($M = 6.07$, $SD = 1.60$) conditions. The latter two conditions did not differ significantly from one another.

**Discussion**

The results of Experiment 2 replicated the findings from the first experiment that witnesses’ memory reports were impaired when they were asked to judge the veracity of the suspect. Moreover, motivation, like suspicion, intensified the memory impairment effect.

As in Experiment 1, judging veracity inflated some testimony-relevant judgments even while reducing the accuracy of witnesses’ memory of the suspect. In comparison with controls, witnesses who judged veracity made higher ratings on five testimony-relevant judgments (appearance certainty, message certainty, line-up certainty, view, and attention). Motivating witnesses to try to judge veracity accurately enhanced some of these effects (appearance certainty, line-up certainty, and attention). Ratings for two testimony-relevant judgments (willingness to testify and clear image) were increased relative to the control condition only if witnesses were motivated when they decided whether the suspect was lying. Additional implications of the results will be discussed in the next section.

**GENERAL DISCUSSION**

**Replication of the basic memory impairment effect**

In two experiments, we replicated Pickel et al.’s (2013) finding that judging the veracity of a suspect impairs witnesses’ subsequent memory for his or her visual appearance and message. We used a new target person, message, and crime scenario, thus adding confidence that the result generalizes across situations. We propose that this basic memory impairment effect can be interpreted within the framework of multiple resource theory. Because attempting to encode information about the suspect and judging his or her veracity are both demanding tasks that collectively require more cognitive resources than the amount available, witnesses found it hard to complete the two tasks simultaneously. Given that the veracity judgment task was prioritized, memory performance declined.

We hypothesized that trying to determine whether the suspect was lying would also decrease line-up identification accuracy. However, although witnesses in both experiments who judged veracity were less likely than controls to identify the suspect, this difference was not significant. The failure to reach significance could be due to the fact that line-up identification, being a dichotomous variable, is a less sensitive measure of memory compared with the ability to describe details related to the suspect.

**The effects of suspicion and motivation while judging veracity**

We replicated the previous result (Pickel et al., 2013) that inducing witnesses to be suspicious of the suspect exaggerates the effect of judging veracity on their ability to remember the suspect’s appearance and message. Moreover, we discovered that suspicious witnesses identified the suspect less accurately than controls in a line-up. We contend that, among witnesses who know they will have to decide whether the suspect is lying, suspicious ones scrutinize the suspect more assiduously as they search for cues they believe are related to deception (Buller, Strzyzewski, & Comstock, 1991; Burgoon et al., 1994; Forrest et al., 2004; Millar & Millar, 1998). This higher level of scrutiny consumes more cognitive resources, leaving fewer resources available for encoding and remembering the suspect. Consistent with this proposed explanation, suspicious versus nonsuspicious witnesses who judged veracity indicated that they expended more effort during this task.

Although suspicious witnesses will study the suspect closely, they monitor cues they associate with deception, such as gaze aversion and speech rate, instead of details that could help police investigators identify and arrest the suspect. Thus, suspicious witnesses are not only allocating a high level of resources to the veracity judgment task, but they are also attending to the wrong details, so they subsequently remember the suspect less accurately.

Following a method commonly used in past studies (Buller et al., 1991; Forrest et al., 2004; McCormack & Levine, 1990; Millar & Millar, 1998; Toris & DePaulo, 1985), we induced suspicion by simply pointing out to witnesses that the suspect could be lying. It seems reasonable to assume that real witnesses observing a suspect could easily become suspicious on their own, without prompts from anyone else, when it occurs to them that the suspect might be deceitful. Therefore, witnesses in actual cases might be more likely to behave like those in the Judge Veracity/Suspicion condition, where stronger effects on memory accuracy emerged, rather than the Judge Veracity condition.

Like suspicion, motivating witnesses to do their best when determining whether the suspect was deceptive amplified the effect of judging veracity on the number of correct and incorrect details reported about the suspect’s appearance and the number of correct (but not incorrect) details reported about his message. Similarly, compared with controls, a smaller proportion of motivated witnesses identified the suspect in a line-up. Given these results, along with the finding that Judge Veracity/Motivation witnesses said they expended more effort than the Judge Veracity witnesses to decide whether the suspect was lying, it appears that motivation may act like suspicion to prompt witnesses to study a suspect particularly closely for deception cues, thus allocating more cognitive resources to the veracity judgment task than they otherwise would (Forrest & Feldman, 2000; Reinhard & Sporer, 2008, 2010) and less to the task of encoding information about the suspect. Combined with the fact that they do not attend to forensically relevant details, this process impairs witnesses’ memory for the suspect.

Witnesses’ motivation in many real-world cases may be naturally high. Although in some daily life circumstances people may prefer not to find out whether a target is lying, for instance when the target compliments one’s unusual new hairstyle, in situations that potentially involve a crime, there is clearly an incentive to uncover the truth (Vrij et al., 2010). For example, observers would like to know whether a politician giving an
interview on television truly has no connection to a bribery scandal, whether a traveler is being honest when he tells an airport security officer before boarding the plane that he has no malicious intentions, whether a street vendor’s merchandise is authentic rather than counterfeit, and whether a stranger asking to borrow a cell phone really has a personal emergency rather than plans to steal the phone. If real witnesses do feel highly motivated to detect deception, the effects of judging veracity on their memory accuracy should be intensified.

Testimony-relevant judgments

Judging veracity inflated several testimony-relevant judgments, even while decreasing the accuracy of witnesses’ memory for the suspect’s appearance and message. In both experiments, witnesses who judged veracity expressed more certainty than controls in their ability to remember the suspect’s message and in their line-up performance, and they reported having a better view of the suspect and paying more attention to his face. Additionally, in Experiment 1, witnesses in the Judge Veracity condition said they had a clearer image of the suspect in their memory compared with controls, and in Experiment 2, they reported more certainty in their description of the suspect’s appearance. For several judgments, inducing suspicion or motivating witnesses increased the ratings even further. For other measures, judging veracity elevated witnesses’ ratings nonsignificantly, but suspicion or motivation amplified the effect enough to make it significant. For example, in both experiments, witnesses in the Judge Veracity condition indicated that they were only slightly more willing than controls to testify in court against the suspect. However, suspicious witnesses (Experiment 1) and motivated witnesses (Experiment 2) were significantly more willing than controls to testify.

We suggest that having to determine whether the suspect was lying led witnesses to scrutinize him carefully as they searched for deception cues. This scrutiny gave them the false sense that they paid close attention to him and encoded various details related to him, including his physical appearance and message, and consequently, several testimony-relevant judgments became inflated. Furthermore, suspicion and motivation magnified this effect on some judgments by prompting even greater scrutiny of the suspect.

Our findings are noteworthy given that, as mentioned earlier, police investigators, prosecutors, and jurors are strongly influenced by witnesses’ expressions of certainty and related statements about their opportunity and ability to observe and remember a perpetrator (Bradfield et al., 2002; Cutler et al., 1990; Luus & Wells, 1994), and misplaced trust in these declarations can lead to wrongful convictions. As other researchers have demonstrated, witnesses’ confidence is malleable and imperfectly correlated with accuracy. Our results imply that judging veracity while concurrently trying to encode information about a suspect can increase witnesses’ certainty while ironically worsening their memory performance.

Five judgments (time to identify the suspect, ability to make out his face, having a good basis for providing information about him, ease of remembering him, and memory for strangers’ faces) did not vary across veracity judgment conditions in either experiment, and it is not clear why. In their meta-analysis, Douglass and Steblay (2006) calculated smaller effect sizes for the influence of confirming feedback (vs. no feedback) for variables that seemed more objective rather than subjective, which led them to conclude that objective measures might be less vulnerable to attempts to manipulate them. However, only one of our nonsignificant results (time to identify the suspect) involved an objective judgment. Additional research is needed to resolve this issue.

Like Bradfield et al. (2002), we found that witnesses who correctly rather than incorrectly identified the suspect in the line-up reported more certainty about their choice, and they estimated that they took less time to decide. These findings seem understandable; as the previous researchers noted, witnesses probably have some ability to evaluate their line-up performance, even if it is not perfect. In making other testimony-relevant judgments, however, witnesses answered general questions about their observation of the suspect or their memory skills rather than about their identification experience specifically, but we nevertheless obtained a main effect of line-up accuracy for three of these. In particular, correct witnesses in both experiments said they had a better view of the suspect and a clearer image of him in their memory, and they claimed to have paid more attention to his face. Bradfield et al. proposed an explanation to account for such findings. Pointing to supporting data obtained by Wells and Bradfield (1999), they argued that witnesses do not make testimony-relevant judgments while observing an event or a target individual. Instead, when asked for these judgments later, they must infer how much attention they paid, how good of a view they had, and so on. The accuracy of the line-up identification feeds into the inferences. For example, witnesses who correctly identified the suspect would notice a strong resemblance between him and their memory of him, so they would likely conclude that they must have had a good view.

Veracity judgments

Among the witnesses who expected to judge veracity in Experiment 1, those who were made suspicious were more likely to decide the suspect was lying. This outcome fits with existing data showing that, although observers surely realize that a target individual might be lying, presenting a brief statement pointing that out reduces the ‘truth bias’ (i.e., the tendency to assume that others are usually truthful; Burgoon et al., 1994; Kim & Levine, 2011; McCormack & Levine, 1990; Stiff, Kim, & Ramesh, 1992; Toris & DePaulo, 1985). Control witnesses’ decisions did not differ from those in the Judge Veracity condition, but their performance cannot be meaningfully compared with past results because, in previous studies of the effects of suspicion, the participants knew while observing the target that they would later judge his or her veracity, whereas our control witnesses did not.

In Experiment 2, the Judge Veracity/Motivation witnesses were more likely than the Judge Veracity witnesses to conclude the suspect was lying. We are aware of only one other study that has analyzed the effect of motivation to detect deception on the truth bias; in contrast with the present result, Porter et al. (2007) reported no change in response bias as a function of motivation. Additional research is needed to clarify the relationship between motivation and the truth bias.
Limitations and future directions

In the present study, witnesses who judged veracity while suspicious (Experiment 1) or while motivated to do their best (Experiment 2) were less likely than controls to identify the suspect in a line-up. However, all of our witnesses viewed a target-present line-up. If some witnesses had been presented with a target-absent version, they would have had to reject the line-up in order to make a correct response. Choosing someone from a line-up and not choosing anyone, however, can represent very different processes, even when both decisions are correct (Sporer et al., 1995). Choosing occurs when the familiarity of a line-up member exceeds some threshold and the witness retrieves contextual details that place that person at the crime scene as the perpetrator. In other words, the witness correctly believes that he or she can remember the perpetrator. In contrast, a witness who rejects the line-up might do so upon realizing that he or she does not recall what the perpetrator looks like. Thus, judging veracity (with or without suspicion and motivation) might not affect witnesses’ performance on target-present and target-absent line-ups in the same way. Future research could explore this question.

Investigators could also attempt to replicate our results using a genuinely deceptive or honest target person rather than an actor playing a role. Although observers typically show little capacity for distinguishing lies from truth, judging the veracity of actual liars or truth tellers could conceivably be more or less cognitively demanding than judging actors, and the level of effort could in turn affect witnesses’ memory for the target.

Another variable that could moderate the effect of judging veracity is the degree to which witnesses interact with the suspect. Our participants merely watched the suspect, but real witnesses might sometimes converse with him or her as they try to detect deception. They may try to formulate diagnostic questions, critically evaluate the suspect’s answers, and then come up with follow-up questions based on the information collected so far while simultaneously monitoring the cues they believe are important. This process probably requires more cognitive effort than merely observing the suspect (Buller, Strzyzewski, & Hunsaker, 1991; Vrij et al., 2010), and it might more severely impair subsequent memory for him or her.

Conclusion

Previous studies have demonstrated that stimuli (e.g., weapons) and tasks (e.g., comprehending accented speech) that divide witnesses’ attention during encoding can have a detrimental effect on their memory of target individuals. The present data extend that research by showing that judging the veracity of a suspect while simultaneously observing that person impairs witnesses’ memory for his or her appearance and message while ironically inflating their certainty and other testimony-relevant judgments. Additionally, making witnesses suspicious about the suspect’s truthfulness and motivating them to judge veracity as accurately as possible augmented the memory impairment effect and further increased several testimony-relevant ratings.

It seems likely that witnesses spontaneously initiate the process of judging the veracity of a suspect or perpetrator based on a number of situational factors. Thus, whether or not the judgment process occurs is outside of the control of police investigators. However, investigators could attempt to take this variable into account by asking witnesses, as part of the interview, about their mental activities during the critical event (e.g., ‘What were you thinking about while this was happening?’). When witnesses disclose that they were wondering whether the target individual was lying or telling the truth, detectives could treat their reports especially cautiously and could make extra effort to find corroborating evidence. Moreover, it could be helpful to appreciate that judging a suspect’s veracity may not only impair witnesses’ memory for the suspect but also lead them to express greater certainty and to claim, for example, that they paid more attention to the suspect and had a better view of him. In this way, investigators and prosecutors could potentially use information provided by witnesses more wisely as they decide how to proceed with a case and whether to file charges against a particular suspect.

ACKNOWLEDGEMENTS

We thank Chris Altman, Breona Barth, Cassie Childress, Brad Root, and Tully Roll for their assistance with this project.

REFERENCES


