Confidence Intervals and Overprecision

BerkeleyHaas

A teaching note by Don Moore

Overprecision in Judgment

Human judgment is routinely overprecise. That is, we are too sure of the accuracy of our knowledge. Our faith in ourselves enables many other biases and errors. Demonstrating to people that they are too sure of themselves may be useful for opening their minds to the possibility that they have something to learn and opening their hearts to new lessons.

Directions for Exercise

1. Select a set of questions with unambiguous quantitative answers. The set can be as large or small as you wish, but ten questions often works well for this exercise. Ten questions are enough for each individual to see a pattern in their own responses, but does not take too long to complete. It can be fun to choose questions relevant to the audience. When presenting to members of an organization, I may try to use numbers specific to that organization such as total revenues, number of employees, or the size of building space.

2. Instruct participants that their assignment will be to write down a confidence interval for each question. That interval should have two numbers: one below the individual's best guess, one above it.

3. You may select a size for the confidence interval. 90% is a good default. At the high end you could ask for a 99% confidence interval, or a 50% confidence interval at the low end. The size of the confidence intervals that people give you will be insufficiently responsive to this confidence level, so their hit rates inside 50% confidence intervals will appear better calibrated. If you ask for 99% confidence intervals, hit rates around 50% will make them look very overprecise indeed.

4. Go through the list of questions one at a time, asking participants to write down their confidence intervals for each as you go.

5. Reveal the answers one at a time, asking participants to score themselves by checking whether the right answer is inside the confidence interval.

6. Ask participants to total up the number of questions on which they "hit" by getting the answer inside their confidence interval. You can ask people to indicate how many they got by show of hands. Ask, "How many of you got ten out of ten?" Count those hands and maybe write the number on the board. "Nine out of ten?" and so on.

Alternative Approaches

It is still possible to conduct this demonstration in a setting where the participants do not have paper and writing implements. In this case, call out the topics one at a time:

- Announce the question. For example, "What is world population?"
- Ask each person in the crowd to think of a 90% confidence interval by specifying an interval: a number below your best guess, and a number above your best guess. This range should be wide enough that you are 90% sure the right answer is inside it.
- Once everyone has a range in mind, announce the right answer to the question.

- By show of hands, ask people to report whether the truth fell inside the interval they had in mind.

You may ask as many questions as you like this way. It is usually apparent, after 3 or 4 questions, that less than 90% of the hands go up each time.

This can serve as a useful demonstration of overprecision in the setting of 90% confidence intervals.

Debrief

1. Begin by noting that hit rates were less than 90%. Ask the class "Why not?"

2. Someone is likely to say something like, "I wanted to show that I knew what I was talking about." To this, you can respond, "To whom were you demonstrating this accurate knowledge? You certainly weren't demonstrating it to me! I didn't get to see your narrow confidence intervals. I just got to see your unimpressive hit rates."

3. Eventually, someone will say, "We were overconfident." Bingo. Pick up on this and run with it. "If you are overly sure of the accuracy of your knowledge, you would be in good company." This is a nice segue into a discussion of the evidence on overprecision in judgment.

Ideas for potential questions

- World population
- Population of your country or state
- Revenues of your organization
- Gross Domestic Products of different countries
- Height of Mount Everest
- Depth of the Mariana Trench, the deepest point in the Pacific Ocean
- Year in which famous people were born or died
- Length of famous books
- Duration of famous movies
- Heights of famous people

Hit rates will be below 90% regardless of whether the topics are familiar or unfamiliar. If you choose unfamiliar topics, such as the height of Everest or the depth of the Mariana Trench, participants will object that they would have done better had you picked a topic closer to their expertise. You can remind them that if they didn't know much about it they simply could have made their confidence interval wider, but not everyone will be persuaded by this response. Therefore, if you can pick questions relevant to the expertise of participants, that will make the exercise more compelling.