Journal of Experimental Social Psychology xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Reports

The benefits of knowing what you know (and what you don't): How calibration affects credibility

Elizabeth R. Tenney [a,*], Barbara A. Spellman [a], Robert J. MacCoun [b]

[a] Department of Psychology, University of Virginia, 102 Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22904-4400, USA
[b] Goldman School of Public Policy and UC Berkeley School of Law, University of California, Berkeley, USA

ARTICLE INFO

Article history:
Received 29 December 2007
Revised 31 March 2008
Available online xxxx

Keywords:
Calibration
Credibility
Confidence
Confidence–accuracy relation
Eyewitness testimony
Metacognition

ABSTRACT

People tend to believe, and take advice from, informants who are highly confident. However, people use more than a mere "confidence heuristic." We believe that confidence is influential because—in the absence of other information—people assume it is a valid cue to an informant's likelihood of being correct. However, when people get evidence about an informant's calibration (i.e., her confidence–accuracy relationship) they override reliance on confidence or accuracy alone. Two experiments in which participants choose between two opposing witnesses to a car accident show that neither confidence nor accuracy alone explains judgments of credibility; rather, whether a person is seen as credible ultimately depends on whether the person demonstrates good calibration. Credibility depends on whether sources were justified in believing what they believed.

Introduction

People who are extremely confident seem to have an advantage in life. They are believed more often than people who express low confidence (e.g., Penrod & Cutler, 1995), have more influence in groups or when giving advice (Van Swol & Sniezek, 2005; Zarnoth & Sniezek, 1997), and appear more knowledgeable (Price & Stone, 2004). Why does having confidence about one's own knowledge garner so much credibility?

There are some subtle psychological explanations for why people rely more heavily on those who exude confidence. People are motivated to reduce uncertainty (e.g., Loewenstein, 1994), and often outcomes predicted with extreme confidence do more to reduce uncertainty than outcomes predicted with moderate confidence (Keren & Teigen, 2001). In addition, people might prefer confident informants for social reasons: it is harder to justify to others a preference for ambiguity (Curley, Yates, & Abrams, 1986), and sometimes it eases interpersonal and group interactions to agree with someone who is sure of herself (Zarnoth & Sniezek, 1997).

All these factors are surely relevant, but there may be a more straightforward explanation. People can rarely independently verify what others tell them, and without that verification they cannot fully assess their sources' reliability. In the absence of any evidence for incompetence or duplicity, it seems functional to assume that sources have some insight into the quality of their own knowledge, so that a high-confidence assertion indicates better, more reliable information than a low-confidence assertion (cf., Bradfield & Wells, 2000; Price & Stone, 2004; Thomas & McFadyen, 1995; Yates, Price, Lee, & Ramirez, 1996; Zarnoth & Sniezek, 1997).

The importance of confidence and of accuracy

Of course, people do not rely solely on confidence to judge credibility. When possible, they also assess accuracy, which is why lawyers try hard during cross-examination to reveal errors in courtroom testimony and why meteorologists who are often wrong get fired. Some previous research has suggested that the effects of confidence and accuracy are additive: there is a main effect of confidence (more confidence/more credibility; Sporer, Penrod, Read, & Cutler, 1995) and a main effect of errors (more errors/less credibility; e.g., Berman & Cutler, 1996) but no interaction (e.g., Brewer & Burke, 2002). Other research has suggested that people initially equate confidence with accuracy, but then fail to modify it despite available cues that the confidence–accuracy relation is weak (Keren & Teigen, 2001; Price & Stone, 2004; Thomas & McFadyen, 1995).

The importance of calibration

Notwithstanding the many benefits of confidence, the notion that more confidence is always better does not hold up under scrutiny. Adjectives like "cocky," "bull-headed," "know-it-all,"—and

* Corresponding author.
E-mail address: tenney@virginia.edu (E.R. Tenney).

0022-1031/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2008.04.006

_indeed, "overconfident"—provide clues that folk psychology has a more nuanced, contingent view of confidence. In this paper, we propose a hypothesis about how confidence and accuracy influence credibility that extends previous research and captures that nuance (called for convenience the "presumption of calibration" hypothesis):

1. People initially presume, in the absence of relevant evidence, that informants are well calibrated (i.e., that informants are good judges of their own knowledge). Thus, high-confident informants should be most persuasive (and it looks as if a simple "confidence heuristic" is at work).
2. People will override that initial presumption when evidence that enables the assessment of the informant's calibration becomes available. Then high-confidence statements from well-calibrated informants should be most persuasive, regardless of the informants' overall confidence.

In other words, the "presumption of calibration"—which is assumed, not inferred from any evidence—gets dropped when actual evidence regarding calibration becomes available. People rarely have the opportunity to accurately assess the overall calibration of others—that usually takes many examples of confidence–accuracy pairings (Lichtenstein, Fischhoff, & Phillips, 1982). Instead, we claim that people will generalize from any useful evidence that calls into question the informant's calibration.[1] People do not trust others simply because they act confident (cf. Price & Stone, 2004; Thomas & McFadyen, 1995); rather, they trust that confident others have a basis for their confidence, unless proven otherwise.

Support for the presumption of calibration hypothesis

Our previous research supports the idea that people will use available calibration information to override an initial presumption that a more confident witness would be more accurate (Tenney, MacCoun, Spellman, & Hastie, 2007). Participants read depositions of two opposing witnesses to a car accident. The "confident witness" claimed to remember everything clearly, whereas the "unconfident witness" claimed to remember some details about the day better than others. The witnesses were both confident about their central claims of who was at fault in the accident. Participants initially were more likely to believe the confident witness and rated that witness as more credible. Later, however, they learned that each witness had made an error. As a consequence, the confident witness showed poor calibration (confident regardless of right or wrong, therefore overconfident), and the unconfident witness showed good calibration (confident when right, unconfident when wrong). With this additional information, there was a switch: more participants believed the central claims of the unconfident witness and participants rated the unconfident one as more credible than the confident one. One interpretation of these results is that participants used the calibration information and gave more credence to the confident statements from the witness who was a good judge of her own knowledge rather than the witness who was confident indiscriminately. However, an alternative explanation, discussed below, is that in the presence of errors, people prefer informants who are more cautious in their testimony.

Participants changed their minds about which witness they preferred after errors were revealed, demonstrating an interaction between confidence and accuracy that did not show up in earlier studies (e.g., Brewer & Burke, 2002). Why? In earlier studies, the witnesses' statements of confidence were about the overall testimony in general, not about specific pieces of information that could be shown to be correct or not. To be able to assess calibration, people must be able to evaluate whether the two variables (confidence and accuracy) vary together. Note that when evidence of calibration can be used, there may still be main effects of confidence and accuracy on perceived credibility; but calibration, which is a function of the confidence–accuracy interaction, provides a better overall explanation of many effects.

Present studies

The present studies demonstrate the explanatory power of the presumption of calibration hypothesis. We use a courtroom context because the confidence–accuracy link has important applications there. But the phenomena we examine presumably occur routinely when executives receive advice from consultants, when travelers receive directions from strangers, and when journalists interview sources.

Participants read competing depositions of confident and cautious witnesses. The witnesses make claims of varying confidence about specific events that are later proven to be true or false. Experiment 1 shows a three-way interaction: the relationship between accuracy and credibility is moderated by confidence and calibration. Experiment 2 shows that not all errors hurt a witness's credibility; rather what matters is whether the witness is justified in holding the mistaken belief. The presumption of calibration hypothesis predicts that calibration can be more important to judgments of credibility than either confidence or accuracy alone. (See Table 1 for the designs of experiments.)

Experiment 1

As described above, Tenney et al. (2007) showed that whereas before an error was revealed participants preferred a high-confidence witness to a low-confidence witness, after an error they preferred a well-calibrated low-confidence witness to a poorly calibrated high-confidence one. Thus, the error eliminates the benefit of high confidence. Although the explanation that people assessed and used calibration is plausible, there is an alternative: perhaps in the absence of errors people prefer high-confidence informants but in the presence of errors people prefer informants who are more modest, or cautious, in their claims overall. When errors are made, participants might be willing to rely on any witness who was not so clearly overconfident. Experiment 1 was designed to unconfound calibration from cautiousness explanations.

All participants read conflicting testimony from one high-confidence ("confident") witness and one lower-confidence (in the present studies called "cautious") witness. The confident witness makes two high-confidence assertions and is shown to be correct about one and wrong about the other (and, thus, is overconfident). The cautious witness makes one high-confidence and one low-confidence assertion and is correct about one and wrong about the other. The conditions vary in how confidence and accuracy for specific pieces of testimony are related for the cautious witness. In the "well-calibrated" condition, the cautious witness is correct about the high-confidence assertion and wrong about the low-confidence assertion (as in Tenney et al., 2007; gamma correlation of 1); in the

[1] We use "calibration" as the general term to refer to the confidence–accuracy relation. Technically, *calibration*, refers to an absolute measure of that relation. For example, for a perfectly calibrated person to be 80% confident that something had happened means that for the set of items stated with 80% confidence, 80% of those would have happened. Calibration allows for the assessment of over- and underconfidence. Another way to assess the confidence–accuracy relation is by measures of relative accuracy (or "resolution"), like the Goodman-Kruskal gamma correlation ("gamma;" see Nelson, 1984), which ranges from –1 to 1. Used here, the gamma correlation provides a measure of people's ability to detect which items are more likely to be accurately remembered. We believe that people might use either the absolute or relative indicia to assess an informant's calibration, depending on what information is available.

Please cite this article in press as: Tenney, E. R. et al., The benefits of knowing what you know (and what you don't): ..., *Journal of Experimental Social Psychology* (2008), doi:10.1016/j.jesp.2008.04.006

Table 1
Experimental conditions

	Description of witness	Time 1 (before error)	Time 2 (after error)	Time 3 (after justification)
Experiment 1	Confident	Confident re: A	Correct	—
		Confident re: B	Incorrect	—
	Cautious well-calibrated	Confident re: C	Correct	—
		Unconfident re: D	Incorrect	—
	Cautious poorly-calibrated	Confident re: C	Incorrect	—
		Unconfident re: D	Correct	—
Experiment 2	Confident	Confident re: A	Correct	—
		Confident re: B	Incorrect	Justified
	Cautious well-calibrated	Confident re: C	Correct	—
		Unconfident re: D	Incorrect	Justified

Note. A, B, C, and D refer to non-critical (i.e., collateral) statements made by the witnesses. The statements were counterbalanced.

"poorly-calibrated" condition, the cautious witness is correct about the low-confidence assertion and wrong about the high-confidence assertion (gamma correlation of −1). In each condition, participants rate the credibility of and choose between the confident witness and either a well- or poorly-calibrated cautious witness.

If cautiousness matters, the two conditions should yield the same results—that is, participants should switch from believing the confident to the cautious witness because only overall confidence in the face of errors is important. If calibration based on specific pieces of testimony matters, the two conditions should yield different results—that is, participants should switch when the cautious witness is well calibrated but not when she is not—demonstrating that the effect of making an error on judgments of witness credibility is moderated not by how confident the witness had been overall, but by how confident the witness had been in the error.

Method

Participants

Sixty-six participants responded to advertisements for a psychology and law study that were posted around a university campus and in a city newspaper (median age = 22, 47 women). Participants were paid $15 and completed an hour of other experiments after finishing this one.

Materials and design

Participants were instructed to act as jurors. First, all participants read the same conflicting depositions about a car accident (based on materials by Borckardt, Sprohge, & Nash, 2003; Tenney et al., 2007). One eyewitness (the "confident witness") was highly confident about everything—her memory of two prior activities she had done on the day of the accident (e.g., she was positive that she went to the post office and took her dog to the veterinarian) and how the accident occurred. The other eyewitness (the "cautious witness") was highly confident about one prior activity (e.g., she knew for a fact that she went to a meeting at work about remodeling) and how the accident occurred, but unconfident about a second prior activity (e.g., she was not sure whether she had lunch with a neighbor that day). The witnesses concluded that different vehicles were at fault in the accident. Participants then (Time 1) rated each witness's credibility on a scale from (1) *not credible* to (6) *credible* and made a binary decision in response to the question, "Assume that only one is correct. Which witness's deposition do you believe?" The order of witnesses and the activities and descriptions of the car accident were counterbalanced across participants. Participants received no information about the witnesses' sex or race.

Next, participants read that lawyers did further research, and records from reliable sources either corroborated or disconfirmed the collateral (i.e., peripheral to the case) aspects of the witnesses' depositions. In all conditions, participants learned that both witnesses (a) had been right about one activity they had claimed occurred on the day of the accident but (b) had been proven wrong about the second activity they had claimed to have done that day—records revealed that those activities had actually been done on different days.

Importantly, for some participants ($n = 31$), the cautious witness showed good calibration: she had been unconfident when wrong and confident when right. For other participants ($n = 35$), this witness showed poor calibration: she had been unconfident when right and confident when wrong. (In both versions, the confident witness also showed poor calibration—confident regardless of whether right or wrong.) With this new information, participants re-answered (Time 2) the same questions.

Therefore, we had a 2 (between-subject: cautious witness shows either good or poor calibration) × 2 (within-subject: confident versus cautious witness) × 2 (within-subject: Time 1 [before error] versus Time 2 [after error]) mixed factorial design.

Results and discussion

Consistent with the presumption of calibration hypothesis, participants used available calibration information to evaluate witnesses; they did not simply trust the witness who was either the most confident or the most cautious. Chi-squares were used to analyze the participants' choice of which witness to believe, and one-way repeated measures analysis of variance, with calibration as a between subjects variable, was used to compare mean credibility ratings. Alpha was set at .05. The pattern of results was not contingent on the sex of participant, the order the witnesses testified, or either of two descriptions of the events surrounding the car accident.

Believability

At Time 1, when the two conditions (well-calibrated and poorly-calibrated) were identical, more participants believed the confident than the cautious witness. (See Fig. 1.)

However, at Time 2, after the errors were revealed, most sided with the cautious witness when that witness was well calibrated, but not when that witness was poorly calibrated. The Time 1–Time 2 increase for the cautious witness was statistically significant when the witness was well calibrated, $X^2(1, N = 31) = 16.96$, $p < .001$, $\phi^2 = .54$, but not when the witness was poorly calibrated, $X^2(1, N = 35) = .54$, *ns*, $\phi^2 = .015$.

Credibility

The pattern of credibility judgments was the same as believability choices and differed across calibration conditions (three-way interaction: $F(1,64) = 8.65$, $p = .005$. See Fig. 2).

Please cite this article in press as: Tenney, E. R. et al., The benefits of knowing what you know (and what you don't): ..., *Journal of Experimental Social Psychology* (2008), doi:10.1016/j.jesp.2008.04.006

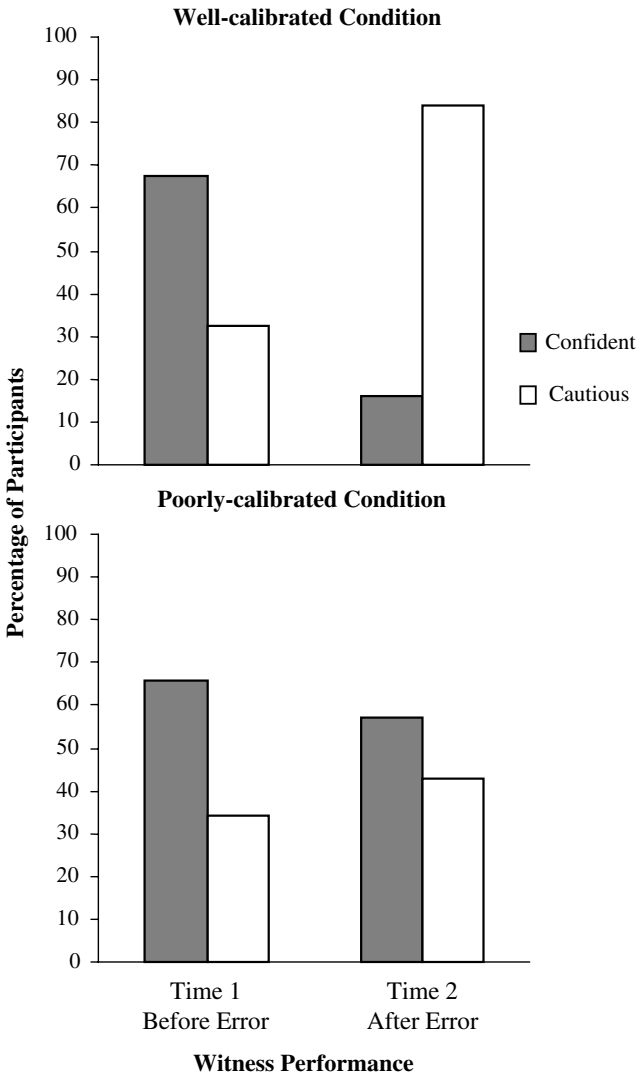

Fig. 1. Participants' choice of which witness to believe at Time 1 (before collateral errors) and Time 2 (after collateral errors) in Experiment 1. Top panel shows cautious well-calibrated witness; bottom panel shows cautious poorly-calibrated witness.

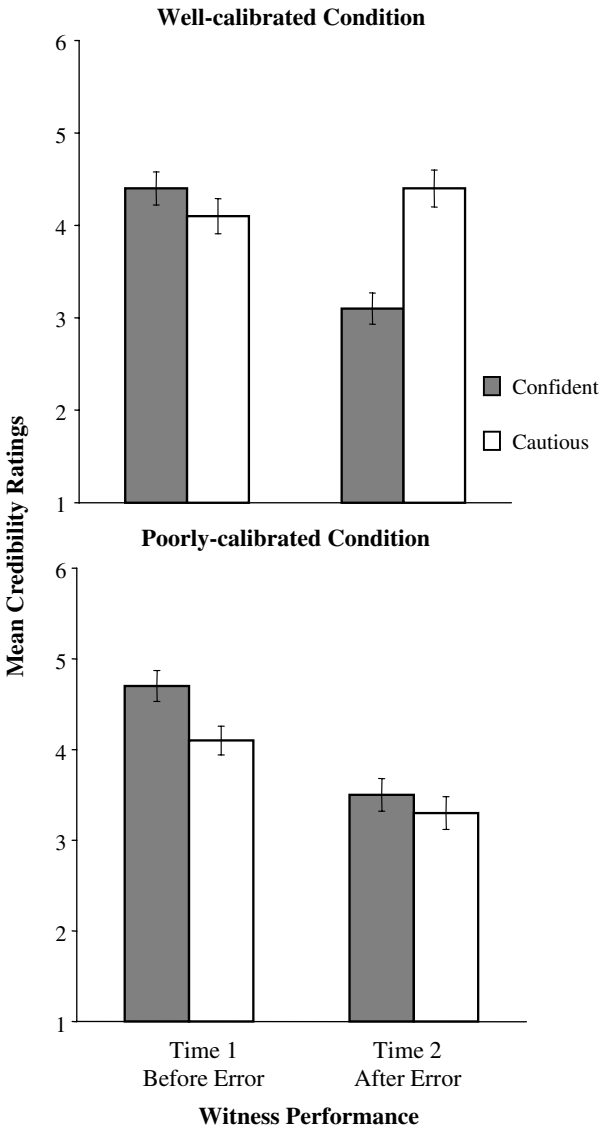

Fig. 2. Mean credibility ratings (1–6) with standard error bars at Time 1 (before collateral errors) and Time 2 (after collateral errors) in Experiment 1. Top panel shows cautious well-calibrated witness; bottom panel shows cautious poorly-calibrated witness.

At Time 1, consistent with previous literature, the confident witness was more credible than the cautious witness, $F(1,64) = 6.05$, $p = .02$, $d = .43$. There was no difference between conditions (i.e., no interaction, $F(1,64) = .73$, *ns*).

However, after learning about the error at Time 2, the pattern differed across conditions. When the cautious witness was well calibrated, she was now rated as significantly more credible than the confident witness. The top panel of Fig. 2 shows the significant Time1–Time2 interaction, $F(1,30) = 19.76$, $p < .001$, $d = .99$. On the other hand, when the cautious witness was poorly calibrated (bottom panel), the credibility of both witnesses dropped by about the same amount (i.e., the Time1–Time2 interaction was not significant, $F(1,34) = 2.34$, *ns*, $d = .36$).

Summary

The three-way interaction cannot be explained by confidence or accuracy alone, or even by a simple interaction that people prefer confident informants when no errors are made but cautious informants when errors are made. Participants cannot have been relying on "average" levels of confidence for the two cautious witnesses. Rather, participants must have been assessing calibration—how confident informants are in the specific pieces of (accurate or erroneous) testimony—in order to generate these results.

Experiment 2

Experiment 1 showed that a high-confidence error was detrimental to overall credibility (whereas a low-confidence error was not). But what if there is a good reason for a high-confidence error? A high-confidence error that is not evidence of poor calibration might not affect perceived credibility.

The Experiment 2 vignettes begin like the well-calibrated condition of Experiment 1: a confident witness (Time 1) is shown to be in error (Time 2) about a high-confidence assertion (the identification of a passenger), whereas a cautious witness is unconfident about the same error. Then, at Time 3, a justification for the error is given: the passenger had an identical twin.[2] The presumption of

[2] We recognize that the mistaken identification of an identical twin sounds like something from a television show; however, the goal of the manipulation was theory testing, not simulating an actual trial (Mook, 1983). We sought an experimental manipulation that would create a justifiable error without also changing the meaning or perceived severity of the offense itself, as might occur with more common realistic examples (e.g., a witness was misinformed, tricked, or lied to).

Please cite this article in press as: Tenney, E. R. et al., The benefits of knowing what you know (and what you don't): ..., *Journal of Experimental Social Psychology* (2008), doi:10.1016/j.jesp.2008.04.006

calibration hypothesis predicts that, as in Experiment 1, at Time 1, when there is no information to assess calibration, the majority of participants will initially side with the confident witness whereas at Time 2, after they learn about the errors, they will switch to side with the cautious witness. But what should happen at Time 3 when the error is justified?

An error that is justified is still technically an error in the absolute sense. If participants care about absolute errors, then they should continue to side with the cautious witness who had been unconfident in the error. However, if errors are important (at least in part) because of what they reveal about calibration, then learning that a high-confidence error is justified—and therefore not indicative of poor calibration—should restore the previously lost credibility of the confident witness. Thus, at Time 3, participants should switch back to side with the confident witness.

Method

Participants

One hundred five undergraduates at the University of Virginia (median age = 18; 81 women; 1 not reported) completed this 20-min experiment for course credit.

Materials and design

Participants were instructed to act as jurors, and they judged the credibility and character of two opposing eyewitnesses to a car accident three times. Time 1 and Time 2 were similar to Experiment 1, but with slightly different facts and additional dependent variables. Participants read that two witnesses had been asked to describe a car accident and identify a passenger in one of the vehicles. (The passenger was said to be suspected of shoplifting, but the shoplifting case would be tried at a later time.) As in Experiment 1, the structure of the two witnesses' depositions was similar, but they differed in how confident they were in collateral aspects of their testimony. The confident witness was highly confident in his memory of all aspects of his testimony—the accident itself, the weather on the day of the accident, and what the passenger looked like. He confidently identified a man sitting in the third row of the courtroom as the passenger. The cautious witness was confident about how the accident occurred and the weather, but not about his identification of the passenger. He acknowledged that he was not sure that he had gotten a good look at the passenger. He identified the same man in the third row, but stated that he was not confident in his identification. Unlike in Experiment 1, the sex of the witnesses was specified as male. At Time 1, participants rated the credibility, likeability, and honesty of each witness on three different scales from (1) *not credible/not likeable/not honest* to (6) *credible/likeable/honest*. Likeability and honesty ratings were included to make sure that our two witnesses started out about equal on those traits despite the difference in confidence (e.g., that people did not find the confident witness too arrogant or the cautious witness too meek). Participants also made a binary decision as to which witness they believed.

At Time 2, participants learned that the weather on the day of the accident was as both witnesses described, but that both witnesses had mis-identified the passenger (who had been away on business). Thus, as in Experiment 1, both witnesses were right about one fact and wrong about another. The confident witness was wrong about something with high confidence (thus poorly calibrated through overconfidence), and the cautious witness was right about something with high confidence and wrong about something with low confidence (thus well calibrated with a gamma correlation of 1). With this new information, participants re-rated each witness as at Time 1.

At Time 3, participants read that there was a good reason for the mistaken identification. After further questioning, the mis-identified passenger admitted that he had an identical twin who was friends with one of the drivers and did not have an alibi. Again, participants re-rated the credibility and character of each witness.

Finally, as a manipulation check, participants rated how confident they thought each witness was in general from (1) *not confident* to (6) *confident*.

Results and discussion

Results provide further evidence that people rely on calibration information when available rather than merely on statements of confidence or the presence of errors. Analyses were similar to Experiment 1 with one-way repeated measures analysis of variance used to compare mean ratings of honesty and liking (in addition to credibility). Three participants did not choose which witness they believed at Time 1 and/or Time 3, but their other judgments were retained in analyses. One of these participants made no Time 3 judgments.

Confidence manipulation check

The confidence manipulation worked. Participants rated the cautious witness ($M = 3.6$, $SD = 1.1$) as significantly less confident than the confident witness ($M = 5.6$, $SD = .73$), $F(1,103) = 210.1$, $p < .001$, $d = 2.2$.

Believability

Participants' decisions about which witness they believed changed over time according to the predicted pattern, $X^2(2, N = 105) = 82.54$, $p < .001$. (See Fig. 3.) At Time 1, more participants sided with the confident witness than the cautious witness. However, at Time 2, after the error was revealed, most participants sided with the cautious witness. This change replicates the well-calibrated condition in Experiment 1. At Time 3, after the error was justified, most participants sided with the confident witness once again.

Credibility

The pattern of results for credibility ratings was the same as for believability. Fig. 4, top panel, shows the significant interaction,

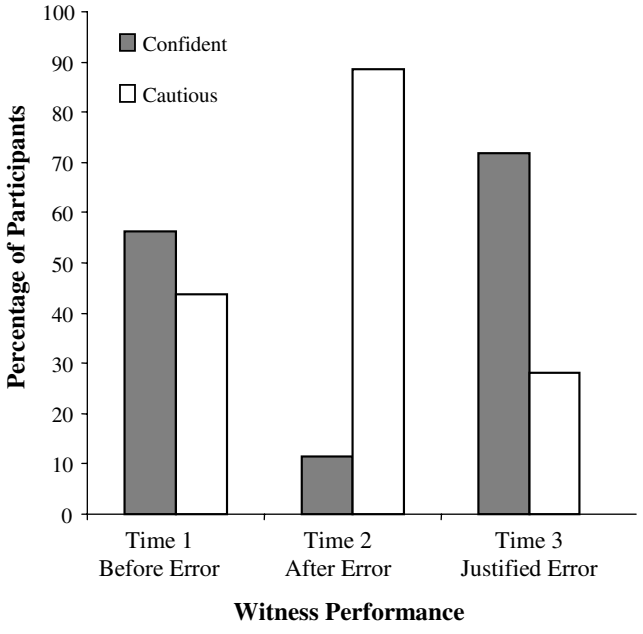

Fig. 3. Participants' choice of which witness to believe at Time 1 (before collateral errors), Time 2 (after collateral errors), and Time 3 (after collateral errors are justified) in Experiment 2.

indicating that the effect of making and justifying an error on ratings of witness credibility were moderated by witness confidence $F(2,102) = 94.4$, $p < .001$. At Time 1, the confident witness was rated as more credible, $F(1,104) = 3.81$, $p = .054$, $d = .21$, whereas at Time 2, the cautious witness was rated as more credible, $F(1,104) = 110.0$, $p < .001$, $d = 1.4$. Then, at Time 3, the confident witness was once again rated as more credible, $F(1,103) = 9.46$, $p = .003$, $d = .35$. This pattern is predicted by the presumption of calibration hypothesis: a justified error is not evidence of poor calibration; thus, at Time 3 participants will assign greater credibility to the confident witness—because they have no information to contradict the presumption that he is well calibrated.

Honesty and likeability

As shown in Fig. 4, honesty and likeability ratings show an overall similar pattern to each other and to credibility ratings. One difference is that at Time 1, credibility ratings are higher for the confident witness whereas honesty ratings are slightly higher for the cautious witness ($F(1,104) = 5.8$, $p = .02$, $d = .24$). Perhaps participants believe that someone who discloses the imperfectness of her memory is exhibiting honesty.

At Time 2, similar to credibility ratings, honesty and likeability ratings for the confident witness plummet whereas those for the cautious witness remain relatively unchanged (Time 1–Time 2 interaction for honesty, $F(1,104) = 102.52$, $p < .001$, $d = .1.4$; for likeability: $F(1,104) = 48.25$, $p < .001$, $d = .65$).

At Time 3, similar to credibility ratings, honesty and likeability ratings for the confident witness return to approximately baseline levels (although now the difference between the confident and cautious witnesses is not significant for honesty, $F(1,103) = .87$, *ns*, $d = .12$, but is for likeability, $F(1,103) = 4.89$, $p = .03$, $d = .22$).

Honesty and likeability ratings correlated with credibility ratings (mean r across times = .58 for honesty and .55 for likeability, $ps \leqslant .01$.).

Summary

The measures all show that a witness who makes an error can be forgiven for it if the error did not indicate that the witness was poorly calibrated. Even though at Time 2 most participants turned against the confident witness who made an error and deemed that witness less credible, honest, and likeable, when the error was explained at Time 3, most participants reconsidered, and the damaging effect of making an error almost disappeared. Thus, the results of Experiment 2 converge with those of Experiment 1 to demonstrate that an error might not be damaging to an informant's perceived credibility if it does not reveal that the person is poorly calibrated.

General discussion

Usually expressing confidence enhances a person's credibility and making a mistake damages it—but not always. Results from the current experiments help explain when and why these effects occur: confidence and errors may help or hinder an informant's credibility in their own right, but their effects will be qualified when, together, they affect perceptions of an informant's calibration.

Experiment 1 showed that the damaging effect of being proved wrong was moderated not by whether a witness had been confident overall, but by whether a witness had been confident *in the specific error* (Experiment 1). Experiment 2 showed that the damaging effect of being proved wrong can disappear if a witness's calibration is not called into question by that error. If it turns out that a witness erred about something she cannot be expected to know, her credibility might remain intact. Thus, calibration appeared to

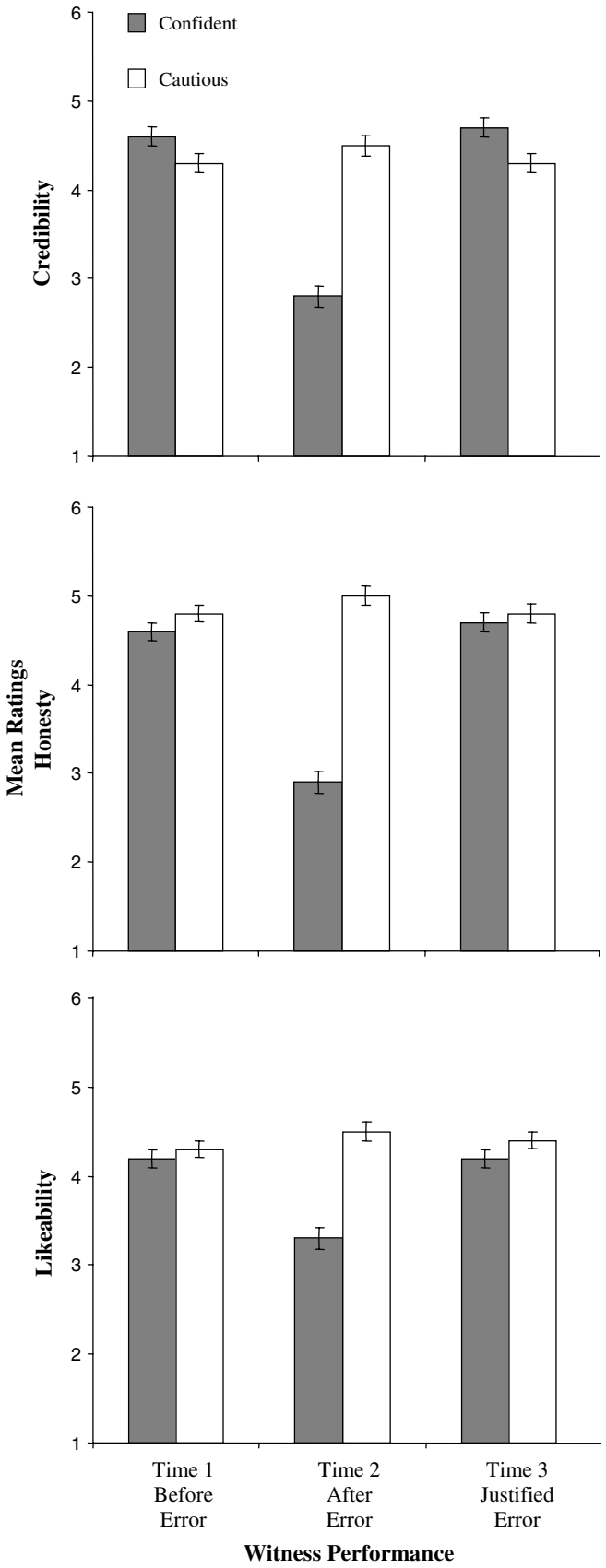

Fig. 4. Mean credibility, honesty, and likeability ratings (1–6) with standard error bars at Time 1 (before collateral errors), Time 2 (after collateral errors), and Time 3 (after collateral errors are justified) in Experiment 2.

matter more than overall confidence or technical errors: confidence was only beneficial when seen as appropriately linked to the likelihood of event outcomes, and errors were only damaging when they indicated that someone was an unreliable judge of her own knowledge.

One potential criticism of these experiments is that participants might be switching which witness they sided with from Time 1 to Time 2 because of demand characteristics. Perhaps participants felt compelled to change their responses because they thought that since the transcript presented new information, they were expected to change their initial inclinations. The significant crossover interaction found in Experiment 1 when the cautious witness was well calibrated but not when poorly calibrated supports our interpretation that whether a witness was calibrated, and not demand characteristics, is driving the effects.[3]

A second potential criticism is that all of the "action" is on the confident witness; perhaps jurors become dissatisfied with a (confident) witness who is poorly calibrated and side with the cautious witness not as a preference for good calibration, but because it is the only other option. As demonstrated in Experiment 1, however, after the errors were revealed the majority of participants sided with the cautious witness only when that witness was well calibrated—suggesting that participants were not simply siding against the confident, poorly-calibrated person indiscriminately. Additionally, although credibility judgments for cautious witnesses remained relatively steady in Experiment 2 and in the well-calibrated condition of Experiment 1, ratings of credibility dropped off for the cautious witness just as much as for the confident witness in the poorly-calibrated condition.

The results provide support for the presumption of calibration hypothesis, which posits that when there is a dearth of information, people presume that others are well calibrated. This presumption could account for why confidence appears to be important: often actual calibration information is not available and confident statements are given the benefit of the doubt. However, when people get new information about an informant's actual calibration, they will update their perceptions of the informant's credibility. People will then prefer the confident statements from a well-calibrated informant regardless of that informant's overall level of confidence or accuracy. Thus, informants might be influential not depending on confidence alone or even on whether they have made a mistake, but ultimately depending on how well calibrated they appear and whether others believe that their statements of confidence are warranted.

An important question that we have so far finessed is how strong the confidence–accuracy relation really is. Many researchers in the psychology and law domain conclude that the confidence–accuracy relationship is weak overall and worry about over reliance on it; however, there are known factors that affect the relationship. For example, when a lineup administrator knows which person is the suspect, when post-identification feedback is given before the witness states her confidence, or when the witness is repeatedly questioned, then confidence is artificially inflated thereby decreasing the confidence–accuracy relationship (see Sporer et al., 1995 and Wells, Memon, & Penrod, 2006 for reviews). Outside of eyewitness-related studies, researchers have determined that individual differences (e.g., tending to be over- or under-confident), the type of task (e.g., whether information about accuracy is obtainable), and the domain of expertise in question (e.g., people are more accurate judging whether they know science than judging whether they know business/law) determine whether confidence is diagnostic (Ackerman, Beier, & Bowen, 2002; Fischer & Budescu, 2005; Soll, 1996; Zarnoth & Sniezek, 1997). Regardless of whether people *should* use confidence as a proxy for accuracy, it is important to know if and when they *do* use it (see Yates et al., 1996, on the importance of studying the habits of "consumers" of probability judgments).

We have demonstrated that people who show that they are good judges of their own knowledge are rewarded, and those who are overconfident in what they know lose face. Although we did not test the effects of calibration of personality attributes, the presumption of calibration hypothesis could explain why self-enhancers, who outwardly exhibit confidence in themselves and their abilities over and above what is warranted, initially make great first impressions but later become less popular (Paulhus, 1998; Robins & Beer, 2001). People may initially be taken with someone who is highly self-confident if people assume high confidence has some basis in reality. However, when people receive concrete evidence that someone else's self-confidence is no indication of that person's ability or performance, that person may appear less credible, likeable, and honest than someone who is better calibrated. Perhaps people find out self-enhancers are poorly calibrated and alter their impressions accordingly. The benefits from exhibiting good calibration, and the damages from poor calibration, might be far reaching.

[3] In another test of demand characteristics, 38 additional participants were assigned to complete Experiment 2 materials but make no Time 1 judgments; thus, the demand for a given response in later judgments was reduced. As in Experiment 2, after the error was revealed, most participants sided with the cautious witness, but after the error was justified, most participants sided with the confident witness; this switch was statistically significant, X^2 (1, $N = 36$) = 39.46, $p < .001$, $\varnothing^2 = 1.0$. The replication suggests that demand characteristics were not driving the effects.

Acknowledgement

We thank Natasha Olinger, Daniel Kuckuck, Steve Yang, and Alexander West for their assistance in data collection. We thank Reid Hastie for thoughtful discussions, and Vikram Jaswal and Timothy Wilson for helpful comments on an earlier draft of the manuscript. This research was part of a master's thesis submitted by the first author under the supervision of the second author to the Department of Psychology at the University of Virginia. It was supported in part by the National Science Foundation under Grant 0437238 to the second author.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 33*, 587–605.

Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology, 81*, 170–177.

Borckardt, J. J., Sprohge, E., & Nash, M. (2003). Effects of the inclusion and refutation of peripheral details on eyewitness credibility. *Journal of Applied Social Psychology, 33*, 2187–2197.

Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior, 24*, 581–594.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26*, 353–364.

Curley, S. P., Yates, J. F., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes, 38*, 230–256.

Fischer, I., & Budescu, D. V. (2005). When do those who know more also know more about how much they know? The development of confidence and performance in categorical decision tasks. *Organizational Behavior and Human Decision Processes, 98*, 39–53.

Keren, G., & Teigen, K. H. (2001). Why is $p = 90$ better than $p = .70$? Preference for definitive predictions by lay consumers of probability judgments. *Psychonomic Bulletin and Review, 8*, 191–202.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty heuristics and biases* (pp. 306–334). New York, NY: Cambridge University Press.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*, 75–98.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38*, 379–387.

Nelson, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin, 95*, 109–133.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*, 1197–1208.

Penrod, S. D., & Cutler, B. L. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1*, 817–845.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17*, 39–57.

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology, 80*, 340–352.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*, 117–137.

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.

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*, 46–50.

Thomas, J. P., & McFadyen, R. G. (1995). The confidence heuristic: A game theoretic analysis. *Journal of Economic Psychology, 16*, 97–113.

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology, 44*, 443–461.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest, 7*, 45–75.

Yates, J. F., Price, P. C., Lee, J., & Ramirez, J. (1996). Good probabilistic forecasters: The 'consumer's' perspective. *International Journal of Forecasting, 12*, 41–56.

Zarnoth, P., & Sniezek, J. A. (1997). The social influence of confidence in group decision making. *Journal of Experimental Social Psychology, 33*, 345–366.

Please cite this article in press as: Tenney, E. R. et al., The benefits of knowing what you know (and what you don't): ..., *Journal of Experimental Social Psychology* (2008), doi:10.1016/j.jesp.2008.04.006