Notice:

The work from which this copy was made included a formal copyright notice. This work is protected by copyright law. Uses may be allowed with permission from the rightsholder, or if the use is "fair use" or within another exemption. The user of this work is responsible for determining lawful uses. (Title 17, US Code 1998.)
Perceived Expertise and Its Effect on Confidence

DAVID TRAFIMOW

Virginia Polytechnic Institute and State University

AND

JANET A. SNIJEZEK

University of Illinois at Urbana–Champaign

Two hypotheses about how people arrive at item confidence judgments were compared. The main difference is that one hypothesis assumes that general considerations are irrelevant; only the information made accessible by the items being judged affects confidence in answers given to those items. The other hypothesis assumes that general information such as one's perceived expertise is chronically accessible, even when one is assessing confidence for a specific item. Findings from two experiments contradicted the first hypothesis, but were consistent with the second. © 1994 Academic Press, Inc.

Several researchers have found that people tend to be overconfident when estimating the probability that their answers are correct (Lichtenstein, Fischhoff, & Phillips, 1982; Ronis & Yates, 1987; Sniezek, Paese, & Switzer, 1990; Yates, Zhu, Ronis, Wang, Shinotsuka, & Toda, 1989). For example, if a subject is given a set of multiple-choice questions, and is also asked to estimate, for each item, the probability that the selected alternative is correct, the average estimate will most likely be greater than the proportion of correct answers. We will call these estimates "item confidences."

The present article will address two issues vis-à-vis item confidences. First, there is the theoretical question about what variables affect item confidences. Second, there is the practical matter of making people's item confidences more accurately reflect their actual performance. In order to fulfill these objectives, we will identify a variable that reduces item confidences and show how people can become better calibrated.

The authors thank Tim Buckley for his insightful comments. We also recognize Sabine Trafimow for inventing the materials used to perform the manipulation in Experiment 2 and Alice Held for helping to process the subjects. Finally, our appreciation is extended to three anonymous reviewers who provide helpful criticisms of an earlier version of this article. Address reprint requests to David Trafimow, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0436.

0749-5978/94 $6.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.
Two Hypotheses about Confidence

Although there are various ways that people could compute item confidences, we will suggest two general possibilities. First, a line of reasoning initiated by Koriat, Lichtenstein, and Fischhoff (1980), elaborated on by others (Paese & Snieszek, 1991; Snieszek et al., 1990), and stated in its most recent form by Snieszek & Buckley (1991), distinguishes between “item” and “aggregate” confidence judgments. Item confidence judgments are assumed to be made on the basis of the information that is either mentioned in the item or is made accessible by what is mentioned in the item. But general considerations, such as how expert a person perceives himself or herself to be, are not used. This “item information processing” hypothesis implies that manipulating the information people have available about the item should affect item confidences, but manipulating people’s perceived expertise about the items in general should not affect item confidences. (Manipulating perceived expertise should affect aggregate confidence judgments. However, the focus of this article is restricted to item confidence judgments.)

A second possibility is that people’s general opinions about their expertise are accessible throughout the performance of the task.\(^1\) Thus, when arriving at a probability judgment, both item information and general information are accessible, and both kinds of information will be influential (Arkes, Christensen, Lai, & Blumer, 1987; Keren, 1987 made similar suggestions). Because the main difference between this hypothesis and the previous one has to do with the role of general opinions in influencing item confidences, we will call it the “general information processing” hypothesis. This hypothesis has received less empirical support than the other one. Although Glenberg and his colleagues (Glenberg & Epstein, 1987; Glenberg, Sanocki, Epstein, & Morris, 1987) reported some findings that indirectly support this hypothesis, Snieszek et al. (1990) found that item and aggregate confidences (which are presumably dependent on general opinions about task-relevant ability) are uncorrelated.

Decreasing Overconfidence

In general, attempts to decrease overconfidence, i.e., to induce people to become well-calibrated, have failed (Fischhoff, 1982; Lichtenstein et al., 1982; Sharp, Cutler, & Penrod, 1988). However, there have been a

\(^1\) Our use of the term “accessibility” is slightly different from the way it has been used by some previous researchers (e.g., Bargh, Bond, Lombardi, & Tota, 1986; Trafimow, Triandis, & Goto, 1991). These researchers used the term to refer to information that is accessible in “real life.” The present authors use the term to refer to information that is accessible as long as the task remains important (e.g., while the experiment is going on). However, if the task is important outside of the experimental setting, this hypothesis would assume that perceived expertise would remain accessible.
couple of attempts that were at least partially successful. Koriat et al. (1980) asked subjects to list reasons why their answers might be incorrect. This manipulation had a small, but significant, effect on subjects' overconfidence relative to subjects who were not asked to do this. However, an attempt by Fischhoff and MacGregor (1982) to replicate the effect achieved consistent, but only modest, improvements in calibration. Furthermore, a study of accuracy in predicting personal life events by Hoch (1985) found that the generation of "con" reasons had an effect only for events with relatively low base rates.

Finally, a recent attempt by Arkes et al. (1987) achieved some success in debiasing overconfidence. They presented different groups of subjects with items that subjects thought were either easy vs. difficult, but were actually of equal difficulty. Subjects were given veridical feedback, and then answered the rest of the items, while also rating their confidence for each item as they proceeded through the questionnaire. Arkes et al. hypothesized that subjects in the "easy" condition would "make a rather low assessment of their own level of knowledge" (p. 135). Consequently, their item confidences should have been, and were, lower than those in the other condition.

From our perspective, however, the Arkes et al. (1987) study has an undesirable feature. Except under circumstances in which an experimenter purposely writes deceptive questions, most questions that seem easy really are relatively easy. The importance of representative item selection in calibration research has been noted by May (1986). Thus, subjects in the Arkes et al. study, upon receiving feedback, learned that they were misled on the initial items and consequently may have assumed that all of the questions would be deceptive. Hence, they might have used anchoring and adjusting to add a negative number to all of their item confidences, thereby resulting in lowered confidence. Such debiasing of overconfidence is not terribly impressive if it merely reflects what Fischhoff (1982) calls mechanically reducing confidence without improving understanding (p. 435). The fact that subjects in the easy condition were actually underconfident on average provides further support for this alternative hypothesis; the initial items—which subjects were deceived to think easy—caused subjects to overadjust in a downward direction on subsequent items. Consequently, the Arkes et al. study does not allow us to distinguish between the hypotheses mentioned earlier, and it does not show us how to induce people to become well-calibrated in general.

EXPERIMENT 1

Experiment 1 was designed to (a) provide a straightforward test of the two hypotheses, and (b) demonstrate how people can be made to be well-calibrated. We used the Koriat et al. (1980) manipulation described earlier to manipulate the item information that was accessible to the sub-
jects. Consistent with Koriat et al., we assumed that telling subjects to write down why they might be wrong should increase the accessibility of information favoring decreased confidence, but telling them to write down why they might be correct should not have this effect. Second, we told subjects that people with a high school class rank similar to theirs were either good or bad at performing the task. According to the item information processing hypothesis, only the Koriat et al. "reasons" manipulation should affect item confidences. In contrast, the general information processing hypothesis predicts that manipulating perceived expertise should affect item confidences.

Method

Subjects. Sixty-one introductory psychology students served as subjects. Participation partially fulfilled a course requirement.

Procedure. Subjects were given the following instructions:

You will be asked to answer 50 multiple-choice questions that refer to general knowledge about the world. Studies have been performed in different places such as Italy, Germany, Canada, Colorado, New Mexico, California, South Carolina, New York, and Florida showing that class standing is highly related to performance on this test. That is, for any given item, people from the top (43% vs 3%) of their high school class were much more likely to be correct than people in the bottom (57% vs 97%). However, none of these studies were performed in the American Midwest. Consequently, this study attempts to replicate the above findings in the American Midwest.

Please write down approximately where you stood with respect to your high school classmates. I was in the top _______% of my high school class.

We are also interested in how people approach this task. After you answer each question by selecting one of the alternatives, please write down, in the booklet provided, one reason that (supports vs contradicts) your chosen answer. Please write the best reason you can think of that either speaks (for vs against) or provides evidence (for vs against) the alternative you have chosen, or speaks (against vs for) or points (against vs in favor of) the alternative that you have rejected.

We assumed that most undergraduates at the University of Illinois would be at least in the top 43% of their high school class (high perceived expertise), but that few would be in the top 3% of their high school class (low perceived expertise).²

² Actually 95% of undergraduates at the University of Illinois were in the top 43% of their high school class and 24% were in the top 3%. However, as will later be mentioned, only two of our subjects reported themselves as being in the top 3% and none of them reported being in the bottom 57%.
Subsequently, subjects were presented with 50 two-alternative multiple-choice questions (based on those used by Sniezek et al. 1990) on a computer screen. After typing in each of their answers, the computer asked for item confidences. Finally, subjects were debriefed and dismissed.

**Results**

An average item confidence score was computed for each subject. In addition, an average overconfidence score was determined for each subject by subtracting the proportion correct from the average item confidence score. Thus, separate analyses were performed for confidence, overconfidence, and proportion of correct responses.

These dependent variables were analyzed as $2 \times 2$ ANOVAs with Reason (write reason why correct vs reason why wrong) and Perceived Expertise (high vs low) as between-subjects factors. Two subjects were eliminated for reportedly being in the top 3% of their class.

**Confidence.** Consistent with Koriat et al. (1980), subjects who wrote a reason why they might be correct were more confident than subjects who wrote a reason why they might be wrong (means are .62 and .60, respectively). But, this difference was not significant, $F(1,57) = 1.45, p > .2$. However, there was a significant main effect for Perceived Expertise, $F(1,57) = 5.36, p < .025$. Subjects who were high in Perceived Expertise were more confident than subjects who were low in Perceived Expertise (means are .64 and .58, respectively). This main effect for Perceived Expertise is inconsistent with the item information processing hypothesis, but is consistent with the general information processing hypothesis. In addition, Sniezek et al. (1990) obtained norms for these items when no manipulations were performed, and the mean item confidence was .67. Item confidences of subjects high in Perceived Expertise were not discernibly different from Sniezek et al.'s control condition, $t(61) = 1.35, p > .1$. In contrast, item confidences of subjects low in Perceived Expertise were significantly different from the control condition, $t(56) = 3.42, p < .001$. Thus, the effect of the Perceived Expertise manipulation on item confidences is due mainly to the low Perceived Expertise condition. Finally, the Reason $\times$ Perceived Expertise interaction was not significant, $F(1,57) < 1$.

**Overconfidence.** These results parallel the confidence findings. There was not a significant main effect for Reason or a significant Reason $\times$ Perceived Expertise interaction, $F(1,57) < 1$ in both cases. However, there was a marginally significant main effect for Perceived Expertise, $F(1,57) = 3.26, p < .08$. Subjects in the low Perceived Expertise condition had appropriate levels of confidence on average ($M = -.0015$), but subjects in the high Perceived Expertise condition were on average overconfident ($M = .046$).
Consistent with the data on means, the proportion of overconfident subjects (i.e., subjects with positive overconfidence means) in the high Perceived Expertise condition (.73) was greater than that in the low Perceived Expertise condition (.54). Thus, subjects in the low Perceived Expertise condition were relatively well-calibrated regardless of whether group means or subject frequencies are considered.

Performance. The Reason manipulation had a marginally significant effect on actual performance (proportion correct), $F(1,57) = 3.48, p < .07$. Subjects who wrote a reason why they might be wrong actually performed less well than the other subjects (means are .57 and .61, respectively). In contrast, the Perceived Expertise manipulation had no effect and there was not a significant Reason × Perceived Expertise interaction, $F(1,57) < 1$ in both cases.

In sum, the findings obtained by manipulating Perceived Expertise contradict the item information processing hypothesis and support the general information processing hypothesis. In addition, subjects in the low Perceived Expertise condition were well-calibrated.

Discussion

An interesting aspect of the data concerns the lack of a main effect for the Reason manipulation on item confidences. There are at least two ways to account for it. First, as Fischhoff and MacGregor (1982) obtained a finding similar to ours, it is easy to argue that the effect is extremely weak or unreliable.

A second possibility is to assume that the Reason manipulation is generally reliable, but that something about our procedure eliminated the effect. Specifically, perhaps our subjects paid attention to accessible impressions of their expertise at the cost of attending to item information when making confidence judgments. This decrease in attention to item information, relative to that given in the Koriat et al. (1980) studies, eliminated the effectiveness of the manipulation. This argument implies that in situations where general impressions would not be expected to influence item confidences (e.g., when there is no general impression) the Reason manipulation should produce an effect, as was found by Koriat et al.

EXPERIMENT 2

Our main goal in Experiment 2 was to replicate the results of Experiment 1, but with a very different manipulation of Perceived Expertise. There are two reasons why we felt this was important. First, it is generally desirable to show that an effect can be reproduced. Second, it is possible that our results were only obtained because a general manipulation was used. One could argue that if subjects had been given a task-specific manipulation then they might not have formed general impressions of
their expertise. Instead, they might simply have encoded exemplars of each item they encountered into memory. Consequently, with no general impressions of their expertise available in memory, such impressions could not have influenced item confidences.

The Arkes et al. (1987) data would seem to contradict this reasoning and to support the general information processing hypothesis. That is, subjects were presented with specific, although misleading, items and yet item confidences on later items were affected. But, to continue the argument against the general information processing hypothesis, subjects in the Arkes et al. study were given immediate feedback which had implications for ability that may have stimulated the formation of a general impression. Suppose subjects had not been given feedback. It is not obvious that the general information processing hypothesis would have been supported. In sum, despite the Arkes et al. data and Experiment 1, the case for this hypothesis is not complete.

Method

Subjects. Forty introductory psychology students served as subjects. Participation partially fulfilled a class requirement.

Procedure. All subjects answered 50 two-alternative multiple-choice questions. Although these items were different from those used in Experiment 1, they were constructed with the intention of being similar and should have been of approximately equal difficulty. The subjects also indicated their item confidences for each question as in Experiment 1. However, Experiment 2 made use of a paper and pencil questionnaire rather than a computer. Theoretically, this should make no difference. However, one way of showing that an effect is robust is to demonstrate that it works across different experimental procedures. The only manipulation was that the first 10 questions were either extremely easy (high Perceived Expertise) or extremely difficult (low Perceived Expertise).

Results

As there were only two conditions, t tests were used to analyze confidence, overconfidence, and actual performance.

The First 10 Items. We wanted to make sure that the manipulation was successful. Statistical analyses showed that the easy items were actually easy and the difficult items were actually difficult (means for proportion are .99 and .43, respectively); the easy items were perceived as being easy and the difficult items as being difficult (mean confidences are .97 and .50, respectively), and subjects were underconfident for easy but overconfident for difficult items (means are -.02 and .07, respectively). These comparisons between easy and difficult items were significant (p < .03 for all comparisons).
TABLE 1
MEAN PROBABILITY ASSESSMENT (CONFIDENCE), OVER/UNDERCONFIDENCE, AND PROPORTION CORRECT OVER EXPERIMENTAL CONDITIONS FOR EXPERIMENTS 1 AND 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean probability assessment</th>
<th>Proportion correct</th>
<th>Over/under confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (from Snieszek et al., 1990)</td>
<td>.67***</td>
<td>.57</td>
<td>.10</td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reason: Correct</td>
<td>.62</td>
<td>.61*</td>
<td>.01</td>
</tr>
<tr>
<td>Reason: Wrong</td>
<td>.60</td>
<td>.57*</td>
<td>.03</td>
</tr>
<tr>
<td>Expertise: High</td>
<td>.64**</td>
<td>.59</td>
<td>.05***</td>
</tr>
<tr>
<td>Expertise: Low</td>
<td>.58***.c.***</td>
<td>.58</td>
<td>.004**</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy (all 40 items)</td>
<td>.64***</td>
<td>.46**</td>
<td>.18</td>
</tr>
<tr>
<td>Difficult (all 40 items)</td>
<td>.57***</td>
<td>.38**</td>
<td>.19</td>
</tr>
<tr>
<td>Easy (nonmisleading)</td>
<td>.66***</td>
<td>.57**</td>
<td>.09</td>
</tr>
<tr>
<td>Difficult (nonmisleading)</td>
<td>.58***</td>
<td>.45**</td>
<td>.13</td>
</tr>
<tr>
<td>Easy (majority correct)</td>
<td>.69***</td>
<td>.71**</td>
<td>.02</td>
</tr>
<tr>
<td>Difficult (majority correct)</td>
<td>.59**</td>
<td>.57**</td>
<td>.02</td>
</tr>
</tbody>
</table>

Note. Pairs of column means with the same superscript differ significantly.
* p < .10
** p < .05
*** p < .01
**** p < .001

The 40 common items. Table 1 contains a summary of these results and those that were obtained in Experiment 1. Consistent with the findings from Experiment 1, subjects in the easy condition were more confident than the other subjects on the 40 items that were the same for both groups (means are .64 and .57, respectively), t(38) = 3.20, p < .004. However, overconfidence was approximately the same for subjects in both conditions, t(38) = -1.16, p > .8. These findings led us to suspect that actual performance might have been affected. Consistent with our suspicions, subjects in the easy condition actually performed better than the other subjects (means are .46 and .38, respectively), t(38) = 2.31, p < .03. Thus, Perceived Expertise had the expected effect on confidence, but also had an unexpected effect on actual performance. Although these data support the general information processing hypothesis, the fact that actual performance was so low might lend a cynic to argue that the easy vs difficult questions manipulation might not have worked had the questions been easier. Certainly, the data indicate the possibility that there were some misleading items, which was something we wanted to avoid.

We addressed this potential problem in two ways. First, we arbitrarily designated all items where 30% or fewer of the subjects gave right answers as being misleading and performed analyses on only those items
that were not misleading (25 items). The results are similar to those previously obtained. Subjects high in Perceived Expertise were more confident than the other subjects (means are .66 and .58, respectively), \( t(38) = 3.20, p < .003 \). There was no significant effect for overconfidence, \( t(38) = -.76, p < .5 \). Finally, subjects high in Perceived Expertise performed better than the other subjects (means are .57 and .45, respectively), \( p < .015 \). Thus, although removing misleading items increases mean performances, the effect of the manipulation on item confidences is similar to that when misleading items are not removed.

One might argue that arbitrarily terming some of the items as being misleading is capricious. Instead, the issue is whether or not the items have a better that 50% chance of being answered correctly. Thus, we performed analyses similar to those above, but only using those (15) items which were answered correctly by more than 50% of the subjects.

In general, the findings were consistent with the previous analyses. Subjects who were high in Perceived Expertise were more confident than those who were low in Perceived Expertise (means are .69 and .59, respectively), \( t(38) = 3.62, p < .002 \). But there was no difference in overconfidence, \( t(38) = -.72, p > .47 \). Finally, subjects high in Perceived Expertise performed better than subjects who were low in Perceived Expertise (means are .71 and .57, respectively), \( t(38) = 2.33, p < .03 \). Thus, regardless of whether we analyzed all of the items, only items that were not misleading, or only items with a greater than 50% chance of being answered correctly, the results were essentially the same. Subjects high in Perceived Expertise are more confident, and perform better, than subjects low in Perceived Expertise.

**Anchor and adjust.** One could argue that putting a person in the low Perceived Expertise condition simply induces him or her to add a negative value when writing down item confidences. That is, subjects form item confidences based on the information accessible to them about the item, but then they "mechanically" add a negative number to all of these confidences in order to be "accurate" (see Fischhoff, 1982). In order to test this, we performed a 2 x 2 repeated measures ANOVA on item confidences with average confidence on easy (more than 50% of the subjects answered correctly) vs difficult (50% or less of the subjects answered correctly) items as the repeated measure. If subjects are simply anchoring and adjusting in the low Perceived Expertise condition, then there should be no interaction between Perceived Expertise and the difficulty of the items. Contradicting this possibility, however, a significant interaction was obtained, \( F(1,38) = 10.64, p < .01 \). When the items were easy, subjects high in Perceived Expertise were much more confident than subjects low in Perceived Expertise (means are .69 and .59, respectively). However, when the items were difficult, this difference was attenuated (means are .61 vs .56, respectively). (Both main effects were also signif-
icant, \( p < .01 \) in both cases.) Thus, the data contradict the possibility that the Perceived Expertise manipulation simply induces subjects to anchor and adjust mechanically downward. \(^3\)

One might argue that the lack of an effect when the items were difficult is due to a floor problem. However, a close look at the data casts doubt on this possibility. The difference between the two conditions when the problems were easy is \( .69 - .59 = .10 \). When the problems were difficult the mean item confidence for subjects who were in the high Perceived Expertise condition was \( .61 \). So, even if we accept \( .50 \) as a lower limit, there was still room to go down by \( .11 \). Despite this, the mean for subjects who were low in perceived expertise was \( .56 \), a difference of only \( .05 \) from \( .61 \). Furthermore, looking back to a previous analysis, subjects who were given 10 difficult items to begin with had a mean item confidence of \( .50 \) on those items. Thus, it is hard to argue that subjects were somehow reluctant to go down that far.

In sum, Perceived Expertise had the hypothesized effect on item confidences, and this was true even when easy items were analyzed separately. However, there was an unexpected effect on actual performance. The data indicate that putting subjects in the low Perceived Expertise condition reduced their actual performance relative to subjects in the other condition.

**DISCUSSION**

The findings from the two experiments show that Perceived Expertise, whether manipulated generally or in a task-specific manner, affects item confidences. Thus, the results contradict the item information processing hypothesis and are consistent with the general information processing hypothesis. From a more practical perspective, the general manipulation of Perceived Expertise successfully induced subjects to become well-calibrated. In addition, unforeseen by either of the present authors, the task-specific manipulation reduced actual performance. \(^4\) Thus, if one

\(^3\) In addition to reducing actual performance, another "side effect" of this task-specific manipulation of Perceived Expertise might be to reduce resolution. However, this is only tangentially relevant to the purpose of the present article.

\(^4\) One possible explanation for the poorer performance of subjects in the low Perceived Expertise condition in Experiment 2 is that they made an external attribution that the items would all be "impossible" and consequently put less effort into answering them. In contrast, subjects in the low Perceived Expertise condition in Experiment 1 may not have made any external attributions (but did, presumably, make a stable, internal attribution that they have low ability on the task). If this explanation is correct, it causes problems for traditional theories of achievement motivation (e.g., Heider, 1958; Weiner, 1974; Weiner, Frieze, Kukla, Reed, Resti, & Rosenbaum, 1972; Weiner, Russell, & Lerman, 1978) that assume that attributions to a stable, internal cause lead to either equal or less effort than do stable, external attributions.
wishes to reduce overconfidence without reducing actual performance, we advise using the general, rather than the task-specific, manipulation.

Although the obtained data indicate that Perceived Expertise can be chronically accessible if primed, and used to form item confidences, we are still in the dark as to the actual process by which people make use of this information. For example, people might form initial item confidences on the basis of Perceived Expertise and then use the item information to make modifications. Another possibility is that they form item confidences on the basis of the item information, and then, in some complicated way (the data in Experiment 2 contradicted a simple anchor and adjust notion), they use Perceived Expertise to make modifications. A third possibility is that they simultaneously integrate Perceived Expertise and item information in some sort of algebraic information integration process where each piece of information is weighted by its salience (this is somewhat analogous to Anderson, 1968, 1974 or to Fishbein & Ajzen, 1975).

There is also a fourth, more complicated, possibility. There may be some questions for which people know that they know the answer or know that they do not know the answer. In these cases, Perceived Expertise is irrelevant, even if accessible. However, when people are unsure of whether or not they know the answer, Perceived Expertise may be treated as simply another cue on which item confidences may be based. This fourth possibility can be made compatible with the information integration notion. One could argue that when people are sure they either know or do not know the answer, Perceived Expertise may not be salient, and therefore receives a weight of zero. Contrary to this, an argument could also be made that people are always very aware of their Perceived Expertise (so it is salient), but they do not use it to form an item confidence when they are sure they know or do not know the answer.

In sum, the present data show that although people do use Perceived Expertise to help them form item confidences, we have not yet determined how people use it. Several possibilities were entertained, and future research should enable us to distinguish between them.

REFERENCES


Received: June 3, 1991