SOCIAL SCIENCE

Predict science to improve science

Systematic collection of predictions of research findings can provide many benefits

By **Stefano DellaVigna[1], Devin Pope[2], Eva Vivalt[3]**

Many fields of research, such as economics, psychology, political science, and medicine, have seen growing interest in new research designs to improve the rigor and credibility of research (e.g., natural experiments, lab experiments, and randomized controlled trials). Interest has similarly grown in efforts to increase transparency, such as preregistration of hypotheses and methods, that seek to allay concerns that improved research designs do not address per se, such as publication bias and p-hacking. Yet, although these efforts improve the informativeness and interpretation of research results, relatively little attention has been paid to another practice that could help to achieve this goal: relating research findings to the views of the scientific community, policy-makers, and the general public. We suggest below three broad ways in which systematic collection of predictions of research results will prove useful: by improving the interpretation of research results, mitigating bias against null results, and improving predictive accuracy and experimental design.

To date, only a relatively small number of studies have collected predictions of research, including recent work predicting original results (*1, 2*) and the replication of academic studies (*3–5*). The limited attention paid to predictions of research results stands in contrast to a vast literature in the social sciences exploring people's ability to make predictions in general (*6–8*), as well as specifically about macroeconomic variables, geopolitical events (*9*), and sporting and political outcomes (*10*), among other variables.

We stress three main motivations for a more systematic collection of predictions of research results. The first ties to the

Tnature 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. What did people predict the study would find? How would knowing this result affect the prediction of findings of future, related studies?

Of course, informally, people routinely evaluate the novelty of scientific results with respect to what is known. However, they typically do so ex post, once the results of the new study are known. Unfortunately, once the results are known,

> ## "…hindsight bias ("I knew that already!") makes it difficult for researchers to truthfully reveal what they thought the results would be."

hindsight bias ("I knew that already!") makes it difficult for researchers to truthfully reveal what they thought the results would be. This stresses the importance of collecting systematic predictions of results ex ante.

For example, consider the debate surrounding the effectiveness of different behavioral factors and nudges to motivate a behavior. Would a gift be more, or less, effective than a modest monetary incentive? To answer this and related questions, an experiment tested how 18 different behavioral and incentive treatments (e.g., gifts, social norms, financial incentives) would motivate participants in a simple task. Notably, the researchers obtained ex ante predictions of the results for each treatment from academic experts and other forecasters such as college students (*11, 12*).

On average, the experts were highly accurate. Furthermore, the rich data set allowed the authors to explore various features of the predictions, such as the strength of the "wisdom of the crowd" phenomenon and the relative accuracy of forecasters with different types of expertise. For example, in this context, highly cited

faculty performed no better than other faculty, and Ph.D. students did best.

Another study provides an example of how predictions can be used to examine and improve belief updating based on research results, in this case in a policy setting. A group of policy-makers made predictions on the effects of conditional cash transfer programs and school meals programs (*13*). Views of policy-makers are of particular interest both because they propose and oversee interventions and because they are the people who would presumably learn from, and use, the results. Policy-makers were found to be more optimistic, but less certain, than researchers and practitioners. Further, policy-makers, practitioners, and researchers all were found to update more on "good" surprising news than on "bad" news and to not respond very differently to results with large confidence intervals as compared to results with small confidence intervals, though there is some evidence that updating can be improved by presenting results differently (*14*).

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.

A third benefit of collecting predictions systematically is that it makes it possible

[1]Department of Economics, University of California, Berkeley, Berkeley, CA, USA. [2]The University of Chicago Booth School of Business, Chicago, IL, USA. [3]Research School of Economics, Australian National University, Canberra, Australia. Email: eva.vivalt@anu.edu.au

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 members' predictions, giving more weight to "superforecasters" (*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.

These three broad uses of predictions highlight two important implications. First, it will be important to collect forecast data systematically to draw general lessons. For example, when do senior researchers make more accurate forecasts than junior researchers, given that the seniors' expertise did not help in the study and forecasts of task performance incentives (*12*)? Under what conditions do policy-makers update in a Bayesian manner from past evidence (*13*)? We will need predictions for a range of settings, including longitudinal predictions by the same forecasters over time, to identify possible superforecasters and to examine whether providing feedback on past forecasts helps improve prediction accuracy.

Second, like preanalysis plans, it is critical to set up the collection of predictions before the results are known, to avoid the impact of hindsight bias. With these features in mind, a centralized platform that collects forecasts of future research results can play an important role. Toward this end, in coordination with the Berkeley Initiative for Transparency in the Social Sciences (BITSS), we have developed an online platform for collecting forecasts of social science research results (www.socialscienceprediction.org). The platform will make it possible to track multiple forecasts for an individual across a variety of interventions, and thus to study determinants of forecast accuracy, such as characteristics of forecasters or interventions, and to identify superforecasters (see the box).

A centralized platform has another advantage. As collecting forecasts grows in popularity, a small number of researchers may receive a disproportionate number of requests. A centralized platform can ensure that this does not happen, analogous to how an editor keeps track of referee requests within a journal, except that a centralized platform could be even better, as editors cannot track referee requests across journals. As a further benefit, the platform provides third-party certification about how forecasts were collected and shared with researchers requesting them (analogous to platforms used for preregistration).

This platform would aim to incorporate lessons learned from other work on forecasts, such as work on replication of experiments in psychology and economics, prediction of geopolitical events in the Good Judgment Project, and forecasts of

macroeconomic indicators in the Survey of Professional Forecasters. The Systematizing Confidence in Open Research and Evidence (SCORE) program is also aiming to develop tools specifically to predict the replicability or reproducibility of social-behavioral science results.

There are many open questions about the details of the platform. For example, should forecasters be paid for participating (just like some journals choose to pay referees)? Should there be incentives for accuracy? We expect that continued work and experimentation will provide more clarity regarding such design questions.

Although here we focus on the benefits of ex ante predictions for improving the interpretation of research results, these predictions have many other potential uses in research and policy. Some researchers may use predictions to explore when forecasts can be trusted or how the accuracy of forecasts can be improved. Others will focus on Bayesian interpretations or learning about belief updating. The forecasts may also have a practical value to policy-makers needing to make a decision in the absence of credible evidence from an academic study. Such a variety of potential uses speaks to the value of making this tool available. ∎

REFERENCES AND NOTES

1. A. Delavande, X. Giné, D. McKenzie, *J. Dev. Econ.* **94**, 151 (2011).
2. A. Cohn, M. A. Maréchal, D. Tannenbaum, C. L. Zünd, *Science* **365**, 70 (2019).
3. A. Dreber *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15343 (2015).
4. C. F. Camerer *et al.*, *Science* **351**, 1433 (2016).
5. D. Benjamin, D. R. Mandel, J. Kimmelman, *PLOS Biol.* **15**, e2002212 (2017).
6. F. Galton, *Nature* **75**, 450 (1907).
7. D. Kahneman, A. Tversky, *Psychol. Rev.* **80**, 237 (1973).
8. R. M. Dawes, D. Faust, P. E. Meehl, *Science* **243**, 1668 (1989).
9. P. Tetlock, D. Gardner, *Superforecasting: The Art and Science of Prediction* (Random House, 2016).
10. J. Wolfers, E. Zitzewitz, *J. Econ. Perspect.* **18**, 107 (2004).
11. S. DellaVigna, D. Pope, *Rev. Econ. Stud.* **85**, 1029 (2018).
12. S. DellaVigna, D. Pope, *J. Polit. Econ.* **126**, 2410 (2018).
13. E. Vivalt, A. Coville, "How Do Policymakers Update?" Working paper (2019).
14. E. Vivalt, A. Coville, "The Implications of Variance Neglect for the Formation and Estimation of Subjective Expectations." Working paper (2019).
15. A. Franco, N. Malhotra, G. Simonovits, *Science* **345**, 1502 (2014).

Science

Predict science to improve science

Stefano DellaVignaDevin PopeEva Vivalt

View the article online
Permissions
https://www.science.org/help/reprints-and-permissions

Use of think article is subject to the Terms of service