

Lay Perceptions of Scientific Findings: Swayed by the Crowd?

First author name¹ & Second author name¹

¹ Insert affiliation here

Abstract

Crowd science is on the rise in behavioral science and aims to increase the credibility of scientific research. Whether it does, however, remains untested. We run an experiment to study whether crowd science improves lay perceptions of scientific findings. We focus on crowdsourced data analysis: a crowd of scientists independently analyzes the same dataset to estimate and report on a parameter of interest.

We compare the effects of providing lay consumers ($N = 1,498$) with a single parameter estimate (the single estimate condition) vs. multiple parameter estimates from a crowd that (a) vary slightly and are all positive, leading to the same qualitative conclusion (the “consistent crowd” condition) or (b) vary widely and are of both signs, leading to differing qualitative conclusions (the “inconsistent crowd” condition). In the single estimate condition, a single team of six researchers reports a 5% estimate; in the consistent crowd condition, six independent researchers report six low variance, high consensus estimates (2%, 4%, 5%, 5%, 6%, 8%; $M = 5\%$, $\sigma = 2\%$); in the inconsistent crowd condition, six independent researchers report six high variance, low consensus estimates (-6%, -2%, 5%, 5%, 12%, 16%; $M = 5\%$, $\sigma = 8.25\%$). We draw from theories on social norms and numerical cognition to hypothesize that consistent (inconsistent) crowd estimates will lead to higher (lower) posterior beliefs and credibility of the results, lower confidence in the precise average parameter estimate, and lower (higher) ratings of bias and error, than a single estimate.

In line with our preregistered hypotheses, we find that lay consumers of inconsistent (high variance, low consensus) crowd estimates have lower posterior beliefs, $b = -22.80$, 95% CI $[-25.59, -20.00]$, $p < .001$; find the results less credible, $b = -6.54$, 95% CI $[-9.31, -3.76]$, $p < .001$; have less confidence in the average effect size estimate, $b = -9.09$, 95% CI $[-12.11, -6.07]$, $p < .001$; and believe the results are more likely to stem from bias, $b = 6.68$,

95% CI [3.48,9.89], $p < .001$, and error, $b = 5.89$, 95% CI [2.98,8.80], $p < .001$, than consumers of a single estimate. Contrary to our hypotheses, we do not find that consuming consistent (low variance, high consensus) crowd estimates improves lay perceptions of scientific findings: instead, to our surprise, it leads to lower posterior beliefs, $b = -5.71$, 95% CI [-8.51, -2.91], $p < .001$, and higher ratings of error, $b = 3.59$, 95% CI [0.68,6.51], $p = .016$, than a single estimate. In the future, it is important for crowd scientists to consider how – e.g., by drawing from research on science communication – to tackle science skepticism and effectively communicate variable crowd estimates to lay consumers.

Keywords: Crowd science, Science skepticism, Variability, Numerical cognition

Word count: Abstract: 400, Project summary: 968

Lay Perceptions of Scientific Findings: Swayed by the Crowd?

Every day, important scientific findings are rejected at large. From man-made climate change to the safety and efficacy of Covid-19 vaccinations, science skepticism has run rampant among lay consumers in modern society (Hornsey & Fielding, 2017). To increase public faith in science, some have proposed the use of crowd science (Silberzahn et al., 2018; Uhlmann et al., 2019). We explore the effects of scientific findings emerging from a crowd of researchers (vs. a typical research collaboration) on lay perceptions of scientific findings. In line with social norms theory (Miller & Prentice, 2016), we expect that observing consensus among a crowd (the consistent crowd condition) will – compared to the conclusion of a single scientist (the single estimate condition) – increase conformity in opinion. Drawing from work on intuitive statistics (Gigerenzer & Murray, 2015), we also expect laypeople to intuitively accord to the logic of “the wisdom of crowds”: the ability of an aggregate of estimates (rather than a single estimate) to reduce noise stemming from individual bias or error (Schweinsberg et al., 2021).

In contrast, when crowd estimates show low consensus and high variance (the inconsistent crowd condition), we predict that observers will be less swayed and more likely to attribute the findings to bias and error. In addition, due to the difficulty of lay reasoning about variation (Garfield & Ben-Zvi, 2004), we predict an aversion to variability: i.e., we expect that observing variable estimates will decrease lay confidence in the precise average parameter estimate in both crowd conditions. In sum, our preregistered hypotheses are as follows: when laypeople observe multiple consistent (inconsistent) estimates from a crowd of independent scientists, we expect – compared to a single estimate and controlling for prior beliefs – higher (lower) posterior beliefs and credibility of the results, lower confidence in the precise average parameter estimate, and lower (greater) ratings of bias and error.

Methods

We report how we determined our sample size, all data exclusions, all manipulations, and all measures. We sampled UK- and US-based participants from the Prolific participant pool (native English speakers, over 18 years old). There was no prior work to base an expected effect size on; due to our expectation of relatively small effect sizes and monetary constraints, we decided on an upper limit of 1,500 participants after exclusions. After participants report their prior beliefs about a research question (“Do religious people report higher well-being?”), we randomly allocate them to one of three conditions. In the single estimate condition, a single team of six researchers reports a 5% increase in well-being among religious people; in the consistent crowd condition, six independent researchers report six low variance, high consensus estimates (2%, 4%, 5%, 5%, 6%, 8%; $M = 5\%$, $\sigma = 2\%$); in the inconsistent crowd condition, six independent researchers report six high variance, low consensus estimates (-6%, -2%, 5%, 5%, 12%, 16%; $M = 5\%$, $\sigma = 8.25\%$). Afterwards, participants rate – on a slider from 0% to 100% – their posterior beliefs, the credibility of the results, their confidence in the aggregate 5% estimate, and the likelihood of researcher bias, error, and discretion (i.e., idiosyncratic choices). We did not preregister any hypotheses for the last measure; rated discretion is thus an exploratory measure.

Results

We recorded 2,019 responses; after excluding 120 observations from participants who attempted to take the survey more than once, 73 participants who were screened out prior to starting the survey or did not consent, and 328 participants who failed the attention check, our total sample size was 1,498 participants. As preregistered, we regressed all outcomes on prior beliefs and condition (with the single estimate condition as the reference category). Figure 1 displays the main findings. Compared to the single estimate condition, inconsistent crowd estimates decreased

posterior beliefs ($b = -22.80, p < .001$, Cohen's $d = -0.72$), ratings of credibility ($b = -6.54, p < .001, d = -0.21$), and confidence in the parameter estimate ($b = -9.09, p < .001, d = -0.26$), and increased ratings of bias ($b = 6.68, p < .001, d = 0.18$), error ($b = 5.89, p < .001, d = 0.18$), and discretion ($b = 9.22, p < .001, d = 0.29$). Compared to the single estimate condition, consistent crowd estimates decreased posterior beliefs ($b = -5.71, p < .001, d = -0.18$) and increased ratings of error ($b = 3.59, p = .016, d = 0.11$) and discretion ($b = 5.14, p < .001, d = 0.16$); we found no significant effects on ratings of credibility ($b = 1.24, p = .380, d = 0.04$), confidence in the average parameter estimate ($b = 0.49, p = .749, d = 0.01$), or ratings of bias ($b = 2.13, p = .192, d = 0.06$). As hypothesized, inconsistent crowd estimates lead to lower posterior beliefs, credibility of the results, and confidence in the average parameter estimate, and higher ratings of bias and error, than a single estimate. Unexpectedly, consistent crowd estimates lead to lower posterior beliefs and higher ratings of error than a single estimate. However, it is worth noting that there is a positive difference between prior and posterior beliefs within the consistent crowd condition ($M_d = 4.75, p < .001, d = 0.19$; see Figure 2). This finding clarifies that a consistent crowd may sway observers — however, compared to a conventional, non-crowdsourced estimate, crowd estimates do not seem to improve lay perceptions of scientific findings.

Discussion

Although crowd science aims to improve the credibility of scientific research, we find that lay consumers resist the variability and lack of consensus that often reflect its reality. To our surprise, even when results generated by a crowd are highly consistent, we find no improvement in lay perceptions. Further research is needed to consider the possible mechanisms (e.g., variability aversion) behind this preliminary finding. In the future, it is important for crowd scientists to consider how to tackle science skepticism and effectively communicate variable estimates.

References

- Garfield, J., & Ben-Zvi, D. (2004). Research on Statistical Literacy, Reasoning, and Thinking: Issues, Challenges, and Implications. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 397–409). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_17
- Gigerenzer, G., & Murray, D. J. (2015). *Cognition as Intuitive Statistics*. Psychology Press.
Retrieved from <http://books.google.com?id=NLJgCgAAQBAJ>
- Hornsey, M. J., & Fielding, K. S. (2017). Attitude roots and Jiu Jitsu persuasion: Understanding and overcoming the motivated rejection of science. *The American Psychologist*, 72(5), 459–473. <https://doi.org/10.1037/a0040437>
- Miller, D. T., & Prentice, D. A. (2016). Changing Norms to Change Behavior. *Annual Review of Psychology*, 67(1), 339–361. <https://doi.org/10.1146/annurev-psych-010814-015013>
- Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., ... et al. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes*, 165, 228–249.
<https://doi.org/10.1016/j.obhdp.2021.02.003>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... Nosek, B. A. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356. <https://doi.org/10.1177/2515245917747646>

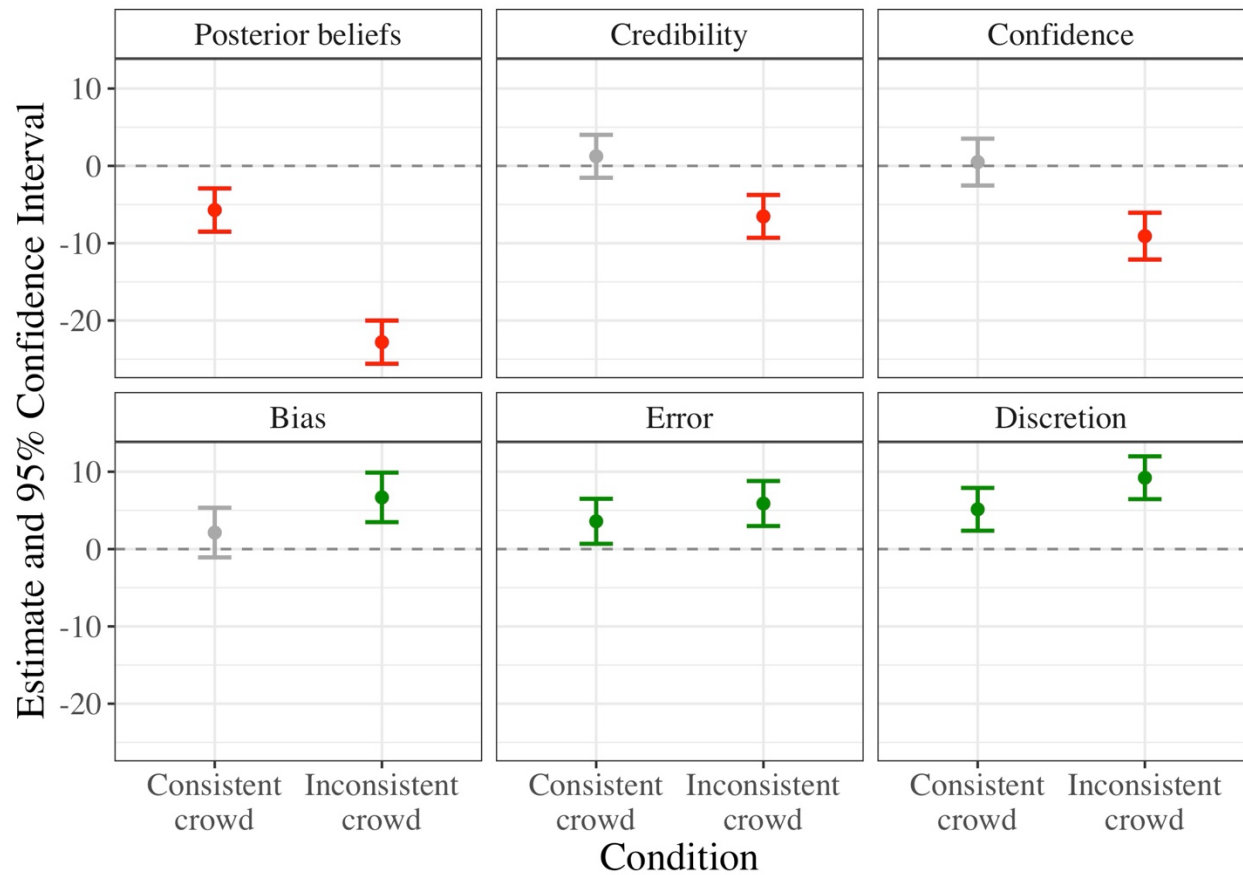
Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., ...

Nosek, B. A. (2019). Scientific Utopia III: Crowdsourcing Science. *Perspectives on Psychological Science*, 14(5), 711–733. <https://doi.org/10.1177/1745691619850561>

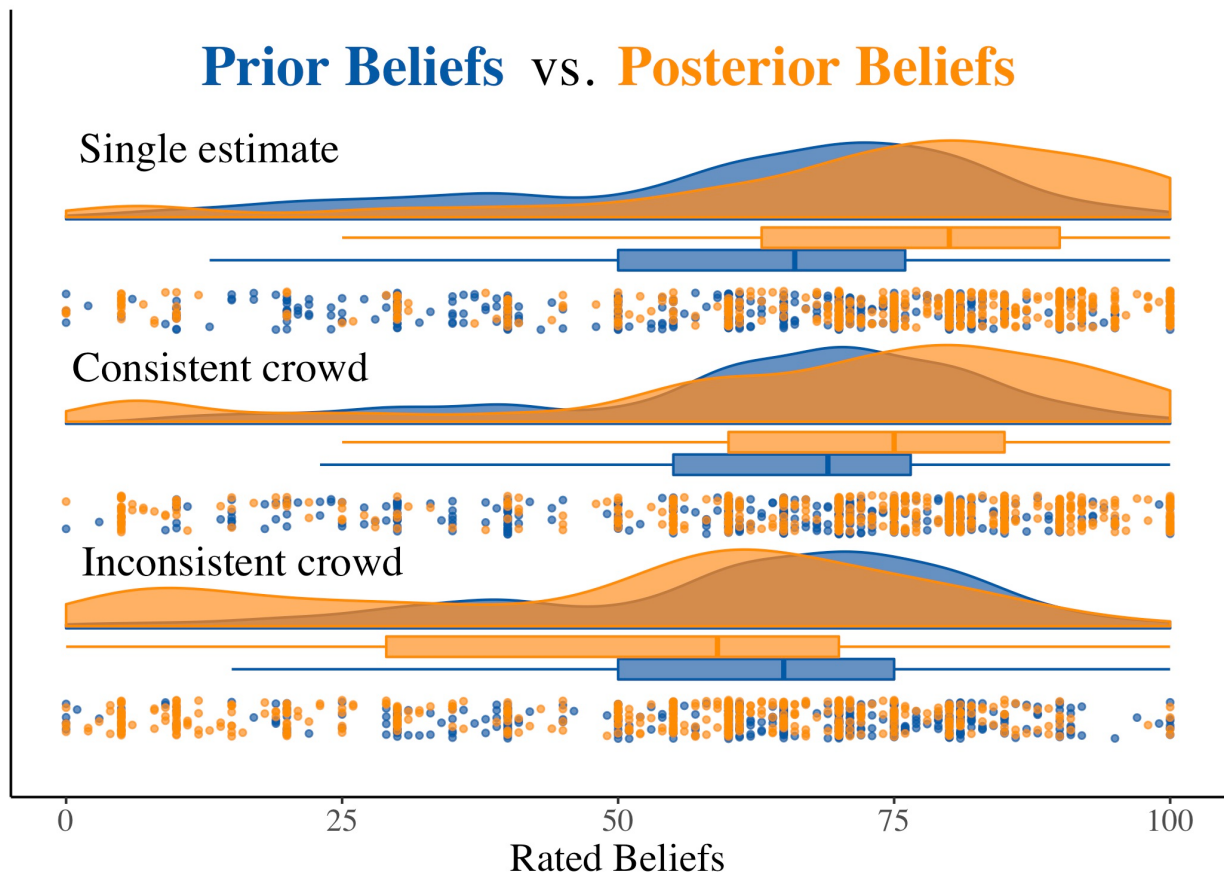
Figures

Figure 1

Estimates of differences with the single estimate condition



Note. Figure 1 displays coefficient estimates (and 95% confidence intervals) of posterior beliefs, credibility, confidence, bias, error, and discretion in the two crowd conditions, compared to the single estimate condition (and controlling for prior beliefs). Green/red/gray error bars indicate positive/negative/insignificant findings, respectively.

Figure 2*Distribution of prior and posterior beliefs by condition*

Note. Prior beliefs are displayed in blue; posterior beliefs are displayed in orange. The respective boxes display the lower quartiles, medians, and upper quartiles of prior and posterior beliefs by condition.