

## Public Eyewitness Confidence Can Be Influenced by the Presence of Other Witnesses

John S. Shaw III · Lauren M. Appio · Tana K. Zerr ·  
Kristin E. Pontoski

Published online: 21 February 2007

© American Psychology-Law Society/Division 41 of the American Psychological Association 2007

**Abstract** Four experiments were conducted to examine whether witnesses' public confidence ratings differ from their private ratings when there are social pressures to use confidence as an impression-management tool. In all four experiments, participants answered questions about a source event (a series of faces in the first three experiments and a simulated crime scene in the fourth). Half of the responses and confidence ratings were given privately and anonymously, and half were given publicly in front of one or more mock jurors. Two central findings emerged from the results. First, public confidence differed from private confidence only when there was more than one witness; when there were no other witnesses, public and private confidence were the same. Second, the direction of the change in public confidence in the multiple-witness settings was influenced by whether or not there was a possibility of being contradicted by the other witnesses. When there was no chance that the participants' responses could be contradicted, they raised their confidence ratings in public; when there was a chance that the other witnesses might contradict them, the participants lowered their public confidence ratings. The results are discussed in terms of self-presentation theory and implications for the legal system.

**Keywords** Eyewitness confidence · Public confidence · Confidence-accuracy relationship

When eyewitnesses testify in court, they are subject to powerful interpersonal social pressures. For example, conformity pressures may lead witnesses to conform their reports to what other witnesses have said (Luus & Wells, 1994; Shaw, Garven, & Wood, 1997). Social pressures may also alter public pronouncements of witness confidence. It is important to isolate and study these effects because witness confidence has a significant impact on jurors' evaluations of the accuracy of eyewitness testimony (Brewer & Burke, 2002; Cutler, Penrod, & Dexter, 1990; Cutler, Penrod, & Stuve, 1988; Lindsay, Wells, & O'Connor, 1989). Lay people, police officers, and attorneys all believe that eyewitness confidence is a reliable predictor of eyewitness accuracy (Brigham & Wolfskeil, 1983; Deffenbacher & Loftus, 1982; Noon & Hollin, 1987). Even the United States

---

J. S. Shaw III (✉) · L. M. Appio · T. K. Zerr · K. E. Pontoski  
Department of Psychology, Lafayette College, Easton, PA 18042-1781, USA  
e-mail: shawj@lafayette.edu

Supreme Court has endorsed the use of witness confidence as an indicator of witness accuracy (Neil v. Biggers, 1972).

For many years, the consensus among researchers was that eyewitness confidence is generally a poor predictor of eyewitness accuracy (Bothwell, Deffenbacher, & Brigham, 1987; Deffenbacher, 1980, 1991; Penrod & Cutler, 1995; Shaw & McClure, 1996). Recently, though, an increasing number of studies have challenged this broad assumption. Some researchers have identified situations in which confidence and accuracy are moderately, or even highly, correlated (Brewer & Wells, 2006; Lindsay, Read, & Sharma, 1998; Read, Lindsay, & Nicholls, 1998; Sporer, Penrod, Read, & Cutler, 1995), while others have noted that measurement and methodological artifacts can affect the magnitude of the confidence-accuracy relationship (Brewer, Keast, & Rishworth, 2002; Brewer, Weber, & Semmler, 2005; Busey, Tunnicliff, Loftus, & Loftus, 2000; Juslin, Olsson, & Winman, 1996). Nonetheless, most experts still agree that eyewitness confidence is highly malleable and that the relationship between eyewitness confidence and eyewitness accuracy can be affected by a wide range of factors (Kassin, Tubb, Hosch, & Memon, 2001).

Of the many factors that might affect eyewitness confidence independently of eyewitness accuracy, one that has received little attention in the literature is the fact that eyewitness confidence is often communicated in public settings, most notably during trials and other courtroom proceedings. Even many witness interviews occur in public or quasi-public settings. Despite the public nature of the eyewitness experience, though, very few studies have collected “public” confidence ratings. Most experimental examinations of eyewitness confidence have utilized anonymous, written data collection methods (for reviews of the eyewitness confidence literature, see Penrod & Cutler, 1995, and Shaw, McClure, & Dykstra, 2007).<sup>1</sup> Although Wells, Ferguson, and Lindsay (1981), Brigham, Maass, Snyder, and Spaulding (1982), and Luus and Wells (1994) all measured “public” confidence, none of these studies compared public and private confidence ratings directly. In one of the few studies to examine public and private confidence together, Kassin, Rigby, and Castillo (1991) found that the confidence-accuracy correlation was higher for participants who were informed that they would have to explain their confidence ratings to the experimenter than for participants who were given no such instruction (cf., Robinson & Johnson, 1998).

More recently, Shaw and his colleagues have demonstrated that confidence ratings shared openly with others can differ from ratings provided privately and anonymously (Shaw, Woythaler, & Zerr, 2001; Shaw, Zerr, & Woythaler, 2001). One possible cause of this effect is that eyewitnesses may use self-presentation strategies in order to manage the impressions they make when speaking in public settings (Schlenker, 2003). Witnesses want to impress judges, jurors, and attorneys with their skills on the witness stand, and one way to achieve this goal is to employ a self-presentation strategy commonly referred to as self-promotion, in which people strive to present themselves as skilled and competent (Arkin, 1981; Jones & Pittman, 1982).

Numerous studies have demonstrated that people engage in self-promotion in a wide range of settings in order to present themselves in a favorable light (for reviews, see Schlenker, 2003, and Schlenker & Pontari, 2000). For example, people will use self-promotion in job interviews (Stevens & Kristof, 1995), for career advancement (Kacmar, Delery, & Ferris, 1992), in everyday social interactions (Nezlek & Leary, 2002), and to make a favorable impression as a talk-show guest (Schütz, 1997). Even young children seem to know how to use self-promotion (Aloise-Young, 1993; Bennett & Yeeles, 1990).

<sup>1</sup>Many articles contain few details about how confidence was measured, so it is difficult to identify all of the studies that have examined public eyewitness confidence.

Thus, it is likely that witnesses engage in self-promotion in the courtroom. For example, they may alter their manner of dress, speech patterns, tone of voice, eye contact, or body language in order to project an impression of competence. Indeed, such strategies can work, for even the simple act of wearing eyeglasses can impart an impression of intelligence (Terry & Krantz, 1993). Witnesses may even adjust their public confidence judgments as an impression-management strategy because confident witnesses are seen as credible and persuasive (Slovenko, 1999). That is, in order to be viewed as skilled and capable, eyewitnesses may artificially inflate their public pronouncements of confidence to impress others with the certainty of their memories.

As with most impression-management strategies, however, the context of the social situation can affect how, or even whether, one engages in self-promotion (Kacmar, Carlson, & Bratton, 2004; Schlenker, 2003). As Schlenker has noted, “self-presentation is an activity that is shaped by a combination of personality, situational, and audience factors” (p. 498). For example, although direct boasting about one’s skills and accomplishments may be an appropriate self-promotion technique during a political campaign or in a job interview, it is inappropriate when consoling a co-worker about a poor performance report.

Therefore, context may play an important role in determining when and how witnesses engage in self-promotion in the courtroom. If there is little likelihood that a witness will be contradicted by other witnesses or additional evidence, that witness may safely inflate her or his public confidence ratings in order to impress others. On the other hand, if there is a substantial possibility that her or his testimony will be inconsistent with other evidence, the witness may take a more cautious approach and provide lower, more conservative, confidence ratings. In such situations, it may be prudent for witnesses to lower their public confidence ratings because the social costs of being overconfident and wrong are potentially higher than the costs of being underconfident and right. One can think of this strategy as the “better safe than sorry” approach to self-promotion.

In previous studies of public confidence, Shaw and his colleagues have simulated the multiple-witness situation in order to test whether participants would indeed follow the “better safe than sorry” strategy and give lower public confidence ratings when faced with the possibility that other participants would know whether their memory reports were accurate. In Shaw, Woythaler, and Zerr (2001), participants engaged in a face-recognition task, and in Shaw, Zerr, and Woythaler (2001) participants answered multiple-choice questions about a videotape of a simulated robbery. In both studies, participants attended in groups of 3 or 4, and all participants viewed the same stimulus materials together as a group. Both studies incorporated a response-privacy manipulation in which half of the responses and confidence ratings were given privately and anonymously, and half were given out loud in the presence of the other participants.

In Shaw, Woythaler, and Zerr (2001), all participants answered the exact same set of test questions, but in Shaw, Zerr, and Woythaler (2001) each participant answered a different set of questions. Because all of the participants viewed the same set of stimulus materials, the participants in both studies were aware that the other participants in the same experimental session would be able to assess the accuracy of their public responses. Consistent with a self-presentation explanation, the confidence ratings were significantly lower in public than in private in both studies, yet there was no difference in public and private response accuracy. These results suggest that the participants provided lower confidence ratings when sharing the ratings aloud with the other participants because they wanted to avoid the possibility of being highly confident about responses that the other participants might know were incorrect.

In contrast to the multiple-witness setting simulated in these two previous studies, many criminal cases involve only a single eyewitness. For example, in many sexual assaults, the victim is the only eyewitness. In such single-witness situations, witnesses might artificially *inflate* their public confidence judgments in court in order to impress judges, jurors, and other courtroom observers, particularly when there is little or no fear of being contradicted by other witnesses. The present study was designed to examine public confidence ratings in such a one-witness setting.

## Experiments 1 and 2

The main goal of Experiments 1 and 2 was to examine whether witnesses inflate their public confidence ratings when there are explicit social pressures to impress others with their witnessing skills and there is little chance that their responses will be contradicted by the responses of other witnesses. To address this question, we utilized a straightforward face-recognition task in the context of a highly simplified mock testimony situation.

In both experiments, participants attended in groups of 3 or 4. In an initial study phase, they each viewed a unique face booklet containing 16 faces. After a brief retention interval, the participants were given a recognition test on a series of 16 faces (8 old and 8 new), and they reported their confidence for each of the 16 decisions. The faces were presented in two sets of 8. For one set, the participants recorded their answers and confidence ratings privately and anonymously (the private condition), and for the other set, they shared their responses and confidence ratings out loud with the other participants in the experimental session (the public condition). These procedures ensured that each participant studied a *unique* set of 16 faces, thus precluding the possibility that the other participants would be able to check the accuracy of their public responses.

When each participant shared her or his responses aloud in the public condition, the other 2 or 3 participants acted as mock jurors and recorded their evaluations of the witness on written forms. This evaluative component of the design, which had not been included in previous studies on public confidence (Shaw, Woythaler, & Zerr, 2001; Shaw, Zerr, and Woythaler, 2001), had two important benefits. First, it helped to create a testifying situation that more closely simulated the evaluative social pressures inherent in courtroom testimony. Each participant knew that the mock jurors were trying to assess the accuracy of her or his “testimony,” just as judges and jurors do in real courtrooms. In effect, this enhanced the possibility that the participants would use self-presentation strategies in order to impress the mock jurors. Second, it allowed us to examine the accuracy of the mock jurors’ evaluations of the participants’ memories.

We expected that the social pressures to impress others would lead the participants to use confidence as an impression-management tool. Because each participant engaged in a recognition task for a unique set of faces, and thus had no fear of being contradicted, we predicted that the pressure to be seen as skilled and capable would result in public confidence ratings that were higher than the confidence ratings reported privately and anonymously.

Given that previous research has demonstrated that jurors rely heavily on witness confidence in judging witness accuracy (Brewer & Burke, 2002; Cutler, Penrod, & Dexter, 1990; Cutler, Penrod, & Stuve, 1988), we also expected that the mock jurors’ ratings of the accuracy of the participants’ memories would be highly correlated with the participants’ public confidence judgments. Furthermore, because we anticipated that public confidence would be artificially inflated due to self-presentation pressures, we expected that the mock jurors’ ratings of participant accuracy would *not* be correlated with the actual accuracy of the participants.

## Method

### *Participants*

Forty-eight undergraduates (34 females and 14 males) and 45 undergraduates (32 females and 13 males) at a small liberal arts college participated in Experiments 1 and 2, respectively, in exchange for extra credit in various psychology courses. Participants' ages ranged from 18 to 21 years, with an overall mean of 18.7.

### *Stimulus Materials*

The stimulus materials were identical for Experiments 1 and 2. For the study phase in each experiment, four different face booklets were created. Each face booklet contained four single-sided pages with 4 faces on each page, for a total of 16 faces per booklet. In each face booklet, 8 of the faces were test faces and 8 were filler faces. The 8 filler faces were identical among the four face booklets, but each booklet contained a unique set of 8 test faces. Although each booklet had some test faces in common with the other three booklets, none of the booklets had the exact same set of 8 test faces.

Each black and white picture in the face booklet depicted the face of a white male college student.<sup>2</sup> No eyeglasses, jewelry, or distinctive pieces of clothing were visible in any of the pictures. Sixteen photographic slides were used in the test phase of the experiment. These slides contained the same images of the faces that were used in the face booklets, although the printed versions of the photographs were slightly darker than the projected slide images.<sup>3</sup> Because of the manner in which the test faces were assigned to the face booklets, for any particular booklet (i.e., for any particular participant) one half of the 16 test faces in the slides were *old* (they did appear in the face booklet), and one half were *new* (they did not appear in the face booklet). The test faces were randomly assigned to two groups of eight slides, with the restriction that each group contained four new and four old faces for each of the four face booklets.

### *Procedures and Design*

Participants attended in groups of 3 or 4, and they sat around a conference table in a small laboratory room.

*Study and retention phases.* The study and retention phases were identical in the two experiments. Each participant was given one of the four unique face booklets, and all participants were instructed to study the 16 faces in the booklet for 30 s. They were also informed that "Some of the faces in your booklet are the same as faces in the other booklets, and some of them are different. Thus, each of you has a unique set of faces." After the study phase, participants engaged in a 4-min filler task in which they answered a few general questions about witnesses.

*Experiment 1 test phase.* In Experiment 1, the test phase was conducted in two segments. In each segment, eight face slides were projected one at a time on a 1.5 m by 1.5 m screen. Each face appeared on the screen for 5 s, with a 5-s interval between faces. The participants were instructed to indicate whether the face had appeared in their personal face booklets by circling "Yes" or "No" in individual anonymous response booklets. Immediately after making

<sup>2</sup>We thank Roy Malpass for permission to use the face pictures.

<sup>3</sup>Busey et al. (2000) found that a mismatch in luminance of target and test photographs can lead to a dissociation of witness confidence and accuracy. Because the difference in luminance was very small in the present study, we do not believe that it had a material impact on the relationship between confidence and accuracy.

their choice for each face, the participants indicated how confident they were in the accuracy of their choice by circling a number on an 11-point scale that ranged from 0 (*totally guessing*) to 10 (*absolutely certain*).

For all participants, one group of eight faces constituted the private condition, and the other group was the public condition. Thus, *Response Privacy* was manipulated within participants, with eight answers shared aloud (public condition) and eight answers given privately and anonymously (private condition). To control for order effects, the public condition was administered first for half of the participants and second for the other half. Also, the order of the two groups of test slides was counterbalanced across participants. These counterbalancing measures resulted in four versions of the test phase ( $N = 12$  for each version).

No special instructions were given prior to the eight faces in the private condition. However, before viewing the test faces in the public condition, the participants were told that they would be sharing their answers and confidence ratings in that segment with the other participants in the session. Specifically, the participants were instructed:

For these eight slides, you are going to share your answers with everyone here. Here is how it will work. For each slide, you will have five seconds to view the slide and answer both questions in your response booklet. When we are done with all eight slides, each of you will read your answers and your confidence ratings aloud to the other people in the room. When it is your turn to read your answers and confidence ratings aloud, the other participants will listen closely and fill out a form in which they rate how good your memory is.

The participants were told that we were asking them to report some answers privately and others publicly “because we are experimenting with different methods of collecting data.” No other justification was given for the response-privacy manipulation. In the public condition, all of the participants made their recognition choices and indicated their confidence ratings on written forms immediately after each slide, just as in the private condition. After all of the written forms had been completed, the participants took turns reading their answers and confidence ratings aloud. One at a time, each participant read all eight of her or his responses and confidence ratings while the other participants acted as mock jurors and listened to the participant’s responses. At the conclusion of each participant’s oral responses, the mock jurors were given evaluation forms on which they rated that participant on three questions. The first question was “In your opinion, how accurate is this witness’s memory for these faces?” which was accompanied by an 11-point scale that ranged from 0 (*Extremely Bad*) to 10 (*Extremely Good*). The second question was “How many of the 8 questions do you think this witness answered correctly?” and the final question was “In your opinion, would this witness be a good witness in a courtroom?” which was followed by a scale that ranged from 0 (*Definitely No*) to 10 (*Definitely Yes*).

After the first participant was done and the mock jurors had completed the evaluation forms for that participant, another participant shared her or his answers aloud. This process continued until all of the participants had given their public responses out loud. Thus, all participants performed once as a witness and two or three times as a mock juror. Because all of the responses and confidence ratings were recorded in the written response booklets *before* any of the participants shared their answers out loud, there was no possibility that a participant’s choices or confidence ratings could be affected by the choices and confidence ratings of the other participants.

*Experiment 2 test phase.* The test phase in Experiment 2 was essentially the same as in Experiment 1, except for a few modifications in the public condition. Instead of Experiment 1’s two-step procedure in which the participants first recorded their responses and confidence ratings on paper and then later shared them with the other participants, in the public condition in Experiment 2 the participants provided their answers and confidence ratings aloud as each test

slide was shown. These changes in the test-phase procedures resulted in a witnessing experience that simulated more closely the experiences of actual witnesses in a courtroom. In the public condition in Experiment 2, the witnesses were seated in a “witness chair” facing the mock jurors. Also, instead of first recording their answers to the recognition questions on a written form, the witnesses in the public condition in Experiment 2 gave their answers and confidence ratings out loud as each test slide appeared on the screen. Thus, each witness “testified” aloud in front of a “jury” of two or three other participants.

One at a time, the participants viewed the test slides during the public condition. As each face appeared on the screen, the participant announced her or his decision aloud (“Yes, it was in my face booklet” or “No, it was not”) and then provided a confidence rating for that decision on the same 11-point scale as in Experiment 1. During each participant’s “testimony,” the mock jurors sat with their backs to the screen and faced the participant. This arrangement helped create an atmosphere in which the presence of the mock jurors was highly salient to the testifying participant, and it also prevented the mock jurors from seeing the test slides before it was their turn to testify. Each participant took a turn in the witness chair, with the other participants acting as jurors for that participant. Although all participants saw the same set of 8 slides during the test phase, they each saw the slides for the first time when “testifying,” because their backs had been to the screen when they were acting as mock jurors. During the test phase, the participants were reminded that they had each studied unique face booklets.

Several safeguards were incorporated to ensure that the participants’ public choices and confidence ratings were not affected by the other participants’ responses. We repeatedly emphasized throughout the session that each participant viewed a unique set of faces; thus, it made no sense for the participants to mimic each other’s recognition decisions or confidence ratings. No talking was allowed during the experimental session (except when the participants gave their responses aloud), and there was no direct interaction among the participants.

As each participant testified, the mock jurors recorded that participant’s responses and confidence ratings on written forms. Not only did this ensure that the mock jurors paid close attention to the “testimony,” but it also provided a means for recording the responses for later analysis. When each participant was finished, the mock jurors completed the same three evaluation questions as in Experiment 1.

Due to the unusual data collection method for the public responses in Experiment 2, the public condition always followed the private condition. This fixed order allowed the participants to become familiar with the recognition task and the confidence scale before they had to speak out loud in front of the other participants. Because there were no testing-order effects in Experiment 1,<sup>4</sup> the fixed order was not expected to compromise the internal validity of Experiment 2.

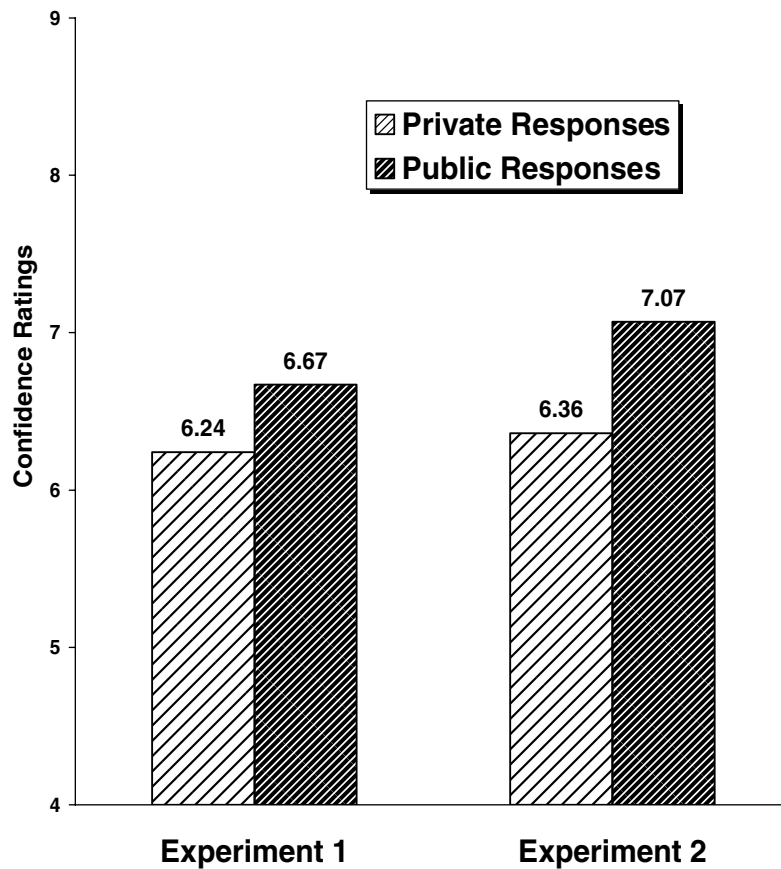
As in Experiment 1, the assignment of the two sets of slides to the private and public conditions was counterbalanced across participants.

## Results

### *Witness Confidence and Accuracy*

As expected, *Response Privacy* had a significant effect on confidence in both experiments. In Experiment 1, confidence for the public responses ( $M = 6.67$ ,  $SD = 1.59$ ) was significantly

<sup>4</sup>In Experiment 1, there was no effect of testing order on confidence,  $F(1, 46) = 1.19$ , *ns*, nor was there a testing order by response privacy interaction,  $F(1, 46) = 0.02$ , *ns*.



**Fig. 1** Mean confidence ratings for the public and private conditions in Experiments 1 and 2

higher than for the private responses ( $M = 6.24$ ,  $SD = 1.52$ ),  $t(47) = 2.08$ ,  $p < .05$ ,  $d = 0.30$ . *Response Privacy* had a similar effect on confidence in Experiment 2. Once again, mean confidence was significantly higher for the public responses ( $M = 7.07$ ,  $SD = 1.19$ ) than for the private responses ( $M = 6.36$ ,  $SD = 1.42$ ),  $t(44) = 3.60$ ,  $p < .001$ ,  $d = 0.54$ . These results are summarized in Fig. 1.

In both experiments, *Response Privacy* had no effect on the accuracy of the participants' responses. There was no significant difference between the response accuracy in the public and private conditions either in Experiment 1 ( $M = 62.2\%$  correct for the public responses,  $SD = 15.8$ ; and  $M = 63.3\%$  correct for the private responses,  $SD = 15.8$ ;  $t(47) = 0.31$ ,  $ns$ ,  $d = 0.04$ ), or in Experiment 2 ( $M = 58.1\%$  correct for the public responses,  $SD = 18.3$ ; and  $M = 61.7\%$  correct for the private responses,  $SD = 16.9$ ;  $t(45) = 1.09$ ,  $ns$ ,  $d = 0.16$ ).

Confidence-accuracy correlations were calculated on both a between-participant and within-participant basis for both experiments. The between-participant confidence-accuracy correlations used the mean confidence and accuracy scores for each participant as the units of analysis, whereas the within-participant correlations were based on the responses to the individual questions.



**Table 1** Between-participant correlations of the mock jurors' ratings with the response accuracy and confidence of the witnesses for the public responses in Experiments 1 and 2

Witness Responses	Jurors' Ratings		
	How accurate is this witness's memory?	How many questions did this witness answer correctly?	How good would this witness be in a courtroom?
<b>Experiment 1</b>			
Response Accuracy	-.22	-.20	-.23
Confidence	.82**	.80**	.68**
<b>Experiment 2</b>			
Response Accuracy	-.13	-.07	-.09
Confidence	.80**	.69**	.74**

\*\*denotes  $p < .01$ .

In Experiment 1, the between-participant confidence-accuracy correlations were  $-.04$  for the private responses and  $-.14$  for the public responses, and the mean within-participant correlations were  $.16$  for both the private and public responses. None of these correlations were significantly different from 0, and there was no significant difference between the between-participant correlations in the public and private conditions, *Fisher's*  $z = 0.21$ , or between the mean within-participant correlations,  $t(45) = 0.01$ , *ns*,  $d = 0.00$ .

In Experiment 2, the between-participant confidence-accuracy correlations were  $.09$  for the private responses and  $-.26$  for the public responses, and the mean within-participant correlations were  $.05$  for the private responses and  $.17$  for the public responses. None of these correlations were significantly different from 0, and there was no significant difference between the mean within-participant correlations,  $t(43) = 1.36$ , *ns*,  $d = 0.21$  or between the two between-participant correlations, *Fisher's*  $z = 1.63$ .

#### *Mock Jurors' Ratings*

For each participant, the average rating of the mock jurors was calculated for each of the three juror evaluation questions. Between-participant correlations for both experiments involving the mean ratings of the mock jurors and the mean response accuracy and mean confidence of the participants for the public condition are displayed in Table 1. As is evident in the table, the pattern of correlations is identical for both experiments. There were large, statistically significant positive correlations between the participants' public confidence and all three of the mock jurors' ratings. For example, the greater the participants' public confidence, the higher the mock jurors' evaluation of the accuracy of the participants' memory. In contrast, there were no significant correlations between the mock jurors' ratings and the response accuracy of the participants. In fact, there was a non-significant trend toward a negative relationship between the jurors' ratings and the participants' response accuracy.

#### Discussion

As predicted, the participants' confidence ratings in Experiments 1 and 2 were significantly higher for responses shared aloud with the other participants than for responses reported privately, yet the accuracy of the participants' answers was unaffected by the response-privacy manipulation. This result was expected for two reasons. First, there were social pressures for the witnesses to "compete" against each other and thus use higher confidence ratings in public as an impression

management tool. Second, there were no social costs stemming from being overconfident about a wrong answer, because none of the other participants could know whether a witness's responses were correct or incorrect.

Experiments 1 and 2 simulated a particular type of single-witness situation. Even though each witness gave responses about a unique source event (a unique set of faces), there were multiple witnesses in each experimental session. This is similar to actual trials in which there are several witnesses, but each witness testifies about a different portion of the event. For example, in a bank robbery trial, a teller insider the bank might testify about the female robber who entered the bank, whereas a passerby on the sidewalk might testify about the male getaway driver who was waiting outside of the bank. In essence, there are two "single witnesses" to the robbery. In such situations, there are likely to be social pressures for witnesses to inflate their public confidence ratings as an impression management tool in an effort to "out-do" other witnesses, and they can do so safely because there is little chance that their answers will be contradicted by the other witnesses.

In contrast, what happens to witness confidence in public if there are no other witnesses? That is, when there is truly only a single witness to an event, such as in a sexual assault case where the victim is the only eyewitness, will that single witness still raise her or his confidence ratings in public in order to impress jurors or a judge, or will the pressure to impress be diminished because there are no other witnesses against whom the witness can be compared? That was the principal question addressed by Experiment 3.

### Experiment 3

In the first two experiments, each of the participants took turns being a witness while the other participants acted as jurors. This turn-taking process likely motivated the participants to try to do a "better job" than the other participants in the experimental session. To eliminate the competitive nature of the witnessing experience in Experiment 3, one participant in each session was randomly chosen to be the witness, and the other two participants were assigned to be jurors. The participants remained in their assigned roles throughout the entire experiment.

Despite the fact that there were no pressures to compete directly with other witnesses in Experiment 3, we still expected that the witnesses would give higher confidence ratings in public than in private because they would want to impress the mock jurors with their witnessing skills. As in the first two experiments, there was no chance that the witnesses could be "caught" being overconfident about wrong answers, because there was no possibility that their responses could be contradicted by the responses of other witnesses. Also, consistent with the results from the first two experiments, we expected that the mock jurors' ratings of the accuracy of the participants' memories would be highly correlated with the witnesses' public confidence judgments but would *not* be correlated with the actual accuracy of the witnesses.

### Method

#### *Participants*

One hundred and thirty two undergraduates (89 females and 43 males) at a small liberal arts college participated in exchange for extra credit in various psychology courses. The mean age of the participants was 19.4 years.

### *Materials and Procedures*

Three participants attended each experimental session. After introductory remarks in which the participants were told that the study was a “simple role-playing experiment about witnesses and jurors,” two of the participants were randomly assigned to be jurors and one to be a witness. In contrast to Experiments 1 and 2, the role of each participant was fixed (as either a witness or a juror) throughout the entire experimental session. Thus, there were 44 witnesses and 88 mock jurors.

The materials and procedures in Experiment 3 were identical to those in Experiments 1 and 2, except as noted here. Once again, the experiment consisted of three phases. In the study phase, the witness had 30 s to study a face booklet that contained 16 faces. The face booklet was identical to “Booklet 2” that was used in Experiments 1 and 2. During the study phase, the two mock jurors worked on a filler task in which they were asked to evaluate several comic strips. After the study phase, there was 3-min retention phase in which all three participants worked on the comic-strip filler task.

During the final test phase, the witness viewed two sets of eight faces via a slide projector (four old and four new in each set) and provided responses and confidence ratings as in the first two experiments. In addition, the witness was asked to give a brief (about 15 s) explanation of her or his choice for each slide (e.g., “I know I saw this person before because he has very large ears” or “I don’t remember this person at all—his face is very skinny”).<sup>5</sup> This modification was included to examine further the large correlations between witness confidence and the mock jurors’ evaluations in the first two experiments. Because these explanations afforded additional information on which the mock jurors could base their evaluations of the accuracy of the witnesses, Experiment 3 provided an even stronger test of the hypothesis that the mock jurors’ evaluations would be more strongly related to witness confidence than to witness accuracy.

*Response Privacy* was again manipulated within participants. In the private condition, the witness provided her or his responses, including the brief explanations, in written anonymous response booklets while the mock jurors worked on the comic-strip filler task. In the public condition, the witness “testified” in the same manner as in Experiment 2, providing responses, confidence ratings, and explanations aloud while facing the two mock jurors. The mock jurors recorded the choices and confidence ratings of the witness in their juror booklets, and they also wrote down the explanations offered by the witness for each of the choices. When the witness was finished with all eight slides in the public condition, the two mock jurors answered the same three witness-evaluation questions as in Experiments 1 and 2. The order of the public and private conditions was counterbalanced across participants, and the assignment of the two sets of eight test slides to the private and public conditions was also counterbalanced, yielding four counterbalancing versions ( $N = 11$  witnesses in each cell).

### Results

#### *Witness Confidence and Accuracy*

In contrast to Experiments 1 and 2, *Response Privacy* did not alter witness confidence in Experiment 3. The mean confidence for the public responses ( $M = 6.25$ ,  $SD = 1.29$ ) was not significantly different than the mean confidence for the private responses ( $M = 6.28$ ,  $SD = 1.24$ ),  $t(43) = 0.23$ ,  $ns$ ,  $d = 0.03$ . Also, there was no significant difference in response

<sup>5</sup>The open-ended explanations of the witnesses were not analyzed.

**Table 2** Between-participant correlations of the mock jurors' ratings with the response accuracy and confidence of the witnesses for the public responses in Experiment 3

Witness Responses	Jurors' Ratings		
	How accurate is this witness's memory?	How many questions did this witness answer correctly?	How good would this witness be in a courtroom?
Response Accuracy	-.14	-.29	-.09
Confidence	.72**	.61**	.52**

\*\*denotes  $p < .01$ .

accuracy between the public and private conditions ( $M = 64.8\%$ ,  $SD = 15.1$  vs.  $61.7\%$ ,  $SD = 18.0$  correct,  $t(43) = 0.86$ ,  $ns$ ,  $d = 0.13$ ).

As in the first two experiments, confidence-accuracy correlations were calculated on both a between- and within-participant basis. The between-participant confidence-accuracy correlations were not statistically significant for either the private ( $r = .14$ ) or public ( $r = -.18$ ) conditions, and they were also not significantly different from each other, *Fisher's z* = 1.45. The mean within-participant correlations were .06 for the private responses and .13 for the public responses. Again, these correlations were not significantly different from 0 or from each other,  $t(42) = 0.84$ ,  $ns$ ,  $d = 0.13$ .

#### Mock Jurors' Ratings

Between-participant correlations were calculated between the mean ratings of the mock jurors on the three evaluation questions and the mean response accuracy and mean confidence of the witnesses for the questions in the public condition.<sup>6</sup> As is apparent in Table 2, the pattern of correlations is identical to that in Experiments 1 and 2. There were no significant correlations between the mock jurors' evaluations of the witnesses and the actual response accuracy of the witnesses, yet the correlations between the witnesses' confidence scores and all three of the mock jurors' ratings were large and statistically significant. Once again, there was a non-significant trend toward a negative relationship between the mock jurors' ratings and the accuracy of the witnesses.

#### Discussion

Contrary to our prediction, public confidence was not higher than private confidence in Experiment 3. Evidently, the opportunity to impress mock jurors was not enough by itself to raise public confidence levels, even though there was no chance that the witnesses' responses might be contradicted by other witnesses. From this result we can infer that witnesses are more likely to inflate their public confidence ratings in situations where there are other witnesses competing for attention (as was the case in Experiments 1 and 2) than they are in pure single-witness situations where there are no pressures to compete with other witnesses (as simulated in Experiment 3).

<sup>6</sup>In three of the experimental sessions, the mock jurors failed to answer the witness-evaluation questions. Thus, the correlations between the mock jurors' evaluations and the witnesses' response accuracy and confidence ratings could be calculated for only 41 of the witnesses.

## Experiment 4

Experiment 4 was designed to add to the results of the first three experiments in two important ways. To begin, Experiment 4 used more complex stimulus materials—a slide show depicting the inside of a student’s apartment—which added to the realism of the participant experience and enhanced the generalizability of the results. Second, Experiment 4 involved a direct comparison of public confidence ratings in a pure single-witness condition with a pure multiple-witness condition.

The single-witness condition served as a conceptual replication of Experiment 3, because in both cases only one witness viewed the stimulus materials and answered questions about those materials. In the multiple-witness condition in Experiment 4, all of the witnesses viewed the same stimulus materials and answered the same set of multiple-choice questions, analogous to real-world situations in which multiple witnesses all testify about the same set of facts. Thus, the witnesses in Experiment 4 could be contradicted by the testimony of the other witnesses in the experimental session because they all answered the same questions about the same source event, in contrast to Experiments 1 and 2, where such public contradiction was not possible. The multiple-witness condition in Experiment 4 was similar in many respects to the multiple-witness settings in Shaw, Woythaler, and Zerr (2001), and Shaw, Zerr, and Woythaler (2001).

We expected that *Response Privacy* and *Number of Witnesses* would interact in their effects on witness confidence in Experiment 4. When there was only a single witness, we expected the results to mirror those of Experiment 3—there would be no difference between the public and private confidence ratings. When there were multiple witnesses, however, we expected that the confidence ratings would be *lower* in public than in private. We believed that the presence of other witnesses would lead the participants to use their confidence ratings as an impression management tool, but because the participants could potentially be contradicted by the other witnesses in the same experimental session, we expected that the participants would *lower* their public confidence ratings to avoid the socially embarrassing possibility of being highly confident about incorrect responses. This “better safe than sorry” approach to public confidence ratings has been demonstrated previously in Shaw, Woythaler, and Zerr (2001) and Shaw, Zerr, and Woythaler (2001).

Finally, consistent with the first three experiments, we expected that the mock judges’ ratings of the accuracy of the witnesses’ memories would be highly correlated with the witnesses’ public confidence judgments but would *not* be correlated with the actual accuracy of the witnesses.

## Method

### *Participants*

One hundred and seventy four undergraduates (131 females and 43 males) at a small liberal arts college participated in exchange for extra credit in various psychology courses. The mean age of the participants was 19.3 years.

### *Materials and Procedures*

Experiment 4 employed a 2 (*Response Privacy*) by 2 (*Number of Witnesses*) mixed-factorial design. *Response Privacy* was manipulated within participants and had two levels—public responses and private responses; *Number of Witnesses* was manipulated between participants and also had two levels—single witness and multiple witnesses.

Participants attended in groups ranging in size from 2 to 5. After being seated between 2 and 4 m from a screen at the front of the laboratory, the participants were randomly assigned to play the role of either a judge or a detective. In the *single-witness condition*, there were two participants, one of whom was assigned to be the judge and the other to be the detective. In the *multiple-witness condition* there were 3, 4, or 5 participants—one participant was randomly selected to be the judge and the other participants were assigned to be detectives. In all cases, each participant played the same role throughout the entire experimental session. There were a total of 105 detectives—49 in the single-witness condition and 56 in the multiple-witness condition.

*Study phase.* After the judge was taken to a soundproof cubicle, the detective(s) watched a series of 31 color slides depicting the inside of a student's apartment. Prior to viewing the slides, the participants were instructed to imagine that they were detectives with a local police department investigating the recent disappearance of Jennifer (the occupant of the apartment). The detectives were told to "view the slides carefully and pay close attention to the kinds of details that might help you figure out what happened to Jennifer."

A Kodak Carousel projector was used to present the slides at the rate of 3 s per slide. The slides depicted various portions of a one-bedroom apartment. Several scenes were contained in the slide series, including a study area, a kitchen, a bathroom, a bedroom, and a hall closet. The images depicted in the slides moved logically through the apartment in order to simulate a detective searching for clues. Eight target items were contained within the stimulus slides—a peach, a kitchen glass, a magazine, a compact disc, a book, a stuffed animal, a playing card, and a soda can. The slides were shown without sound or narration, and all of the detectives watched the slide show together.

*Test phase.* The test phase was conducted in two segments. In each segment, the detectives answered four multiple-choice questions about the target items in the slides. For each question, the experimenter asked the detectives to identify which of four objects was actually in the apartment (e.g., "Which of these magazines did you see in the apartment?"). The four response alternatives were presented via a photographic slide that contained four objects, labeled A, B, C, and D, arranged in a single row. For all questions, the four objects in the slide were of the same general category (e.g., four magazines, four soda cans, or four pieces of fruit). The slide for each question remained on the screen for 5 s, during which time the detective gave her or his response. After each question, the experimenter asked the detective to provide a confidence rating "as a percentage number that can range anywhere from 0% confident to 100% confident." The eight multiple-choice questions were randomly assigned to two groups of four questions.

For all participants, one group of four multiple-choice questions constituted the *private condition*, and the other group of four questions was the *public condition*. Thus, four multiple-choice responses and the associated confidence ratings were shared aloud with the judge (public condition) and four responses were given privately and anonymously (private condition). To control for order effects, the public condition was administered first for half of the participants and second for the other half.

In the private condition, the detectives provided their multiple-choice answers and confidence ratings in writing on individual response sheets, and they were assured that their answers would remain completely private and confidential. The judge remained in the soundproof cubicle while the detectives answered the four questions in the private condition, and their written responses were collected as soon as they were done. The private condition was conducted exactly the same for both the single-witness and multiple-witness conditions.

In the public condition, the detectives gave their multiple-choice responses and confidence ratings aloud in front of the judge. Prior to answering the questions, each detective was reminded that the judge did not see the slide show that the detectives had seen. As each detective answered

the four questions aloud in the public condition, the judge recorded the detective's responses and confidence ratings on a written form.

The questioning of the detectives in the public condition was essentially the same in both the single-witness and multiple-witness conditions—each detective was questioned one at a time, with the judge recording that detective's responses and confidence ratings. In the multiple-witness condition, all of the detectives were placed in individual soundproof cubicles, and they were retrieved one at a time to answer the multiple-choice questions. When they were brought out, they were informed that each detective would be answering the same set of four questions in front of the judge. Thus, the possibility that the other detectives might provide contradictory responses distinguished this pure multiple-witness situation from the hybrid single-witness/multiple-witness setting in Experiments 1 and 2.

When each detective was finished in the public condition, the mock judge completed the same three evaluation questions as in the first three experiments—"How accurate is this detective's memory?" "How many of the questions do you think this detective answered correctly?" and "In your opinion, would this detective be a good witness in a courtroom?". Then the detective returned to her or his soundproof cubicle, and the next detective was retrieved to answer the questions about the slides.

## Results

### *Witness Confidence and Accuracy*

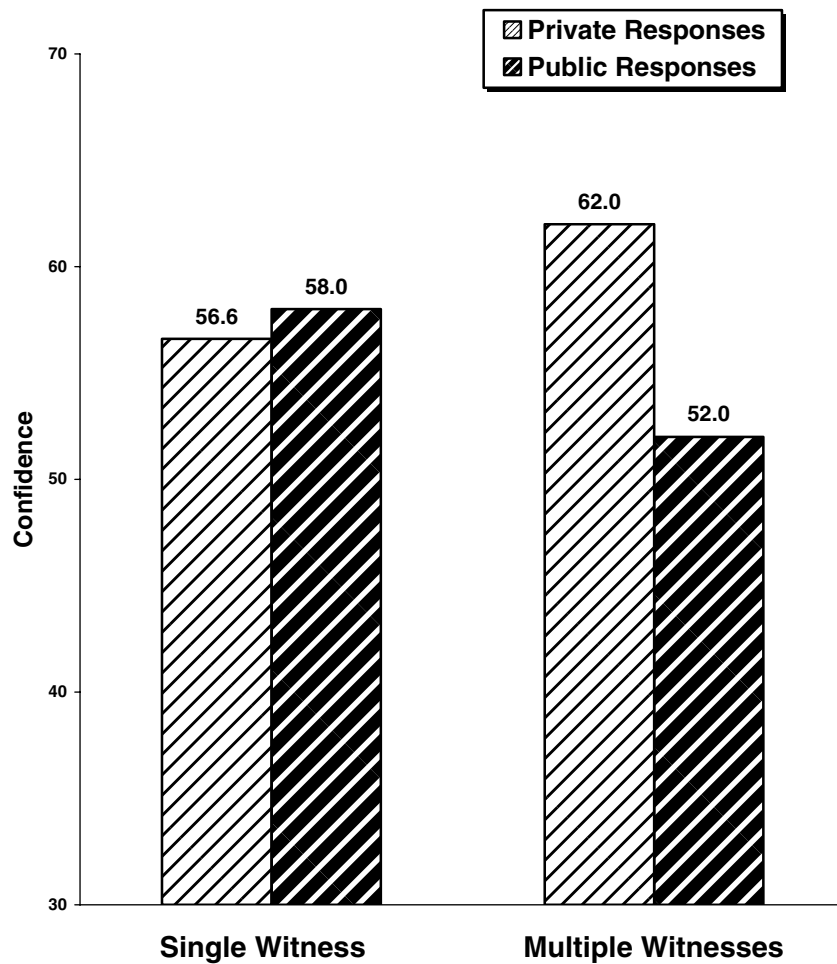
As expected, there was a significant *Response Privacy* by *Number of Witnesses* interaction on witness confidence,  $F(1,103) = 8.89, p < .01, \eta^2 = 0.08$ . As is apparent in Fig. 2, for the multiple-witness condition, confidence was significantly lower in public ( $M = 52.0, SD = 17.4$ ) than in private ( $M = 62.0, SD = 19.5$ ),  $t(55) = 3.78, p < .001, d = .51$ , yet there was no significant difference between public ( $M = 58.0, SD = 13.9$ ) and private ( $M = 56.6, SD = 17.7$ ) confidence in the single-witness condition,  $t(48) = 0.52, ns, d = .08$ .

Consistent with the other three experiments, *Response Privacy* did not alter witness accuracy,  $F(1,103) = 0.58, ns, \eta^2 = 0.01$ . Also, there was no main effect of *Number of Witnesses* on witness accuracy,  $F(1,103) = 0.64, ns, \eta^2 = 0.01$ , nor was there a *Response Privacy* by *Number of Witnesses* interaction,  $F(1,103) = 1.88, ns, \eta^2 = 0.02$ .

In contrast to the first three experiments, there were moderately large, statistically significant, between-participant confidence-accuracy correlations for all four cells of the experimental design in Experiment 4. In the single-witness condition, the confidence-accuracy correlation was .40 for the private responses ( $p < .01$ ) and .49 for the public responses ( $p < .001$ ), and these correlations were not significantly different from each other, *Fisher's z* = 0.51. Similarly, for the multiple-witness condition the confidence-accuracy correlation was .38 for the private responses ( $p < .01$ ) and .50 for the public responses ( $p < .001$ ). Once again, these two correlations were not significantly different from each other, *Fisher's z* = 0.71. In addition, there was no significant difference between the overall confidence-accuracy correlations for the private ( $r = .39, p < .001$ ) and public ( $r = .50, p < .001$ ) responses, *Fisher's z* = 0.98.

Within-participant confidence-accuracy correlations were calculated for the participants who had at least one correct and one incorrect answer to the multiple-choice questions.<sup>7</sup> As was the case for the between-participant correlations, the within-participant correlations were moderately

<sup>7</sup>Of course, within-participant confidence-accuracy correlations could not be calculated for participants with accuracy scores of 100% or 0%, which reduced the number of participants for whom within-participant correlations could be calculated to 34 in the single-witness condition and 40 in the multiple-witness condition.



**Fig. 2** Mean confidence ratings as a function of *Response Privacy* and *Number of Witnesses* in Experiment 4.

large for the private ( $r = .52, p < .001$ ) and public ( $r = .53, p < .001$ ) responses in the single-witness condition as well as for the private ( $r = .61, p < .001$ ) and public ( $r = .56, p < .001$ ) responses in the multiple-witness condition. None of these four correlations were significantly different from one another (all  $t$ 's  $< 1.00$ ).

#### *Judges' Ratings*

As in the first three experiments, correlations were calculated between the mean ratings of the judges on the three evaluation questions and the mean response accuracy and mean confidence of the detectives in the public condition (see Table 3). For the single-witness condition, the pattern of correlations was similar to the pattern in the first three experiments. The correlations between the judges' evaluations of the detectives and the actual response accuracy of the detectives were generally small and not significant (although the correlation was statistically significant for the second evaluation question), whereas the correlations between the detectives' confidence ratings



**Table 3** Between-participant correlations of the mock judges' ratings with the response accuracy and confidence of the detectives for the public responses in the single-witness and multiple-witness conditions in Experiment 4

Detective Responses	Judges' Ratings		
	How accurate is this detective's memory?	How many questions did this detective answer correctly?	How good would this detective be in a courtroom?
<b>Single-witness Condition</b>			
Response accuracy	.25	.28*	.21
Confidence	.62***	.45***	.61***
<b>Multiple-witness Condition</b>			
Response accuracy	.51***	.57***	.54***
Confidence	.83***	.76***	.79***

\*denotes  $p < .05$ ; \*\*\*denotes  $p < .001$ .

and all three of the judges' evaluations were large and statistically significant, with  $r$ 's ranging from .45 to .62.

For the multiple-witness condition, though, a different pattern emerged. As in the single-witness condition, and similar to the other three experiments, there were large, statistically significant correlations between the judges' evaluations of the detectives and the detectives' public confidence ratings ( $r$ 's ranging from .76 to .83). In contrast to the first three experiments, though, there were also large, significant correlations between the judges' ratings of the detectives and the accuracy of the detectives' public responses, with all  $r$ 's greater than .50 (see Table 3).

## General Discussion

### Public Confidence Ratings

The results from these four experiments support the proposition that witnesses may alter their public confidence ratings when there are social pressures to use confidence as an impression-management tool. The overall pattern of results is intriguing—sometimes public confidence ratings were higher than private ratings, sometimes they were lower, and sometimes they were the same.

Public confidence was significantly *higher* than private confidence in Experiments 1 and 2. These two studies simulated a hybrid of single-witness and multiple-witness situations. Although there were several witnesses in each experimental session, which allowed the witnesses to “compete” against one another, each witness saw a unique source event and thus acted as a sole witness for that event, which meant that each witness's responses could not be contradicted by the other witnesses in the experimental session.

Public confidence was significantly *lower* than private confidence in the multiple-witness condition in Experiment 4. Unlike Experiments 1 and 2, Experiment 4 incorporated a pure multiple-witness situation in which all of the witnesses viewed the same source event and thus could potentially contradict each other's responses.

Public confidence was the *same* as private confidence in Experiment 3 and in the single-witness condition in Experiment 4. These two experiments simulated a pure single-witness situation in which there was only one witness. Thus, there was no possibility that the witness could be contradicted by the testimony of other witnesses, and there were no other witnesses against whom to “compete.”

From these results, we can draw several general conclusions about how and when the presence of other witnesses affected the participants' public confidence reports in these four experiments. First, public confidence differed from private confidence only when there were other witnesses in the same experimental session (Experiments 1, 2, and the multiple-witness condition in Experiment 4). When there were no other witnesses present (Experiment 3 and the single-witness condition in Experiment 4), public and private confidence were the same. This implies that witnesses are more likely to use confidence as an impression-management tool when there are other witnesses present than when they are the sole witness. Originally, we expected that public confidence would differ from private confidence even in the pure single-witness situations, because we believed that the sole witnesses would alter their public confidence in order to impress the mock jurors or judges. It appears, however, that confidence is a more potent impression-management tool when it can be used to differentiate a witness from other witnesses. In retrospect this makes sense, because confidence is a salient dimension on which witnesses can compare themselves to other witnesses. Thus, the motivating factor for the participants, at least in these experiments, was not simply to appear to be "good witnesses," but rather to differentiate themselves positively from the other witnesses in the experimental session. Such a finding is consistent with the well-established principle that people are often motivated to appear superior to other people (for a complete review of the social comparison literature, see Suls & Wheeler, 2000).

Although the presence of other witnesses motivated the participants to alter their public confidence ratings, the direction of the change was determined by a second factor—whether or not there was a possibility of being contradicted by the other witnesses. As we expected, when there was no chance that the participants' responses could be contradicted by the responses of the other witnesses in the multiple-witness situations (Experiments 1 and 2), the participants *raised* their confidence ratings in public, without the fear that they would be "caught" being highly confident about incorrect memory reports. In contrast, when there was a chance that the other witnesses might contradict them because they all answered the same questions about the same source event (the multiple-witness condition in Experiment 4), the participants *lowered* their public confidence ratings to avoid the possibility of being overconfident about incorrect responses. This "better safe than sorry" approach is consistent with previous studies involving similar multiple-witness settings (Shaw, Woythaler, & Zerr, 2001; Shaw, Zerr, & Woythaler, 2001).

The overall pattern of confidence ratings is consistent with research showing that the use of self-presentation strategies is highly dependent on the context of the situation. Schlenker (2003) summarized the dilemma inherent in public self-presentations:

"Public self-presentations, on the one hand, thereby offer possible opportunities to impress others, but on the other hand, they pose a risk of appearing egotistical or even being discredited if the audience knows of publicly available, contradictory information. These competing pressures explain why public performances sometimes produce more, sometimes less, and sometimes about the same levels of self-glorification as private responding." (p. 496)

Finally, it is important to note that all of these changes in witness confidence occurred independent of any changes in witness accuracy. In fact, there were no differences between public and private accuracy in any of the four experiments. Thus, the pressures to present themselves favorably affected witnesses' public confidence ratings but not their public accuracy, which has important implications for the criminal justice system, as we discuss below.

The large body of research concerning the effects of accountability on people's judgments (for a review, see Lerner & Tetlock, 1999) provides another perspective from which to interpret the pattern of public confidence ratings in these four experiments. Several previous studies have

demonstrated that people provide confidence ratings that are more accurately calibrated when they are told that they will have to explain their responses to others (Kassin et al., 1991; Siegel-Jacobs & Yates, 1996; Tetlock & Kim, 1987). More specifically, these studies have found that overconfidence is diminished when people are informed before they provide their confidence ratings that they will be held accountable for those ratings.

As described by Lerner and Tetlock (1999), experimental manipulations of accountability usually involve several “submanipulations.” The two that are most relevant here are the *mere presence* of others and the *potential for evaluation*. As already mentioned, these were the two key factors that determined whether, and in what direction, public confidence ratings differed from private ratings in the present study. That is, public confidence differed from private confidence only when there were other witnesses in the same experimental session (mere presence), and the direction of the change in confidence was determined by whether or not there was a possibility of being contradicted by the other witnesses (potential for evaluation). Thus, the present results add to the literature on how accountability can affect cognitive judgments such as confidence ratings.

#### Confidence-accuracy Correlations

Consistent with many previous studies (e.g., Bothwell, et al., 1987; Luus & Wells, 1994; Shaw, 1996; Shaw & McClure, 1996; Shaw, McClure, & Wilkens, 2001), all of the confidence-accuracy correlations in Experiments 1, 2 and 3 (whether calculated on a between- or within-participant basis) were very small, and none of them were significantly different than 0. In Experiment 4, though, the confidence-accuracy correlations were large and statistically significant in all conditions [ $r = .40$  (between) and  $r = .52$  (within) for the private responses and  $r = .49$  (between) and  $r = .53$  (within) for the public responses in the single-witness condition, and  $r = .38$  (between) and  $r = .61$  (within) for the private responses and  $r = .50$  (between) and  $r = .56$  (within) for the public responses in the multiple-witness condition]. These confidence-accuracy correlations are considerably larger than those usually reported in the literature (e.g., Bothwell et al., 1987, reported an average correlation of .25), but they are consistent with those studies that suggest that the confidence-accuracy correlation may not be as small as has generally been assumed (Lindsay, Read, & Sharma, 1998; Read, Lindsay, & Nicholls, 1998). For example, in their meta-analysis of eyewitness identifications from lineups, Sporer et al. (1995) found a weighted average confidence-accuracy correlation of .37 for participants who actually make a choice from a lineup.

One important difference between Experiment 4 and the other three experiments was the nature of the memory task. Whereas the first three experiments involved a simple yes-no face-recognition task, Experiment 4 involved much richer stimulus materials and four-alternative forced-choice questions. Because of these differences, one might expect that the variability of the accuracy and confidence data would be greater in Experiment 4 than in the other three experiments, and that is exactly what occurred. Table 4 depicts the between-participant confidence-accuracy correlations and the standard deviations for accuracy and confidence for the private and public conditions in all four experiments.<sup>8</sup> As is evident in the table, the standard deviations of both measures were consistently larger in Experiment 4 than in the other three experiments. Given that both the accuracy and confidence data were more widely dispersed in Experiment 4,

<sup>8</sup> Because the confidence-accuracy correlations and standard deviations were virtually identical for the single-witness and multiple-witness conditions in Experiment 4, the data for these two conditions have been combined in Table 4 for simplicity of presentation. Also for simplicity, only the data for the between-participant correlations are presented because the pattern was very similar for the between- and within-participant correlations.

**Table 4** Between-participant confidence-accuracy correlations and standard deviations for all four experiments

	Experiment 1	Experiment 2	Experiment 3	<b>Experiment 4</b>
<b>Private Condition</b>				
Confidence-accuracy Correlation	$r = -.04$	$r = .09$	$r = .14$	$r = .39^{***}$
Accuracy Standard Deviation	$SD = 15.8$	$SD = 16.9$	$SD = 18.0$	$SD = 22.8$
Confidence Standard Deviation	$SD = 15.2$	$SD = 14.2$	$SD = 12.4$	$SD = 18.8$
<b>Public Condition</b>				
Confidence-accuracy Correlation	$r = -.14$	$r = -.26$	$r = -.18$	$r = .50^{***}$
Accuracy Standard Deviation	$SD = 15.8$	$SD = 18.3$	$SD = 15.1$	$SD = 27.4$
Confidence Standard Deviation	$SD = 15.9$	$SD = 11.9$	$SD = 12.9$	$SD = 16.1$

\*\*\*denotes  $p < .001$ .

it is not surprising that the confidence-accuracy correlations were larger there than in the other three experiments, where the range of those variables was more restricted. As noted both by Lindsay et al. (1998) and Read et al. (1998), there is a greater likelihood of a large confidence-accuracy correlation when both measures have substantial variation. Thus, the present findings lend credence to Read and Lindsay's claim that the generally small confidence-accuracy correlations reported in the eyewitness literature may be due in part to the restricted range of the data in many of the eyewitness laboratory studies.

#### Mock Jurors' Evaluations of the Witnesses

In all four experiments, the mock jurors' (or mock judges, as they were called in Experiment 4) ratings of the witnesses on the three evaluation questions were highly correlated with the witnesses' public confidence ( $r$ 's ranging from .45 to .83, as displayed in Tables 1, 2, and 3). This finding is consistent with much previous research demonstrating that jurors weigh witness confidence heavily (Brewer & Burke, 2002; Cutler, Penrod, & Dexter, 1990; Cutler, Penrod, & Stuve, 1988; Lindsay, Wells, & O'Connor, 1989; Penrod & Cutler, 1995; Wells, 1993), but the magnitude of the correlations was unexpectedly high.

Given the paucity of other information available to the mock jurors in the present study (i.e., the "testimony" of the witnesses consisted only of a series of responses to questions and the associated confidence ratings, as well as some brief explanatory remarks in Experiment 3), the witnesses' confidence ratings were particularly salient to the mock jurors, which likely contributed to the large correlations involving witness confidence. Thus, we must be careful about generalizing this finding to real jurors and witnesses. Nonetheless, it is clear that the mock jurors placed a great deal of weight on the witnesses' confidence judgments when evaluating the accuracy of the witnesses' memories. Because the mock jurors' evaluations were completely unrelated to the *actual accuracy* of the witnesses' answers in the public condition in the first three experiments (Tables 1 and 2), those jurors' reliance on witness confidence is clearly problematic for the legal system.

In the multiple-witness condition in Experiment 4, though, the mock judges' evaluations were highly correlated with witness accuracy for all three evaluation questions and moderately correlated with witness accuracy for one evaluation question in the single-witness condition (see Table 3). This unexpected finding was likely due to the fact that witness confidence and witness accuracy were highly correlated in Experiment 4. It is not clear why the magnitude of the correlations between the judges' evaluations and witness accuracy were so much higher in the

multiple-witness condition than in the single-witness condition, and this finding merits further investigation in future research.

#### Potential Limitations of the Study

We designed the witnessing experiences in these four experiments to be as simple as possible so that we could isolate the effects of the response-privacy manipulation on the witnesses' confidence ratings. For eyewitnesses in the real world, many other factors complicate the public-private distinction. For example, public confidence ratings tend to be given in the presence of many strangers in formal settings such as courtrooms, whereas private ratings are often given in familiar, less formal settings. By reducing or eliminating many of these possible confounds, we enhanced the internal validity of our study.

The primary cost of using such a simple eyewitness paradigm was that the experiences of the witnesses in our study were dissimilar in many respects to the experiences of witnesses in real courtrooms. All four experiments were conducted in a laboratory setting, the stimulus materials were fairly simple, the retention interval was very brief, and the memory tasks were straightforward and uncomplicated. The witnesses did not have to face either direct or cross-examination, and the content of their "testimony" was limited to yes-no answers or responses to multiple-choice questions and simple confidence ratings. Also, the mock jurors had very little information on which to base their evaluations of the witnesses. Because of all of these limitations, we have to be careful about generalizing the present results to real courtrooms.

In addition, the procedures in the first two experiments were somewhat unusual because of the turn-taking aspect of the data collection. Although witnesses would never act as jurors in a real courtroom, we employed the turn-taking procedure in order to maximize the competitive nature of the witnessing experience for our participants in the first two experiments. Despite the unrealistic nature of this procedure, there are clearly times in real courtrooms when witnesses do compete in order to "out-do" one another, and we tried to capture such social pressures here.

One might think that the turn-taking procedures in the first two experiments created dependencies in the data among the various participants in each experimental session, thereby compromising the statistical analyses, but there is no evidence that such dependencies actually occurred. In fact, we took every precaution to ensure that the answers and confidence ratings from each participant were completely independent from the responses of the other participants. Of course, because of the completely anonymous and confidential nature of the responses in the private condition, all of the private responses were completely independent. Also, in Experiment 1 there was no risk of any data dependency in the public condition, because even in the public condition all of the participants recorded their choices and confidence ratings in their individual response booklets *before* any of the participants began sharing their responses. Thus, there was no chance that a participant would adjust her or his responses because of the responses of other participants. Furthermore, in both Experiments 1 and 2 there was no possibility that the participants' decisions (concerning whether the faces in the test slides were the ones they had seen before) could have been affected by the other participants' decisions, because each participant studied a unique set of faces during the study phase. Only for the public confidence ratings in Experiment 2 was there a chance that the participants' responses might have affected each other. In order to minimize this possibility, we repeatedly emphasized throughout the session that each participant was testifying about a unique set of faces. There was no talking allowed during the experimental session, except when the participants were testifying, and there was no direct interaction among the participants. To test for the possibility of order effects, one-way ANOVAs were conducted on the confidence ratings for the four witness positions in Experiments 1 and 2.

These analyses confirmed that there were no significant differences among the mean public confidence ratings for the four witness positions in Experiment 1,  $F(3, 44) = 0.27, ns$ , or in Experiment 2,  $F(3, 41) = 0.79, ns$ .

Of course, it is possible that the witnesses in the first two experiments may have altered some non-verbal characteristics of the way that they delivered their confidence ratings in the public condition after hearing other witnesses give their answers. For example, some witnesses might have altered their language inflection, body posture, or eye gaze to indicate their feelings about their own confidence ratings as they said them out loud. There was no mechanism for assessing such non-verbal cues in the present study, so we cannot speculate whether any changes in non-verbal behavior actually occurred, or if they did occur, what effect they might have had on the mock jurors. This is an issue that merits attention in future studies.

Another potential criticism of our study is that the experimental procedures created substantial demand characteristics that led to unmistakable social pressures to be seen as skilled and capable. But that was indeed our intention - to demonstrate that public confidence can be affected by just such social pressures. It is important to emphasize that at no point in any of the four experiments did we suggest, either directly or indirectly, that the witnesses should “adjust” their confidence ratings in the public condition to impress the mock jurors or judges.

We did not endeavor to create truly realistic eyewitness experiences; indeed, the pressures inherent in courtroom testimony are far more complex than we could possibly simulate in this simple laboratory study. But all three of our witnessing situations—(a) the *pure single-witness setting* (Experiment 3 and the single-witness condition in Experiment 4), (b) the *pure multiple-witness setting* (the multiple-witness condition in Experiment 4), and (c) the *hybrid single-witness/multiple-witness setting* (Experiments 1 and 2)—have common real-world analogues. Certainly we must be very careful about generalizing the present results to real witnesses in real courtrooms, but our study represents an intriguing first step. Future studies can take the next few steps by exploring the limits of our findings and addressing the generalizability of our results.

#### Concluding Comment

Although previous studies have shown that public confidence ratings can differ from those given privately, this is the first study to demonstrate that in certain situations, social pressures can lead to *inflated* public confidence without any corresponding changes in response accuracy. Thus, our findings add to the accumulating evidence that people may adjust their confidence ratings in public as an impression-management strategy, particularly when there are strong social pressures to impress others. The story is not a simple one, though, given that the context of the memory test may dictate whether people *inflate* or *deflate* their public confidence ratings.

#### Acknowledgement

This research was supported in part by faculty research funds provided to the first author by Lafayette College and by EXCEL scholar funding provided by Lafayette College to the second, third, and fourth authors.

#### References

- Aloise-Young, P. A. (1993). The development of self-presentation: Self-promotion in 6- to 10-year-old children. *Social Cognition, 11*, 201–222.
- Arkin, R. M. (1981). Self-presentation styles. In J. T. Tedeschi (Ed.), *Impression management theory and social psychological research* (pp. 311–333). New York: Academic Press.

- Bennett, M., & Yeeles, C. (1990). Children's understanding of the self-presentational strategies of ingratiation and self-promotion. *European Journal of Social Psychology*, 20, 455–461.
- Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, 72, 691–695.
- Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, 26, 353–364.
- Brewer, N., & Wells, G. L. (2006). The Confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12, 11–30.
- Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied*, 8, 46–58.
- Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness identification. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 177–221). New York: Guilford Press.
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin and Review*, 7, 26–48.
- Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology*, 42, 673–681.
- Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identification. *Law and Human Behavior*, 7, 337–349.
- Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior*, 14, 185–191.
- Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior*, 12, 41–55.
- Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior*, 4, 243–260.
- Deffenbacher, K. A. (1991). A maturing of research on the behaviour of eyewitnesses. *Applied Cognitive Psychology*, 5, 377–402.
- Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior*, 6, 15–30.
- Jones, E. E., & Pittman, T. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 1, pp. 231–262). Hillsdale, NJ: Erlbaum.
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 1304–1316.
- Kacmar, K. M., Carlson, D. S., & Bratton, V. K. (2004). Situational and dispositional factors as antecedents of ingratiation behaviors in organizational settings. *Journal of Vocational Behavior*, 65, 309–331.
- Kacmar, K. M., Delery, J. E., & Ferris, G. R. (1992). Differential effectiveness of applicant impression management tactics on employment interview decisions. *Journal of Applied Social Psychology*, 22, 1250–1272.
- Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy-confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology*, 61, 698–707.
- Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the general acceptance of eyewitness testimony research: A new survey of the experts. *American Psychologist*, 56, 405–416.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125, 255–275.
- Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science*, 9, 215–218.
- Lindsay, R. C. L., Wells, G. L., & O'Connor, F. (1989). Mock juror belief of accurate and inaccurate eyewitnesses: A replication. *Law and Human Behavior*, 13, 333–340.
- Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology*, 79, 714–723.
- Neil v. Biggers, 409 U.S. 188, 34 L.Ed.2d 401 (1972), cert. denied, 444 U.S. 909, 100 S.Ct. 221, 62 L.Ed.2d 144 (1979).
- Noon, E., & Hollin, C. R. (1987). Lay knowledge of eyewitness behaviour: A British survey. *Applied Cognitive Psychology*, 1, 143–153.
- Nezlek, J. B., & Leary, M. R. (2002). Individual differences in self-presentational motives in daily social interaction. *Personality and Social Psychology Bulletin*, 28, 211–223.
- Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, 1, 817–845.

- Read, J. D., Lindsay, D. S., & Nicholls, T. (1998). The relation between confidence and accuracy in eyewitness identification studies: Is the conclusion changing? In C. P. Thompson, D. J. Herrmann, J. D. Read, & D. Bruce (Eds.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Robinson, M. D., & Johnson, J. T. (1998). How not to enhance the confidence-accuracy relation: The detrimental effects of attention to the identification process. *Law and Human Behavior*, 22, 409–428.
- Schlenker, B. R. (2003). Self-presentation. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 492–518). New York: Guilford.
- Schlenker, B. R., & Pontari, B. A. (2000). The strategic control of information: Impression management and self-presentation in daily life. In A. Tesser, R. Felson, & J. Suls (Eds.), *Perspectives on self and identity* (pp. 199–232). Washington, DC: American Psychological Association.
- Schütz, A. (1997). Self-presentational tactics of talk-show guests: A comparison of politicians, experts, and entertainers. *Journal of Applied Social Psychology*, 27, 1941–1952.
- Shaw, J. S., III (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied*, 2, 126–146.
- Shaw, J. S., III, Garven, S., & Wood, J. M. (1997). Co-witness information can have immediate effects on eyewitness memory reports. *Law and Human Behavior*, 21, 503–523.
- Shaw, J. S., III, & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, 20, 629–653.
- Shaw, J. S., III, McClure, K. A. & Dykstra, J. A. (2007). Eyewitness confidence from the witnessed event through trial. In M. P. Toglia, J. D. Reed, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology: Volume 1, Memory for events* (pp. 371–397). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Shaw, J. S., III, McClure, K. A., & Wilkens, C. E. (2001). Recognition instructions and recognition practice can alter the confidence-response time relationship. *Journal of Applied Psychology*, 86, 93–103.
- Shaw, J. S., III, Woythaler, K. A., & Zerr, T. K. (2001). Courtroom pressures can alter eyewitness confidence. In R. Roesch, R. R. Corrado, & R. J. Dempster (Eds.), *Psychology in the courts: International advances in knowledge* (pp. 223 – 240).- Amsterdam: Harwood Academic.
- Shaw, J. S., III, Zerr, T. K., & Woythaler, K. A. (2001). Public eyewitness confidence ratings can differ from those held privately. *Law and Human Behavior*, 25, 141–154.
- Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, 1, 1–17.
- Slovenko, R. (1999). Testifying with confidence. *Journal of the American Academy of Psychiatry and the Law*, 27, 127–131.
- Sporer, S. L., 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, 315–327.
- Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology*, 80, 587–606.
- Suls, J., & Wheeler, L. (2000). *Handbook of social comparison: Theory and research*. New York: Kluwer Academic/Plenum Press.
- Terry, R. L., & Krantz, J. H. (1993). Dimensions of trait attributions associated with eyeglasses, men's facial hair, and women's hair length. *Journal of Applied Social Psychology*, 23, 1757–1769.
- Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, 52, 700–709.
- Wells, G. L., Ferguson, T. J., & Lindsay, R. C. L. (1981). The tractability of eyewitness confidence and its implications for triers of fact. *Journal of Applied Psychology*, 66, 688–696.
- Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, 48, 553–571.