Markets for replication

Alec Brandon and John A. List[1]

Department of Economics, University of Chicago, Chicago, IL 60637

Since 1955, *The Journal of Irreproducible Results* has offered a satirical view on academic research, publishing pieces such as, "A double blind efficacy trial of placebos, extra strength placebos, and generic placebos" (1). Scientists are now taking a less satirical look at the reproducibility of results, with questions emerging about the robustness of findings in fields as diverse as epidemiology, genetics, neuroscience, and the social sciences (2–4). The aftermath has seen myriad efforts to close the gap in reproducibility.

In PNAS, Dreber et al. (5) take the innovative approach of considering whether markets can play a crucial role. The Dreber et al. study focuses on the replicability of recent publications in top psychology journals. For each study they created a market wherein contracts were traded that pay out \$1 if the study was successfully replicated and \$0 if it was not. Traders in the market were professors and graduate students of psychology and each trader was endowed with \$100. Trading occurred over a 2-wk period, after which a replication of the study was conducted.

In a very promising set of results, Dreber et al. (5) report that markets capture the dynamics of replicable science quite well. For example, performing the simple exercise of using market prices above (below) \$0.50 as an indicator that the study is replicable (not likely replicable) successfully predicted 71% of replications. Furthermore, beyond the simple \$0.50 dividing line, the market price was strongly correlated with the success of replications, and it outperforms the predictive ability of a survey conducted with traders before actual trading started. In this way, the market aggregated individual information in an efficient manner.

The miraculous ability of markets to coordinate preferences is well known to many experimental economists. In fact, following in the footsteps of Harvard economist Edward Chamberlin, Vernon Smith's experimental studies on laboratory markets (6) provided early important glimpses of the powers of markets. Smith's study, and subsequent research, act as a proof-of-concept of sorts, with the experimenter inducing the demand (supply) curve for each buyer (seller) in the market and then observing whether prices and quantities differ from the intersection of market supply and demand. The main finding is that economic theory predicts outcomes stunningly well in this setting, with samples ranging from children to college students to amateur pin collectors across very different market types (6, 7).

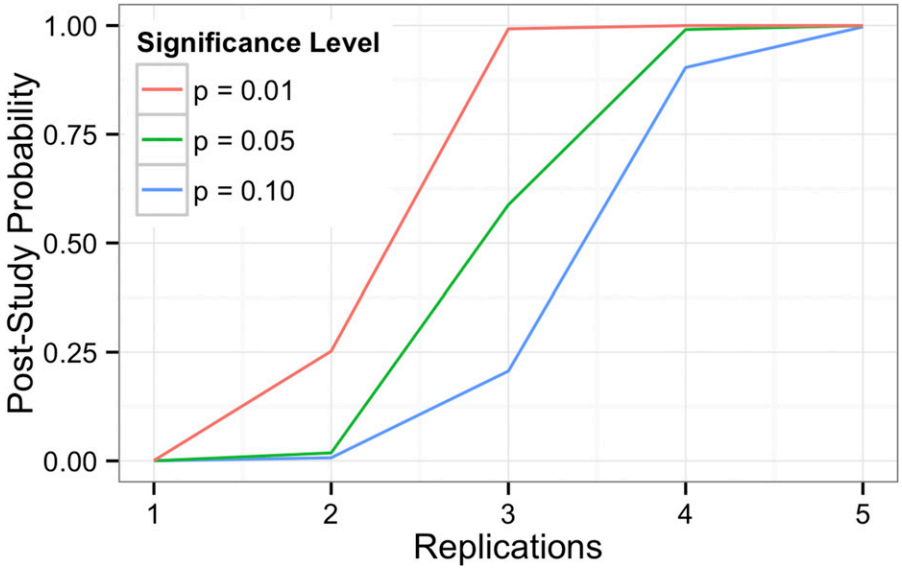

Fig. 1. Probability of a true scientific finding as a function of number of successful replications. Poststudy probability as the number of successful replications given a fixed number of replications. Power is assumed to be 0.8. Significance is 0.1, 0.05, or 0.01, depending on the line plotted.

Recent Trends in Increasing Replicability

The crux of just about every empirical study is the *P* value. Researchers pose a null hypothesis meant to capture the status quo line of thinking. Data are then analyzed and if the *P* value is small enough (typically less than 0.05) then the researcher rejects the null hypothesis and an alternative hypothesis is evident. However, the mechanics of the inference problem call into question this simple approach. As Maniadis et al. (8) show, inference not only relies on reported *P* values, but also priors and the power of the test.

In this spirit, the power of replications, as initially proposed within the experimental tripod of Fisher (9), is evident. Fig. 1 illustrates the importance of replications by plotting the poststudy probability (i.e., the probability that the proposed relationship is true) as a function of the number of successful replications (2, 8). Each line plotted in Fig. 1 assumes the prior of a skeptic (e.g., very low prior that the relationship is a true relationship) for different levels of statistical significance. The story that emerges is simple: if we want to identify true findings then we need replications. Even in those cases when we allow only a small false-positive rate (a *P* value of 0.01), we need three successful replications before we can be very confident that the observed relationship is a true relationship. Furthermore, as we allow a larger false-positive rate, more replications are necessary.

Can Markets Incentivize Replications?

How can the scientific community increase the supply of replications? Dedicating existing journal space or creating new journals is one option. For example, the Economic Science Association—the main experimental economics association—has started a new journal that lists publishing replications as one of its objectives. Another idea is to

Author contributions: A.B. and J.A.L. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 15343.

[1]To whom correspondence should be addressed. Email: jlist@uchicago.edu.

change the incentives for the authors whose work is being replicated (10). Although these ideas can work on the margin, unless there is a sweeping change in academic culture, the returns to publishing original work will always dominate work on replications.

Prediction markets, such as those used in Dreber et al. (5), offer a different type of incentive for replications: financial returns. Imagine a market wherein academics can trade on the outcome of replications and a small cut on transactions funds the work of actually conducting the replications. Such a market may suffer the liquidity problems that have doomed other prediction markets (e.g., Intrade), but in light of the ideas currently on the table, this one is worth strong consideration.

1 Stanfield J (2005) A double blind efficacy trial of placebos, extra strength placebos, and generic placebos. *The Journal of Irreproducible Results* 49(5):13.

2 Ioannidis JPA (2005) Why most published research findings are false. *PLoS Med* 2(8):e124.

3 Bennett CM, Baird AA, Miller MB, Wolford GL (2010) Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: An argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results* 1(1):1–5.

4 Open Science Collaboration (2015) PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* 349(6251):aac4716.

5 Dreber A, et al. (2015) Using prediction markets to estimate the reproducibility of scientific research. *Proc Natl Acad Sci USA* 112:15343–15347.

6 Smith VL (1962) An experimental study of competitive market behavior. *J Polit Econ* 70(2):111–137.

7 List JA (2004) Testing neoclassical competitive theory in multilateral decentralized markets. *Journal Policy Economy* 112(5):1131–1156.

8 Maniadis Z, Tufano F, List JA (2014) One swallow doesn't make a summer: How economists (mis-) use experimental methods and their results. *Am Econ Rev* 104(1):277–290.

9 Fisher RA (1935) *The Design of Experiments* (Macmillan, London).

10 Maniadis Z, Tufano F, List JA (2015) How to make experimental economics research more reproducible: Lessons from other disciplines and a new proposal. *Replication in Experimental Economics*, eds Deck CA, Fatas E, Rosenblat T (Emerald Group Pulishing, Bingley, UK), pp 215–230.