

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

Word count: 904

## Introduction

A recent movement towards crowd science has emerged in the behavioral sciences, with the aim of increasing the credibility of scientific research. The extent to which this goal is reached, however, remains untested: does crowd science improve lay perceptions of scientific findings?

We focus on crowdsourced data analysis, also known as the multi-analyst approach: different teams of scientists independently analyze the same dataset to estimate a parameter of interest. We compare the effects of providing lay consumers with a single parameter estimate (the single-analyst condition) vs. multiple parameter estimates that show (a) low variance and high consensus (the “multi-consistent” condition) or (b) high variance and low consensus (the “multi-inconsistent” condition).

Table 1 reports our preregistered hypotheses, which compare the two multi-analyst conditions to the single-analyst condition (while controlling for prior beliefs). When laypeople observe several scientists independently come to consistent (inconsistent) qualitative conclusions, we expect them to have higher (lower) posterior beliefs, find the results more (less) credible, and perceive lower (higher) degrees of bias and error. In addition, we expect that the act of providing multiple (slightly to widely varying) parameter estimates decreases confidence in the exact, aggregate parameter estimate in both multi-analyst conditions.

## Methods

After reporting their prior beliefs about a research question (“Do religious people report higher well-being?”), participants are randomly allocated to one of three experimental conditions in which they learn about the approach and findings of a scientific study. 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 consistent estimates (2%, 4%, 5%, 5%, 6%, and 8%, respectively) that average to 5% and have low variance ( $\sigma^2 = 4$ ); and in the multi-inconsistent condition, six independent researchers report six inconsistent estimates (-6%, -2%, 5%, 5%, 12%, and 16%, respectively) that average to 5% and have high variance ( $\sigma^2 = 68$ ).

Afterwards, participants rate their posterior beliefs, the credibility of the results, their confidence in the effect size estimate, and how likely it is that the estimate was influenced by bias, error, and researcher degrees of freedom. All questions were answered on a slider from 0% to 100%. For all measures, we run linear regression models with condition as the independent variable (with the single-analyst condition as the reference category) and prior beliefs as a covariate. All hypotheses, statistical models, and code were preregistered.

## Results

We sampled 1,498 participants from the Prolific participant pool. We included participants from the United States or the United Kingdom who were over 18 years old and spoke English as a first language. Figure 1 displays the main findings of our linear regression models, which compare each condition to the single-analyst condition and control for prior beliefs. For the multi-inconsistent condition, we found significantly lower posterior beliefs, ratings of credibility, and confidence in the effect size estimate (standardized effect size

estimates  $b = -22.80, -6.54, \text{ and } -9.09$ , respectively, all  $ps < .001$ ), and significantly higher ratings of bias, error, and discretion ( $b = 6.68, 5.89, \text{ and } 9.22$ , respectively, all  $ps < .001$ ). For the multi-consistent condition, we found significantly lower posterior beliefs ( $b = -5.71, p < .001$ ), significantly higher ratings of error ( $b = 3.59, p = .016$ ) and discretion ( $b = 5.14, p < .001$ ), and no significant effects on ratings of credibility ( $b = 1.24, p = .380$ ), confidence in the effect size estimate ( $b = 0.49, p = .749$ ), or ratings of bias ( $b = 2.13, p = .192$ ).

In line with our hypotheses, lay consumers of multi-analyst studies with inconsistent results (compared to single-analyst studies) have lower posterior beliefs, find the results less credible, have less confidence in the average effect size estimate, and believe the results are more likely to stem from bias and error. Contrary to our hypotheses, multi-analyst studies with consistent results perform worse or no better *relative* to single-analyst studies: lay consumers of multi-analyst studies with consistent results (compared to single-analyst studies) report significantly lower posterior beliefs and are more likely to believe the results stem from error. However, it is worth noting (on the basis of Figure 2, which displays the distribution of prior and posterior beliefs across conditions) that, there is a significant, positive *absolute* effect on posterior beliefs within the multi-consistent condition ( $M_d = 4.75, p < .001$ ). This finding clarifies that multi-analyst studies can potentially be convincing in their own right – however, when comparing 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 have unintended consequences and impose high costs. 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 and researcher degrees of freedom. 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.