

Criminal Justice and Behavior

<http://cjb.sagepub.com>

Using Witness Confidence can Impair the Ability to Detect Deception

Veronica S. Tetterton and Amye R. Warren
Criminal Justice and Behavior 2005; 32; 433
DOI: 10.1177/0093854805276406

The online version of this article can be found at:
<http://cjb.sagepub.com/cgi/content/abstract/32/4/433>

Published by:



<http://www.sagepublications.com>

On behalf of:

[International Association for Correctional and Forensic Psychology](#)

Additional services and information for *Criminal Justice and Behavior* can be found at:

Email Alerts: <http://cjb.sagepub.com/cgi/alerts>

Subscriptions: <http://cjb.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://cjb.sagepub.com/cgi/content/refs/32/4/433>

USING WITNESS CONFIDENCE CAN IMPAIR THE ABILITY TO DETECT DECEPTION

VERONICA S. TETTERTON

AMYE R. WARREN

University of Tennessee, Chattanooga

Prior research has shown that jurors rely on confidence in discriminating between accurate and inaccurate testimonies despite the weak relationship between the two. The purpose of this study is to learn if truth seekers also use confidence in judging truthfulness. In two studies, participants were either not given instructions regarding witness confidence or were told not to use witness confidence, and then they were asked to rate the believability of the videotaped testimony of four witnesses who varied in confidence and truthfulness. Regardless of the instructions, participants did rely on confidence and rated highly confident testimonies as more believable. They also rated false testimonies as significantly more believable than true statements.

Keywords: detecting deception; judgment; eyewitness testimony; confidence; juries; nonverbal communication; verbal communication

Many researchers have attempted to determine the cues jurors use as indicators of accuracy in eyewitness testimony. Laypeople assume that there is a strong relationship between witness confidence

AUTHOR NOTE: *This research was supported by a grant from the University of Alabama. The authors would like to thank Jack Brigham of Florida State University for allowing us to use the videotaped stimulus materials. The authors would also like to thank Richard Metzger and David Ross for their valuable suggestions and Brandi Mangan, Laura Zmuda, Toni Carpenter, and Carnita McKeithen for their assistance in data collection. Some of these data were previously presented at the American Psychology–Law Society Biennial Conference (in Austin, Texas, March 2002). Correspondence concerning this article should be addressed to Veronica S. Tetterton, Department of Psychology, University of Alabama, Box 870348, Tuscaloosa, AL 35487-0348; e-mail: arnol031@bama.ua.edu.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 32 No. 4, August 2005 433-451

DOI: 10.1177/0093854805276406

© 2005 American Association for Correctional and Forensic Psychology

and accuracy (Whitley & Greenberg, 1986). Furthermore, the confidence expressed by an eyewitness is the most influential feature used by jurors to determine the accuracy of the testimony (Leippe, Romanczyk, & Manion, 1992).

Jurors' heavy reliance on confidence as a predictor of accuracy is troublesome given that evidence suggests that confidence and accuracy are only weakly related. In a meta-analysis, Sporer, Penrod, Read, and Cutler (1995) found that the overall correlation between confidence and accuracy was .29. Despite the lack of a robust relationship, laypeople continue to rely on confidence when judging accuracy. Expert testimony often serves to improve jurors' sensitivity to factors that influence eyewitness memory with hopes to reduce jurors' baseless reliance on confidence (Devenport, Penrod, & Cutler, 1997).

Although it is clear that the use of witness confidence is potentially problematic when judging witness accuracy, does this hold true when witnesses are knowingly and intentionally inaccurate rather than honestly mistaken? Consciously fabricated accounts are fundamentally different from events that are erroneously believed to be true (Pezdek & Taylor, 2000). In other words, someone who is honestly mistaken about an event that he or she witnessed is likely to engage in different cognitive processes while testifying than someone who is intentionally lying about an incident. When witnesses are motivated to lie at trial, they may purposely appear highly confident in their testimony in an attempt to sway the jury to believe their version of events. Given this, are jurors capable of detecting a confident but deceptive witness?

According to Ekman and O'Sullivan (1991), the past 20 years of research have demonstrated that people perform poorly in judging if an individual is being dishonest. In general, when adults are instructed to distinguish between lies and true statements, the overall accuracy approaches chance levels (50%) and rarely exceeds 60% (Ekman & O'Sullivan, 1991). It is surprising that this has also been the finding among most trained groups of individuals, whom one would expect to have a greater ability to detect deception.

In light of the difficulty in detecting deception, researchers have attempted to develop criteria that reliably distinguish true and false statements. Statement Validity Analysis (SVA) is used as a diagnostic

technique to determine the validity of statements (Steller, 1989). Criteria-Based Content Analysis (CBCA) is a subsection of the SVA that explicitly defines 19 criteria by which the validity of an individual's statement can be evaluated (Steller, 1989).

Landry and Brigham (1992) assessed the validity of CBCA with college students. First, 60 students served as witnesses. Each witness videotaped a 1- to 2-min description of two personal emotional incidents; one incident was true and the other was invented. The witnesses were given 2 days (prior to videotaping) to invent their stories but were given no instructions other than the time restrictions.

Seven raters evaluated these videotaped descriptions according to three criteria: how traumatic the incident appeared to be, how emotional the witness appeared while describing the event, and how great a loss of control over events appeared to be involved in the incident. Twelve of the original stories given were actually selected based on the mean ratings given by the evaluators. Six of the stories chosen were true, and six were false testimonies. The topics of the events discussed ranged from the death of a grandmother to being a near victim of date rape.

A second sample of college students was selected to serve as raters of witness truthfulness. They were randomly assigned to one of four conditions: either trained to detect deception with CBCA or not and exposed to the testimonies via videotape or written transcripts. Training consisted of a 45-min session in applying 14 of the original 19 CBCA criteria. Trainees were told that finding more than 5 criteria present in a statement is a good indication of truthfulness. The trainees were then given a checklist of the CBCA criteria to use in evaluating the 12 statements. After viewing the videotaped testimony or reading the statements, the trained participants checked on their evaluation list if each criterion was absent, present, or strongly present. Finally, both trained and untrained participants made a judgment regarding the probability that the incident was truthful on a 9-point Likert-type scale (1 = *false* and 9 = *true*).

Both training and mode of presentation significantly affected accuracy. Those who were trained were accurate 55% of the time, whereas untrained participants were accurate 47% of the time. Furthermore,

participants that viewed the videotape were accurate 55% of the time; those who read transcripts were accurate 47% of the time. With regard to accuracy in the training, the videotape condition was the best (58%) and the only condition significantly above chance (50%). Untrained individuals who read transcripts scored significantly below chance.

Landry and Brigham (1992) concluded that both verbal and visual cues must be provided for optimal lie detection performance, as truth seekers rely on more than the content of the witness's message. What cues, verbal and nonverbal, are useful? DePaulo et al. (2003) conducted a meta-analysis of 120 independent samples in search of a list of behavioral cues to deception. Overall, lies were less detailed, less plausible, less logical, more likely to display internal conflict, and more likely to have repeated words and phrases. Furthermore, liars seemed evasive and impersonal.

If a witness lies during court proceedings, he or she may experience different pressure than a person who lies about a mundane day-to-day activity. Frank and Ekman (1997) distinguished between low-stakes lies and high-stakes lies. There are serious consequences (e.g., prison) associated with getting caught in high-stakes lies; therefore, they evoke emotions (e.g., guilt and fear) that are not observable in low-stakes lies. Frank and Ekman (1997) proposed that the strong emotions betray the liar, leaking through nonverbal behavior, such as facial expressions, and thereby providing cues to those seeking the truth. Prior research has provided support for the notion that attention to the visible displays of emotions (nonverbal cues) predicts higher accuracy in detecting deception (Ekman & O'Sullivan, 1991).

In prior studies, researchers have varied the levels of accuracy in eyewitness accounts in an attempt to determine the effects of witness confidence on credibility. However, in these studies, the eyewitnesses were not being deceptive but honestly mistaken. The purpose of the present study was to examine the effect of confidence on believability when witnesses are intentionally lying. We predicted that mock jurors would rely on confidence in discriminating between true and invented accounts. Furthermore, we expected that instructions explicitly admonishing the use of confidence would reduce the use of confidence as an indicator of truthfulness and improve performance in detecting deceit.

EXPERIMENT 1

METHOD

Participants

Forty-one undergraduate students (32 women and 9 men) at a Southern university received extra credit in a psychology course for their involvement. Seventy-eight percent of the participants were Caucasian, and 22% were African American. Their ages ranged from 18 to 56 ($M = 21$, $SD = 5.81$), and none had ever served on a jury.

Design

The study had a $2 \times 2 \times 2$ factorial design. There were two within-subjects variables: witness confidence (high or low) and nature of witness testimony (true or false). The between-subjects variable was instructions (NC = instructions that do not mention the use of confidence in detecting deception, and DC = instructions that explicitly caution against the use of confidence in detecting deception).

Materials

Videotaped testimony. The videotaped materials used in this study were provided by Jack Brigham (Landry & Brigham, 1992). The videotapes included 60 Florida State University undergraduate students, 4 of whom served as witnesses in the present study. The specifics of Landry and Brigham's (1992) development of the testimonies were outlined in the literature review.

The four witnesses were selected based on the results of a pilot study, in which 43 undergraduate students reviewed and rated various subsets of the 60 witnesses from Landry and Brigham (1992). On a 9-point Likert-type scale (1 = *false* and 9 = *true*), they rated level of confidence, likeability, believability, and attractiveness. We asked participants to rate attractiveness and likeability because they were potential extraneous variables that could have influenced the measure of believability. The raters were not given any instructions concerning cues and

behaviors to look for in rating each characteristic, nor were they provided with operational definitions or examples. Therefore, the raters used their own discretion and definitions when they evaluated the witnesses' level of confidence.

Based on the mean confidence ratings, we selected two high-confidence witnesses and two low-confidence witnesses. The mean rating for the witnesses who displayed low confidence was 2.78 ($SD = 1.2$), compared to 6.82 ($SD = 1.5$) for the high-confidence witnesses. Each witness described both a true and invented personal incident. The chosen witnesses exhibited comparable levels of confidence in their true and false testimonies. Thus, witnesses rated high in confidence in detailing a true incident subsequently received similar ratings of confidence for their false testimony. For the entire sample of 60 witnesses, the average confidence rating for the true testimonies was 5.2 ($SD = .97$), and for the false testimonies it was 5.3 ($SD = .90$).

To control for ethnicity and gender, two of the selected witnesses were White females, and two were White males. Initially, our intent was to select four witnesses matched on their likeability and attractiveness ratings. However, ratings of attractiveness and likeability for the low-confidence witnesses (for attractiveness, $M = 3.7$, $SD = 1.7$; for likeability, $M = 3.7$, $SD = 1.7$) were lower than that of the high-confidence witnesses (for attractiveness, $M = 6.5$, $SD = 1.1$; for likeability, $M = 6.6$, $SD = 1.2$). We found significant correlations (ranging from about .34 to .40 on average) between attractiveness, likeability, and confidence, which made it impossible to hold attractiveness and likeability ratings constant while manipulating levels of confidence.

With our four witnesses selected, we then constructed different versions of the tapes to present only one testimony from each witness (either the witness's true or false story) and to deal with potential order effects. We counterbalanced the order in which participants saw the witnesses, alternating the presentation of confidence (high, low, high, low vs. low, high, low, high) and the nature of the testimonies (true, false, false, true vs. false, true, true, false). These manipulations lead to the creation of four videotapes that included the same witnesses but in different order and providing different testimonies. This meant that the order of the witnesses' gender also varied across the tapes (male,

female, female, male vs. female, male, male, female). Ultimately, each participant viewed one of the four videotapes in which he or she observed a highly confident truth teller and a highly confident liar as well as a truth teller low in confidence and a liar low in confidence.

Given the importance of the reliability of the initial confidence ratings, we conducted another pretest. Independent raters (separate from the raters who initially provided confidence and truthfulness ratings) viewed one of the two tapes and provided believability and confidence ratings, again using 9-point Likert-type scales. A repeated measures ANOVA revealed a significant difference in the confidence ratings for Tape 1, $F(1, 16) = 26.15, p < .001$. Confidence ratings for the high-confidence witnesses (6.4) were significantly higher than the confidence ratings for the low-confidence witnesses (3.75). A pretest was also necessary for the alternate version of Tape 1 (Tape 2), in which the four witnesses described different personal incidents. A repeated measures ANOVA revealed a significant difference in the ratings for the low-confidence stories and the high-confidence stories, $F(1, 12) = 12.20, p = .004$. The overall confidence rating for the high-confidence testimonies was 6.7, and for the low-confidence testimonies it was 4.9. Therefore, no changes were made to the stimuli.

Experiment 1 used only one of the four videotapes (Tape 1, Version A). It depicted a male exhibiting low confidence when describing a false story and was followed by a female high in confidence telling a true story, a female low in confidence describing a true story, and finally, a male high in confidence describing a false story.

Written instructions. Two dissimilar instruction sets were employed in Experiment 1. Both instruction sets included the following statements:

You are about to observe four unrelated witness statements. The statements vary in topic. However, all four testimonies are descriptions of traumatic incidents. Some of the statements are true, whereas others are fabricated (false). You will be asked to put yourself in the place of a juror and try to determine which statements are true (events that really happened) and which are false (events that did not happen). Research has shown that there are few factors that can reliably distinguish between the truth and a lie. Because we lack techniques that discrimi-

nate between accurate and fabricated testimonies, we are interested in seeing which factors you use in attempting to detect lies and which factors seem to be the most successful.

The instructions varied in one respect: One of the instruction sets explicitly mentioned the use of confidence in making their judgments (don't use confidence, or DC). The DC instruction set stated the following:

The confidence the witness displays is *NOT* a reliable indicator of true or false statements. Memories for traumatic events are often inaccurate or deceitful despite the confidence the person exhibits in their recollection of the incident. Therefore, confidence should *NOT* influence your decision.

The NC (no confidence mentioned) instruction set did not include any statement related to the use of witness confidence in detecting deception. Thus, depending on which set of instructions they were given, the participants were either advised to avoid using confidence (DC), or they were not given any warning regarding the use of confidence (NC) in detecting deception.

Procedure

Participants were tested in a group but were randomly assigned written packets that included either DC or NC instructions. They were informed that the objective of the study was to determine how well people tell the difference between true and invented stories and what factors they use in doing so. Participants completed an informed consent and a demographics questionnaire. The informed consent cautioned that the witnesses' topics were highly emotional and traumatic in nature, and provided examples of the more sensitive topics (e.g., near date rape or death of a loved one). Furthermore, it encouraged participants to refrain from participating if they believed that the subject material would elicit a negative response, but nobody dropped out. The participants were then told to read their respective written instructions carefully. Because each participant received written instructions, those in different instruction conditions were tested simultaneously, thereby reducing the chance of experimenter bias.

After reading the instructions, participants viewed the videotape of four testimonies. Immediately following each testimony, the tape was stopped while participants provided a judgment of the likelihood that the incident was truthful on a 9-point scale (1 = *false* and 9 = *true*). This was followed by the question, "If forced to choose, would you say that this was a true or false testimony?"

RESULTS

The ratings of truthfulness were examined using a repeated measures ANOVA. There was a significant main effect of witness confidence. The high-confidence testimonies ($M = 6.6$, $SD = 1.89$) were rated more truthful than the low-confidence stories ($M = 3.6$, $SD = 1.86$), $F(1, 39) = 83.63$, $p < .001$. There was also a significant interaction between confidence and instructions, $F(1, 39) = 4.92$, $p = .032$. There was a smaller difference between the high- and low-confidence means in the DC conditions than there was in the NC condition (see Table 1). Nevertheless, participants in the DC condition continued to rate high-confidence witnesses higher than low-confidence witnesses. A significant interaction between confidence and truth was also discovered. High-confidence true stories were rated less believable than the high-confidence false stories, yet low-confidence false stories were rated less believable than low-confidence true stories $F(1, 39) = 17.743$, $p < .001$.

The accuracy in detecting deception was established by the response to the dichotomous follow-up question. Participants were given a score of 1 if they accurately classified a statement and 0 for inaccurate assessments. A proportion accuracy score was calculated by dividing the number of accurate assessments provided by each participant by 4 (the total number of statements judged). Overall, the accuracy for the sample was 45%. Separate examination of accuracy proportions for judging true stories (45%), false stories (45%), high-confidence stories (30%), and low-confidence stories (60%) were also calculated. Participants were significantly more accurate in detecting deception for the low-confidence witnesses than for the high-confidence witnesses. Further analysis revealed that the high-confidence proportion was the only one significantly different from

TABLE 1: Experiment 1: Believability and Overall Accuracy Analysis (N = 41)

	M	SD	df	F	p <	d
Confidence			1, 39	83.63	.000*	.682
High	6.6	1.89				
Low	3.6	1.85				
Truth			1, 39	3.46	.071	.081
True	4.8	2.27				
False	5.3	1.48				
Instructions			1, 39	.267	.609	.007
NC	5.2	1.78				
DC	5.0	1.91				
Confidence instructions			1, 39	4.92	.032*	.112
High confidence						
NC	7.1	1.78				
DC	6.1	1.91				
Low confidence						
NC	3.3	1.77				
DC	3.9	1.91				
Confidence truth			1, 39	17.74	.000*	.313
High confidence						
True	5.8	2.56				
False	7.4	1.24				
Low confidence						
True	3.9	1.99				
False	3.3	1.72				
Instruction truth			1, 39	.267	.609	.007
True						
NC	5.0	2.16				
DC	4.7	2.27				
False						
NC	5.4	1.39				
DC	5.3	1.56				
Accuracy instructions			1, 39	.174	.679	.116
NC	44%	.179				
DC	46%	.228				

Note. NC = no confidence mentioned in instructions; DC = don't use confidence instructions.

* $p < .05$.

chance. The participants were significantly below chance in judging the high-confidence testimonies.

The overall accuracy scores were analyzed using a one-way ANOVA, with the instruction conditions as the independent variable. There was no significant effect of instruction on overall accuracy. The

difference in accuracy for those in the DC condition (46%, $n = 21$) compared to those in the NC condition (44%, $n = 20$) was negligible. Separate examination of independent accuracy proportions for true stories, false stories, high-confidence stories, and low-confidence stories also revealed no significant differences across instruction conditions.

DISCUSSION

The results revealed that participants did use confidence in their decisions of truthfulness and therefore rated high-confidence witnesses as more truthful. The participants who were told not to use confidence rated the high-confidence witnesses less believable and low-confidence witnesses more believable compared to those in the NC condition. Thus, the difference in the truthfulness ratings between the high- and low-confidence testimonies was reduced in the DC condition. However, even participants in this condition continued to use confidence, rating high-confidence testimonies more truthful than the low-confidence testimonies regardless of actual truthfulness. Furthermore, when forced to choose between truth and lying, the instructions cautioning against the use of confidence did not significantly increase accuracy in detecting deception. It is interesting that high confidence seemed to significantly impair the participants' ability to detect deception, yet participants found the high-confidence stories most believable. To see if many of the trends would become significant with a larger sample size and would generalize to additional testimonies, we conducted a second study.

EXPERIMENT 2

METHOD

Participants

Three months following data collection in the initial study, 217 undergraduate students enrolled at the same Southern university participated in the second study, again receiving extra credit in one of

their psychology courses for their participation. The ages of the participants ranged from 16 to 54 ($M = 20$, $SD = 4.29$). The sample was 72% ($n = 157$) female, 27% ($n = 58$) male, and 1.8% ($n = 4$) did not specify gender. Fifty-nine percent of the participants were Caucasian ($n = 130$), 36% were African American ($n = 78$), 2% self-identified as Asian ($n = 4$), 1% indicated other ($n = 3$), and 2 participants did not provide their race. The majority of the participants (97%) had never served on a jury.

Design

The design was a $2 \times 2 \times 3$ factorial. The within-subjects variables were witness confidence (high or low) and nature of witness testimony (true or false). The second study differed in that it had three sets of instructions: no confidence instructions (NC, $n = 71$), don't use confidence instructions (DC, $n = 75$), and use confidence instructions (UC, $n = 71$). As in Experiment 1, a truthfulness rating scale (9-point Likert-type scale, with 1 = *false* and 9 = *true*) and a dichotomous (true and false) follow-up question measured the ability to detect deception.

Materials

Written instructions. There were three dissimilar sets of instructions. In addition to the DC and NC instructions identical to those in Experiment 1, we included the UC condition, in which participants were advised to use confidence when evaluating the testimonies: "The confidence the witness displays *IS* a reliable indicator of true or false statements. Therefore, confidence *SHOULD* influence your decision." This condition was added to determine whether participants spontaneously use confidence as the biggest factor in evaluating testimony (in the NC condition) or whether we could further increase their reliance on confidence.

Videotaped personal testimonies. All four videotapes, counterbalancing for order and content, were used in this experiment. Participants were randomly assigned a tape to view.

Procedure

The participants followed the same procedure discussed in the previous study. One addition to the second study was that after the entire tape was viewed, the participants were asked to list the characteristics they had used in making their judgments. We did this to see how frequently confidence was mentioned and whether participants were aware of using confidence in their decisions.

RESULTS

Preliminary analysis of order and tape effects. Including all four tapes, the design was a 3 (instructions) \times 2 (story version) \times 2 (orders) \times 2 (witness confidence) \times 2 (truth) factorial design. The between-subjects factors were the written instructions (DC, UC, and NC), the version of the witnesses' statements (Tape Version 1 vs. Tape Version 2), and the order in which the witnesses were presented (Tape Version A vs. Tape Version B). The within-subjects factors (confidence and truth) each had two levels, resulting in high- and low-confidence true stories and high- and low-confidence false stories. Repeated measures analysis of variance (ANOVA) was performed on the 9-point Likert-type credibility ratings. There were no significant main effects of order or tape version, but there were significant interactions involving these variables. However, the general pattern of the findings was the same across tapes, thus the interactions result from slight differences in effect size for each tape version.

Truthfulness ratings. Consistent with Experiment 1, there was a main effect of confidence, $F(1, 203) = 182.93, p < .001$. The high-confidence testimonies were rated as significantly more truthful ($M = 6.2, SD = 2.18$) than the low-confidence testimonies ($M = 3.9, SD = 2.35$). However, unlike Experiment 1, there was no significant interaction between confidence and instructions (see Table 2).

A main effect of truthfulness was also discovered, $F(1, 203) = 29.86, p < .001$. Participants rated the false testimonies ($M = 5.4, SD = 2.21$) significantly more believable than the true stories ($M = 4.7, SD = 2.03$). The general trend for higher ratings for the false testimonies was consistent across the different tape versions.

TABLE 2: Experiment 2: Believability and Overall Accuracy Analysis (N = 217)

	M	SD	df	F	p <	d
Confidence			1, 203	182.94	.000*	.436
High	6.2	2.19				
Low	3.9	2.05				
Truth			1, 203	29.86	.000*	.115
True	4.7	2.03				
False	5.4	2.21				
Instructions			2, 203	.405	.668	.003
NC	5.1	2.14				
DC	5.0	2.13				
UC	5.0	2.11				
Confidence instructions			2, 203	.108	.898	.001
High confidence						
NC	6.4	2.19				
DC	6.2	2.22				
UC	6.2	2.17				
Low confidence						
NC	3.9	2.08				
DC	3.9	2.05				
UC	3.9	2.05				
Confidence truth			1, 203	.166	.685	.001
High confidence						
True	5.9	2.30				
False	6.6	2.07				
Low confidence						
True	3.6	1.76				
False	4.2	2.35				
Instructions truth			2, 203	.756	.471	.007
True						
NC	4.8	2.13				
DC	4.6	2.00				
UC	4.8	1.98				
False						
NC	5.5	2.14				
DC	5.4	2.27				
UC	5.2	2.24				
Accuracy instructions (in percentages)			1, 203	.336	.715	.002
NC	44	.203				
DC	46	.228				
UC	45	.172				

Note. NC = no confidence mentioned in instructions; DC = don't use confidence instructions; UC = use confidence instructions.

* $p < .05$.

There was a significant interaction of the Confidence \times True \times Tape \times Version, $F(1, 203) = 10.16, p = .002$. There were lower ratings of truthfulness for the low-confidence true testimonies. In five out of eight cases, false testimonies were rated higher than the true testimonies. There were three exceptions to this pattern in which either the direction was reversed (higher ratings for true testimonies) or the ratings were nearly equal.

There was also a significant effect of order in which the participants viewed truthfulness, $F(1, 203) = 13.69, p < .001$. Videotaped versions that began with the presentation of a false testimony were rated more truthful ($M = 5.3, SD = 1.93$) than those that began with a witness describing a true personal incident ($M = 4.8, SD = 2.23$).

Proportion accuracy. A one-way ANOVA revealed no significant difference in overall accuracy across the three written instruction conditions ($p = .715$). In addition, there were no significant differences between the instruction conditions in the independent calculations of accuracy for the high-confidence, low-confidence, true, or false testimonies.

In comparing the independent accuracy proportions to one another (high vs. low and true vs. false), no significant differences were discovered. This counters findings from Experiment 1, wherein accuracy was lower for high-confidence testimonies. In this study, none of the accuracy proportions—overall accuracy (45%), high-confidence accuracy (43%), low-confidence accuracy (47%), true accuracy (44%), and false accuracy (46%)—were significantly different from chance (50%).

Using overall accuracy and the independent accuracy proportions, an ANOVA was conducted with tape as the between-subjects variable. The results revealed a significant difference in the accuracy proportion for the high-confidence testimonies, $F(1, 2) = 19.64, p < .001$, with participants being significantly more accurate in Tape 2 (52%) than in Tape 1 (35%). Likewise, there was a significant difference in accuracy for the low-confidence testimonies across the tapes, $F(1, 1) = 9.48, p = .002$. Participants were significantly more accurate in Tape 1 (53%) than in Tape 2 (40%) in judging deception for the low-confidence testimonies.

Mention of confidence. The frequency with which the participants indicated that they used the witnesses' confidence in judging veracity was calculated and compared across conditions. A chi-square analysis revealed that the frequency of mentioning confidence differed across instructions, $\chi^2(2, N = 217) = 25.15, p < .001$. Few people actually mentioned using confidence as an aid to deception, but as expected, confidence was mentioned more frequently when people had been told to use it (16 out of 71 in the UC condition) than when they had not (1 out of 71 in the NC condition and 2 out of 75 in the DC condition).

DISCUSSION

In two studies, we attempted to determine if truth seekers use confidence in determining if a statement is truthful or deceptive and whether the use of confidence improves or impairs their ability to detect deception. Furthermore, we wanted to see if participants were consciously aware of using confidence as a guide to deception and whether we could influence their use of confidence through explicit instructions.

The fact that high-confidence testimonies were rated more truthful than low-confidence testimonies suggests that these college student judges were indeed using confidence but not in an effective manner. Similar to the effect of witness confidence on ratings of eyewitness testimony accuracy, these raters seemed to equate high confidence with truthfulness and low confidence with deception. Even more disturbing was the finding that the participants generally rated false stories as more truthful than the stories that were actually true. It appears that witness confidence and other cues the participants were using led them to find the deceptive testimonies more believable.

Instructions advising against the use of confidence did not have consistent effects across our two studies. In Experiment 1, participants instructed to ignore witness confidence did rate low-confidence witnesses as more believable and high-confidence witnesses as less believable compared to participants given no warnings about using confidence. However, the warning did not significantly increase their accuracy in detecting deception when they were forced to categorize each testimony as true or false. In Experiment 2, on the other hand, the explicit written instructions to ignore confidence or even to use confi-

dence did not significantly influence the truthfulness ratings or accuracy in detecting deception. This finding was counter to our hypothesis but somewhat in keeping with prior researchers' findings that expert testimony was rather unsuccessful in diminishing the jurors' use of confidence as an indicator of accuracy (Fox & Walters, 1986).

In our first study, participants were significantly better at detecting deception for low-confidence witnesses than for high-confidence witnesses and worse than chance in detecting deception in high-confidence witnesses. We thought that perhaps high confidence overwhelms all other cues and impairs the ability to identify deception. However, this finding was not replicated in Experiment 2, when we included other testimonies presented in different orders. Further research using various types of lies and truthful accounts will be necessary to determine whether high witness confidence can dominate other potentially more reliable but subtle cues to deception.

Consistent with Ekman and O'Sullivan's (1991) and Landry and Brigham's (1992) findings, lie detection ability in our untrained college students was no better than chance (50%). Prior research would have predicted that there would be higher accuracy in judging the true stories compared to the false stories (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). However, we found no support for that prediction; instead, our findings revealed no significant differences in accuracy rates for true and false testimonies.

Participants were most likely to mention that they had used confidence when making their decision if they had been explicitly told to do so. Even then, very few said that they used confidence as a guide to truthfulness. This result suggests that participants are not terribly aware of using confidence *per se* in making their judgments. Perhaps they used cues that are related to confidence (e.g., attractiveness), but they do not describe them in the same way that we (the experimenters) would. Further examination of the characteristics listed by the participants revealed that across conditions, they mentioned other potentially beneficial cues, such as eye movement, facial expressions, tone of voice, and body language. This, however, had little bearing on their overall ability to detect deception, because it remained at chance levels.

Although instructions did not influence lie detection performance, the findings of these two studies have important implications. Clearly,

laypeople (in this case, college students) rely on confidence in judging truthfulness; the high-confidence testimonies were rated most truthful. As stated previously, if a witness is attempting to persuade a juror to believe his or her version of events, chances are he or she will exhibit high confidence in his or her testimony. Given long delays between initiating events and trials, deceptive witnesses have many opportunities to rehearse their stories and increase their apparent confidence. The fact that our participants typically rated false stories as more believable than true stories further illustrates the difficulties that novices may face in recognizing dishonesty.

Limitations. One of the limitations of the experiments is that the participants were college students, many of whom had never served on a jury. In addition, it is possible that there were not any good cues to deception present in the testimonies used in this study. First, by controlling for confidence, we rendered it useless as a cue, even though it may help in some settings. But our four witnesses, and the overall sample of 60 witnesses from which they were drawn, appeared equally confident when telling the truth and when lying. Second, our lies were low stakes, so the emotions necessary to betray the liars through nonverbal behavior may not have been present (Frank & Ekman, 1997). Thus, perhaps confidence was the only cue that the truth seekers could pick up, but it did not assist them in distinguishing between true and false testimonies. Finally, our legal system was devised to identify liars through mechanisms such as confrontation and cross-examination, which our study did not include. In future studies, by using more complete and realistic lies and lie detection scenarios, perhaps we can determine whether jurors can effectively use cues available across various sources to improve their ability to detect deceit.

REFERENCES

- DePaulo, B., Charlton, K., Cooper, H., Lindsay, J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*, 346-357.

- DePaulo, B., Lindsay, J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74-112.
- Devenport, J., Penrod, S., & Cutler, B. (1997). Eyewitness identification evidence evaluating commonsense evaluations. *Psychology, Public Policy, and Law, 3*, 338-361.
- Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913-920.
- Fox, S. G., & Walters, H. A. (1986). The impact of general versus specific expert testimony and eyewitness confidence upon mock juror judgment. *Law and Human Behavior, 10*, 215-228.
- Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stakes lies. *Journal of Personality and Social Psychology, 72*, 1429-1439.
- Landry, K. L., & Brigham, J. C. (1992). The effects of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior, 16*, 663-675.
- Leippe, L. R., Romanczyk, A., & Manion, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports. *Journal of Personality and Social Psychology, 63*, 181-197.
- Pezdek, K., & Taylor, J. (2000). Discriminating between true and false events. In D. Bjorkland (Ed.), *False-memory creation in children and adults* (pp. 69-91). Mahwah, NJ: Lawrence Erlbaum.
- Sporer, S., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*, 315-324.
- Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135-154). Boston: Kluwer Academic.
- Whitley, B. E., & Greenberg, M. S. (1986). The role of eyewitness confidence in juror perceptions of credibility. *Journal of Applied Social Psychology, 16*, 387-409.