TRAINING NOVICES TO DETECT DECEPTION IN ORAL NARRATIVES AND EXCHANGES – PART II

R. Edward Geiselman, Ph.D., Emiline Musarra, Natalya Berezovskaya, Cory Lustig, and Sandra Elmgren

This research was conducted to identify a limited set of indicators of truthfulness and deception in oral narratives and exchanges, as well as to train novices to use the indicators. Participants attempted to judge whether a storyteller in a video was being truthful or deceptive about an autobiographical event. The narrative story was followed by a brief question and answer period with an interviewer. Differences in verbal, vocal, and behavioral components were quantified and were used as training materials. With limited training on the indicators, laypersons improved their rate of accuracy relative to a control group by accurately detecting more of the false stories. This improvement in performance was associated with a significant shift toward reliance on verbal indicators of deception rather than vocal and behavioral indicators. Practical implications for detecting deception and implementing training protocols are discussed.

Research has shown that distinguishing truthful from deceptive oral narratives and exchanges is not an easy task (1-5). One reason why both laypersons and professionals have difficulty distinguishing truthful from deceptive narratives is that some of the more reliable indicators of deception are not consistent with “common sense” (6). For example, one might logically believe that a person who is in a position of needing to be deceptive would choose to offer an elaborate account so as to convince the listener that s/he is being truthful. S/he might try to “sell the story” to the listener. While some will make this choice, research indicates that most persons who find themselves in a position of needing to be deceptive choose to offer limited accounts that are lacking in details (7, 8). In one research study, participants who were randomly assigned to fabricate an autobiographical event stated that they believed offering too many details would appear deceptive (7).

Truthful persons are able to tell their stories with the benefit of a memory record that contains sensory details. To the extent that these details are absent from a narrative account, suspicion might be warranted as to the
truthfulness of the story (4, 9, 10). Deceptive persons must tell their stories based on their imaginations and rationalizations because they have no actual perceptual experiences for their stories. Nevertheless, the hypothesis that content-based criteria can be used to reliably discriminate truthful from confabulated stories has received limited support in the literature (1, 11). More generally, most researchers have acknowledged that any proposed set of indicators should be viewed as imperfect clues (“red flags” or “hot spots”) that only sometimes discriminate deceptive statements from truthful statements in a forensic context (3, 5, 12).

McCormack et al. (5) set out to identify a limited set of the more reliable verbal and vocal indicators of truthfulness and deception. They found that confabulated stories about autobiographical events, as compared with true ones, were lacking in contextual details (time and place) and interactions with others, and contained spontaneous rationalizations for the limited details that were given. Furthermore, the bare bones confabulated stories were presented with relatively low speech fluidity (variable speech rate, sentence fragments, stops and starts). Given that these findings were consistent with past results from studies that employed different methodologies (7, 13), McCormack et al. concluded that these indicators of deception are likely to be the more reliable ones.

As a practical application, McCormack et al. (5) reasoned that law enforcement personnel might be able to rely on this limited set of indicators, in addition to behavioral cues, to assess the likelihood of deception in consensual conversations with persons who have already raised suspicion. These brief evaluations would likely be more accurate if a greater number of indicators were used to make the evaluations, but for most day-to-day operations in sensitive areas such as mass-transportation, security personnel would likely need to rely on a limited set of the more promising indicators. Otherwise, the evaluation task could overwhelm the evaluators’ cognitive resources in real time. A more comprehensive analysis might be impractical in many real-world contexts.

The present work was designed to refine a limited, manageable set of verbal, vocal, and behavioral indicators to discriminate truthful from deceptive oral narratives and exchanges, and then to evaluate the ability of layper-
sons to accurately judge the veracity of oral statements both before and after training compared to a control (no training) group.

VIDEO STORY DEVELOPMENT

The volunteers in this experiment were asked to be truthful or deceptive about significant autobiographical events. This element of realism could affect which factors are found to best discriminate truthful from deceptive narratives (2, 14). In addition, the volunteers were asked a few follow-up questions from a research assistant about their life events at the conclusion of their narratives. Their verbal, vocal, and behavioral responses to these questions were quantified toward better discriminating the truthful and deceptive exchanges (15).

METHOD

Participants

The participants for the video development were 15 undergraduate students who were enrolled at the University of California, Los Angeles. Each student participated in the study in exchange for course credit by enrolling in an experiment titled “telling stories.” Each volunteer signed a form agreeing to allow his/her videos to be used in future research studies.

Procedure

Each participant first was presented with a list of possible life-events and was asked to choose two of them, one with which to tell a true story (the event actually happened within the past three months) and one with which to tell a confabulated story (the event has not happened even in part). The list of life events included: party, pet, trip, relationship, accident, purchase, anniversary, promotion, award, concert, and wedding. The order in which the true and confabulated stories were to be told to a second research assistant was counterbalanced across participants.

Once the two topics were chosen, the participant was told, “The first story I want you to tell is the one about (insert topic selected). You will be video recorded while you tell the story to a research assistant in the other room. While I prepare the tape recorder, you will have about five minutes to think about what you want to say in the story. When you tell it, we would like you to tell it in about as much detail as you would if you were telling it to a friend
over coffee. When you are finished, the research assistant will ask you a few follow-up questions about your story. The assistant will not know whether you are telling the truth or not.” The follow-up questions asked the participants what else was going on in their lives at the time of this event, who was present at the event, what this event meant to them, as well as any questions of clarification about the story.

After the first story was told, the experimenter repeated the instructions and the participant was given five minutes to prepare the second story while the experimenter again readied the video recorder. The procedure for the second story was the same as for the first story. As before, the research assistant to whom the story was told was blind as to whether the participant’s story was true or false.

**Analysis**

The video-recorded stories were reviewed independently by five research assistants. First, the truthful and deceptive arguments from each volunteer were compared for differences. Then, patterns were identified across volunteers. The three authors then met at a joint session to compare notes on each volunteer and on patterns across the volunteers. These analyses of the video protocols produced the following summary and guidelines for use as training materials in the follow-up experiment.

The following is a set of considerations that can be used to decide whether a person is being truthful or deceptive when telling a story about a life event.

**A. Something unnatural about the story:**

1. Starts with inappropriate word(s) such as “so” or “seriously.”
2. Ends with summaries such as “that’s all I have” or “that’s my story” or “yea, so that’s the time we…”
3. Starts with a central detail, then repeats the detail verbatim, adding something each time, with conflicting details.
4. Filler details are told simply to make it longer, not central to the theme of the story, with repetition of details verbatim.
5. Elements of the story are told out of chronological order.
6. Inappropriate descriptor terms are used to introduce a part of the story (such as “scary,” but the story element was not scary).

7. The story is told without reference to emotions, or without relating the story to his/her present life.

8. The storyteller wants the interviewer to control the interview with specific questions. Watch for statements such as, “I could tell you more if you ask me something, but I cannot recall anything else right now.”

B. Behavior in an exaggerated, strained, or peculiar manner:
   1. Sits upright as if ready to do battle with the interviewer (stiff posture).
   2. Emotions are mentioned, but not displayed.
   3. Appears to want to end the conversation as soon as possible—with short narratives and answers, fewer add-ons.
   4. Increased bodily movements during specific questions, sometimes just in the extremities (hands).
   5. Upward inflections with words; covering mouth.
   6. Duping delight is shown (smirks, half smiles).

C. Unusual eye contact or eye movements:
   1. Searches around room with his/her eyes.
   2. Often looks down or at some object when thinking about the response.
   3. Squints as though trying hard to give a correct answer.
   4. Long blinks and closing eyes in response to questions.

Try to base your decision on the sum total of these considerations. Just because you find one indicator of deception does not necessarily mean the person is being deceptive about his/her life experience.

D. The major summary areas indicating deception are:
   1. Something unnatural about the story: few details, ends abruptly, contradictions, lacking chronology, vague or illogical story line, rationalizations, and awkward use of terms.
2. Unusual eye movements or contact: blinking, squinting, exaggerated movements, and looking down or around the room.
3. Exaggerated behavior: smiling, shrugging, or grooming.

THE EXPERIMENT ON TRAINING

The purpose for the present experiment was to evaluate the ability of laypersons to discriminate the truthful and deceptive stories either with or without receiving training on the guidelines derived above. Previous attempts to train laypersons to improve their ability to discriminate truthful and deceptive narratives have met with mixed success (1, 3, 4, 12, 16). The design consisted of a pre-test versus post-test comparison for a training group and a control (no instruction) group. This design allowed for the evaluation of training effects independently from simple experiential effects. The main dependent variables were accuracy in discriminating true from false stories and self-reported confidence in those judgments. The participants also completed some short-answer items about perceived indicators of deception and truthfulness.

METHOD

Participants

The participants in the experiment were 51 undergraduate students who were enrolled in introductory psychology courses at the University of California, Los Angeles. Each of the participants volunteered in exchange for course credit. Twenty-six participants served in the training condition and the remaining 25 participants served in the control condition.

Design

The data formed a 2x2x2 matrix, with the factors being Group (training, control), Test Session (day one, day two), and Story Type (true, false). Group was the only between-subject factor. One dependent variable was the accuracy of participants’ judgments of the storytellers for truth and deception. A second dependent variable was the participants’ confidence judgments in their abilities to detect deception. The participants also were asked during each test session to list cues that they considered to be indicators of truth and deception.
Materials and Apparatus

Twenty-four video clips were selected for use in the experiment, twelve showing a student describing a true autobiographical event and twelve showing a student describing a confabulated autobiographical event. Participants viewed a string of six video clips during test session one and another string of six video clips during test session two. The six clips included in each of four video strings were randomly selected from the pool of clips, with the only restriction that there must be three true stories and three false stories in each string. The presentation of the four video strings was counterbalanced so that each one appeared in test session one and in test session two for some participants. Each video clip lasted between two and three minutes, and each string of six clips lasted twenty-two minutes on average.

Participants recorded a judgment of true or false for each video clip on a response sheet. The clips were numbered one through nine on the answer sheet even though each participant would view only six clips during each session. This was done so that the participants could not manipulate their judgments based on the instruction that approximately one half of the stories would be true and half would be false.

The participants were told that they would view a few 2-3 minute video clips of students who had been asked to tell a story about a life experience, and that approximately half would be true and half would be false. They were told that the students in the videos would be sitting opposite an interviewer who would be sitting off camera and would not know whether the story was true or false. The interviewer would ask the students a few questions about their stories when the students were finished. It also was explained that the students in the videos had been told that they would win 50 dollars if they could convince the interviewer that their stories were true.

At the end of test session one, the participants in the training condition were sent away with a one and one-half page guide on detecting deception. The training guide was divided into three main categories of indicators, with specific cues provided for each category. The categories were: unnatural aspects of the story, exaggerated physical behaviors, and unusual eye movements. There were a total of 18 cues, which are listed at the end of the Introduction section above.
During both test session one and test session two, the participants were asked to list what they considered to be signs of truth and deception. For test session two, the participants in the training condition also indicated how many times they studied the training materials as well as the approximate total duration of study.

**Procedure**

The participants first completed a brief questionnaire requiring them to list some cues that they considered to be indicators of truth and deception. The participants also were asked to estimate their ability to detect deception on a Likert scale of one to five. Then, the participants were given a copy of the procedural instructions and were asked to follow along as the researcher read the instructions verbatim. It was explained that there would be approximately 15 seconds following each video clip during which time the participants were to make their true-false judgments. The viewing of the string of six video clips required approximately twenty-two minutes. After the viewing, the participants in the training condition were given the training guide to take with them for study before they returned two days later.

At test session two, the participants in the training condition completed a brief questionnaire to indicate how many times they had studied the training materials and for approximately how many minutes in total. They were reminded that their answers were anonymous, that it was important to answer honestly, and that the researcher would not look at their answers upon completing the questionnaire. The participants then were given ten minutes to look over the training guide one last time. At the conclusion of this study period, the researcher read the training guide verbatim to the participants as they followed along.

All participants were re-read the procedural instructions from test session one as a reminder. The participants then viewed a new string of six video clips using the same procedure as in test session one. After the video string, both the training and control participants completed a brief questionnaire asking them to list cues of truth and deception, to indicate how confident they were in their ability to detect deception, and to indicate their sex.
RESULTS AND DISCUSSION

Prior to beginning the second test session, the participants in the training group followed along as the experimenter reviewed the training guidelines. When asked anonymously how many times they studied the guidelines prior to arriving for the second session, they reported that they had studied the guidelines on their own two times for an estimated total time of 15 minutes. Neither study variable was found to be correlated significantly with the performance variables.

<table>
<thead>
<tr>
<th></th>
<th>Test Session</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
<td>Two</td>
</tr>
<tr>
<td>Training Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Stories</td>
<td>.52</td>
<td>.49</td>
</tr>
<tr>
<td>False Stories</td>
<td>.51</td>
<td>.67</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Stories</td>
<td>.51</td>
<td>50</td>
</tr>
<tr>
<td>False Stories</td>
<td>.52</td>
<td>.38</td>
</tr>
</tbody>
</table>

The average proportion correct judgment is presented in Table 1 as a function of Group, Test Session, and Story Type. For the true autobiographical stories, there were no differences between groups or between groups and test sessions. However, for the false (deceptive) autobiographical stories, the control group believed more of them to be true during session two than during session one (62% versus 48%, p < .05) whereas the training group accurately detected more of them to be false during session two than during session one (67% versus 51%, p < .05), F(1, 49) = 5.71, MSe = 33.83, p < .01. Thus, with simple experience the participants in the control group started believing more false stories whereas with the limited training the participants in the training group stated detecting more false stories. Overall then, the training group showed a modest eight percent improvement in performance, whereas the control group showed a six percent decrease in performance.
Table 2. Average Self-Confidence Rating (4-pt Scale) as a Function of Group and Test Session

<table>
<thead>
<tr>
<th></th>
<th>Test Session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
</tr>
<tr>
<td>Training Group</td>
<td>1.50</td>
</tr>
<tr>
<td>Control Group</td>
<td>1.56</td>
</tr>
</tbody>
</table>

The average self-confidence rating is presented in Table 2 as a function of Group and Test Session. Both the training and control groups indicated that they believed they improved their ability to detect deception during test session two \(F(1, 49) = 133.70, MSe = .48, p < .001\), with the self-perceived improvement being significantly greater for the training group, \(F(1, 49) = 3.70, MSe = 0.48, p < .05\). Thus, participants felt they were doing significantly better with experience even though they were doing worse without training and only modestly better with training. Also, during test session one, males believed that they were doing better than did the females \(p < .05\), but they were not.

Each participant was asked to list factors that they consider signs of deception and each was asked to do this at each test session immediately following viewing the video string. The results in terms of frequency counts are presented in Table 3. At the first test session, the participants listed more behavioral indicators (e.g., grooming) than vocal indicators (e.g., irregular speech) and more vocal indicators than verbal indicators (e.g., unnatural story elements). At the second test session, the participants listed more verbal indicators than behavioral indicators and more behavioral indicators than vocal indicators. This shift toward a greater emphasis on verbal indicators was observed only for the participants in the training group, with \(\chi^2(2) = 10.82, p < .01\). The participants who received the training materials reliably modified their basis for identifying likely false stories by focusing on the content of the stories rather than non-verbal indicators. In contrast, the control participants continued to rely on non-verbal indicators. Thus, the modest increase in detection of false stories following the limited training can reasona-
bly be attributed to an increased focus on the content of the stories rather than on non-verbal indicators (11). The most frequently cited verbal indicator was lesser or irrelevant details tied with an irregular chronology of the events. The most frequently cited behavioral indicator was gaze aversion tied with grooming and/or fidgeting. The most frequently cited vocal indicator was unusual pausing and stammering.

Table 3. Percentage of Participants Listing Behavioral, Vocal, and Verbal Indicators of Deception as a Function of Group and Test Session

<table>
<thead>
<tr>
<th>Test Session</th>
<th>One</th>
<th>Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>.54</td>
<td>.34</td>
</tr>
<tr>
<td>Vocal</td>
<td>.38</td>
<td>.15</td>
</tr>
<tr>
<td>Verbal</td>
<td>.08</td>
<td>.51</td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>.60</td>
<td>.52</td>
</tr>
<tr>
<td>Vocal</td>
<td>.32</td>
<td>.36</td>
</tr>
<tr>
<td>Verbal</td>
<td>.08</td>
<td>.12</td>
</tr>
</tbody>
</table>

At the first test session, the participants also were asked to list factors that they consider signs of truthfulness. The pattern was effectively the same as for indicators of deception in that non-verbal indicators were listed more frequently than verbal indicators (45% behavioral, 35% vocal, and 20% verbal). The indicators were the same as those listed as signs of deception, but opposite (e.g., eye contact instead of gaze aversion; calmness instead of fidgeting).
GENERAL DISCUSSION

The results of the present experiment are consistent with some past findings in that laypersons without training rely mainly on non-verbal indicators of deception and do not discriminate truthful from deceptive oral narratives and exchanges better than chance (1, 4). The results also demonstrate that laypersons with limited training do not discriminate truthful and deceptive narratives and exchanges very well either, although performance was improved reliably for false stories (67% versus chance) with a greater reliance on verbal indicators of deception. The statistically significant improvement in the participants’ ability to detect false stories with training was accompanied by self-reports that their performance was indeed more accurate. To a lesser extent, participants who received no training also believed that their performance was more accurate during a second session even though their ability to detect false stories actually decreased (38% versus chance).

Research on training law enforcement officers to detect deception has shown that plenty of practice with video clips and feedback is necessary to improve the accuracy of their judgments reliably (17, 18). The present results suggest that the same is likely to be true with laypersons (19, 20). Future research will explore more substantial protocols for training, based on the more reliable indicators of truthfulness and deception.

The limited set of indicators compiled by McCormack et al. (5), Geiselman et al. (17), and the present video development work provide a partial basis for judging truthfulness and deception in oral narratives and exchanges. It is important to view these indicators as “red flags” or “hot spots” that only sometimes discriminate truthful from deceptive statements. For example, gaze aversion alone cannot be taken as a sign of deception because it also can be a sign of increased concentration (21). Judgments of truthfulness and deception must be based on a collective pattern of indicators (22), especially verbal indicators, and the results of present experiment demonstrate that more than limited training is required to master the ability to consider multiple indicators. Nonetheless, the present results confirm that improving deception-detection performance is possible (23). With limited training, the participants showed a significantly greater reliance on the content and structure of the stories, and that shift was associated with significantly more accurate judgments for false stories (11).
It is noteworthy that sometimes the interviewer is not left simply with monitoring indicators of deception, but rather the interviewer can utilize statement-evidence consistency cues. Comparing the subject’s statements with the case evidence as a litmus test has been found in one study to increase deception detection accuracy by nearly 30 percent, with an overall accuracy rate of 85 percent (24). Even if such evidence does exist, we recommend using a cognitive-interview, information-gathering type format (25, 26) at least for the initial phase of the investigative interview, as opposed to a confrontational-type format (13, 27). Use of the cognitive interview for suspects (CIS) should maximize the amount of information obtained from the subject that then can be compared with the case evidence (28) as well as monitored for indicators of deception.

REFERENCES

6. Levine TR, Serota KB, Shulman HC: The impact of Lie to Me on Viewers’ actual ability to detect deception. Communication Research 2010; 37:847-856


ABOUT THE AUTHORS

Dr. R. Edward Geiselman is Professor of Psychology at UCLA and has authored more than one hundred research papers and six books. He teaches classes on investigative interviewing at the request of law enforcement agencies, and has testified in more than 330 criminal trials as an expert witness. The four junior authors were students at UCLA at the time of this research. They contributed equally at different stages of the project.