Organizational Behavior and Human Decision Processes 131 (2015) 110–120

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

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

Wide of the mark: Evidence on the underlying causes of overprecision in judgment ☆

Don A. Moore [a,*], Ashli B. Carter [b], Heather H.J. Yang [c]

[a] University of California, Berkeley, United States
[b] Columbia University, United States
[c] Massachusetts Institute of Technology, United States

ARTICLE INFO

Article history:
Received 9 March 2015
Revised 9 September 2015
Accepted 27 September 2015
Available online 9 November 2015

Keywords:
Overconfidence
Overprecision
Probability
Epistemic uncertainty
Judgment

ABSTRACT

Overprecision is the most robust and least understood form of overconfidence. In an attempt to elucidate the underlying causes of overprecision in judgment, the present paper offers a new approach – examining people's beliefs about the likelihood of chance events drawn from known probability distributions. This approach allows us to test the assumption that low hit rates inside subjective confidence intervals arise because those confidence intervals are drawn too narrowly. In fact, subjective probability distributions are systematically too wide, or insufficiently precise. This result raises profound questions for the study of overconfidence.

1. Introduction

Overprecision is the excessive faith that one's beliefs are correct. It is simultaneously the most robust and the least understood form of overconfidence (Moore, Tenney, & Haran, 2015). The typical overprecision study asks people questions with quantitative answers (e.g., "How much does Barack Obama weigh?") and asks them to estimate 90% confidence intervals around these answers. However, these 90% confidence intervals routinely contain the correct answer less than 50% of the time (Alpert & Raiffa, 1982). This overprecision effect is one of the most dramatic and impressive in the decision making literature, and has been replicated in many paradigms and populations (Block & Harper, 1991; Harvey, 1997; Mamassian, 2008).

However, one of the impediments to the study of overprecision has been the difficulty specifying the relevant knowledge that individuals possess and, consequently, whether they use that information effectively. Most prior research approaches have not made it easy to compare research participants' beliefs with the normatively correct beliefs at the level of the individual question (see Lawrence, Goodwin, O'Connor, & Önkal, 2006). Instead, when researchers observe that hit rates inside 90% confidence intervals are below 90%, they quite sensibly assume that this is because subjects are underestimating the uncertainty around their beliefs across a set of questions (Tversky & Kahneman, 1974). However, without being able to compare how sure someone *is* with how sure they *ought to be* of something in particular, we cannot know whether judgmental overprecision is, in fact, always due to overly narrow subjective probability distributions.

This limitation imposes several problematic constraints. For one thing, it obscures the cause of overprecision because it cannot tell us whether or when confidence intervals have such low hit rates because they are too narrow or because they are centered on the wrong estimate. Many researchers write about overprecision as if it occurs because people have overly narrow subjective probability distributions (Harvey, 1997; Jain, Mukherjee, Bearden, & Gaba, 2013; Mannes & Moore, 2013; Soll & Klayman, 2004). Others posit that overprecision is a consequence of people relying on available, but potentially biased, information. For example, Juslin, Winman, and Hansson (2007) characterized people as "naïve intuitive statisticians." The naïveté of the intuitive statistician is the uncritical reliance on *sample* properties and mistaking them for *population* properties. Here, the underlying assumption is that people's beliefs are centered on sensible estimates gathered from experience, but that individuals fail to appreciate the fact that their small samples underestimate error variance—a mistake that leads to overly precise beliefs.

Hit rates below 90% appear as *prima facie* evidence that 90% confidence intervals are drawn too narrowly relative to the individual's own error distribution. Prior research has, however, relied

☆ Materials and data: http://learnmoore.org/BDE/.

* Corresponding author.
E-mail address: don.moore@alumni.carleton.edu (D.A. Moore).

http://dx.doi.org/10.1016/j.obhdp.2015.09.003
0749-5978/

on the untested assumption that, if only we could specify the error distribution (rather than inferring it post hoc from observed error rates), we would observe that individuals' self-reported confidence intervals are too narrow. In this paper, we test this assumption. We can specify what participants *should* believe for a full range of outcomes and whether subjective probability distributions are in fact too narrow relative to this benchmark. Using a novel experimental paradigm, this paper questions whether low hit rates necessarily imply too much certainty in a particular belief as measured by confidence interval width. Furthermore, our results shed light onto the process that people use to make judgments in the face of uncertainty and provide evidence against some common assumptions in overconfidence research.

1.1. Explanations for overprecision

The results will inform three of the most prominent explanations for overprecision in judgment: anchoring, conversational norms, and naïve intuitive statistics. The anchoring explanation holds that overprecision is the result of people first making some best estimate and then adjusting insufficiently from it (Block & Harper, 1991; Plous, 1995). If this explanation is right, then helping set the anchor by eliciting a best guess should lead to subject probability distributions centered even more tightly around the best-guess judgment. We do not find that it does.

The conversational norms explanation holds that people express overprecision because they are trying to provide informative judgments, even when that comes at the expense of accuracy. Imagine I ask my friend for the location of Stanford University. If my friend tells me it is in the city of Palo Alto, that would be informative but inaccurate, given that the university is, in fact, in the municipality of Stanford, California (next to the larger city of Palo Alto). However, my friend's response is more useful than if she would have said that Stanford is somewhere in northern California, a response that would have been accurate at the expense of being informative. Indeed, many people express a preference to get informative over accurate advice (Yaniv & Foster, 1995). For this to be able to account for overprecision, it must arise from overly precise subjective distributions. That is not what we observe in our data.

The naïve intuitive statistician argument holds that subjective error distributions are smaller than actual error distributions because our minds take a small sample of relevant facts. Our minds are limited to thinking of about 7 (plus or minus 2) facts at once (Juslin et al., 2007). This small sample will have a smaller variance than the actual population of relevant facts, leading people to underestimate the uncertainty around their knowledge. This explanation also posits that subjective probability distributions are overly narrow, and it applies best to epistemic uncertainties that arise due to the imperfections in our own knowledge. It does a poor job explaining why we observe underprecision in subjective probability distributions, regardless of whether uncertainties are framed as either epistemic or aleatory.

1.2. Overview of the studies

Experiment 1 tests the traditional method of eliciting confidence intervals against our new approach. Experiment 2 attempts to reconcile results from our new approach with apparently contradictory conclusions in the research literature. For Experiment 2 and the remaining experiments, rather than ask for a confidence interval, we use the Subjective Probability Interval Elicitation (SPIES) measure introduced by Haran, Moore, and Morewedge (2010) whereby participants estimate the full probability distribution of outcomes—providing a probability estimate (from 0% to 100%) for the likelihood of each possible outcome.

In Experiment 3, we examine the degree to which novel results from our new approach are due to its lack of familiarity. In Experiment 4, we explore the degree to which the expression of uncertainty in our new paradigm is moderated by its conceptualization as uncertainty around a repeatable event with many different possible outcomes based upon rules of chance (aleatory uncertainty) or as lack of knowledge regarding a unique event with a particular outcome (epistemic uncertainty), (Fox & Ülkümen, 2011). We explore this possibility because past work shows that individuals express less certainty when making judgments of events that are unknown due to chance factors (aleatory frame) compared to events that are unknown due simply to lack of knowledge (epistemic frame), (Fox & Ülkümen, 2011). Finally, Experiment 5 examines the robustness of our results using a behavioral measure of precision in judgment.

We report how we determined our sample size, all data exclusions (if any), all conditions, and all measures for all studies. Data and materials are available online: http://learnmoore.org/BDE/.

2. Experiment 1

2.1. Method

In order to compare our new approach to traditional methods, we conducted a survey with three general knowledge questions of the sort that have consistently produced overprecision in prior research: Estimating the increase in the value of a stock, estimating someone's weight, and estimating someone's age (Gino & Moore, 2007). To these, we added two questions with known probability distributions, shown in Table 1 (estimating the final location of a jumping bean and estimating the winnings of a lottery).

2.1.1. Design

All participants answered questions about each of the five topics, presented in random order. For the lottery topic questions, following Jain et al. (2013), we tried to help our participants understand the lottery's random nature by providing a picture of 10 random paths of winnings that were possible for the first 150 days of the lottery. However, because we were concerned about participants anchoring their judgments on this sample of 10 random paths, we generated 20 such pictures (each with 10 different paths) and randomly presented each participant with one of them. We were also concerned that the figure's scale would provide an implicit possible range of possible winnings, so we generated two versions: ten of the pictures had a scale that ran to $100 (where the 10 paths were easy to distinguish), and the other ten had a scale that ran to $500, the theoretical maximum. The pictures presented to participants appear in this paper's online supplement. These precautionary variations did not end up producing any significant effects on our results, so we do not dwell on them.

For each of the five topics, we asked two questions that have been used in prior research on overprecision in judgment:

(1) *90% confidence interval:* "Please give us two numbers: a 'lower bound' and an 'upper bound'. The 'lower bound' is a number so low that there is only a 5% probability that the right answer is less than that. Similarly, an 'upper bound' is a number so high that there is only a 5% probability the right answer is more than that. In other words, you should be 90% sure that the answer falls between the lower and upper bounds."
(2) An *item-confidence judgment*, which had two parts:
    a. A "best estimate" of the right answer,
    b. A confidence question: "How confident are you that your answer is within 5% of the right answer?"