

Overconfidence in ignorant experts

JAMES V. BRADLEY

New Mexico State University, Las Cruces, New Mexico 88003

When asked true-false questions that were too difficult to be answered correctly at better than a chance frequency, subjects often guessed in preference to admitting ignorance and often expressed certainty in the correctness of their answers. These tendencies were strongly correlated with the degree of expertise the subject believed himself to have in the area in which the question was asked. Under certain presumably ego-threatening conditions, subjects contradicted by answering "false" appreciably more often than they agreed by answering "true," although half the questions were, in fact, true and half were false.

Experts are often called upon to confirm or deny statements of alleged fact in court, and referees presumed to have at least peer-level expertise are often required to confirm or deny the statements of authors of manuscripts submitted for publication. Sometimes, however, the expert's knowledge must be inadequate for the task. When it is, will the expert guess rather than admit ignorance, will he express certainty in the correctness of his guess, and will the guess be a negativistic denial rather than an acquiescent agreement? It was conjectured that such tendencies would exist to an appreciable degree and would be strongest in areas of greatest claimed expertise.

METHOD

Two experiments were conducted, having in common the following characteristics. Subjects were equal numbers of male and female students in an introductory psychology course who earned class credit by serving in the experiment. (There were 144 and 184 subjects in the first and second experiments, respectively.) Each subject was asked to rank 12 areas of knowledge (logic, mathematics, science, literature, history, spelling, music, travel or geography, warfare, food, clothes, and sports) according to the subject's expertise in these areas. Subjects were also given a true-false test containing three questions in each of the 12 areas. Half of the 36 questions were true and half were false, and in each of the 12 areas, there was at least 1 true and at least 1 false question. There were two forms of the test. For each question that was true (or false) in Form A (e.g., "Andorra is a small country in the Pyrenees"), there was a highly similar counterpart question that was false (or true) in Form B (e.g., "Andorra is a small country in the Alps"). Half the subjects of each sex received Form A and half received Form B. Thus, for the subjects of each sex in each experiment, half of the questions in each area of knowledge were true and half were false. Since the true and false questions were liberally intermingled, their equal numbers should have tended to give those questions whose answers the subject did not know a probability of .5 of being answered "true" (or "false") and of being answered correctly (or wrongly).

In Experiment 1, subjects were run individually, and the test was presented in the form of a booklet. There was a separate page in the booklet for each of the 36 true-false questions.

The confidence rating procedure described for Experiment 2 was inspired by discussions with Darwin P. Hunt.

Above each question, in capital letters, was printed the area in which it fell (e.g., SCIENCE). Beneath each question were printed the words "true" and "false," and the subject was instructed that these were the only acceptable answers and that he should respond by circling the correct one. The three questions in each of the 12 areas were presented consecutively. One of the questions in each area was so difficult as to be deemed virtually "impossible" (e.g., "The boiling point of mercury is 357°C"). The other two were designed to be answerable by a person who was reasonably well informed in the area involved (e.g., "Asteroids are moderately unstable chemical compounds"), one being true and the other false. For each subject, half the impossible questions were true and half were false.

The sequential position at which the block of questions about a given area appeared was very carefully counterbalanced, as was the position of the impossible question within such a block.

After completing the booklet and returning it to the experimenter, the subject was given another booklet containing exactly the same questions in exactly the same order. The following was printed on the first page: "In the test you have just completed you may have done some guessing, since only two answers 'true' and 'false' were permitted. This booklet contains the same test. But this time three different answers are permitted: 'true,' 'I don't know,' and 'false.'" The subject was instructed to answer "I don't know" by writing a question mark in the space between the "true" and the "false."

Ninety-six of the subjects were asked to rank the areas of knowledge according to their believed expertise before answering both test booklets, and 48 were asked to do so afterward. For the first group, the instructions to the first test booklet were made intentionally ego threatening. (It was conspicuously titled "Expertise Study," and the first sentence of the instructions was the unnerving statement, "Now we are going to test your knowledge and expertise in the 12 subjects you have just ranked.") For the second group, all overt ego threats were removed or replaced by innocuous verbiage. Results for the two groups were highly similar, so their data were combined. This was also the case for the two sexes, in both this and the following experiment.

Experiment 2 investigated the influence of expertise upon confidence in one's answer rather than upon willingness to answer "I don't know." Subjects were run in huge groups, rather than individually, and none of the questions were designed to be impossible for a really well informed layman. The quasi-anonymity of the subjects, the generally easier questions, and the absence of any overt ego threat (e.g., the test was labeled "Information Survey") should have made this experiment the least ego-threatening situation with which the subjects were confronted.

Each subject was given a three-page test. The following was

printed on the first page: "This is a survey of students' general information. Each question is preceded by: 'True False 1 2 3 4 5.' Circle the correct answer, True or False. Then rate your confidence in the correctness of your answer by circling one of the numbers from 1 to 5. If you are completely uncertain—so that you believe your answer is no more likely to be correct than if you had arrived at it by the toss of a coin—circle the 1. If you are completely certain that your answer is the correct one, circle the 5. If your degree of certainty about your answer is somewhere between complete uncertainty and complete certainty, i.e., if you feel your chance of being right is more than 50-50 but less than 100 percent, circle the appropriate number on the five-point scale." The following descriptions were provided for the points on the scale: (1) completely uncertain, a toss-up, (2) slightly more likely right than wrong, (3) probably right, (4) very probably right, (5) completely certain, a sure thing. Thirty-six questions followed, in each third of which "true" was the correct answer for a randomly located six questions and "false" was the correct answer for the other six. The counterpart questions in Forms A and B had the same ordinal position within the test.

After the subject handed the completed test to the experimenter, he was given a sheet listing the 12 areas covered by the "information survey" and asking him to write the names of these areas in 12 vertically arranged blocks labeled "Best subject (know most about), 2nd best subject, . . . , Worst subject (know least about)," the same procedure used in Experiment 1.

Of the 72 questions in Forms A and B of the test, 18 were answered incorrectly more often than correctly. Of these 18, 3 had a suspiciously large surplus of correct over incorrect high-confidence answers in the area of greatest claimed expertise. Eliminating these 3 leaves 15 questions that appear to have been answered correctly only by chance. These 15 will be termed "overly difficult" questions. They are analogous to the "impossible" questions of Experiment 1, except that (1) the latter were designed, a priori, to be answerable correctly only by chance, whereas the former simply turned out empirically, a posteriori, to be so, (2) the "overly difficult" questions lack

the unreasonable quality of the "impossible" questions, and (3) they do not consist of pairs of counterpart questions, as the impossible questions do.

RESULTS

Admissions of Ignorance and Uncertainty

Results for Experiments 1 and 2 are shown in Figures 1 and 2, respectively. Except for "correctness" in certain cases, all of the curves in these figures represent variables having a significant rank-order correlation with degree of claimed expertise.

Consider first the impossible and overly difficult questions. Despite the fact that these questions were answered correctly only at or below a chance frequency and that their correctness was not correlated with degree of claimed expertise, subjects were increasingly unwilling to admit their incompetence as claimed expertise increased. They decreasingly admitted that they did not know the answer and decreasingly admitted to any degree of uncertainty whatever (i.e., they increasingly expressed complete certainty in their answers). Furthermore, when their answers to these questions were, in fact, wrong, subjects in the first experiment increasingly rejected the opportunity to admit ignorance (or to switch to the right answer) in the second booklet and perseverated in giving the wrong answer, and subjects in the second experiment increasingly expressed certainty in the accuracy of their wrong answers.

The really crucial data are those for the impossible and overly difficult questions; however, the data for all types of question provide supportive evidence. If we

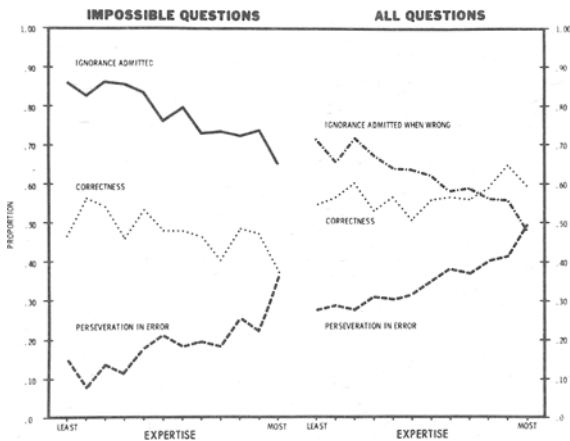


Figure 1. For each level of claimed expertise in Experiment 1, (1) the proportion of questions (of the specified type) that were answered correctly in the first booklet ("correctness"), (2) the proportion of impossible questions that were answered "I don't know" in the second booklet ("ignorance admitted"), (3) for all those questions that were answered wrongly in the first booklet, the proportion that were answered "I don't know" in the second booklet ("ignorance admitted when wrong"), and (4) for those questions (of the specified type) that were answered wrongly in the first booklet, the proportion that received the same wrong answer, rather than "I don't know" or the correct answer, in the second ("perseveration in error").

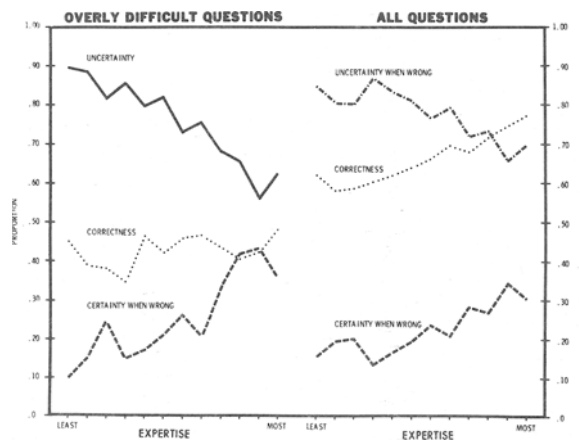


Figure 2. For each level of claimed expertise in Experiment 2, (1) the proportion of questions (of the specified type) that were answered correctly ("correctness"), (2) the proportion of overly difficult questions about whose answer the subject was less than completely certain, having chosen a confidence rating of 1, 2, 3, or 4 ("uncertainty"), (3) for all those questions that were answered wrongly, the proportion of answers about which the subject felt any degree of uncertainty ("uncertainty when wrong"), and (4) for those questions (of the specified type) that were answered wrongly, the proportion of answers about whose correctness the subject expressed complete certainty ("certainty when wrong"), having circled Confidence Rating 5.

consider all questions, rather than simply a very difficult subset of them, we find that the questions were answered correctly at a greater than chance frequency (and that correctness is correlated with expertise for the totality of questions in Experiment 2 and for the non-impossible questions in Experiment 1). This makes it desirable to confine the analysis to wrong answers. The "ignorance admitted when wrong" curve can be compared with the "ignorance admitted" curve for impossible questions, since, if plotted, an "ignorance admitted when wrong" curve for impossible questions would have intertwined with the "ignorance admitted" curve and would have had almost exactly the same slope and average height. Likewise, for analogous reasons, the "uncertainty when wrong" curve for all questions can be meaningfully compared with the "uncertainty" curve for overly difficult questions. The curves for "uncertainty when wrong" and "certainty when wrong" are, of course, mutually redundant; the reason for including both is to facilitate comparisons with their counterparts in other graphs.

In Experiment 2, except for "correctness," the curves for "all questions" were highly similar to their counterparts for "overly difficult questions." Much the same can be said for the slopes of the curves in Experiment 1, but not for the heights of the curves. There was less willingness to admit ignorance when wrong and more perseveration in error when all questions are considered than just for impossible questions.

Other indexes would have told the same story as the ones actually used. For example, admissions of complete uncertainty (selection of Confidence Rating 1) diminished with increasing expertise both in general and when the answer was wrong.

Negativism

A sort of diffuse negativism appeared in Experiment 1 but not in Experiment 2. Although exactly half the questions were false, the proportion of times the answer "false" was chosen (in the first booklet) by the first and second groups, respectively, in Experiment 1 was .549 ($p < .01$) and .533 (n.s.) for impossible questions and .560 ($p < .001$) and .569 ($p < .001$) for the nonimpossible ones. Thus, in general, the tendency to contradict was highly significant (by Wilcoxon signed-ranks test). However, although negativism clearly exceeded chance, it did not correlate significantly with degree of claimed expertise. Indeed, a curve plotting negativism against degree of claimed expertise in Experiment 1 would merely show chance fluctuations about a nearly horizontal line.

DISCUSSION

The least complicated and most easily interpretable results are those obtained when difficulty level remained essentially constant at a level so high that correctness could be attributed to chance alone. Not only were the impossible questions of Experiment 1 so difficult that they were answered correctly at

less than a chance frequency (48% of the time), but they were so "impossibly" difficult it would seem that the subjects must have realized that they had no inkling as to the correct answer. Yet, despite this, they admitted their ignorance only 78% of the time, the percentage declining from 86% at least claimed expertise to 65% at greatest. And if we confine the analysis to wrong answers, thereby excluding any subjects who might really know the answer, results are almost identical: The percentage of subsequent admissions of ignorance declined from 84% at least expertise to 60% at most, averaging 78%. The results for all questions combined are similar, except that ignorance when wrong was admitted relatively less often and perseveration in error occurred relatively more often than was the case for impossible questions alone. This might be attributable to the very impossibility of the impossible questions. Perhaps the nonimpossible, but still difficult, questions were more ego threatening precisely because it was not totally unreasonable to expect a truly well informed layman to answer them correctly. If this interpretation is correct, subjects can be expected to become more unreasonable as questions become less so. In any event, the conclusion seems inescapable that totally ignorant subjects were often unwilling to admit their ignorance and that this unwillingness increased substantially with the degree of expertise they believed themselves to have.

The fact of the relationship seems established. One possible reason for it might be that some subjects, although having no delusions about their ignorance, simply prefer to guess rather than to say "I don't know" and that this preference becomes stronger with increasing expertise, or perhaps interest, in the general subject area in which the question lies. If this were the proper explanation (and if subjects were candid), however, one would not expect the subjects to express certainty about their blind guesses, as Experiment 2 suggests that they do. The remarkable similarity of the "uncertainty" and "certainty when wrong" curves from Experiment 2 to the "ignorance admitted" and "perseveration in error" curves, respectively, from Experiment 1 suggests that refusal to admit ignorance is accompanied by an unwillingness to admit the slightest uncertainty.

These results are in harmony with the discovery by Mahoney and DeMonbreun (Mahoney, 1976; Mahoney & DeMonbreun, 1977) that in a problem solving situation, scientists showed "tenacious fidelity" to hypotheses already declared wrong and students often "remained virtually certain of the veracity of their hypothesis," despite disconfirmatory feedback.

Of particular interest is the case in which the person being consulted is a journal referee. Mahoney (1976, 1977) sent manuscripts "identical in topic and procedure" but reporting results that either supported or contradicted or did neither to the viewpoint of the psychologist to whom it was sent for refereeing. He found that "reviewers were strongly biased against manuscripts which reported results contrary to their theoretical perspective." The present experiment suggests that this type of negativism may be accompanied by a negativistic tendency to contradict when in doubt about esoteric matters and by a reluctance to admit to ignorance when one should, the reluctance being most pronounced in one's area of greatest expertise.

REFERENCES

- MAHONEY, M. J. *Scientist as subject: The psychological imperative*. Cambridge, Mass: Ballinger, 1976.
- MAHONEY, M. J. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1977, 1, 161-175.
- MAHONEY, M. J., & DEMONBREUN, B. G. Psychology of the scientist: An analysis of problem-solving bias. *Cognitive Therapy and Research*, 1977, 1, 229-238.

(Received for publication January 20, 1981.)