

# Lay Perceptions of Scientific Findings: The Risks of Variability and Lack of Consensus

Word count: 699

## Introduction

Crowd science is on the rise in behavioral science and aims to increase its credibility. Whether it does, however, remains untested: does crowd science improve lay perceptions of scientific findings?

We focus on crowdsourced data analysis: independent analysts estimate a parameter from the same dataset. We compare the effects of providing lay consumers with a single parameter estimate vs. multiple consistent (inconsistent) parameter estimates that show low (high) variance and high (low) consensus.

Table 1 reports our preregistered hypotheses: when laypeople observe multiple consistent (inconsistent) estimates from independent scientists, we expect – compared to one estimate from a single team – an increase (decrease) in posterior beliefs and credibility of the results, and a decrease (increase) in ratings of bias and error. In addition, we expect that the act of providing multiple estimates decreases confidence in the aggregate parameter estimate in both multi-estimate conditions.

## Methods

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-analyst condition, a single team of six researchers reports a 5% increase in well-being among religious people; in the multi-consistent condition, six independent researchers report six low variance, high consensus estimates (2%, 4%, 5%, 5%, 6%, 8%;  $M = 5\%$ ,  $\sigma = 2\%$ ); in the multi-inconsistent 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.

## Results

We sampled 1,498 participants from Prolific (native English speakers from the US and UK, over 18 years old). As preregistered, we regressed all outcomes on prior beliefs and condition (with the single-analyst condition as the reference category). Figure 1 displays the main findings. The multi-inconsistent condition decreased posterior beliefs, ratings of credibility, and confidence in the parameter estimate (standardized effect size  $b = -22.80$ ,  $-6.54$ , and  $-9.09$ , respectively, all  $ps < .001$ ), and increased ratings of bias, error, and discretion ( $b = 6.68$ ,  $5.89$ , and  $9.22$ , respectively, all  $ps < .001$ ). The multi-consistent condition decreased posterior beliefs ( $b = -5.71$ ,  $p < .001$ ) and increased ratings of error ( $b = 3.59$ ,  $p = .016$ ) and discretion ( $b = 5.14$ ,  $p < .001$ ); we found no significant effects on ratings of credibility ( $b = 1.24$ ,  $p = .380$ ), confidence in the parameter estimate ( $b = 0.49$ ,  $p = .749$ ), or ratings of bias ( $b = 2.13$ ,  $p = .192$ ).

As hypothesized, consuming multiple, inconsistent estimates decreases lay posterior beliefs, credibility of the results, confidence in the average parameter estimate, and increases ratings of bias and error. Unexpectedly, consuming multiple, consistent estimates lowers lay posterior beliefs and increases ratings of error. Although the multi-consistent condition performs worse or no better than the single-analyst condition for all outcomes, there is a significant, positive *absolute* effect on posterior beliefs within the multi-consistent condition ( $M_d$

= 4.75,  $p < .001$ ; see Figure 2). This finding clarifies that multi-analyst studies with consistent outcomes can potentially be convincing in their own right – however, relative to conventional, single-analyst research, crowdsourced data analysis does not seem to improve lay perceptions of scientific findings.

## Discussion and Impact

From the proliferation of big team science and large-scale replication initiatives to preregistration and registered reports, several scientific fields have undergone significant reform with the well-intended goal of improving the reliability of scientific research. However, as with any real-world intervention, scientific reform can impose high costs, lack effectiveness, or have unintended consequences.

While partly instituted with the goal of improving the credibility of scientific research, lay consumers appear to resist the inherent variability and lack of consensus that often reflect the reality of multi-analyst research. To our surprise, even when results generated by independent analysts are largely consistent, lay consumers are less likely to believe in the reported phenomenon and more likely to think that the findings stem from error or idiosyncratic choices. Given the resource-intensity of crowdsourced data analysis (Silberzahn & Uhlmann, 2015), our findings suggest that scientists who hope to improve lay perceptions of scientific findings may need to think twice before implementing multi-analyst studies.