# UC Irvine
## UC Irvine Previously Published Works

### Title
Interpreting eyewitness confidence: Numeric, verbal, and graded verbal scales

### Permalink
https://escholarship.org/uc/item/65t947rc

### Journal
Applied Cognitive Psychology, 38(1)

### ISSN
0888-4080

### Authors
Greenspan, Rachel Leigh
Loftus, Elizabeth F

### Publication Date
2024

### DOI
10.1002/acp.4151

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Received: 10 June 2023 | Revised: 27 September 2023 | Accepted: 31 October 2023

DOI: 10.1002/acp.4151

RESEARCH ARTICLE

WILEY

# Interpreting eyewitness confidence: Numeric, verbal, and graded verbal scales

Rachel Leigh Greenspan[1] | Elizabeth F. Loftus[2]

[1]Department of Criminal Justice and Legal Studies, University of Mississippi, University, Mississippi, USA

[2]Department of Psychological Science, University of California, Irvine, Irvine, California, USA

**Correspondence**
Rachel Leigh Greenspan, 313 Mayes Hall, University, MS 38677, USA.
Email: rlgreen1@olemiss.edu

**Funding information**
National Science Foundation, Grant/Award Number: DGE-1321846

## Abstract

In empirical research, eyewitnesses typically report their confidence numerically (e.g., "I'm 90% sure"). In contrast, in the field, lineup administrators typically ask witnesses to explain their confidence verbally, in the witness' own words (e.g., "I'm quite sure"). Across three studies, we explored how evaluators assess verbal confidence statements: both freely reported and reported using a graded verbal scale. Results showed wide variability in the interpretation of both kinds of confidence statements. Even when evaluating seemingly very strong statements of verbal confidence (e.g., "completely certain") participants did not necessarily translate these statements into the strongest levels of numeric confidence. Variability in the interpretation of verbal confidence was particularly pronounced for low confidence statements. Moreover, participants preferred to report their confidence numerically rather than verbally. These results indicate the importance of documenting confidence verbatim at the time of the lineup so that the meaning of the witness' original confidence statement is preserved.

**KEYWORDS**
eyewitness confidence, graded verbal scales, numeric confidence, verbal confidence

## 1 | INTRODUCTION

In everyday life, people commonly use verbal statements to express uncertainty. Someone might text a friend that they will *probably* be late for dinner or that it is *unlikely* to rain that day. In these everyday situations, there are few consequences to misunderstanding verbal uncertainty statements. However, this is not always the case. For instance, a public health official might advise that it is *improbable* a new illness will result in hospitalization. If people who hear this statement believe that they will not be hospitalized if they contract this illness, they might engage in fewer preventive behaviors than those who interpret this statement as indicating a greater amount of risk.

Within the criminal justice system, examples abound of both verbal and numeric expressions of uncertainty. Legal thresholds like probable cause or reasonable doubt are essentially verbal labels designating degrees of certainty. Forensic evidence examiners use a variety of numeric and verbal scales to explain the likelihood of evidence matches to jurors (Eldridge, 2019). One key area in which expressions of uncertainty help indicate the diagnostic value of evidence is in the domain of eyewitness identifications. The confidence with which an eyewitness makes their initial identification can, under unbiased circumstances, relate to accuracy (Wixted & Wells, 2017).

Several best practice reports recommend how police should document eyewitness confidence. A 1999 report by the U.S. Department of Justice recommended that investigators ask witnesses to state "in his/her own words, how certain he/she is of any identification" (Technical Working Group for Eyewitness Evidence, 1999, p. 19). The National Research Council (2014) echoed this recommendation for verbal confidence in their 2014 report. However, in 2020, the American Psychology-Law Society scientific review paper on eyewitness identifications advised confidence "be collected on a graded scale

Portions of this project were presented at the 2020 American Psychology-Law Society Conference in New Orleans, LA and in the first author's dissertation.

 © 2023 John Wiley & Sons, Ltd. 

using words (e.g., 'positive', 'probably', 'maybe')" (i.e., graded verbal scales) or numbers (Wells et al., 2020, p. 22).

In the field, officers predominantly ask witnesses to explain their confidence in their own words and often discourage numeric confidence statements (Police Executive Research Forum, 2013; Smalarz et al., 2021). The current studies explore how evaluators interpret verbal confidence statements. Do evaluators have a common understanding of what witnesses mean when they explain their confidence using verbal or graded verbal scales? For confidence to potentially serve as a cue to accuracy, evaluators must first understand the meaning conveyed in verbal confidence statements. Moreover, empirical data on evaluator understanding of verbal confidence is needed to help develop effective, accurate graded verbal confidence scales.

## 1.1 | Confidence in a social context

In considering questions about evaluating an eyewitness' confidence statement, it is important to note that a confidence statement is not produced in a vacuum. Rather, lineup procedures inherently occur in a social context and the interaction between a lineup administrator and a witness can influence confidence (Kovera & Evelo, 2021).

One framework for the social interaction inherent in a lineup is to consider the production of a confidence statement as a type of interpersonal communication. Communication frameworks for conveying uncertainty address "who communicates what, in what form, to whom and to what effect" (van der Bles et al., 2019, p. 3). This framework proposes that factors at each stage of the communication process (e.g., the person conveying the message, the way the message is conveyed, and the person receiving the message) impact the effect of the communicated message (i.e., the witness' confidence) on the receiver's cognition, emotions, and behavior.

Much of the existing research on interpreting eyewitness confidence focuses on the interaction between the communicator and receiver (i.e., witness and officer). For instance, we know that confirming feedback from a lineup administrator can inflate witness confidence (Steblay et al., 2014). Lineup administrators are also affected by the social context of a lineup procedure. The way an officer interprets a confidence statement can be affected by factors like the officer's pre-existing expectations or the way the witness justifies their identification decision (Dodson & Dobolyi, 2015; Grabman & Dodson, 2019). However, less research has focused on the form of the confidence statement itself. How confidence is expressed can impact both the witness' ability to convey their confidence and the way the receiver interprets that confidence statement.

Notably, communication of uncertainty in the eyewitness context differs in one important way from that previously studied in the literature. Much of the existing research on communications of uncertainty studies the expert (e.g., doctor, scientist, or intelligence analyst) as the communicator and a member of the general public as the receiver (Gurmankin et al., 2004; Ho et al., 2015). In the eyewitness context, the communicator is typically a member of the general public who has no experience with expressing verbal confidence and the receiver is the police officer who likely has experience with lineup confidence. This difference is especially important as evidence suggests that context affects the interpretation of verbal uncertainty statements (Brun & Teigen, 1988). This highlights the need to explore perceptions of verbal uncertainty statements in the eyewitness context particularly.

## 1.2 | Verbal and numeric expressions of eyewitness confidence

Even though officers primarily ask witnesses to report confidence in their own words, most laboratory research studying eyewitness confidence uses numeric scales (Arndorfer & Charman, 2022). One reason officers may ask for a verbal confidence statement is that prosecutors prefer freely reported verbal confidence. Anything less than 100% on a numeric scale (or less than the top option on a graded verbal scale) can easily suggest reasonable doubt (Newirth, 2021).

Moreover, verbal confidence can be problematic if observers do not accurately interpret a witness' freely reported confidence statement in the way the witness intends. When participants report their confidence in their own words and then translate this response into a number, results show that the verbal and numeric estimates are generally highly correlated. However, discrepancies emerge between how witnesses explain their confidence and how trained coders interpret these statements (Kenchel et al., 2021).

Recent computerized approaches represent a novel way of understanding the meaning of verbal confidence statements. Seale-Carlisle et al. (2021) used machine-learning techniques to analyze the content of verbal confidence statements for cues to accuracy. This approach revealed that verbal confidence was predictive of accuracy and that the content of verbal confidence statements contained additional diagnostic cues beyond information provided by numeric confidence. However, these confidence statements included both verbal statements of confidence as well as verbal statements of justification which officers currently do not typically collect in the field.

There is clearly a disconnect between the method of documenting confidence in lab studies and in the field. While some recent research has explored the relationship between verbal confidence and accuracy (e.g., Arndorfer & Charman, 2022), this relationship may not be useful to practitioners if observers cannot consistently interpret what witnesses mean when they explain their confidence in their own words. Research outside the eyewitness domain may provide a useful framework for understanding the relationship between verbal and numeric statements of uncertainty.

## 1.3 | Interpreting verbal uncertainty statements

One common way to study the relationship between verbal and numeric statements of uncertainty involves participants reading

verbal uncertainty statements and then translating these statements into a numeric estimate. Overall, studies have shown tremendous variability in participants' numeric translations of verbal uncertainty statements (Theil, 2002). While there is some between-group consistency (e.g., most people judge "very certain" as expressing high confidence), there is significant within-group variability in the numeric translation of these statements by individuals (Teigen & Brun, 2003). This is problematic given that participants generally underestimate the degree of variability in interpreting verbal confidence (Brun & Teigen, 1988).

The most common explanation for the variability in interpretations of verbal uncertainty is that words are inherently vague or ambiguous (Teigen & Brun, 2003). On the other hand, numeric confidence is precise and easily interpretable. If someone reports they are 90% certain, this is generally interpreted to mean they believe the event will happen 9 out of 10 times. Verbal uncertainty statements do not have this property. Indeed, interpretations of verbal uncertainty statements are more influenced by base rates (Wallsten et al., 1986) and framing effects (Kuhn, 1997). However, verbal confidence does contain features that allow it to express additional information beyond numeric confidence. In particular, only verbal uncertainty statements contain directional information (e.g., the likelihood of an event happening or not happening; Teigen & Brun, 2003).

## 1.4 | Modality preference for uncertainty statements

Given these differences between verbal and numeric uncertainty, understanding which method people prefer can have important downstream consequences (Juanchich & Sirota, 2020). Some researchers have found that people prefer expressing uncertainty using verbal phrases (Druzdzel, 1989). Several reasons for this preference have been proposed including that people simply understand words better than numbers, that words developed earlier evolutionarily than numbers, and that words are less precise than numbers (Budescu et al., 1988). This preference for words replicates in samples of experts (Dhami et al., 2015) and has been codified as the recommended communication method by leading scientific organizations (e.g., Intergovernmental Panel on Climate Change, 2013).

Although police most often use and thus seem to prefer verbal confidence (Police Executive Research Forum, 2013; Smalarz et al., 2021), it remains an open question about how witnesses themselves prefer to report their confidence. When asked about their confidence using an open-ended question, witnesses most often use words and not numbers suggesting a preference for verbal confidence (Dobolyi & Dodson, 2018). However, the prompt for this question begins by asking witnesses to explain their confidence "in your words" possibly biasing witnesses towards verbal confidence. On the other hand, when witnesses are asked directly about their preferences between reporting their confidence verbally or numerically, most witnesses prefer numeric confidence (Kenchel et al., 2021). Notably though, this confidence preference was assessed after participants completed a lineup task where they were randomly assigned to report their confidence either using words, numbers, or both words and numbers potentially priming their preferences. We know of no research that has explored preferences for graded verbal confidence.

Thus, preference for either verbal or numeric confidence may not be absolute. Windschitl and Wells (1996) proposed that verbal confidence corresponds more with intuitive thinking and numeric confidence with deliberate thinking. The variants of uncertainty account theorizes that uncertainty preference depends on the judgment task (Juanchich & Sirota, 2020). This account proposes that verbal statements will be preferred for epistemic and dispositional uncertainty whereas numbers will be preferred for distributional uncertainty (e.g., uncertainty due to "the stable physical properties of the world," like flipping a coin; p. 2331).

## 1.5 | The current studies

The current studies explored whether observers have a common understanding of what eyewitnesses mean when they explain their confidence in their own words or using a graded verbal scale. In Studies 1 and 2, participants translated common freely reported verbal confidence statements into a numeric percentage. In Study 3, we also tested numeric translations of confidence statements reported using a graded verbal scale. A secondary goal of this research was to explore preferences for confidence format. We explored this question from multiple perspectives (e.g., witness, juror). While the goal of this study was largely descriptive, we predicted, in line with past research, that results would show some evidence of consistency in comparing different confidence statements (e.g., most participants would rate phrases like "completely confident" as displaying higher confidence than phrases like "somewhat confident") but large within-item variability in the numeric translations translation of specific confidence statements (Teigen & Brun, 2003).

## 1.6 | Open practices statement

All sample sizes in this study were determined a priori by the first author. We did not conduct traditional power analyses to determine sample size as the main results from this study are descriptive and we do not conduct hypothesis testing. We report all data exclusions and all measures in these studies. The data for these studies are available from the corresponding author upon reasonable request. The analysis code for these studies is not publicly available. The materials for these studies are publicly available through the online supplementary materials. These studies were not formally preregistered.

## 2 | METHOD

### 2.1 | Study 1 Method

#### 2.1.1 | Participants

Participants ($N = 75$) were recruited on Amazon Mechanical Turk using the CloudResearch platform. Participants were eligible if they were 18 years of age or older, had never completed a survey by the first author before, had a HIT approval rate of 80% or above, and had completed at least 50 HITs. Three participants were removed due to duplicate location information leaving a final sample of 72 participants. Participants were mostly female (53%, $M_{age} = 35.7$, $SD_{age} = 11.9$). The majority identified as White/Caucasian (68%) with a minority identifying as Black/African-American (15%), Hispanic/Latino (7%), or Asian-Asian/American (6%). Full demographics for each study are available in Table S1.

#### 2.1.2 | Materials

Study 1 assessed comprehension of freely reported confidence statements. To create these statements, we transformed the graded verbal scale originally developed by Windschitl and Wells (1996) by taking each phrase from the scale and framing it to participants, in an ostensibly freely reported manner, as: "I am [*level of certainty*]." This graded verbal is commonly used in the eyewitness literature (e.g., Arndorfer & Charman, 2022; Weber et al., 2008) and has been pilot-tested such that participants, in general, rate phrases at the top of the scale as displaying higher confidence than phrases at the bottom of the scale.

We tested 11 freely reported confidence statements all framed: "I am [*level of certainty*]." We chose not to include elaborate confidence statements that contained justification information to avoid any featural justification effects (Dodson & Dobolyi, 2015). The confidence phrases tested here included: completely certain, extremely certain, quite certain, rather certain, somewhat certain, as certain as uncertain, somewhat uncertain, rather uncertain, quite uncertain, extremely uncertain, and completely uncertain (Windschitl & Wells, 1996). We refer to the top half of the scale as reflecting high confidence statements and the lower half of the scale as reflecting low confidence statements.

#### 2.1.3 | Procedure

After completing the informed consent, we told participants that the purpose of the study was to understand how people view eyewitness evidence. Participants then read a brief description of a typical identification procedure. We created a cover story that informed participants that we conducted a previous study in which witnesses made an identification from a lineup and gave their confidence in their identification decision. Current participants were tasked with reading two of the confidence statements given by these fictitious previous participants and deciding which statement indicated a greater level of confidence.

Participants evaluated nine pairs of confidence statements shown in a random order. The pairs contrasted each phrase on the scale with the option two scale points away (e.g., "completely certain" and "quite certain"). Participants read each pair of confidence statements and answered "which of these witness statements shows more certainty" using response options: [statement 1], [statement 2], and "the two statements show equal certainty."

After completing this paired comparison judgment task, participants then read instructions for a second task. In this task, they translated confidence statements supposedly given by the previous participants into a percentage from 0 to 100. Each of the 11 certainty statements was displayed on a separate page and in a random order. The study concluded with participants completing demographic questions.

These paired comparison and numeric translation tasks provide differing information about how participants evaluate verbal confidence statements. The paired comparison task provides relative judgments about the meaning of verbal confidence statements in context of other confidence statements and the numeric translation task provides an absolute judgment about the meaning of verbal confidence statements in isolation (Jaffe-Katz et al., 1989).

### 2.2 | Study 1 Results

#### 2.2.1 | Paired comparison judgments

For each pair, responses were coded as consistent, inconsistent, or equal. Consistent responses were those in which the participant selected the response that was higher on the scale (i.e., displayed more confidence) and inconsistent responses were those in which the participant selected the response lower on the scale (i.e., displayed less confidence). Equal judgments were those in which the participant responded that the two statements indicated equal confidence.

Table 1 displays the full results of the paired comparison judgment task. Participants correctly identified the higher confidence statement most of the time (67%). However, a sizable portion of participants believed that these two responses on average indicated witnesses displaying the same amount of confidence (15%) or believed that the statement lower on the scale displayed greater confidence (18%). Participants more often made consistent judgments with high confidence pairs (78%) than low confidence pairs (46%) suggesting particular challenges with the comprehension of low confidence statements. A paired sample proportion test confirmed that participants made more consistent judgments for high confidence pairs than low confidence pairs, $p = .013$.

#### 2.2.2 | Numeric translations

While the paired comparison judgments provide initial evidence about whether participants discriminate between verbal confidence

**TABLE 1** Paired comparison judgments—Study 1.

| Confidence pair | Consistent (%) | Inconsistent (%) | Equal (%) |
|---|---|---|---|
| Completely certain: quite certain | 74 | 19 | 7 |
| Extremely certain: rather certain | 83 | 10 | 7 |
| Quite certain: somewhat certain | 78 | 10 | 13 |
| Rather certain: as certain as uncertain | 78 | 14 | 8 |
| Somewhat certain: somewhat uncertain | 82 | 6 | 13 |
| As certain as uncertain: rather uncertain | 36 | 26 | 38 |
| Somewhat uncertain: quite uncertain[a] | – | – | – |
| Rather uncertain: extremely uncertain | 53 | 35 | 13 |
| Quite uncertain: completely uncertain | 50 | 25 | 25 |
| Top 4 average | 78 | 13 | 9 |
| Bottom 4 average | 46 | 29 | 25 |
| Total average | 67 | 18 | 15 |

*Note*: Numbers may not equal 100% due to rounding.
[a]Due to survey error, no data for this comparison was collected.

**TABLE 2** Numeric translation of verbal confidence—Study 1.

| Confidence phrase | *M* | Mdn | SD | *N* | % rated as highly confident |
|---|---|---|---|---|---|
| Completely certain | 93.4 | 100.0 | 15.8 | 68 | 91.2 |
| Extremely certain | 92.1 | 95.5 | 12.4 | 68 | 88.2 |
| Quite certain | 72.8 | 80.5 | 23.8 | 72 | 56.9 |
| Rather certain | 67.4 | 74.5 | 20.8 | 72 | 31.9 |
| Somewhat certain | 59.0 | 59.0 | 19.7 | 72 | 15.2 |
| As certain as uncertain | 45.3 | 50.0 | 20.4 | 72 | 5.6 |
| Somewhat uncertain | 32.0 | 33.0 | 15.8 | 71 | 0 |
| Rather uncertain | 33.2 | 26.0 | 23.1 | 72 | 5.6 |
| Quite uncertain | 31.3 | 20.0 | 28.7 | 72 | 9.7 |
| Extremely uncertain | 23.5 | 5.5 | 33.3 | 72 | 11.1 |
| Completely uncertain | 28.4 | 3.5 | 40.9 | 72 | 22.2 |
| Top 5 average | 75.7 | – | 15.9 | 72 | |
| Bottom 5 average | 29.8 | – | 23.1 | 72 | |
| Total average | 52.2 | – | 10.4 | 72 | |

statements, it does not reveal how evaluators interpret an individual confidence statement. After responding to the paired comparison judgments, participants translated each of the 11 confidence statements to a numeric percentage on a sliding scale from 0% to 100% (see Table 2). We excluded outlying responses greater than three standard deviations from the mean. The full results without removing outliers are reported for all studies in the supplemental materials.

Similar to the results of the paired comparison judgment task, results indicate that participants performed well at discriminating between different confidence statements. Median results show that participants did rank the verbal confidence phrases in the same order as the scale from Windschitl and Wells (1996). However, results also indicated great variability in participants' understanding of individual confidence statements (Figure 1) with standard deviations ranging from 12.4 to 40.9. Overall, non-overlapping confidence intervals show variability among low confidence statements (SD = 23.1, 95% CI [19.8, 27.6]) was larger than variability among high confidence statements (SD = 15.9, 95% CI [13.7, 19.0]).

Exploring extreme values further, results reveal that even when participants express statements that seemingly convey extremely high levels of confidence (i.e., "completely certain"), participants do not consistently translate this to extreme labels of numeric confidence. Only 60% of participants translated a witness stating they were completely certain to 100% numeric certainty and 15% of participants assigned a numeric label of less than 90%. Arguably, in the field, the key task for law enforcement is not to directly translate a witness' freely reported confidence statement into a number but rather to identify whether that witness is highly confident (and thus, under unbiased conditions, likely to be accurate; Wixted & Wells, 2017).

The numeric bins by which eyewitness researchers define low, medium, and high confidence vary across studies (Sauer et al., 2019;

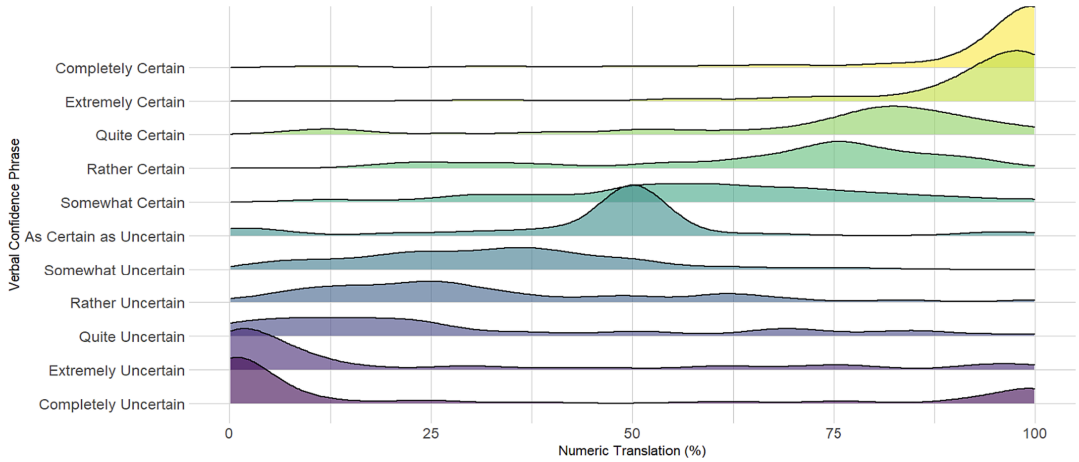


**FIGURE 1** Numeric translation of verbal confidence—Study 1.

Smalarz et al., 2021). We defined high confidence identifications as those assigned a numeric value of 80% or greater (Table 2). These results further highlight the fact that even when witnesses use verbal phrases that seemingly express very high confidence (e.g., "completely" or "extremely certain") a substantial minority of evaluators do not interpret these statements as highly confident. This indicates challenges with the use of verbal confidence as a potential cue to accuracy.

Potential concerns with the interpretation of extreme values of verbal confidence are more notable for phrases that seemingly express very low levels of confidence. Forty-three percent of participants translated the phrase completely uncertain to 0% numeric certainty whereas 14% translated this phrase to 100% certainty.[1] This could reflect participants misreading or misinterpreting the *uncertain* base word. However, it could also reflect participants noting the strong assertion made by the witness and using the 100% label to reflect this strong conviction, even if the assertion indicated strong uncertainty rather than strong certainty. Overall, these results match those of the paired comparison task indicating particular difficulties with comprehension of low confidence statements.

## 2.3 | Study 1 Discussion

Results show that participants largely, although not always, discern differences between high and low confidence statements. However, in the field, officers likely primarily focus on the strength of the identification of a single witness rather than comparing the confidence of multiple witnesses. For confidence statements to have value, the investigating officers (as well as other downstream criminal justice actors) must interpret that confidence statement consistently. If one officer interprets a confidence statement as reflecting 90% confidence and another as reflecting 60% confidence, this could have significant and lasting consequences on the investigation and prosecution of a case. The current results suggest that such a scenario is probable. Interpretations of freely reported verbal confidence statements varied widely between participants, particularly for statements expressing low confidence.

In Study 1, we tested confidence phrases on a bipolar, mirrored scale to capitalize on the unique aspect of verbal confidence containing directional information. Results indicated participants particularly struggled with low confidence phrases that contained the *uncertain* base word. This may be because comprehending degrees of uncertainty requires additional cognitive effort than evaluating degrees of certainty (Ferguson et al., 2013). Moreover, freely reported verbal confidence from real eyewitnesses primarily describe degrees of certainty rather than uncertainty (Behrman & Richards, 2005). Thus, for Study 2, we tested evaluations of freely reported verbal confidence on a unipolar scale containing statements only describing degrees of certainty. We were particularly interested in exploring whether the decreased comprehension of low confidence statements was an artifact of scale type or indicative of true difficulties in interpreting low confidence statements.

## 2.4 | Study 2 Method

### 2.4.1 | Participants

Participants ($N = 75$) were recruited in the same manner as in Study 1. Nine participants were removed due to duplicate IP addresses or location information leaving a final sample of 66 participants (see Table S1 for participant demographics).

### 2.4.2 | Materials

The confidence phrases tested in this study were derived from Windschitl and Wells (1996). We tested 10 statements all framed to participants as: "I am [*level of certainty*]." All statements described differences in degrees of certainty: completely certain, almost totally certain, extremely certain, very certain, quite certain, rather certain,

**TABLE 3** Paired confidence judgments—Study 2.

| Confidence pair | Consistent (%) | Inconsistent (%) | Equal (%) |
|---|---|---|---|
| Completely certain: extremely certain | 36 | 38 | 26 |
| Almost totally certain: very certain | 27 | 50 | 23 |
| Extremely certain: quite certain | 76 | 12 | 12 |
| Very certain: rather certain | 79 | 14 | 8 |
| Quite certain: fairly certain | 65 | 15 | 20 |
| Rather certain: somewhat certain | 56 | 18 | 26 |
| Fairly certain: slightly certain | 61 | 11 | 29 |
| Somewhat certain: not at all certain | 85 | 5 | 11 |
| Top 4 average | 55 | 29 | 17 |
| Bottom 4 average | 67 | 12 | 21 |
| Total average | 61 | 20 | 19 |

fairly certain, somewhat certain, slightly certain, and not at all certain. It is important to note all these phrases were selected from the high confidence half of the original scale developed by Windschitl and Wells (1996), whereas the phrases from Study 1 were representative of the full scale.

### 2.4.3 | Procedure

The procedure for Study 2 was identical to Study 1. After the informed consent, participants completed the paired comparison judgment task, the numeric translation task, and demographic questions.

## 2.5 | Study 2 Results

### 2.5.1 | Paired comparison judgments

The results from the unipolar scale tested here matched that of the bipolar scale tested in Study 1. On average, participants made a consistent judgment 61% of the time (Table 3). The unipolar scale did improve judgments on low confidence pairs (67% consistent judgments in Study 2 compared to 46% in Study 1). However, consistent judgments made for high confidence pairs in Study 2 (55%) decreased compared to Study 1 (78%). One reason for this could be that the true difference between the high confidence pairs in Study 2 was smaller than the distance between the pairs in Study 1. These results provide additional evidence that evaluators can largely distinguish between the certainty expressions of different witnesses.

### 2.5.2 | Numeric translations

We excluded outlying responses by removing any response greater than three standard deviations from the mean. Similar to Study 1, median results indicate that participants ranked the phrases in the same order as the original Windschitl and Wells (1996) scale. The numeric translations in Study 2 were also highly variable (see Table 4). However, even with the unipolar scale, significant variability remained in participants' numeric translations of individual confidence statements. Non-overlapping confidence intervals show variability was again larger for low confidence statements (SD = 17.6, 95% CI [15.0, 21.3]) than for high confidence statements (SD = 11.9, 95% CI [10.2, 14.4]).

Variability in interpretation remained for extreme values (see Figure 2). Fifty-three percent of participants translated the phrase "*I am completely certain*" to 100% numeric certainty and 71% translated this phrase to mean 90% confidence or greater. On the other hand, ratings of low confidence statements became more consistent. Twenty-eight percent of participants translated the phrase "*I am not at all certain*" to 0% numeric certainty and only 11% of participants translated the phrase to 75% certainty or greater. Thus, we found some evidence that the unipolar scale tested here seemed to reduce confusion compared to the bipolar scale tested in Study 1.

Like Study 1, there was significant variability in whether participants rated each confidence phrase as highly confident (i.e., numeric translation of 80% or greater). For instance, only 65% of participants identified a witness stating they were "almost totally certain" in their identification as being highly confident (Table 4). This provides further evidence that even when witnesses express seemingly very high confidence levels, this high level of confidence is not consistently interpreted by evaluators.

## 2.6 | Study 2 Discussion

The results of Study 2 largely replicate Study 1. While participants generally agree on the rank ordering of common confidence phrases, there is wide variability in the interpretation of verbal confidence statements by independent evaluators. Moreover, results indicate that freely reported verbal confidence statements struggle to convey extreme values of confidence. Even when a witness states they are "completely certain" in their identification, only 77% of evaluators rate this as a statement expressing high confidence.

**TABLE 4** Numeric translation of verbal confidence—Study 2.

| Scale point | *M* | Mdn | SD | *N* | % rated as highly confident |
|---|---|---|---|---|---|
| Completely certain | 89.4 | 100.0 | 16.0 | 65 | 76.9 |
| Almost totally certain | 81.9 | 85.0 | 14.7 | 65 | 64.6 |
| Extremely certain | 87.9 | 93.0 | 15.1 | 65 | 80.0 |
| Very certain | 81.0 | 81.5 | 11.8 | 64 | 59.3 |
| Quite certain | 71.2 | 76.0 | 15.1 | 65 | 35.4 |
| Rather certain | 63.9 | 69.0 | 18.8 | 66 | 25.8 |
| Fairly certain | 57.3 | 60.0 | 19.0 | 66 | 12.1 |
| Somewhat certain | 54.8 | 55.0 | 19.8 | 66 | 10.6 |
| Slightly certain | 44.4 | 46.0 | 21.3 | 66 | 4.5 |
| Not at all certain | 21.4 | 8.5 | 27.3 | 66 | 1.5 |
| Top 4 average | 85.0 | – | 11.9 | 65 | |
| Bottom 4 average | 44.5 | – | 17.6 | 66 | |
| Total average | 64.9 | – | 10.2 | 66 | |

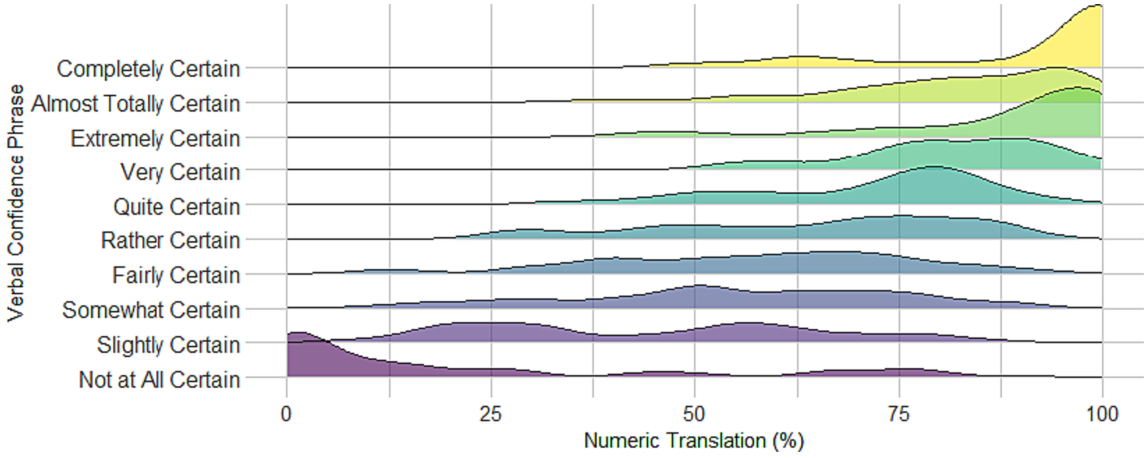


**FIGURE 2** Numeric translation of verbal confidence—Study 2.

Freely reported verbal confidence is not the only type of non-numeric confidence statement. Another form of non-numeric confidence is graded verbal confidence where participants select from a series of set scale options with verbal labels. The most recent American Psychology-Law Society scientific review paper advocated for graded verbal confidence (Wells et al., 2020). One reason for this recommendation could be to avoid confusion in the interpretation of freely reported verbal confidence. Yet, despite this recommendation, to date, no research has explored how observers interpret this alternate form of verbal confidence. Study 3 aims to answer this question.

## 2.7 | Study 3 Method

### 2.7.1 | Participants

Participants ($N = 149$) were recruited in the same manner as in the prior studies. Three participants were removed due to duplicate location information leaving a final sample of 146 participants (see Table S1 for participant demographics).

### 2.7.2 | Materials

Study 3 compared how participants interpreted both freely reported verbal confidence and confidence reported on a graded verbal scale. We created two graded verbal scales, a 5-point and a 3-point scale. The 5-point scale included the phrases: not at all confident, slightly confident, moderately confident, very confident, and extremely confident. The 3-point scale included the phrases: not at all confident, moderately confident, and extremely confident. The graded verbal scales displayed five (or three) blank response bubbles horizontally with the confidence phrase centered above each bubble (Figure S1). To compare perceptions of graded verbal and freely reported confidence, we also tested the five phrases from the 5-point

**TABLE 5** Numeric translation of verbal confidence phrases—Study 3.

| Scale point | *M* | Mdn | SD | *N* | % rated as highly confident |
|---|---|---|---|---|---|
| 5-Point graded verbal | | | | | |
| Extremely confident | 90.5 | 96.0 | 14.5 | 143 | 81.8 |
| Very confident | 77.5 | 77.0 | 11.9 | 145 | 42.8 |
| Moderately confident | 62.3 | 54.0 | 17.2 | 146 | 22.6 |
| Slightly confident | 45.1 | 30.0 | 27.2 | 146 | 21.2 |
| Not at all confident | 32.3 | 5.5 | 38.4 | 146 | 21.2 |
| Top 2 average | 83.7 | – | 11.5 | 145 | |
| Bottom 2 average | 38.7 | – | 31.6 | 146 | |
| 3-Point graded verbal | | | | | |
| Extremely confident | 90.4 | 95.0 | 12.8 | 144 | 82.6 |
| Moderately confident | 62.4 | 55.0 | 16.5 | 146 | 19.2 |
| Not at all confident | 30.9 | 9.5 | 37.1 | 146 | 17.1 |
| Freely reported | | | | | |
| Verbal confidence | | | | | |
| Extremely confident | 92.1 | 96.0 | 10.1 | 142 | 88.0 |
| Very confident | 86.3 | 88.0 | 10.6 | 144 | 79.2 |
| Moderately confident | 62.5 | 56.0 | 17.8 | 146 | 21.9 |
| Slightly confident | 48.9 | 44.0 | 27.2 | 146 | 19.2 |
| Not at all confident | 31.8 | 8.5 | 37.7 | 146 | 20.5 |
| Top 2 average | 88.9 | – | 9.3 | 144 | |
| Bottom 2 average | 40.4 | – | 30.7 | 146 | |

graded verbal scale presented as freely reported confidence statements (e.g., "I am extremely confident in my answer").

### 2.7.3 | Procedure

After completing the informed consent, participants read a short overview of how police conduct lineups and ask witnesses for a confidence statement. Participants then read a brief description and example of police asking witnesses for their confidence using words, numbers, and a graded verbal scale. Following this, participants were given the same cover story as the prior studies.

Participants first completed the numeric translation task for all eight graded verbal confidence statements provided by the fictitious prior participants (i.e., all five options on the 5-point scale and all three options on the 3-point scale). Statements were shown in a random order. Following this, participants completed the same task for the fictitious prior participants who ostensibly explained their confidence in their own words. They translated 11 freely reported confidence statements shown in a random order. Five statements used the same confidence phrases as the 5-point graded verbal scale but in free response form (e.g., "*I'm very confident in my choice*"). The remaining six were filler statements designed to obscure the focus on the five confidence phrases used in the graded verbal scale (e.g., "*Fairly confident but not positive*"; see Table S5).

A secondary goal of this study was to explore participants' preferences for reporting their confidence. Participants read a brief scenario and were asked to imagine they witnessed a robbery, viewed a lineup, and made an identification. They then reported how they would prefer to report their confidence: using numbers, words, or a graded verbal scale. Participants then explained the reasons for their choice using a free response box. In addition, participants responded to several follow-up questions about their confidence preference in different scenarios (e.g., if they were a juror at trial evaluating witness evidence). Finally, participants answered a series of demographic questions before completing the study.

## 2.8 | Results

### 2.8.1 | Numeric translations

Outliers more than three standard deviations from the mean were removed for this analysis. Participants' numeric translations of graded verbal confidence statements followed largely the same pattern as translations of freely reported confidence statements (Table 5). The average ratings matched the scale order with statements on the higher end of the scale rated as displaying higher confidence than statements on the lower end of the scale. For the five-point graded verbal scale, there was greater variability for low confidence (SD = 31.6, 95% CI

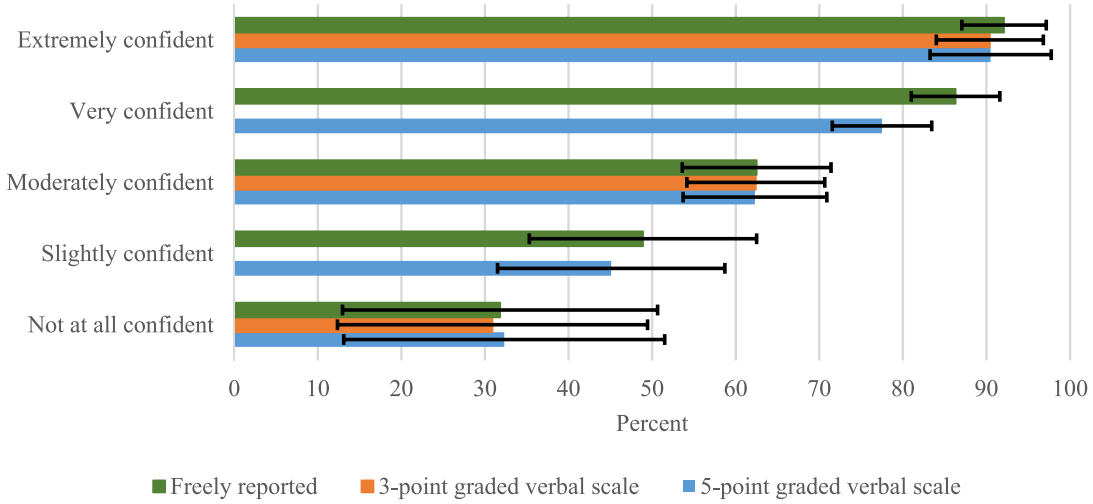


**FIGURE 3** Numeric translation of verbal confidence–Study 3. Errors bars represent one standard deviation.

[28.3, 35.7]) than for high confidence statements (SD = 11.5, 95% CI [10.3, 13.0]). Variability for low confidence was also greater for the lowest option (SD = 37.1, 95% CI [33.3, 41.9]) on the 3-point graded verbal scale than the highest option (SD = 12.8, 95% CI [11.5, 14.5]).

Visual inspection of Figure 3 reveals no notable patterns differentiating the 3-point and 5-point graded verbal scales. Paired sample *t*-tests confirm that participants did not differ in their translation of the same confidence phrases between the 3-point and 5-point graded verbal scales (Table S7). This coheres with previous work showing no differences in the confidence-accuracy relationship depending on the number of scale points (Dodson & Dobolyi, 2016; Tekin et al., 2018).

Next, we compared the numeric translation of confidence phrases on the 5-point graded verbal scale to the same phrases framed as freely reported confidence (see Table S8). Participants translated the phrases *extremely confident*, *moderately confident*, and *not at all confident* similarly if they were presented as a part of a graded verbal scale or as freely reported confidence statements. However, participants interpreted the phrase *very confident* as expressing a greater amount of certainty if it was presented as a freely reported confidence statement than as a part of the 5-point graded verbal scale, $t$ (142) = 8.02, $p < .001$, $d = 0.67$, 95% CI $d$ [0.49, 0.85]. This same pattern occurred for the phrase *slightly confident*, $t$ (145) = 2.45, $p = .015$, $d = 0.20$, 95% CI $d$ [0.04, 0.37]. It is unclear why this pattern emerged for some statements and not others. These results reveal some potential challenges with converting freely reported confidence statements to response options on graded verbal scales.

### 2.8.2 | Confidence preferences

When asked to imagine they were a witness to a robbery, most participants (69%) reported preferring numeric confidence to explain their certainty in their lineup identification, followed by freely reported verbal confidence (25%) and graded verbal confidence (6%). Preference for confidence format did not differ based on gender, race, education, political affiliation, or favorability towards the police (Table S9).

After reporting their preferred confidence method, participants explained the reasons for their choice using a free response box. The first author read these responses and identified common themes. Two independent coders then evaluated each response and identified whether each theme was mentioned in that response (all $\alpha$s > .74). Disagreements were resolved via discussion. More details about the coding scheme, categories, reliability, and sample responses can be found in the supplemental materials.

Table S10 displays the full results of the coding scheme. The most common explanation for preference for freely reported verbal confidence was that participants felt it would allow them to provide more information (22%). One participant wrote that "Because I can explain why I was or was not confident (some of them looked the same) or I can say I know for certain it was him if I did." The most common explanation for preference for numeric confidence was that it would allow the participant to be more precise (25%; e.g., "feels like I can be more precise in my answer. I prefer this one because it will give the police a good idea on how much to weigh my confidence."). The most common explanation for preference for graded verbal confidence was that it was easier for participants (44%, e.g., "I think it is easier to identify a rating that corresponds to my feelings.").

In addition to reporting how they would prefer to explain their confidence if they witnessed a robbery, participants also reported their confidence preference in three other scenarios. In the first two, participants chose how they would prefer to explain their confidence if they were highly confident they picked the person who robbed them and if they felt they did not pick the robber. In the third scenario, they reported what kind of confidence statement they would prefer if they were a juror evaluating witness testimony during a trial. Across all these scenarios, most participants preferred numeric confidence (see Figure 4).

In order to obtain some information about eyewitness confidence in the field, participants reported about their own experience as an

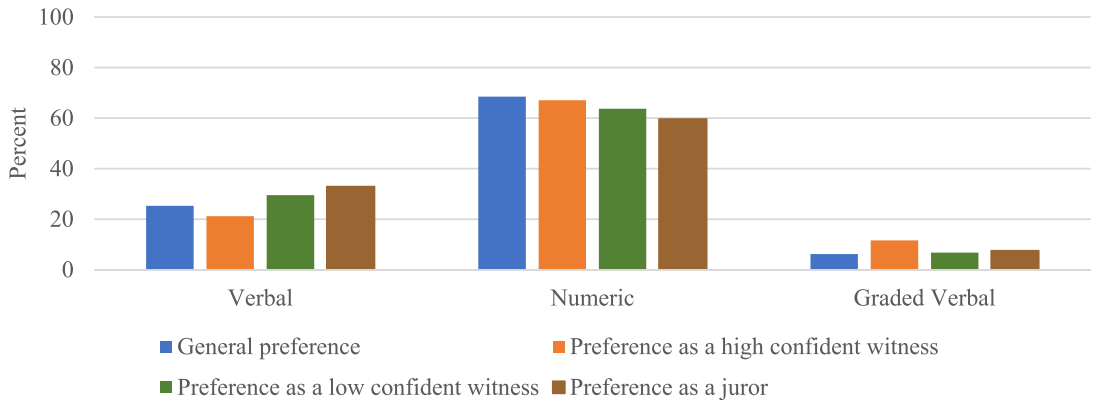


**FIGURE 4** Confidence reporting preferences.

eyewitness (Figure S2). Sixty-three percent of participants reported that they had once been an eyewitness to a crime. Two-thirds of these participants stated they were interviewed by police about their memory of the crime and made an identification from a lineup. Participants most commonly reported police asked for confidence using graded verbal ($N = 16$) or freely reported verbal confidence ($N = 12$). It is important to note that memory and self-report biases should be considered when evaluating these results. Participants might not be able to accurately recall the type of confidence question asked in a stressful situation or may be overreporting their witness experience given the study's explicit focus on eyewitness identifications. Nonetheless, these data provide at least some initial information about the type of confidence questions witnesses experience outside the lab.

## 3 | GENERAL DISCUSSION

Across three studies, we tested whether people have a common understanding of the meaning of verbal eyewitness confidence statements (see Table S11 and Figure S3 for combined results across studies). Results from the numeric translation tasks show wide variability in how different people interpret the same verbal confidence statement: both when confidence is freely reported and when it is reported on a graded verbal scale. The variability in interpretations of verbal confidence indicates that two people assessing a confidence statement from the same witness may come away with very different impressions of that witness' certainty.

There are distinct downstream consequences to this variability. Officers who view a witness as more certain in their identification may then lessen the investigation of alternate suspects or end an investigation altogether. Prosecutors who view an eyewitness as expressing low confidence may then view this evidence as less probative, potentially impacting charging decisions. Evaluation of eyewitness evidence by jurors is strongly driven by confidence (Slane & Dodson, 2022). The probative value of an identification itself may depend on the interpretation of verbal confidence.

Study 3 provides some of the first data exploring perceptions of graded verbal confidence in an eyewitness context. Results indicate that variability in the interpretation of confidence remains even when a graded verbal scale is used. When a witness indicates they are extremely confident, and this is clearly the strongest option they can select, significant variability remains in the interpretation of this confidence statement. Thus, if recommendations for graded verbal confidence are based on the assumption that this confidence form reduces ambiguity in evaluator interpretation, then our results challenge this notion.

One key question regarding the use of graded verbal confidence centers on the labeling and number of options for this scale. The current results provide some initial evidence regarding how participants interpret various verbal confidence phrases as well as which phrases have more or less variability in their interpretation. For instance, our results would suggest against including phrases that use an "uncertain" base word and highlight that some phrases, like "very certain," result in less ambiguity than others, like "quite certain." Results such as these can help form the foundation for the creation of evidence-based practices for the creation of graded verbal scales.

Questions about the design of graded verbal scales as well as the selection of confidence type (i.e., numeric, verbal, and graded verbal) can also be informed by existing literature on scale design in surveys. There are known response biases like edge aversion and advantage (i.e., preference or avoidance of first and last response options) depending on the presentation of options in a simultaneous display (Bar-Hillel, 2015). Specifically relevant to this study is the body of work on the use of slider scales in online surveys (Rivera-Garrido et al., 2022). Slider scales typically take more time and may require greater cognitive ability or effort than scales with set response options (Funke, 2016; Funke et al., 2011). Results from slider scales are also influenced based on the default position of the slider (Liu & Conrad, 2019). In the current studies, the slider default position was in the middle of the scale (i.e., at 50), potentially biasing the pattern of our results. Research has shown that witnesses' reports of their confidence are informed not only by the strength of their memory but also by the social interaction between the witness and the lineup administrator (Greenspan & Loftus, 2020; Kovera & Evelo, 2021). However, it is also important to consider how the framing and design of the confidence question itself can additionally influence reports of confidence.

Consistent across all three studies, results indicate difficulty in interpreting low confidence statements. Extant research shows that participants have greater difficulty processing uncertain events (Ferguson et al., 2013). In the case of confidence, uncertainty can also be thought of as a kind of linguistic negation and negation tends to slow down mental processing (Just & Carpenter, 1971; Wason & Jones, 1963). One explanation for this decreased processing speed is that mental representations of negation are more difficult than representations of affirmations (Khemlani et al., 2012). It is particularly important to consider the context of the current studies (Arroyo, 1982). Participants were informed that these witnesses made an affirmative identification from a lineup before giving their confidence statements. Thus, an expectation likely formed that the witness recognized the suspect from the video. Because of this, a low confidence statement would be a deviation from their pre-conceived expectations (Kaup et al., 2007). Future research is needed to explore perceptions of verbal confidence from witnesses who reject the lineup.

Most participants preferred reporting their confidence numerically over verbal and graded verbal scales. We note that participants reported their confidence preference after the numeric translation tasks. It is possible that participants' views on their confidence reporting were affected by completing the numeric translation task, possibly biasing participants towards preferring numeric confidence. The variants of uncertainty account theorizes that confidence preference depends on context (Juanchich & Sirota, 2020). Specifically, it predicts that people will prefer numeric confidence for distributional uncertainty. The current results thus suggest that eyewitness confidence can be considered a type of distributional uncertainty. In a lineup task, witnesses choose from a limited number of photographs and there is an objectively correct answer. This differentiates lineups from other tasks that generate epistemic uncertainty, such as estimating the likelihood a person has answered a trivia question correctly. Characterizing eyewitness confidence as a type of distributional uncertainty advances theoretical models of eyewitness confidence and links eyewitness confidence with research about distributional uncertainty in other domains.

## 3.1 | Limitations

The confidence phrases tested in the current studies were short, written statements. While many witnesses do indeed provide such statements after a lineup, others provide much more complicated, detailed statements. Past research has shown that when witnesses provide a justification for their confidence, this can impair interpretations (Dodson & Dobolyi, 2015). Non-verbal information including tone and body language when witnesses give an auditory, rather than written, confidence statement may also affect interpretations (Walker, 1977). We chose to use the verbal confidence statements from Windschitl and Wells (1996) as they are well-validated as well as commonly used in the eyewitness literature (e.g., Arndorfer & Charman, 2022; Weber et al., 2008). However, the phrases used here vary in how often witnesses use them in spontaneously describing their confidence. Phrases like "very confident" occur quite frequently compared to phrases like "rather confident" (Kenchel et al., 2021). Future research is needed to explore how evaluators interpret the complex, auditory verbal confidence statements spontaneously reported by real witnesses.

Participants in the current study were community members recruited online. Officers who interpret witness confidence statements may have years or decades of experience in this task which may optimize their performance. On the other hand, officers, particularly non-blind officers, may be motivated to interpret a confidence statement in line with their pre-existing expectations of a case (Kovera & Evelo, 2017). Thus, they may interpret greater confidence from a statement from a witness who has identified the suspect and interpret lesser confidence from a witness who makes a filler identification (Grabman & Dodson, 2019).

The current study focused on consensus in interpreting verbal confidence statements. However, another essential element to the study of verbal confidence statements is accuracy. That is, are evaluators able to accurately glean the level of confidence the witness intended to express? Even if evaluators consistently assigned the same numeric label to a verbal confidence statement, that consensus is only so useful if it differs from the value the witness would assign. Witnesses themselves do not always assign the same numeric label to similar verbal confidence statements (Kenchel et al., 2021). One witness who states they are "very certain" may mean 80% certainty and another witness who says they are "very certain" may mean 100% certainty. Thus, the variability in interpretations by evaluators may be accurate to the variability inherent in the intended meaning given by witnesses. The goal of this study was to provide an initial picture of how multiple evaluators assess the confidence statement of a single witness. Future research is needed to explore questions surrounding whether the level of confidence a witness intends to express is accurately detected by evaluators (Kenchel et al., 2021).

## 3.2 | Conclusion

The current studies show a lack of consensus in interpreting what witnesses mean when they explain their lineup confidence using verbal and graded verbal scales. This impairs the value of confidence as a potential investigative tool and raises questions about evidence-based guidelines for law enforcement for collecting eyewitness confidence statements. If choosing a confidence methodology is primarily driven by what makes witnesses themselves most comfortable, then our results indicate numeric confidence. On the other hand, if the goal of confidence is primarily to help obtain a conviction, then the vagueness in the interpretation of verbal and graded verbal confidence statements may be beneficial. Nevertheless, any recommendations to collect confidence using verbal or graded verbal scales must carefully consider how these confidence statements will then be interpreted by others. Indeed, there may be no one best method for how eyewitness confidence should be documented. Rather, different methodologies likely meet different goals for different criminal justice actors and at different stages of a case.

The results of the current studies do highlight one clear recommendation for lineup administration. Existing best practices advise confidence be documented immediately after an identification (Wells et al., 2020). Our results demonstrate that, regardless of whether lineup confidence is gathered using numbers, words, or a graded verbal scale, the confidence statement given by a witness should be documented *verbatim*, exactly as the witness expresses it at the time of the lineup. An officer who hears a witness state that they are "completely certain" and records this identification as "100% certainty" is quite possibly distorting the intended meaning of the witness when they explained their confidence. Given the variability in the interpretation of verbal statements shown in the current studies, it is essential that the confidence statement reported for the record is identical to what the witness states at the time of the lineup to reduce any distortions in this valuable piece of evidence. One way to ensure verbatim confidence is documented would be to video record the lineup procedure (Wells et al., 2020). Among other benefits, video recording an identification procedure ensures that both the exact language of the questions asked by the lineup administrator and the response of the witness are accurately documented.

## ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1321846. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interests to report.

## DATA AVAILABILITY STATEMENT

Data from these studies is available from the first author upon reasonable request.

## ORCID

*Rachel Leigh Greenspan* https://orcid.org/0000-0001-8395-945X

## ENDNOTE

[1] Even if we exclude the 14% of participants who translated the phrase *completely uncertain* to 100% numeric certainty, variability among low confidence statements (SD = 22.2, 95% CI [19.1, 26.6]) is still larger than variability among high confidence statements (SD = 15.9, 95% CI [13.7, 19.0]).

## REFERENCES

Arndorfer, A., & Charman, S. D. (2022). Assessing the effect of eyewitness identification confidence assessment method on the confidence-accuracy relationship. *Psychology, Public Policy, and Law.*, *28*, 414–432. https://doi.org/10.1037/law0000348

Arroyo, F. V. (1982). Negatives in context. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 118–126. https://doi.org/10.1016/S0022-5371(82)90500-X

Bar-Hillel, M. (2015). Position effects in choice from simultaneous displays: A conundrum solved. *Perspectives on Psychological Science*, *10*(4), 419–433. https://doi.org/10.1177/1745691615588092

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior*, *29*(3), 279–301. https://doi.org/10.1007/s10979-005-3617-y

Brun, W., & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, *41*(3), 390–404. https://doi.org/10.1016/0749-5978(88)90036-2

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 281–294. https://doi.org/10.1037//0096-1523.14.2.281

Dhami, M. K., Mandel, D. R., Mellers, B. A., & Tetlock, P. E. (2015). Improving intelligence analysis with decision science. *Perspectives on Psychological Science*, *10*(6), 753–757. https://doi.org/10.1177/1745691615598511

Dobolyi, D. G., & Dodson, C. S. (2018). Actual vs. perceived eyewitness accuracy and confidence and the featural justification effect. *Journal of Experimental Psychology: Applied*, *24*(4), 543–563. https://doi.org/10.1037/xap0000182

Dodson, C. S., & Dobolyi, D. G. (2015). Misinterpreting eyewitness expressions of confidence: The featural justification effect. *Law and Human Behavior*, *39*(3), 266–280. https://doi.org/10.1037/lhb0000120

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and eyewitness identifications: The cross-race effect, decision time and accuracy. *Applied Cognitive Psychology*, *30*(1), 113–125. https://doi.org/10.1002/acp.3178

Druzdzel, M. J. (1989). *Verbal uncertainty expressions: Literature review. Technical report CMU-EPP-1990-03-02*. Department of Engineering and Public Policy, Carnegie Mellon University.

Eldridge, H. (2019). Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy*, *1*, 24–34. https://doi.org/10.1016/j.fsisyn.2019.03.001

Ferguson, H. J., Tresh, M., & Leblond, J. (2013). Examining mental simulations of uncertain events. *Psychonomic Bulletin & Review*, *20*(2), 391–399. https://doi.org/10.3758/s13423-012-0348-1

Funke, F. (2016). A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*, *34*(2), 244–254. https://doi.org/10.1177/0894439315575477

Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*, *29*(2), 221–231. https://doi.org/10.1177/0894439310376896

Grabman, J. H., & Dodson, C. S. (2019). Prior knowledge influences interpretations of eyewitness confidence statements: 'The witness picked the suspect, they must be 100% sure'. *Psychology, Crime & Law*, *25*(1), 50–68. https://doi.org/10.1080/1068316X.2018.1497167

Greenspan, R. L., & Loftus, E. F. (2020). Eyewitness confidence malleability: Misinformation as post-identification feedback. *Law and Human Behavior*, *44*(3), 194–208. https://doi.org/10.1037/lhb0000369

Gurmankin, A. D., Baron, J., & Armstrong, K. (2004). The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. *Medical Decision Making*, *24*(3), 265–271. https://doi.org/10.1177/0272989X04265482

Ho, E. H., Budescu, D. V., Dhami, M. K., & Mandel, D. R. (2015). Improving the communication of uncertainty in climate science and intelligence analysis. *Behavioral Science & Policy*, *1*(2), 43–55. https://doi.org/10.1353/bsp.2015.0015

Intergovernmental Panel on Climate Change. (2013). *Climate change 2013: The physical science basis. Summary for policymakers*. www.climatechange2013.org

Jaffe-Katz, A., Budescu, D. V., & Wallsten, T. S. (1989). Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory & Cognition*, *17*(3), 249–264. https://doi.org/10.3758/BF03198463

Juanchich, M., & Sirota, M. (2020). Do people really prefer verbal probabilities? *Psychological Research*, *84*(8), 2325–2338. https://doi.org/10.1007/s00426-019-01207-0

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244–253. https://doi.org/10.1016/S0022-5371(71)80051-8

Kaup, B., Ludtke, J., & Zwaan, R. A. (2007). The experiential view of language comprehension. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 255–288). Lawrence Erlbaum Associates. https://www.researchgate.net/publication/237805089

Kenchel, J. M., Greenspan, R. L., Reisberg, D., & Dodson, C. S. (2021). "In your own words, how certain are you?" Post-identification feedback distorts verbal and numeric expressions of eyewitness confidence. *Applied Cognitive Psychology*, *35*(6), 1405–1417. https://doi.org/10.1002/acp.3870

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, *24*(5), 541–559. https://doi.org/10.1080/20445911.2012.660913

Kovera, M. B., & Evelo, A. J. (2017). The case for double-blind lineup administration. *Psychology, Public Policy, and Law*, *23*(4), 421–437. https://doi.org/10.1037/law0000139

Kovera, M. B., & Evelo, A. J. (2021). Eyewitness identification in its social context. *Journal of Applied Research in Memory and Cognition*, *10*(3), 313–327. https://doi.org/10.1016/j.jarmac.2021.04.003

Kuhn, K. M. (1997). Communicating uncertainty: Framing effects on responses to vague probabilities. *Organizational Behavior and Human Decision Processes*, *71*(1), 55–83. https://doi.org/10.1006/obhd.1997.2715

Liu, M., & Conrad, F. G. (2019). Where should I start? On default values for slider questions in web surveys. *Social Science Computer Review*, *37*(2), 248–269. https://doi.org/10.1177/0894439318755336

National Research Council. (2014). *Identifying the culprit: Assessing eyewitness identification*. The National Academies Press.

Newirth, K. A. (2021). The change to make change: A call for a moratorium on the admissibility of eyewitness identification evidence. *Journal of Applied Research in Memory and Cognition*, *10*(2), 225–228. https://doi.org/10.1016/j.jarmac.2021.05.004

Police Executive Research Forum. (2013). *A national survey of eyewitness identification procedures in law enforcement agencies*. https://www.ojp.gov/pdffiles1/nij/grants/242617.pdf

Rivera-Garrido, N., Ramos-Sosa, M. P., Accerenzi, M., & Brañas-Garza, P. (2022). Continuous and binary sets of responses differ in the field. *Scientific Reports*, *12*(1), 14376. https://doi.org/10.1038/s41598-022-17907-4

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law*, *25*(3), 147–165. https://doi.org/10.1037/law0000203

Seale-Carlisle, T. M., Grabman, J. H., & Dodson, C. S. (2021). The language of accurate and inaccurate eyewitnesses. *Journal of Experimental Psychology: General*, *151*, 1283–1305. https://doi.org/10.1037/xge0001152

Slane, C. R., & Dodson, C. S. (2022). Eyewitness confidence and mock juror decisions of guilt: A meta-analytic review. *Law and Human Behavior*, *46*(1), 45–66. https://doi.org/10.1037/lhb0000481

Smalarz, L., Yang, Y., & Wells, G. L. (2021). Eyewitnesses' free-report verbal confidence statements are diagnostic of accuracy. *Law and Human Behavior*, *45*(2), 138–151. https://doi.org/10.1037/lhb0000444

Steblay, N. K., Wells, G. L., & Douglass, A. B. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law*, *20*(1), 1–18. https://doi.org/10.1037/law0000001

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. http://www.ojp.usdoj.gov.ny

Teigen, K. H., & Brun, W. (2003). Verbal expressions of uncertainty and probability. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgement and decision making* (pp. 125–145). John Wiley & Sons, Ltd.

Tekin, E., Lin, W., & Roediger, H. L. (2018). The relationship between confidence and accuracy with verbal and verbal + numeric confidence scales. *Cognitive Research: Principles and Implications*, *3*(1), 41. https://doi.org/10.1186/s41235-018-0134-3

Theil, M. (2002). The role of translations of verbal into numerical probability expressions in risk management: A meta-analysis. *Journal of Risk Research*, *5*(2), 177–186. https://doi.org/10.1080/13669870110038179

van der Bles, A. M., van der Linden, S., Freeman, A. L. J., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, *6*(5), 181870. https://doi.org/10.1098/rsos.181870

Walker, M. B. (1977). The relative importance of verbal and nonverbal cues in the expression of confidence. *Australian Journal of Psychology*, *29*(1), 45–57. https://doi.org/10.1080/00049537708258726

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, *25*(5), 571–587. https://doi.org/10.1016/0749-596X(86)90012-4

Wason, P. C., & Jones, S. (1963). Negatives: Denotation and connotation. *British Journal of Psychology*, *54*(4), 299–307. https://doi.org/10.1111/j.2044-8295.1963.tb00885.x

Weber, N., Brewer, N., & Margitich, S. (2008). The confidence-accuracy relation in eyewitness identification: Effects of verbal versus numeric confidence scales. In K. H. Kiefer (Ed.), *Applied psychology research trends* (pp. 103–118). Nova Publishers.

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology*, *2*(4), 343–364. https://doi.org/10.1037/1076-898X.2.4.343

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10–65. https://doi.org/10.1177/1529100616686966

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Greenspan, R. L., & Loftus, E. F. (2023). Interpreting eyewitness confidence: Numeric, verbal, and graded verbal scales. *Applied Cognitive Psychology*, 1–14. https://doi.org/10.1002/acp.4151