PSYCHOLOGY

Pulling guesses up by their bootstraps

Multiple guesses from one individual, like guesses from a crowd, yield a better estimate when averaged. How far can such solipsistic polling take us in real, high-stakes settings? Now 1.2 million incentivized, real-world guesses show just how much people can improve their judgements by reconsidering their own estimates.

Edward Vul

A crowd of people gains wisdom when averaged: stadiums sing on key; the aggregate market yields higher risk-adjusted returns than individual investors; and the average of many inexpert guesses about the weight of an ox is surprisingly accurate[1]. But how much gain can be had by averaging guesses from a single person? Writing in *Nature Human Behaviour*, Dennie van Dolder and Martijn van den Assem tackle this question with a dataset that is large in every way: 1.2 million estimates from casino patrons trying to guess how many pearls were in a gigantic novelty champagne glass to win a €100,000 prize[2]. They find consistent benefits from averaging a single individual's guesses. However, they are a tiny fraction of the gains made from averaging guesses by different people.

Practically all decisions rely on some estimate of what the world is like or forecasts of how it will be in the future. Regrettably, such estimates are hard for all but the most trivial variables, so we tend to be quite wrong. Fortunately, other people's guesses are often wrong in different ways, so we can average these incorrect guesses and come up with a much better estimate. Discovering this 'wisdom of crowds'[3] in a competition to guess an ox's weight bolstered Francis Galton's[1] confidence in democracy, which involves the aggregation of many inexpert opinions.

Unfortunately, we don't often have the luxury of polling many people for their opinions, either because of pragmatic considerations, or because we are trying to estimate something subjective, private or personal, such as how long it will take to write a News & Views piece. Might a single person be able to improve her estimates by repeatedly 'polling' herself and averaging these estimates? It's not at all obvious that this would work as it requires the guesses to have independent error, and thus additional information, as though not all of the available information was used when forming the first guess. Surprisingly, this seems to be the case: we can generate better estimates by polling the same person many times and averaging; and even better estimates by spreading these guesses out over a period of time[4].

Credit: Panther Media GMBH / Alamy Stock Photo

This wisdom of the 'crowd within' has been studied in small-scale laboratory experiments, with few participants, few guesses per person, short timescales and inconsequential incentives for getting the right answer. Thus, it may be the case that the independent error in guesses from the same person is just a consequence of participants not caring. This suggests that if participants were sufficiently motivated, they might produce their best-possible guess the first time, eliminating the benefit of averaging in subsequent guesses. Furthermore, averaging guesses from one person bestows a small fraction of the benefit of averaging guesses from multiple people, but small-scale laboratory experiments can not identify why the inner crowd is not nearly as wise as a real crowd. Do individuals have their own idiosyncratic biases, such that no amount of guesses from one person could yield an estimate as good as the average of multiple people? Or do people just tend to stick to their initial guess? If so, could a sufficient delay between guesses make many estimates from a single person more informative?

Van Dolder and van den Assem resolved all of these questions by using a rare, real-world dataset, in which visitors to a casino in the Netherlands guessed how many pearls, diamonds or chips were in a gigantic novelty champagne glass. In each of the three contests, those who guessed closest to the right answer would share a non-trivial prize of €100,000, so there was unambiguously sufficient incentive to provide the best possible guess. The approximately 160,000 casino patrons who made guesses in a given contest could do so repeatedly over a two-month window, yielding roughly 400,000 guesses per contest.

In this high-stakes case, van Dolder and van den Assem found that an individual can reduce her error by making more guesses and averaging them, and this benefit increases with more guesses and with greater delays between guesses. Moreover, the richness of the data allowed the authors to measure whether averaging guesses from the same person helps less than different people because each person has their own idiosyncratic bias or because people tend to stick to their initial guesses.

The authors could estimate the different sources of error by virtue of the correlation structure across guesses. Some error applies to every guess (for example, on average everyone overestimates), some error applies to all guesses from the same person (for example, Steve reliably overestimates even more), some error applies to a person but varies over time (for example, Steve tends to err in the direction of his initial guess until he forgets it a few weeks later) and the remaining error arises because each guess is just a bit different. This study gives us the first precise breakdown of these sources of errors: about 7% is from an overall population bias, 50% is from consistent individual biases, 27% from time-varying individual biases and 16% from random fluctuation. This means that gathering many guesses over large spans of time from one person would reduce error by 43%, but averaging just two guesses from different people cuts error by nearly 47%. In this case, it's clearly inefficient to obtain many estimates from a single person over a long period of time to eek out — at most — a benefit comparable to just having asked one other person.

In the casino task, averaging guesses from a single individual is inefficient because half of the total error arises from stable, individual biases and so cannot be corrected by averaging. However, there are cases where we are explicitly interested in such individual biases. For example, consider estimates of subjective qualities, such as how much you would like one or another vacation. Your interest in these cases is precisely in forecasting your own preferences and so it can be quite useful and actionable to know that a two-month delay will offer a fresh, nearly independent, sample of your idiosyncratic preferences, which according to the casino study, might reduce your error by almost 90%.

However, it would be imprudent to leap to such conclusions based on the casino estimation task, since these errors are unlikely to generalize across settings. Some settings produce large biases that are consistent across all observers (like perceptual illusions and some judgement errors), in which case averaging guesses from many people would not reduce the error. Other settings instead have large, stable, individual differences. For example, only 3% of decided US voters changed their preferred presidential candidate in the year preceding the 2016 election[5] (while the magnitude of stable individual biases in casino estimation would suggest a figure closer to 25%). Even within a given domain, the stability of individual biases seems to vary across contexts. In New Zealand, for instance, the rate of opinion change in the month before an election was nearly 30% (ref. [6]). The current study provides a rigorous framework for characterizing sources of errors, using an objective, high-stakes and largely unbiased estimation task. It remains for future work to see how far these estimates generalize, and the framework laid out in the study allows other researchers to do exactly that. ❒

Edward Vul
Department of Psychology, University of California, San Diego, La Jolla, CA 92037, USA.
e-mail: evul@ucsd.edu

Published online: 22 December 2017
https://doi.org/10.1038/s41562-017-0280-5

References

1. Galton, F. *Nature* **75**, 450–451 (1907).
2. van Dolder, D. & van den Assem, M. J. *Nat. Hum. Behav*. https://doi.org/10.1038/s41562-017-0247-6 (2017).
3. Surowiecki, J. *The Wisdom of Crowds* (Random House, New York, NY, 2004).
4. Vul, E. & Pashler, H. *Psychol. Sci.* **19**, 645–647 (2008).
5. Hopkins, D. Most voters haven't changed their minds all year. FiveThirtyEight http://fivethirtyeight.com/features/most-voters-havent-changed-their-minds-all-year/ (2016).
6. Blais, A. *PS Polit. Sci. Polit.* **37**, 801–803 (2004).

Competing interests

The author declares no competing interests.