Judging veracity impairs eyewitnesses’ memory of a perpetrator

Kerri L. Pickel, Teresa C. Kulig, and Heather M. Bauer

Department of Psychological Science, Ball State University, Muncie, IN, USA

Witnesses to crimes sometimes perform cognitively demanding tasks while simultaneously observing a perpetrator. This division of attentional resources can cause witnesses to remember the perpetrator less accurately. We hypothesised that judging the veracity of a target individual can impair subsequent memory for his or her appearance and message. In Experiment 1, we demonstrated that the veracity judgement task is cognitively demanding by having participants perform a concurrent secondary task. In three additional experiments, we confirmed that witnesses who judged the veracity of a target remembered his or her appearance and message less accurately than witnesses who simply observed the target. We also extended this result by showing that suspicion amplified the memory impairment effect, apparently by inducing witnesses to allocate even more resources to the judgement task (Experiments 2a and b), and that witnesses’ memory was less accurate when they used a cue within the message content rather than a nonverbal cue to judge veracity (Experiment 3). Contrary to our prediction, however, witnesses who monitored two cues versus one did not display worse memory performance.

Keywords: Eyewitness Memory; Deception Detection; Divided Attention.

Dividing attention at encoding interferes with eyewitnesses’ subsequent memory for information related to a perpetrator. For example, when a visible weapon captures their attention, witnesses allocate fewer attentional resources to the task of processing the perpetrator’s features and clothing (Hope & Wright, 2007; Loftus, Loftus, & Messo, 1987; Pickel, Ross, & Truelove, 2006). Consequently, the witnesses tend to describe the perpetrator less accurately than they would without the weapon’s presence (for a review, see Fawcett, Russell, Peace, & Christie, 2013). As a second example, witnesses trying to comprehend a message spoken by a perpetrator with a foreign accent (as opposed to no accent) provide poorer physical descriptions and are more likely to misidentify the perpetrator’s voice (Pickel & Staller, 2012). In both of these examples, reduced memory performance can be attributed to the fact that some attentional resources are diverted away from the perpetrator, so that details associated with him or her are not encoded or are encoded less elaboratively or accurately.

In addition to these processes, there are others that might occur simultaneously while witnesses observe a perpetrator. In situations in which the perpetrator converses with a victim, bystander, or collaborator, witnesses might want to determine whether the perpetrator is lying. For instance, when evaluating whether to cooperate versus attempt to activate an alarm, a bank supervisor might assess the truthfulness of a robber who claims to have accomplices and weapons. Or, a store manager might have to decide whether to believe a customer who denies shoplifting. The manager might let the customer go, only to
discover evidence of the shoplifting later. In each case, at some point after the interaction the witness would need to remember the perpetrator’s appearance for a police report.

As when confronted with a weapon or when listening to an accented message, judging veracity may also demand considerable attentional resources and lead to less accurate memory for the perpetrator. Previous research suggests that judging veracity is cognitively effortful (Porter, McCabe, Woodworth, & Peace, 2007; Reinhard & Sporer, 2008, 2010; Vrij, Granhag, & Porter, 2010). To detect lies, people inspect cues they notice while observing or interacting with the target individual as well as information they gather or come across later, such as a report from a third party or physical evidence (Park, Levine, McCormack, Morrison, & Ferrara, 2002). Previous research has focused primarily on lie detectors’ use of cues observed while the target is speaking, rather than information obtained afterwards. This research has shown that most people believe certain cues can distinguish truth from deception. Many of these cues are nonverbal or paralinguistic, such as gaze aversion, posture shifts, hand movements, fidgeting, and speech pauses (The Global Deception Team, 2006; Mann, Vrij, & Bull, 2004; Taylor & Hick, 2007; Vrij, Akehurst, & Knight, 2006), but others involve the content of the target’s message (e.g., logical consistency, negative statements; Mann et al., 2004; Vrij et al., 2006, 2010). Most cues people use are not diagnostic, and in fact there is no indicator that uniquely signals deception across individual communicators and circumstances (DePaulo et al., 2003; Vrij et al., 2010). As a result, attempts to detect deception using these cues are usually not very successful (Bond & DePaulo, 2006). Nevertheless, people do try to monitor and evaluate the cues they associate with deceit, and this task may be difficult to perform while simultaneously observing the target’s physical appearance and interpreting his or her message.

Multiple resource theory (Wickens, 2002, 2008), which was developed to understand dual-task performance, can be applied to this situation (see Figure 1). The model assumes that cognitive resources are limited, allocatable, and organised into distinct dimensions, each with two levels (e.g., the “perceptual modalities” dimension,

![Multiple Resource Theory Diagram](image)

**Figure 1.** Illustration of multiple resource theory. The two easy tasks can be performed simultaneously with relatively little interference because they mainly draw resources from different perceptual modality-specific pools. However, the demanding tasks will likely deplete the resources in their respective modality-specific pools and require supplementation from the general pool. Because they compete for general resources, it may be impossible to complete the two concurrently, regardless of whether they primarily utilise distinct perceptual modalities. We contend that judging an individual’s veracity and encoding details about him or her are both very effortful, although probably neither task is exclusively visual or auditory, in contrast to the two demanding tasks in this example.
with a visual and an auditory level). Two tasks can be performed concurrently only if there are enough available resources to support both. Success is more likely if they draw resources from different rather than the same levels along a particular dimension. However, the tasks can still compete for common resources, as when very difficult tasks deplete their respective level-specific resource pools and require supplementation from the larger common pool. Thus, whether the tasks involve different levels or not, it may be impossible to perform them simultaneously if they are sufficiently taxing. Based on this framework, we developed three hypotheses. First, judging veracity requires a high level of cognitive effort. Second, because the task is demanding, it should interfere with concurrent attempts to encode details associated with the target (which is itself a cognitively effortful task; Pickel & Staller, 2012) so that witnesses’ subsequent reports are less accurate than those of control witnesses. Third, additional factors that increase the amount of cognitive resources allocated to the veracity judgement task should further impair witnesses’ memory performance.

We examined these hypotheses in the current study. To test the first one, we verified that judging veracity is more cognitively demanding than merely observing a target person by measuring participants’ performance on a secondary task (Experiment 1). In the subsequent experiments, participants watched a video of a target while adopting the role of witnesses who were instructed either simply to watch the target or to decide whether he or she was lying. Afterwards, the witnesses attempted to remember the target’s appearance and message so that we could determine whether judging veracity would reduce memory accuracy, as predicted by our second hypothesis.

In relation to our third hypothesis, we investigated certain variables associated with the veracity judgement task that we thought would affect the amount of resources allocated to that task. In Experiments 2a and 2b, we induced some witnesses to be suspicious about the target’s truthfulness. Previous researchers have suggested that suspicion increases the cognitive demands of the veracity judgement task by leading observers to scrutinise the target more closely for cues ostensibly related to deception (e.g., Buller, Strzyzewski, & Comstock, 1991; Burgoon, Buller, Ebesu, & Rockwell, 1994; Forrest, Feldman, & Tyler, 2004; Millar & Millar, 1998). For example, Forrest et al. (2004) found that suspicion increased participants’ self-reported level of attention to the target individuals. Moreover, while interacting with targets, Buller, Strzyzewski, and Comstock’s (1991) suspicious participants displayed behaviours indicative of greater cognitive effort relative to controls; they spoke more slowly and less fluently, their statements were less clear, and they had longer response latencies. In sum, suspicion seems to make witnesses expend more cognitive effort as they carefully and deliberately search for indicators of deception. Note, however, that deception cues (e.g., gaze aversion, speech pauses) are not details that police investigators normally want from eyewitnesses (e.g., hair colour or height). We therefore predicted that, when asked for details relevant to a police investigation, control witnesses would describe the target most accurately. Furthermore, of the two groups that were instructed to judge veracity, we expected suspicious witnesses to remember the target less accurately.

In Experiment 3, we tested the hypothesis that the number and type of cues witnesses use to judge veracity would influence the amount of resources consumed by the veracity judgement task and thus witnesses’ memory for information related to the target. Previous research suggests that processing cues within the content of a target’s message, as opposed to nonverbal cues, requires more cognitive resources. Reinhard and Sporer (2008; see also Reinhard & Sporer, 2010) manipulated cognitive load by asking some participants to perform a distracting task while observing a target and judging her veracity. Unlike control participants, those working under a high cognitive load were unable to utilise content cues that would have entailed analysing the plausibility and consistency of the target’s statements. However, both groups of participants used nonverbal cues (gaze aversion, adaptors, and posture shifts). Based on these results, we predicted that witnesses who used content cues to judge veracity would subsequently recall the target less accurately than witnesses using nonverbal cues.

Regarding the number of cues used, it seems logical that attempting to monitor and evaluate multiple cues versus a single cue might be more demanding. Although observers sometimes intuitively zero in on one particular cue and rely on it nearly exclusively (The Global Deception Team, 2006; Vrij et al., 2010), experts and training manuals recommend using multiple cues to detect deception. Moreover, a strategy that involves examining a combination of cues in a flexible
manner probably does increase detection accuracy (Porter & ten Brinke, 2010; Vrij et al., 2010). Given that people interested in improving their lie detection skills may seek out and try to use this expert advice, it is important to discover whether its implementation will produce poorer memory for a target. We anticipated that witnesses using multiple cues rather than a single cue would remember the target less accurately.

**EXPERIMENT 1**

The purpose of this experiment was to verify that judging veracity demands significantly more attentional resources than simply observing a target person. Participants in different conditions either merely watched a video of a target delivering a short monologue or tried to determine whether he was lying as they watched him. All participants were asked to prioritise their assigned task while completing a simultaneous secondary task. We predicted that, compared to control participants, those who judged veracity would have fewer resources available for the secondary task, thus performing that task more poorly.

**Method**

*Participants.* Introductory psychology students (*N* = 85) at a Midwestern US university participated to fulfil a course requirement. They ranged in age from 18 to 61 years (*M* = 19.71, *SD* = 5.57); 79% were women and 88% self-identified as White.

*Materials and procedure.* In groups of up to six, the participants watched a video (Pickel & Staller, 2012; approximate running time 1 min 45 s) of a male actor portraying a robber standing in a bank manager’s office and instructing the manager to withdraw a specific amount of cash from the bank’s vault and hand it over. He warns that he will harm the manager or other people in the bank if his detailed directions are not followed precisely. The robber’s head, torso, and arms can be seen, and he is facing the camera. Although he claims to have a gun and two accomplices, no weapon is visible, and no other individuals are shown. The participants were instructed to imagine themselves as the bank manager in the video and that, in keeping with this role, they should attend to the robber carefully. They were also told to work on a secondary task that required them to remember a word list. Before watching the video, they listened to an audio recording of nine 3-syllable words (musician, telephone, avenue, foreigner, radio, destroyer, socialist, photograph, convention; Jalbert, Neath, Bireta, & Surprenant, 2011) with the goal of maintaining them in memory until after the video ended. We chose this secondary task for two reasons. First, it measures memory performance and is therefore similar to the dependent variable we thought would be impaired by judging veracity. Second, we wanted to avoid asking participants to complete a task (such as monitoring visual stimuli) that would inhibit encoding of the target by requiring them to look away from the video. Our hypothesis was that judging veracity should harm memory even when participants are free to observe the target individual as much as they wish.

We manipulated whether the participants judged the veracity of the robber. Participants randomly assigned to the Judge Veracity condition were told they should decide whether he was lying or telling the truth. Control participants were not given this instruction. All participants were asked to prioritise their assigned task of judging veracity or simply watching the video. After viewing the video, participants completed a written form that asked them to recall the list words and to provide demographic information.

**Results and discussion**

As predicted, participants instructed to judge veracity recalled fewer words (*M* = 3.62, *SD* = 1.31) than controls (*M* = 4.33, *SD* = 1.25), *F*(1, 83) = 6.50, *p* = .01, *η*² = .07. Most participants did not report any words that were not on the studied list, but those who judged veracity (*M* = 0.31, *SD* = 0.52) were slightly more likely than controls to commit this error (*M* = 0.14, *SD* = 0.35), *F*(1, 83) = 3.16, *p* = .08, *η*² = .04. In sum, participants’ performance on the secondary task supported the hypothesis that judging veracity requires significantly more cognitive resources than merely observing a target.

**EXPERIMENTS 2a AND 2b**

Experiment 1 provided direct evidence that the veracity judgement task is cognitively effortful. In the next pair of experiments, our objective was to
extend that result by demonstrating that witnesses who are trying to determine whether a target is being truthful or lying will subsequently encode and remember information associated with the target less accurately compared to witnesses who are not trying to judge veracity. We also wanted to test the hypothesis that, among witnesses who judged veracity while observing the target, those who were induced to be suspicious would allocate relatively more cognitive resources to this task and would therefore remember the information about the target even more poorly.

In both experiments, participants attending the same university as those in Experiment 1 watched a video depicting a target individual who delivered a brief message and later tried to remember his or her appearance and message. Beforehand, the witnesses were instructed simply to watch the video, to judge the veracity of the target, or to judge veracity after being made suspicious about the target’s truthfulness. As explained earlier, we expected suspicion to make the veracity judgement task even more demanding. Therefore, we anticipated that control witnesses would most accurately remember the information, followed by those who judged veracity, and we thought those who judged veracity while suspicious would exhibit the worst memory performance.

Besides reporting information about the target, the witnesses also provided two other responses. First, they judged the target’s veracity. In most situations, people tend to assume others are being truthful (Bond & DePaulo, 2006; Buller, Strzyzewski, & Hunsaker, 1991; Granhag & Stromwall, 2001; Vrij et al., 2010), but suspicion reduces the truth bias (Burgoon et al., 1994; Kim & Levine, 2011; McCormack & Levine, 1990; Stiff, Kim, & Ramesh, 1992; Toris & DePaulo, 1985). Accordingly, we predicted that, among witnesses who were instructed to judge veracity, inducing suspicion would increase the likelihood that they would decide the target was lying. In previous studies about suspicion, the participants realised while observing the target that they would have to make a veracity judgement. Our control witnesses did not. Therefore, we had no basis for a specific prediction about how the control witnesses’ judgements would compare with those made by witnesses in the other two conditions.

Second, we asked witnesses to identify the cues they used to judge veracity. For those who were instructed to judge veracity, suspicion should induce them to expend more cognitive effort in scrutinising the target thoroughly and critically. Therefore, we hypothesised that suspicious witnesses might attempt to use more cues in general (i.e., either nonverbal cues or cues within the message content). However, we particularly expected a difference in terms of content cues. Because monitoring and evaluating content cues requires a relatively large amount of cognitive resources (Reinhard & Sporer, 2008, 2010), non-suspicious witnesses should be less motivated to utilise cues of this type.

This prediction that suspicious witnesses would attend closely to cues within the content of the target’s message may seem to contradict the hypothesis (described earlier) that these witnesses would remember the message less accurately than those in other conditions. However, in Experiments 2a and 2b the messages included not only statements that witnesses were asked to evaluate for veracity but also information that was not directly relevant to these statements. For example, the witnesses in Experiment 2a, who watched the bank robber video, had to decide whether the target actually had accomplices and a weapon as he claimed, but they also heard him make several comments about the cloth bags he wanted the manager to fill with cash. Thus, even if witnesses chose to utilise content cues, it does not follow that they should necessarily recall the entire message accurately.

Thus far, the discussion of cue usage has focused on witnesses who were asked in advance to judge the target’s veracity. Control witnesses, in contrast, found themselves in a substantially different situation. They had to make a veracity judgement retrospectively, relying on their memory of the cues that were present within the video. Given that they did not anticipate making a veracity judgement, it seems unlikely that these witnesses would monitor and encode many cues, especially ones that are more challenging to use. Therefore, we thought control witnesses would report fewer cues (especially content cues) than the suspicious witnesses who were instructed to judge veracity, and perhaps also fewer cues than those who were not made suspicious.

Experiments 2a and 2b followed the same design and procedure, but they used different stimulus videos, with different target individuals, messages, and crime scenarios. Thus, we intended Experiment 2b to be a systematic replication of Experiment 2a.
Experiment 2a

Method

Participants. Introductory psychology students (*N* = 110) participated to fulfill a course requirement. They ranged in age from 18 to 31 years (*M* = 19.03, *SD* = 1.68). Most (75%) were women, and most (90%) self-identified as White.

Materials and procedure. In groups of up to 10 students, the participants watched the video used in Experiment 1. Before starting the video, the experimenter asked the participants to adopt the role of a bank manager and explained that the video would depict a man who has entered the manager’s office demanding cash. The experimenter emphasised that, like a real manager would in this situation, they should pay close attention. At this point, the witnesses were randomly assigned to one of three conditions. Control witnesses were given no additional instructions. Those in the Judge Veracity condition were told they would need to decide whether the robber was being truthful or deceptive. Witnesses in the Judge Veracity/Suspicion condition were given the same task but were also told that bank robbers sometimes “bluff”, or lie, in order to gain cooperation from their victims. Previous researchers (Buller, Strzyzewski, & Comstock, 1991; Forrest et al., 2004; McCornack & Levine, 1990; Millar & Millar, 2008; Toris & DePaulo, 1985) have similarly induced suspicion in participants by pointing out that target individuals might lie.

After watching the video, the witnesses completed a written questionnaire with several sections. One section asked witnesses to remember the robber’s appearance, including his physical features (e.g., hair colour and body build) and his clothing (e.g., jacket colour and style). Most questions offered alternatives and requested elaboration. For example, one item asked witnesses to indicate whether the robber was or was not wearing glasses and, if yes, to describe the type (e.g., sunglasses) and appearance (e.g., black frames). Witnesses were free to write that they did not know the answer. A second section of the questionnaire required witnesses to recall the robber’s message in as much detail as possible.

In a third section, the witnesses made a binary judgement regarding whether the robber was truthful or lying when he claimed to have a weapon in his possession and accomplices, and they rated their confidence in that decision on an 11-point scale, where higher numbers represented greater confidence. In addition, witnesses responded to an open-ended item asking them to list the cues they used to make the judgement. The questionnaire ended with a demographics section.

Results

Except where noted, all dependent variables in this section were analysed using a simple analysis of variance. When the main analysis was significant, we used a post hoc Ryan REGWQ procedure with alpha set at .05 to determine which conditions differed significantly. This procedure utilises the Studentised range distribution and maximises power to examine all pairwise comparisons while holding the familywise error rate at alpha (Howell, 2007).

Memory for the robber’s appearance. Using a scoring key that identified the correct details related to the robber’s appearance, two coders independently determined the number of correct and incorrect details witnesses reported on the questionnaire (any detail not identified on the scoring key was counted as incorrect). As an example, the robber wore a black jacket, so “black” was one correct detail a witness could have reported. Interrater reliability was high; based on a sample of 50 questionnaires, *r* = .96. We analysed the scores calculated by one arbitrarily chosen coder.

Regarding the number of correct details, we found a significant difference, *F*(2, 107) = 45.67, *p* < .001, η² = .46 (see Table 1). The Judge Veracity/Suspicious witnesses remembered fewer details than those in the Judge Veracity condition, and the latter witnesses remembered fewer details than controls.

We also obtained a significant difference in incorrect details, *F*(2, 107) = 32.02, *p* < .001, η² = .37. Witnesses who judged veracity while suspicious made more errors than those who performed this task without suspicion, followed by control witnesses.

Memory for the robber’s message. As with memory for appearance, two independent coders used a scoring key to count the number of correct and incorrect details witnesses remembered about the content of the message. Interrater reliability was high (*r* = .97 with a sample of 50 questionnaires). Regarding correct details, we obtained no
significant differences ($p = .56$; see Table 1). However, the number of incorrect details varied across conditions, $F(2, 107) = 19.59$, $p < .001$, $\eta^2 = .27$. For witnesses who judged veracity, those who were suspicious reported more incorrect details than those who were not. Control witnesses reported the fewest incorrect details.

### Veracity judgements

We used a chi-square test to examine the proportion of witnesses who judged the robber as deceptive. There was a difference across the three groups, with the proportion being highest in the Judge Veracity/Suspicion condition and the lowest in the control condition, $\chi^2(2, N = 110) = 7.73, p = .02$, Cramer’s $V = .27$ (see Table 2). We had predicted that, for witnesses who expected to judge veracity, those who were suspicious would be more likely to decide the robber was lying. Therefore, we conducted an additional chi-square analysis using only the two experimental conditions. The prediction was supported, $\chi^2(1, N = 75) = 4.24, p = .04$, Cramer’s $V = .24$.1

The witnesses rated their confidence in the veracity judgement. This variable did not differ across conditions ($p = .19$).

### Reported use of cues

Witnesses were asked to list the cues they considered in deciding whether the robber was being deceptive. Our two coders worked independently to classify each cue as visual (e.g., gaze aversion), auditory (e.g., speech pauses), or relating to message content (e.g., the

#### TABLE 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Appearance details</th>
<th>Message details</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Exp. 2a (bank robber)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>21.94</td>
<td>5.00</td>
</tr>
<tr>
<td>Judge veracity</td>
<td>17.86</td>
<td>6.70</td>
</tr>
<tr>
<td>Judge Veracity/Suspicion</td>
<td>15.08</td>
<td>8.55</td>
</tr>
<tr>
<td>Exp. 2b (shoplifter)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>16.76</td>
<td>3.62</td>
</tr>
<tr>
<td>Judge veracity</td>
<td>14.42</td>
<td>5.05</td>
</tr>
<tr>
<td>Judge Veracity/Suspicion</td>
<td>12.88</td>
<td>6.93</td>
</tr>
</tbody>
</table>

For each variable, means are reported with standard deviations in parentheses. Means in the same column that do not share the same alphabetical subscript differ significantly, $p < .05$.

#### TABLE 2

<table>
<thead>
<tr>
<th>Cues</th>
<th>Condition</th>
<th>Judgement</th>
<th>Confidence</th>
<th>Visual</th>
<th>Auditory</th>
<th>Content</th>
<th>Total</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 2a (bank robber)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.49</td>
<td>6.09</td>
<td>0.60</td>
<td>0.37</td>
<td>0.23</td>
<td>1.20</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Judge veracity</td>
<td>0.57</td>
<td>6.95</td>
<td>0.68</td>
<td>0.35</td>
<td>0.22</td>
<td>1.24</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Judge Veracity/Suspicion</td>
<td>0.79</td>
<td>6.50</td>
<td>0.79</td>
<td>0.76</td>
<td>0.95</td>
<td>2.50</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Exp. 2b (shoplifter)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.43</td>
<td>6.68</td>
<td>0.43</td>
<td>0.49</td>
<td>1.00</td>
<td>1.92</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Judge veracity</td>
<td>0.47</td>
<td>6.03</td>
<td>0.39</td>
<td>0.63</td>
<td>0.92</td>
<td>1.95</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Judge Veracity/Suspicion</td>
<td>0.73</td>
<td>6.58</td>
<td>1.05</td>
<td>0.70</td>
<td>1.70</td>
<td>3.45</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>

For veracity judgements, the table shows the proportion of witnesses who judged the target as deceptive. For other variables, means are reported with standard deviations in parentheses. Confidence ratings were made on an 11-point scale, with higher numbers representing more confidence. Values in the same column that do not share the same alphabetical subscript differ significantly, $p < .05$.1 The veracity judgement was not significantly related to the number of correct or incorrect details reported about the target’s appearance or message in any of the experiments included in this research ($ps \geq .12$).
robber made contradictory statements). For a sample of 50 questionnaires, interrater reliability was high ($r = .96$).

Suspicious witnesses listed more cues in total than witnesses in the other two conditions, $F(2, 107) = 19.43$, $p < .001$, $\eta^2 = .27$ (see Table 2). This difference emerged because suspicious witnesses reported using more auditory, $F(2, 107) = 5.20$, $p = .007$, $\eta^2 = .09$, and content cues, $F(2, 107) = 11.60$, $p < .001$, $\eta^2 = .18$. The three groups did not vary in their use of visual cues ($p = .63$).

**Experiment 2b**

**Method**

**Participants.** Introductory psychology students ($N = 115$) participated to fulfil a course requirement. They ranged in age from 18 to 45 years ($M = 19.48$, $SD = 2.71$). Most participants (61%) were women and most (88%) self-identified as White.

**Materials and procedure.** In groups of up to 10, participants watched a video (approximate running time 1 min 30 s) of a female actor portraying a customer who was present in a music store when a man shoplifted some merchandise and escaped. The participants were asked to adopt the role of the store manager. They were to imagine that they stopped the woman from leaving the store because she might have been an accomplice or, if she was innocent, she might provide information to help identify and catch the shoplifter.

The video showed the woman’s upper body as she faced the camera. The woman denied being an accomplice but said she saw the man in the store and noticed that he looked very nervous. She related what she remembered about him, explained her purpose for being in the store, and provided a few additional, irrelevant details, such as a description of an advertisement in one of the aisles.

Before watching the video, participants were randomly assigned to one of three conditions. As in Experiment 2a, control witnesses were told to watch the video carefully. Witnesses in the judge veracity condition were told to decide whether the woman was being truthful or lying. Witnesses in the Judge Veracity/Suspicion condition were given the same task, but were also asked to keep in mind that shoplifters may lie rather than confess. After watching the video, the witnesses completed a questionnaire that asked for the same information as in Experiment 2a.

**Results**

Unless noted, all measures in this section were analysed using a simple analysis of variance. If a follow-up test was needed, we used a Ryan REGWQ procedure with alpha set at .05 to determine which conditions differed significantly.

**Memory for the suspect’s appearance.** As in Experiment 2a, two coders independently used a scoring key to determine the number of correct and incorrect details witnesses reported on the questionnaire. Interrater reliability based on a sample of 50 questionnaires was high ($r = .97$).

The number of correct details varied significantly, $F(2, 112) = 31.69$, $p < .001$, $\eta^2 = .36$ (see Table 1). Witnesses who judged veracity while suspicious reported fewer details than those who performed this task without suspicion, followed by control witnesses.

In addition, we found a significant difference for incorrect details, $F(2, 112) = 35.37$, $p < .001$, $\eta^2 = .39$. Witnesses in the Judge Veracity/Suspicion condition reported more incorrect details than those in the Judge Veracity condition, who made more errors than controls.

**Memory for the suspect’s message.** As before, the coders used a scoring key to figure the number of correct and incorrect details witnesses recalled about the content of the message. With a sample of 50 questionnaires, interrater reliability was $r = .98$.

For correct details, we observed a significant difference, $F(2, 112) = 3.95$, $p = .02$, $\eta^2 = .07$ (see Table 1). Control witnesses reported more correct details than those in the Judge Veracity/Suspicion condition. The mean for witnesses in the Judge Veracity condition fell between the other two and did not differ significantly from either.

We also found a difference for incorrect details, $F(2, 112) = 10.69$, $p < .001$, $\eta^2 = .16$. Witnesses who judged veracity while suspicious reported more incorrect details than witnesses who judged veracity without suspicion, and controls reported the fewest incorrect details.

**Veracity judgements.** A chi-square test revealed a difference across the three conditions in the proportion of witnesses who judged the suspect as deceptive, with the Judge Veracity/Suspicion witnesses being most likely to decide she was lying
and the control witnesses being least likely, $\chi^2(2, N=115) = 7.90$, $p = .02$, Cramer’s $V = .26$ (see Table 2). We conducted an additional chi-square analysis using only the two experimental conditions and found that, as predicted, suspicious witnesses were more likely to conclude the suspect was lying than witnesses who were not suspicious, $\chi^2(1, N=78) = 5.14$, $p = .02$, Cramer’s $V = .26$. We found no effect for confidence ratings ($p = .33$).

**Reported use of cues.** As in Experiment 2a, the two coders examined each cue witnesses said they considered in deciding whether the suspect was being deceptive and classified each as a visual, auditory, or content cue. For a sample of 50 questionnaires, interrater reliability was $r = .95$.

Judge Veracity/Suspicion witnesses listed a greater number of cues than witnesses in the other two conditions, $F(2, 112) = 32.15$, $p < .001$, $\eta^2 = .36$ (see Table 2). These witnesses reported using more visual, $F(2, 112) = 10.98$, $p < .001$, $\eta^2 = .16$, and content cues, $F(2, 112) = 11.66$, $p < .001$, $\eta^2 = .17$. There was no difference in the use of auditory cues ($p = .41$).

**Discussion—Experiments 2a and 2b**

We obtained the same pattern of results in two separate experiments using different target individuals (one of each sex), messages, and crime scenarios. The data generally support the hypothesis that the act of judging veracity causes witnesses to remember a target’s appearance and message less accurately.

We also found results consistent with the hypothesis that suspicion induces witnesses to allocate more cognitive resources to the veracity judgement task. Suspicion not only exaggerated the effect of the veracity task on witnesses’ memory for the target but also encouraged witnesses to monitor and evaluate a greater number of cues as they tried to determine whether the target was deceptive. To some extent, the type of cue suspicious witnesses add to the set of those being utilised may depend on the situation. For instance, the suspicious witnesses in Experiment 2a used more auditory but not visual cues than other witnesses, but the reverse was true in Experiment 2b. Variations in the two targets’ mannerisms and speech patterns may have led to these different results. For example, the Experiment 2b target displayed more illustrators (hand or arm movements that modify or supplement a statement; Mann et al., 2004) than her Experiment 2a counterpart, so witnesses searching for cues could have found more occurrences of this one in the Experiment 2b video. In both experiments, however, suspicious witnesses used more content cues compared to witnesses in other conditions, perhaps because of their increased motivation to expend cognitive effort in scrutinising the target (Buller, Strzyzewski, & Comstock, 1991; Burgoon et al., 1994; Forrest et al., 2004; Millar & Millar, 1998). As noted previously, content cues require more resources to use than nonverbal cues (Reinhard & Sporer, 2008, 2010).

**EXPERIMENT 3**

In the Experiment 2 series, we showed that inducing suspicion led witnesses who were instructed to judge veracity to remember information about the target less accurately than they otherwise would have. We contend that the suspicious witnesses put more cognitive effort into the veracity judgement task, as evidenced by their reported use of more cues, particularly content cues. In Experiment 3, we extended these findings by examining whether manipulating the number and type of cues used to judge veracity influences witnesses’ memory for the target. Following the same procedure as in the prior experiments, witnesses watched a video of an individual communicating a brief message and later tried to remember her appearance and message. We instructed some witnesses to judge her veracity and to use either one nonverbal cue, one cue within the content of the message, or two cues (one of each type). We also included a control group that simply watched the target. As explained previously, we hypothesised that monitoring content cues versus nonverbal cues would be more cognitively demanding and would result in less accurate memory for the target. We expected that using two cues rather than a single cue would similarly require more resources and would produce a memory impairment. In sum, we predicted that the control group should most accurately remember the target, because this group did not anticipate being asked to judge veracity and therefore should not have monitored any deception cues while watching the video. Witnesses who monitored one nonverbal cue should perform second best, and better than those who used one content cue. Finally, we predicted that the number of cues matters as well as the
type, so that utilising two cues would result in the worst performance among the four conditions. On the other hand, if the type but not the number affects memory, then witnesses using two cues should remember the target about as well as those using one content cue.

As in the previous experiments, witnesses made a veracity judgement. Although we expected most witnesses in the control condition to judge the target as truthful (replicating the Experiment 2b results), we made no other specific predictions regarding this variable due to a lack of prior research on the effect of cue number or type on the truth bias.

We also asked witnesses in the experimental conditions to rate the difficulty of using their assigned cue(s) to judge veracity. We expected these ratings to reflect the cognitive resources allocated to the judgement task, so that using two cues versus one cue and using a content cue versus a nonverbal cue should produce higher ratings.

Method

Participants. Introductory psychology students (N = 139) attending the same university as those in the previous experiments participated to fulfil a course requirement. They ranged in age from 18 to 66 years (M = 20.37, SD = 4.22). Most participants (57%) were women and were White (81%).

Materials and procedure. In groups of up to 10, participants watched the video used in Experiment 2b. As before, they were asked to imagine themselves as the store manager questioning a woman suspected of being involved in shoplifting.

We manipulated the cues participants used when making their veracity judgements. Witnesses in the three experimental conditions were asked to determine whether the woman was lying or telling the truth, and to use certain cues while doing so. The experimenter provided definitions of each cue as well as examples. The specific cues were chosen because they are commonly believed to reveal deception (The Global Deception Team, 2006; Mann et al., 2004; Taylor & Hick, 2007; Vrij et al., 2006).

Witnesses in one experimental condition were instructed to monitor one nonverbal cue (either pause length or illustrators). Witnesses were randomly assigned to one of the two subconditions. Witnesses in a second experimental condition were told to monitor one content cue (either perceptual details or contextual details) and were randomly assigned to one of the two subconditions. Witnesses in the third experimental condition were asked to monitor two cues (one nonverbal and one content cue). They were randomly assigned to one of the four subconditions created by crossing the two nonverbal cues with the two content cues. We included the subconditions within the three experimental groups to increase our confidence that any observed effects could be attributed to the general type of cue (i.e., nonverbal or content) rather than to a specific cue’s idiosyncratic characteristics. In addition to the experimental conditions, we included a control group that was simply instructed to watch the video.

After watching, witnesses completed a written questionnaire that requested the same information as in Experiments 2a and b, except that two additional items were included. First, as manipulation checks, witnesses were asked to identify the cue(s) they were assigned to use and to indicate whether they actually used the cue(s). Second, witnesses rated on a 10-point scale how difficult it was to attend to their assigned cue(s) while watching the video. Higher numbers on the scale reflected greater difficulty.

Results

Except where noted, the measures below were analysed using a simple analysis of variance, with a follow-up Ryan REGWQ procedure with alpha equal to .05 when needed. Preliminary analyses revealed no variation as a function of monitoring illustrators versus pauses in the nonverbal condition (p > .10), monitoring contextual versus perceptual details in the content condition (p > .20), or among the four possible combinations of two cues (p > .15). Therefore, within each of the three experimental conditions we combined data from witnesses who used different cues.

Manipulation checks. All witnesses in the experimental conditions correctly identified the cues they were asked to use to judge veracity. In addition, all witnesses verified that they did use their assigned cue(s).

Memory for the suspect’s appearance. As before, two coders figured the number of correct and incorrect details witnesses reported on the
questionnaires. Interrater reliability calculated on a sample of 50 questionnaires was $r = .96$.

The number of correct details differed significantly, $F(3, 135) = 51.22, p < .001$, $\eta^2 = .53$ (see Table 3). The control witnesses reported the most details, followed by those who used a nonverbal cue. Witnesses who used a content cue and those who used two cues reported the least number of details and did not differ from each other.

The number of incorrect details also varied across conditions, $F(3, 135) = 18.68, p < .001$, $\eta^2 = .29$. Witnesses who used two cues or a content cue made the most errors, followed by those who used a nonverbal cue, followed by controls.

Memory for the suspect’s message. Once again the coders determined the number of correct and incorrect details witnesses recalled about the content of the message. With a sample of 50 questionnaires, interrater reliability was $r = .98$.

We discovered a significant difference in correct details, $F(3, 135) = 3.85, p = .01$, $\eta^2 = .08$ (see Table 3). Control witnesses reported more details than those in the other three conditions, which did not differ from each other.

The incorrect details also varied significantly, $F(3, 135) = 15.63, p < .001$, $\eta^2 = .26$. Using two cues or a content cue led to the most errors, followed by using a nonverbal cue. Controls reported the fewest incorrect details.

Veracity judgements and difficulty of monitoring cues. A chi-square test indicated that witnesses in the four conditions made similar judgements ($p = .97$; see Table 3). Overall, 39% of the witnesses decided the suspect was lying. Confidence ratings were also nonsignificant ($p = .73$).

The experimental witnesses rated the difficulty of monitoring their assigned cues. As expected, these ratings varied across conditions, with the mean being lowest for witnesses who used a nonverbal cue, $F(3, 135) = 6.11, p = .003$, $\eta^2 = .11$ (see Table 3). Witnesses who used a content cue or two cues rated the difficulty as higher, and these two conditions did not differ.

Discussion

As in the previous experiments, control witnesses remembered detailed related to the target more accurately than those who judged veracity. Moreover, in support of our hypothesis concerning the effect of cue type, witnesses who judged veracity using a content cue rather than a nonverbal cue performed worse on three memory measures (correct and incorrect details about the target’s appearance and incorrect details within the message), and they also reported that using their assigned cue was more difficult. These results imply that the high cognitive demand associated with monitoring the content of a target’s message as opposed to nonverbal information leaves too few resources available for encoding and remembering details about the target.

We did not, however, find evidence that using two cues versus one cue requires more effort or impairs memory. On one hand, this result suggests that experts’ advice to monitor multiple cues (Porter & ten Brinke, 2010; Vrij et al., 2010) is reasonable, in that the cost of considering more than one cue may be low and unlikely to interfere with concurrent tasks. On the other hand, our two-cue condition required witnesses to use one nonverbal and one content cue, and we did not examine situations in which witnesses used two cues of the same type. It is possible that monitoring two content cues would impair memory for the target more than using just one. Moreover, we
included only conditions with one and two cues, and we recognize that the outcome might be different when witnesses try to monitor three or more cues. We consider the current experiment to be a first step in exploring the consequences of using multiple cues to detect deception, and we leave it to future research to investigate more extensively the number of cues that observers can realistically examine.

**GENERAL DISCUSSION**

In Experiment 1, a secondary memory task provided evidence that judging the veracity of a target individual demands more cognitive resources than merely observing the target. We extended this finding in the subsequent experiments, demonstrating that judging veracity versus simply observing led witnesses to remember the target’s physical appearance less accurately. Compared to controls, witnesses instructed to determine whether the target was lying also made more errors in remembering his or her message and (in two of three experiments) reported fewer correct message details. This basic memory impairment effect occurred with two different targets, messages, and crime scenarios. To account for it, we point to multiple resource theory (Wickens, 2002, 2008). We contend that both judging veracity and attempting to encode information about a target are cognitively effortful tasks, collectively requiring more general resources than the amount available. Consequently, witnesses found it difficult to complete the assigned tasks concurrently. Because the veracity judgement task was prioritised, memory performance suffered.

We further discovered that suspicion amplified the memory impairment effect. Among witnesses who expected to judge veracity, inducing suspicion caused them to describe the target’s appearance less accurately and to report more incorrect details when recalling the message. Previous research suggests that suspicious versus nonsuspicious witnesses scrutinise the target more carefully as they search for cues they believe will reveal deception (e.g., Burgoon et al., 1994; Forrest et al., 2004; Millar & Millar, 1998), and this intense scrutiny requires a higher level of cognitive effort (Buller, Strzyzewski, & Comstock, 1991), which in turn diverts resources away from the task of encoding information about the target. Complementing these results, we found that suspicious witnesses were more likely than nonsuspicious ones to choose to monitor content cues, which require more effort to use than nonverbal cues (Reinhard & Sporer, 2008, 2010). Moreover, although suspicious witnesses attend closely to the target, they are looking for deception cues such as gaze aversion, illustrators, and speech rate. It is possible that memory for such behaviours might improve as a result of witnesses’ focus on them, but they are not forensically relevant details that would help police investigators identify and apprehend a suspect. Thus, because suspicious witnesses devote more cognitive resources to the veracity judgement task than nonsuspicious witnesses and because they are not attending to the right details, they will less accurately remember the target’s appearance and message.

In addition to suspicion, we examined two other variables that could influence witnesses’ allocation of resources to the veracity judgement task. First, we obtained support for the hypothesis that processing cues within the content of the target’s message rather than processing nonverbal cues is more demanding (Reinhard & Sporer, 2008, 2010) and leads to less accurate memory for the target and the message. These results are important because this type of cue may be appealing to deception detectors; some content cues are commonly believed to be valid indicators of deception (Mann et al., 2004; Vrij et al., 2006), and in fact some (e.g., number of details) are (DePaulo et al., 2003).

Second, because experts advocate using multiple cues to judge veracity rather than relying on just one (Vrij et al., 2010), we explored whether the number of cues affects memory performance. Although logically one might expect cognitive difficulty to increase with the number of cues, monitoring two cues did not impair memory for the target’s appearance and message more than using a single content cue. As stated earlier, this result should be interpreted cautiously because we did not include a condition in which witnesses used two content cues or conditions requiring them to use three or more cues. Nevertheless, the experiment provides some preliminary information about the costs involved in monitoring multiple cues during deception detection.

In addition to witnesses’ memory reports, we also analysed their veracity judgements. Among witnesses who anticipated this task, suspicious witnesses were more likely than nonsuspicious ones to decide the target was being deceptive.
This finding is consistent with previous results that suspicion reduces the truth bias (Burgoon et al., 1994; Kim & Levine, 2011; McCornack & Levine, 1990; Stiff et al., 1992; Toris & DePaulo, 1985). Previous researchers have demonstrated that this effect is simply a shift in the decision criterion; suspicious witnesses do not detect deception more accurately.

The current study is the first to show that judging veracity while observing a target impairs subsequent memory for the target. As noted previously, there are various situations in which a victim or witness might wish to determine whether an individual is lying or being truthful. Perpetrators might issue commands interlaced with threats that the victim must decide whether to believe, or they might give potentially false information about their identity. Or, suspects may converse with a witness, trying to persuade him or her that they have done nothing wrong or that their future intentions are honourable. In such cases, it may become important later for the witness to recall the target’s characteristics. Our results fit with others demonstrating the negative impact of divided attention at encoding. For example, prior research has shown that certain tasks performed during a crime event, such as trying to comprehend an accented message, can degrade memory for the perpetrator (Pickel & Staller, 2012). Perhaps there are other tasks that witnesses might perform that would be sufficiently demanding and that would have a similar effect. For example, victims confronted by a perpetrator might weigh the odds of escape, or witnesses might try to figure out whether the perpetrator is someone they have encountered before. Extending this principle beyond a forensic context, one could predict that making any sort of thoughtful judgement about an individual, such as an assessment of personality or intellect, could affect the ability to remember him or her.

In the present study, the participants observed actors portraying individuals who ostensibly lied or told the truth. It would be beneficial to attempt to replicate our results using targets who are genuinely being truthful or deceptive. Additionally, our participants acted as passive third-party observers rather than directly interacting with the targets. Actively engaging with a target as part of the veracity judgement process may demand more cognitive resources than merely watching and listening to him or her (Buller, Strzyzewski, & Hunsaker, 1991; Vrij et al., 2010). For instance, in situations in which deception detectors are able to interrogate a target, they must work to formulate diagnostic questions, present the questions in a logical sequence, critically evaluate the responses, and then devise new questions based on the target’s replies, all while attempting to monitor the cues they believe are relevant and make a veracity decision. Future research could address the influence of the participant’s level of engagement.

Other questions also remain for future study. For example, some people, such as police detectives, customs officials, and airport security personnel, must judge veracity as part of their jobs. Although they typically do not perform this task more accurately, these “professional lie detectors” are more likely to have received training than the average person (or the college students in our experiments), to have had more practice, and to express more confidence in their judgements (Vrij et al., 2010). As a result, it is possible that they allocate fewer cognitive resources to the veracity judgement task, believing that they have learned how to approach it and relying on similar cues and tactics every time. If so, these individuals may not display the memory impairment effect demonstrated in the current research.

Another variable that might interact with the veracity judgement task is the witnesses’ level of anxiety. Observing a crime event can be traumatic, especially if one is a victim-witness rather than a bystander or if the perpetrator threatens the witness with harm. Elevated anxiety in a witness can cause a defensive stress response (e.g., increased heart rate, physiological arousal, and cognitive anxiety), diminishing attentional capacity and making fewer cognitive resources available (Deffenbacher, 2008; Deffenbacher, Bornstein, Penrod, & McGorty, 2004). Additional research could clarify the effects of judging veracity while observing a suspect or perpetrator in different contexts.

REFERENCES


