Predict science to improve science

Description

Tweet

DellaVigna, Stefano, Devin Pope Vivalt, and Eva Vivalt. 2019. â??Predict Science to Improve Scienceâ??. *Science* 366 (6464): 428â??29.

Selected quotes follow:

The limited attention paid to predictions of research results stands in contrast to a vast literature in the social sciences exploring peopleâ??s ability to make predictions in general

We stress three main motivations for a more systematic collection of predictions of research results. ***1. The nature of scientific progress. A new result builds on the consensus, or lack thereof, in an area and is often evaluated for how surprising, or not, it is****. In turn, the novel result will lead to an updating of views. Yet we do not have a systematic procedure to capture the scientific views prior to a study, nor the updating that takes place afterward.*

2. A second benefit of collecting predictions is that they can not only reveal when results are an important departure from expectations of the research community and improve the interpretation of research results, but they can also potentially help to mitigate publication bias. It is not uncommon for research findings to be met by claims that they are not surprising. This may be particularly true when researchers find null results, which are rarely published even when authors have used rigorous methods to answer important questions (15). However, if priors are collected before carrying out a study, the results can be compared to the average expert prediction, rather than to the null hypothesis of no effect. This would allow researchers to confirm that some results were unexpected, potentially making them more interesting and informative because they indicate rejection of a prior held by the research community; this could contribute to alleviating publication bias against null results.

3. A third benefit of collecting predictions systematically is that it makes it possible to improve the accuracy of predictions. In turn, this may help with experimental design. For example, envision a behavioral research team consulted to help a city recruit a more diverse police department. The team has a dozen ideas for reaching out to minority applicants, but the sample size allows for only three treatments to be tested with adequate statistical power. Fortunately, the team has recorded forecasts for several years, keeping track of predictive accuracy, and they have learned that they can combine team membersâ?? predictions, giving more weight to â??superforecastersâ?• (9). Informed by its longitudinal data on forecasts, the team can elicit predictions for each potential project and weed out those interventions judged to have a low chance of success or focus on those interventions with a higher value of information. In addition, the research results of those projects that did go forward would be more impactful if accompanied by predictions that allow better interpretation of results in light of the conventional wisdom.

Rick Davies comment: I have argued, for years, that evaluators should start by eliciting client, and other stakeholders, predictions of outcomes of interest that the evaluation might uncover (e.g. Bangladesh, 2004). But I canâ??t think of any instance where my efforts have been successful, yet. But I have an upcoming opportunity and will try once again, perhaps armed with these two papers.

See also Stefano DellaVigna, and Devin Pope. 2016.*â??Predicting Experimental Results: Who Knows What?â??* NATIONAL BUREAU OF ECONOMIC RESEARCH.

ABSTRACT
Academic experts frequently recommend policies and treatments. But how well do they anticipate the impact of different treatments? And how do their predictions compare to the predictions of non-experts? We analyze how 208 experts forecast the results of 15 treatments involving monetary and non-monetary motivators in a real-effort task. We compare these forecasts to those made by PhD students and non-experts: undergraduates, MBAs, and an online sample. We document seven main results. First, the average forecast of experts predicts quite well the experimental results. Second, there is a strong wisdom-of-crowds effect: the average forecast outperforms 96 per cent of individual forecasts. Third, correlates of expertiseâ??citations, academic rank, field, and contextual experienceâ??do not improve forecasting accuracy. Fourth, experts as a group do better than non-experts, but not if accuracy is defined as rank-ordering treatments. Fifth, measures of effort, confidence, and revealed ability are predictive of forecast accuracy to some extent, especially for non-experts. Sixth, using these measures we identify `superforecastersâ?? among the non-experts who outperform the experts out of sample. Seventh, we document that these results on forecasting accuracy surprise the forecasters themselves. We present a simple model that organizes several of these results and we stress the implications for the collection of forecasts of future experimental results.

See also: The Social Science Prediction Platform, developed by the same authors.

Twitter responses to this post:

Howard White@HowardNWhite Ask decision-makers what they expect research findings to be before you conduct the research to help assess the impact of the research. Thanks to @MandE_NEWS for the pointer. https://socialscienceprediction.org

Marc Winokur@marc_winokur Replying to @HowardNWhite and @MandE_NEWS For our RCT of DR in CO, the child welfare decision makers expected a â??no harmâ?• finding for safety, while other stakeholders expected kids to be less safe. When we found no difference in safety outcomes, but improvements in family engagement, the research impact was more accepted

Category

1. Journal article

Date
19/11/2025
Date Created
27/01/2020
Author
admin