Available online at www.sciencedirect.com
ScienceDirect
Organizational Behavior and Human Decision Processes 107 (2008) 179–191
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

Overconfidence in interval estimates: What does expertise buy you? ☆

Craig R.M. McKenzie [a,*], Michael J. Liersch [b], Ilan Yaniv [c]

[a] *Rady School of Management and Department of Psychology, University of California, San Diego, 9500 Gilman Drive, MC 0553, La Jolla, CA 92093-0553, USA*
[b] *Department of Psychology, University of California, San Diego, 9500 Gilman Drive, MC 0553, La Jolla, CA 92093-0553, USA*
[c] *Hebrew University, Jerusalem, Israel*

Received 14 July 2005
Available online 28 March 2008

Accepted by Robyn Dawes

Abstract

People's 90% subjective confidence intervals typically contain the true value about 50% of the time, indicating extreme overconfidence. Previous results have been mixed regarding whether experts are as overconfident as novices. Experiment 1 examined interval estimates from information technology (IT) professionals and UC San Diego (UCSD) students about both the IT industry and UCSD. This within-subjects experiment showed that experts and novices were about equally overconfident. Experts reported intervals that had midpoints closer to the true value—which increased hit rate—and that were narrower (i.e., more informative)—which decreased hit rate. The net effect was no change in hit rate and overconfidence. Experiment 2 showed that both experts and novices mistakenly expected experts to be much less overconfident than novices, but they correctly predicted that experts would provide narrower intervals with midpoints closer to the truth. Decisions about whether to consult experts should be based on which aspects of performance are desired.

Keywords: Overconfidence; Expertise; Interval estimates

People often express uncertain values in terms of an interval, such as when they estimate their arrival time ("Between 5:00 and 5:30"), another person's age ("35 to 40"), or next year's inflation rate ("3% to 5%"). The accuracy of such estimates is usually measured in terms of hit rate: How often do the intervals contain the true value? Hit rates are often compared to the degree of confidence reported in the intervals. For example, participants might be asked to report low and high values for the populations of various cities such that they are 90% confident that each resulting interval contains the city's true population. If people were well calibrated, 90% of their 90% confidence intervals would contain the true value. However, true values typically fall within such intervals between 30% and 60% of the time, indicating extreme overconfidence (e.g., Alpert & Raiffa, 1982; Juslin, Wennerholm, & Olsson, 1999; Klayman, Soll, González-Vallejo, & Barlas, 1999; Lichtenstein, Fischhoff, & Phillips, 1982; Soll & Klayman, 2004; Teigen & Jørgensen, 2005; Yaniv & Foster, 1997).[1]

☆ This research was supported by National Science Foundation Grant SES-0551225 and by Israel Science Foundation Grant 344/05. Some of the results were presented at the 2004 Annual Meeting of the Society for Judgment and Decision Making in Minneapolis, MN.
* Corresponding author. Fax: +1 858 534 7190.
E-mail address: cmckenzie@ucsd.edu (C.R.M. McKenzie).
URL: http://psy.ucsd.edu/~mckenzie/ (C.R.M. McKenzie).

[1] Although we will usually refer to intervals associated with a particular level of confidence as "subjective confidence intervals" (or "confidence intervals" for short), readers should be aware that these intervals are sometimes referred to in the literature as "credible intervals", "uncertainty intervals", "probabilistic prediction intervals", and "fractile assessments" (Teigen & Jørgensen, 2005).

0749-5978/$ - see front matter
doi:10.1016/j.obhdp.2008.02.007

Attempts to overcome overconfidence in interval estimates have been only modestly successful, indicating that the degree of overconfidence is not only large, but also robust (Alpert & Raiffa, 1982; Lichtenstein et al., 1982). Such overconfidence has practical, as well as theoretical, significance. Russo and Schoemaker (1992) described a leading US manufacturer that elicited a projected range of sales from its marketing staff in order to plan the production capacity of a new factory. The range turned out to be too narrow, and the new factory was incapable of meeting the unexpected demand.

Most confidence interval studies ask undergraduate students about values they are unlikely to know much about (e.g., "What is the gestation period of an Asian elephant?"), and one might wonder whether being more knowledgeable reduces overconfidence. Several studies have examined how well experts assign probabilities to events, and the results have been mixed (e.g., Christensen-Szalanski & Bushyhead, 1981; Keren, 1987; Lichtenstein & Fischhoff, 1977; Murphy & Winkler, 1977; Oskamp, 1965; for reviews, see Camerer & Johnson, 1997; Koehler, Brenner, & Griffin, 2002). Only a few studies have examined experts' interval estimates, and those results have been mixed as well. Russo and Schoemaker (1992) asked advertising, petroleum, and money management professionals (among others) for interval estimates in their domain of expertise, and these experts' 90% and 95% confidence intervals were typically accompanied by hit rates of 40–60%. Although these results seemed to suggest that experts are just as overconfident as novices, no direct comparison was made between them. Önkal, Yates, Simga-Mugan, and Öztin (2003) did make direct comparisons by studying experts' and (sophisticated) novices' ability to predict foreign exchange rates. The experts tended to outperform the novices in terms of point predictions and predicting direction of change, but there were no differences in hit rates for 90% confidence intervals. When predicting exchange rates 1 day and 1 week ahead, hit rates for the two groups ranged between 40% and 56%.

Yates, McDaniel, and Brown (1991) found that graduate students in finance classes ("semi-experts") were *more* overconfident than undergraduate finance students when predicting changes in stock prices, suggesting that expertise can even exacerbate overconfidence (see also Staël von Holstein, 1972). However, these participants assigned probabilities to six fixed, nonoverlapping intervals for each stock ("increase in price greater than 10%", "increase in price between 5% and 10%", and so on), which is different from generating a single (high) confidence interval. Furthermore, although there was a reliable difference in accuracy, the graduate and undergraduate students nonetheless provided very similar predictions.

Finally, Tomassini, Solomon, Romney, and Krogstad (1982) studied professional auditors' subjective probability distributions for financial statement account balances. They concluded that these experts tended to be *under*confident, but only for relatively low-confidence intervals (50% and 80% intervals). For high-confidence intervals (98%), the auditors were overconfident, but much less so than the typical findings in the literature (they had hit rates of 93% for their 98% confidence intervals). However, there was no control (i.e., novice) condition, the participants reported multiple fractiles (rather than a single high-confidence interval), and they were given "detailed training" by the experimenters about subjective probability distributions (and given opportunities to change their responses after further review and consultation with the experimenters), making it difficult to know if their auditing expertise was responsible for the diminished overconfidence.

In short, although it is clear that overconfidence in (high) confidence interval estimates is large, robust, and can have important consequences, little is known about differences between experts and novices. It seems safe to say that experts are overconfident, but it is unclear how they compare to novices, and knowing how they compare is important. If experts are just as overconfident—or more overconfident—than novices, why consult them at all?

In this article, we investigate three related topics. First, we examine hit rates for high-confidence intervals provided by information technology (IT) professionals and University of California, San Diego (UCSD) undergraduate students to questions about both the IT industry and UCSD. Thus, for questions about UCSD, UCSD students are experts and IT professionals are novices, while for questions about the IT industry, expertise is reversed. This within-subjects examination of expertise and interval estimates is the first of its kind, as far as we know, and allows us to make direct comparisons between expert and novice hit rates while controlling for differences between the groups of participants and the sets of questions. The "naïve" prediction is that experts will be less overconfident simply because they know more, but, as mentioned, there is little evidence to support this prediction and, depending on how relevant the Yates et al. (1991) study is perceived to be, there is even evidence supporting the opposite prediction.

Second, we examine more than just hit rates by evaluating both interval width and error, where error is defined as the absolute distance between the midpoint of an interval and the true value (Yaniv & Foster, 1995, 1997). Such measures can provide insight into why hit rates are relatively high or low. For example, wider intervals will generally increase hit rate, all else equal. If experts have higher hit rates than novices, it may be because they know more about the limits of their knowledge (i.e., have more metaknowledge) and therefore provide wider intervals. But intervals better centered on the truth will also lead to higher hit rates, holding interval width constant. Instead of reporting wider intervals, experts may make more accurate point predictions (e.g., Önkal et al., 2003), or best guesses of the true value. This could lead