

# Using prediction markets to estimate the reproducibility of scientific research

Anna Dreber<sup>a,1,2</sup>, Thomas Pfeiffer<sup>b,c,1</sup>, Johan Almenberg<sup>d</sup>, Siri Isaksson<sup>a</sup>, Brad Wilson<sup>e</sup>, Yiling Chen<sup>f</sup>, Brian A. Nosek<sup>g,h</sup>, and Magnus Johannesson<sup>a</sup>

<sup>a</sup>Department of Economics, Stockholm School of Economics, SE-113 83 Stockholm, Sweden; <sup>b</sup>New Zealand Institute for Advanced Study, Massey University, Auckland 0745, New Zealand; <sup>c</sup>Wissenschaftskolleg zu Berlin—Institute for Advanced Study, D-14193 Berlin, Germany; <sup>d</sup>Sveriges Riksbank, SE-103 37 Stockholm, Sweden; <sup>e</sup>Consensus Point, Nashville, TN 37203; <sup>f</sup>John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; <sup>g</sup>Department of Psychology, University of Virginia, Charlottesville, VA 22904; and <sup>h</sup>Center for Open Science, Charlottesville, VA 22903

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved October 6, 2015 (received for review August 17, 2015)

**Concerns about a lack of reproducibility of statistically significant results have recently been raised in many fields, and it has been argued that this lack comes at substantial economic costs. We here report the results from prediction markets set up to quantify the reproducibility of 44 studies published in prominent psychology journals and replicated in the Reproducibility Project: Psychology. The prediction markets predict the outcomes of the replications well and outperform a survey of market participants' individual forecasts. This shows that prediction markets are a promising tool for assessing the reproducibility of published scientific results. The prediction markets also allow us to estimate probabilities for the hypotheses being true at different testing stages, which provides valuable information regarding the temporal dynamics of scientific discovery. We find that the hypotheses being tested in psychology typically have low prior probabilities of being true (median, 9%) and that a "statistically significant" finding needs to be confirmed in a well-powered replication to have a high probability of being true. We argue that prediction markets could be used to obtain speedy information about reproducibility at low cost and could potentially even be used to determine which studies to replicate to optimally allocate limited resources into replications.**

reproducibility | replications | prediction markets

The process of scientific discovery centers on empirical testing of research hypotheses. A standard tool to interpret results in statistical hypothesis testing is the  $P$  value. A result associated with a  $P$  value below a predefined significance level (typically 0.05) is considered "statistically significant" and interpreted as evidence in favor of a hypothesis. However, concerns about the reproducibility of statistically significant results have recently been raised in many fields including medicine (1–3), neuroscience (4), genetics (5, 6), psychology (7–11), and economics (12, 13). For example, an industrial laboratory could only reproduce 6 out of 53 key findings from "landmark" studies in preclinical oncology (2) and it has been argued that the costs associated with irreproducible preclinical research alone are about US\$28 billion a year in the United States (3). The mismatch between the interpretation of statistically significant findings and a lack of reproducibility threatens to undermine the validity of statistical hypothesis testing as it is currently practiced in many research fields (14).

The problem with inference based on  $P$  values is that a  $P$  value provides only partial information about the probability of a tested hypothesis being true (14, 15). This probability also depends on the statistical power to detect a true positive effect and the prior probability that the hypothesis is true (14). Lower statistical power increases the probability that a statistically significant effect is a false positive (4, 14). Statistically significant results from small studies are therefore more likely to be false positives than statistically significant results from large studies. A lower prior probability for a hypothesis to be true similarly increases the probability that a statistically significant effect is a false positive

(14). This problem is exacerbated by publication bias in favor of speculative findings and against null results (4, 16–19).

Apart from rigorous replication of published studies, which is often perceived as unattractive and therefore rarely done, there are no formal mechanisms to identify irreproducible findings. Thus, it is typically left to the judgment of individual researchers to assess the credibility of published results. Prediction markets are a promising tool to fill this gap, because they can aggregate private information on reproducibility, and can generate and disseminate a consensus among market participants. Although prediction markets have been argued to be a potentially important tool for assessing scientific hypotheses (20–22)—most notably in Robin Hanson's paper "Could Gambling Save Science? Encouraging an Honest Consensus" (20)—relatively little has been done to develop potential applications (21). Meanwhile, the potential of prediction markets has been demonstrated in a number of other domains, such as sports, entertainment, and politics (23–26).

We tested the potential of using prediction markets to estimate reproducibility in conjunction with the Reproducibility Project: Psychology (RPP) (9, 10). The RPP systematically replicated studies from a sampling frame of three top journals in psychology. To investigate the performance of prediction markets in this context, a first set of prediction markets were implemented in November 2012 and included 23 replication studies scheduled to be completed in the subsequent 2 mo, and a second set of prediction markets were implemented in October 2014 and included

## Significance

**There is increasing concern about the reproducibility of scientific research. For example, the costs associated with irreproducible preclinical research alone have recently been estimated at US\$28 billion a year in the United States. However, there are currently no mechanisms in place to quickly identify findings that are unlikely to replicate. We show that prediction markets are well suited to bridge this gap. Prediction markets set up to estimate the reproducibility of 44 studies published in prominent psychology journals and replicated in The Reproducibility Project: Psychology predict the outcomes of the replications well and outperform a survey of individual forecasts.**

Author contributions: A.D., T.P., J.A., B.A.N., and M.J. designed research; A.D., T.P., J.A., S.I., B.W., Y.C., B.A.N., and M.J. performed research; A.D., T.P., J.A., and M.J. analyzed data; A.D., T.P., J.A., and M.J. wrote the paper.

Conflict of interest statement: Consensus Point employs B.W. and provided the online market interface used in the experiment. The market interface is commercial software.

This article is a PNAS Direct Submission.

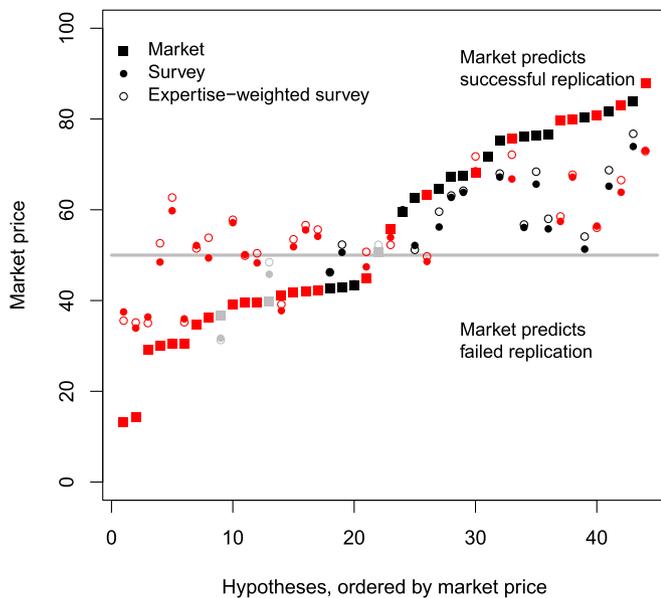
Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Open Science Framework database, <https://osf.io/yjmht>.

<sup>1</sup>A.D. and T.P. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: [anna.dreber@hhs.se](mailto:anna.dreber@hhs.se).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516179112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516179112/-DCSupplemental).



**Fig. 1.** Prediction market performance. Final market prices and survey predictions are shown for the replication of 44 publications from three top psychology journals. The prediction market predicts 29 out of 41 replications correctly, yielding better predictions than a survey carried out before the trading started. Successful replications (16 of 41 replications) are shown in black, and failed replications (25 of 41) are shown in red. Gray symbols are replications that remained unfinished (3 of 44).

21 replication studies scheduled to be completed before the end of December 2014. The prediction markets were active for 2 wk at each of these occasions.

For each of the replication studies, participants could bet on whether or not the key original result would be replicated. Our criterion for a successful replication was a replication result, with a  $P$  value of less than 0.05, in the same direction as the original result. In one of the studies, the original result was a negative finding, and successful replication was thus defined as obtaining a negative (i.e., statistically nonsignificant) result in the replication. Information on the original study and the setup of the replication were accessible to all participants.

In the prediction markets, participants traded contracts that pay \$1 if the study is replicated and \$0 otherwise. This type of contract allows the price to be interpreted as the predicted probability of the outcome occurring. This interpretation of the price is not without caveats (27) but has an advantage of being simple and reasonably robust (28), especially in settings where traders' initial endowments are the same and traders' bets are relatively small. Invitations to participate in the prediction markets were sent to the email list of the Open Science Framework, and for the second set of markets also to the email list of the RPP collaboration. Participants were not allowed to bet in those markets where they were involved in carrying out the replication. In the first set of prediction markets, 49 individuals signed up and 47 of these actively participated; in the second set, 52 individuals signed up and 45 of these actively participated. Before the markets started, participants were asked in a survey for their subjective probability of each study being replicated. Each participant was endowed with US\$100 for trading.

## Results

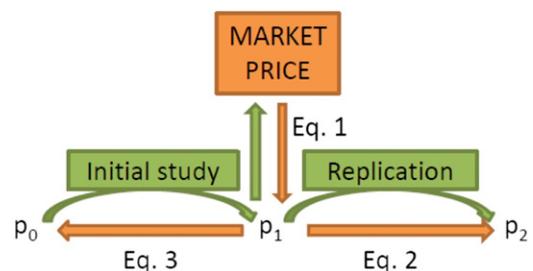
The prediction markets functioned well in an operational sense. Participation was broad, i.e., trading was not dominated by a small subset of traders or concentrated to just a few of the markets. In total, 2,496 transactions were carried out. The number of

transactions per market ranged from 28 to 108 (mean, 56.7), and the number of active traders per market ranged from 18 to 40 (mean, 26.7). We did not detect any market bias regarding bets on success ("long positions") or failure ("short positions") to replicate the original results. In the final portfolios held at market closing time (*Supporting Information*), we observed approximately the same number of bets on success and failure.

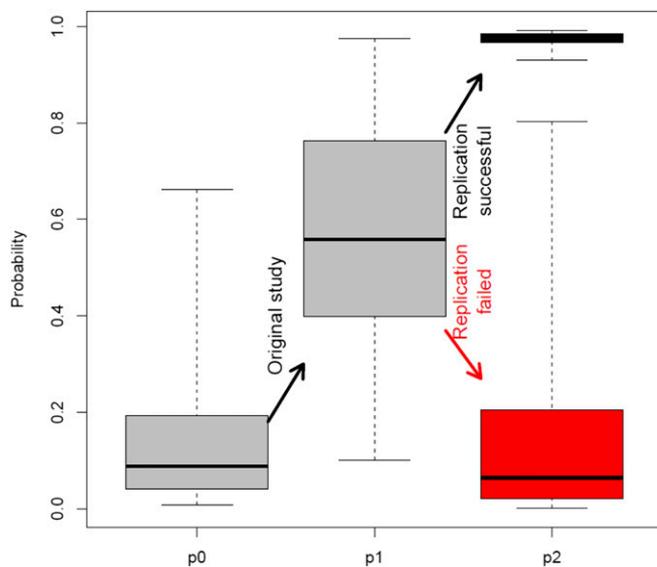
The mean prediction market final price is 55% (range, 13–88%), implying that about half of the 44 studies were expected to replicate. Out of the 44 scientific studies included in the prediction markets, the replications were completed for 41 of the studies, with the remaining replications being delayed. Of the 41 completed, 16 studies (39%) replicated and 25 studies (61%) did not replicate according to the market criterion for a successful replication (*Supporting Information*).

We evaluate the performance of the markets in three ways. We test whether the market prices are informative; if the market prices can be interpreted as probabilities of replication; and if the prediction markets predict the replication outcomes better than a survey measure of beliefs. When interpreting a market price larger than 50% as predicting successful replication and a market price smaller than 50% as predicting failed replication, informative markets are expected to correctly predict more than 50% of the replications. We find that the prediction markets correctly predict the outcome of 71% of the replications (29 of 41 studies; Fig. 1), which is significantly higher than 50% (one-sample binomial test;  $P = 0.012$ ).

Interpreting the prediction market prices as probabilities means that not all markets with a price larger (smaller) than 50% are expected to correspond to successful (failed) replications. The expected prediction rate of the markets depends on the distribution of final market prices, which in our study implies that 69% of the outcomes are expected to be predicted correctly. This is very close to the observed value of 71%. To formally test whether prediction market prices can be interpreted as probabilities of replication, we estimated a linear probability model (with robust SEs) with the outcome of the replication as a function of the prediction market price. If market prices equal replication probabilities, the coefficient of the market price variable should be equal to 1 and the constant in the regression should be equal to zero. The coefficient of the market price variable is 0.995, which is significantly different from zero ( $P = 0.003$ ), but not significantly different from 1 ( $P = 0.987$ ). The constant ( $-0.167$ ) is not significantly different from zero ( $t = -1.11$ ,  $P = 0.276$ ).



**Fig. 2.** Relationship between market price and prior and posterior probabilities  $p_0$ ,  $p_1$ , and  $p_2$  of the hypothesis under investigation. Bayesian inference (green arrows) assigns an initial (prior) probability  $p_0$  to a hypothesis, indicating its plausibility in absence of a direct test. Results from an initial study allows this prior probability to be updated to posterior  $p_1$ , which in turn determines the chances for the initial result to hold up in a replication, and thus the market price in the prediction market. Once the replication has been performed, the result can be used to generate posterior  $p_2$ . Observing the market price, and using the statistical characteristics of the initial study and the replication, we can thus reconstruct probabilities  $p_1$ ,  $p_2$ , and  $p_0$ . Detailed calculations are presented in *Supporting Information*.



**Fig. 3.** Probability of a hypothesis being true at three different stages of testing: before the initial study ( $p_0$ ), after the initial study but before the replication ( $p_1$ ), and after replication ( $p_2$ ). “Error bars” (or whiskers) represent range, boxes are first to third quartiles, and thick lines are medians. Initially, priors of the tested hypothesis are relatively low, with a median of 8.8% (range, 0.7–66%). A positive result in an initial publication then moves the prior into a broad range of intermediate levels, with a median of 56% (range, 10–97%). If replicated successfully, the probability moves further up, with a median of 98% (range, 93.0–99.2%). If the replication fails, the probability moves back to a range close to the initial prior, with a median of 6.3% (range, 0.01–80%).

The prediction market can also be compared with the pretrading survey of participants’ beliefs about the probability of replication. A simple average of the survey correctly predicts 58% of outcomes (23 of 40; Fig. 1; survey data are missing for one market), which is not significantly different from 50% (one-sample binomial test;  $P = 0.429$ ). A weighted average, using self-reported expertise as weights, correctly predicts 50% (20 of 40) of outcomes, which is not significantly different from 50% (one-sample binomial test;  $P = 1.00$ ). The absolute prediction error is significantly lower for the prediction market than for both the pretrading survey (paired  $t$  test,  $n = 40$ ,  $t = -2.558$ ,  $P = 0.015$ ) and the weighted survey (paired  $t$  test,  $n = 40$ ,  $t = -2.727$ ,  $P = 0.010$ ; see [Supporting Information](#) for a more detailed comparison of the prediction market and survey responses). The prediction market thus outperforms the survey measure of beliefs.

The above results suggest that the prediction markets generate good estimates of the probability that a published result will be replicated. Note that the probability of successful replication is not the same thing as the probability of a tested hypothesis being true. The probability of a tested hypothesis being true, also referred to as the positive predictive value or PPV (4), can however be estimated from the market price (Fig. 2). Using information about the power and significance levels of the original study and the replications (see [Supporting Information](#) for details), it can be estimated for three stages of the testing process: the prior probability ( $p_0$ ) before observing the outcome of the initial study; the probability after observing the result of the initially published study ( $p_1$ ); and the probability after observing the outcome of the replication ( $p_2$ ). A summary of the results of these estimations are shown in Fig. 3; a more detailed breakdown is given in [Supporting Information](#).

Our analysis reveals priors ( $p_0$ ) for the 44 studies ranging from 0.7% to 66% with a median (mean) of 8.8% (13%). This relatively low average prior may reflect that top psychology journals

focus on publishing surprising findings, i.e., positive findings on relatively unlikely hypotheses. The probability that the research hypothesis is true after observing the positive finding in the first study ( $p_1$ ) ranges from 10% to 97% with a median (mean) of 56% (57%) for the 44 studies. This estimate implies that about 43% of statistically significant research findings published in these top psychology journals can be expected to be false positives.

For the 41 studies replicated so far, we can also estimate the posterior probability that the research finding is true contingent on observing the result of the replication ( $p_2$ ). This probability ranges between 93.0% and 99.2% with a median (mean) of 98% (97%) for the 16 studies whose result was replicated, and between 0.1% and 80% with a median (mean) of 6.3% (15%) for the 25 studies that were not replicated.

These results show that prediction markets can give valuable insights into the dynamics of information accumulation in a research field. Eliciting priors in this manner allows us to evaluate whether hypotheses are tested appropriately in a given research field. A common, but incorrect, interpretation of a published result with a  $P < 0.05$  is that it implies a 95% probability of the research hypothesis being true. Interestingly, our findings imply that to achieve such a high probability of the research hypothesis being true, a “statistically significant” positive finding needs to be confirmed in a well-powered replication. This illustrates the importance of replicating positive research findings before they are given high credibility. It remains to be studied how psychology compares in this aspect to other fields.

## Discussion

The RPP project recently found that more than one-half of 100 original findings published in top psychology journals failed to replicate (10). Our prediction market results suggest that this relatively low rate of reproducibility should not come as a surprise to the profession, as it is consistent with the beliefs held by psychologists participating in our prediction market.

As can be seen in Fig. 1, original findings for which the market prices indicated a low probability of replication were indeed typically not replicated. However, there were also some findings that failed to replicate despite high market prices that indicated that participants had less doubts about those findings. An interesting hypothesis is that in some of these cases it was the replication itself, rather than the original finding, that failed. It would thus be particularly interesting to carry out additional replications of these studies.

Although our results suggest that prediction markets can be used to obtain accurate forecasts regarding the outcome of replications, one limitation of the approach we used in this study lays in the necessity to run replications so that there is an outcome to trade on. Some studies such as large field experiments may be very costly to replicate (29). One way to mitigate this would be to run prediction markets on a number of studies, from which a subset is randomly selected for replication after the market closes (20). Such an approach could provide quick information about reproducibility at low cost. Moreover, prediction markets could potentially be used as so-called “decision markets” (30, 31) to prioritize replication of some studies, such as those with the lowest likelihood of replication. This would generate salient and informative signals about reproducibility, and help optimizing the allocation of resources into replication.

## Materials and Methods

The RPP by the Open Science Collaboration (10) sampled papers in the 2008 issues of three top psychology journals: *Journal of Personality and Social Psychology*, *Psychological Science*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*. In the case of several studies in one paper, typically the last study of each paper was selected for replication.

We chose 23 studies for the first set of prediction markets and 21 studies for the second set of prediction markets, where the chosen studies were

scheduled to be replicated within 2 mo after the completion of the prediction market. For each replication, the hypothesis of the original study was summarized by one of the authors of this paper and submitted to the replication team for comments and final approval. In 1 of the 23 studies in the first prediction market, the chosen experiment was changed by the replicating researcher after the survey had been performed but before the trading started (SI ref. 34 in [Supporting Information](#)); we thus lack survey data for this study. One of the 21 studies in the second prediction market was later changed for a different experiment to be replicated (SI ref. 59 in [Supporting Information](#)), but for completeness we still include the prediction market and survey data for this study (although there are no current plans to replicate this study).

Participants in the prediction market were researchers in various fields of psychology, ranging from graduate students to professors. Fourteen participants were directly involved in one or several replication studies (15 studies in total) and were not allowed to make trades on the outcomes of these specific studies. Sixteen participants participated in both sets of prediction markets. Before the prediction market, the participants filled out a survey. For each study, participants were asked two questions. One was meant to capture their beliefs of reproducibility: "How likely do you think it is that this hypothesis will be replicated (on a scale from 0% to 100%)?" Participants were also asked about their expertise in the area: "How well do you know this topic? (not at all, slightly, moderately, very well, extremely well)." We transformed this latter measure into a 1–5 scale, and it was used to construct the weighted average belief measure from the survey.

Trading in the prediction market took place through a web-based market interface in collaboration with Consensus Point ([www.consensuspoint.com/](http://www.consensuspoint.com/)), a leading provider of prediction market research technology. Before starting to trade, participants received information about the trading procedure as well as logins. Trading accounts were initially endowed with \$100 (expressed as 10,000 "points"). These points were used to make predictions of successful replication. Predictions were made by buying and selling stocks on the hypotheses on an interface that highlighted the forecasting functionality of the market ([Supporting Information](#)). In the prediction market, participants traded contracts that pay \$1 (i.e., 100 points) if the study is replicated and \$0 otherwise. This type of contract allows the price to be interpreted as the predicted probability of the outcome occurring. For each hypothesis, participants could see the current market prediction for the probability of successful replication.

The trading platform used an automated market maker implementing a logarithmic market scoring rule (32). This algorithm offers a buying price and

a selling price at all times, ensuring that there is always a counterpart with which to trade. More specifically, the algorithm uses the net sales ( $s$ ) the market maker has done so far in a market to determine the prices for a (infinitesimally small) trade as  $P = \exp(s/b)/(\exp(s/b) + 1)$ . To buy stocks, participants chose YES on the trading interface and entered how many points they would like to invest. For each additional point invested in a YES position, the price (and the predicted probability for successful replication) increased. To sell stocks, participants chose NO on the trading interface and entered how many points they would like to invest. For each additional point invested in a NO position, the price decreased. Participants could also buy (sell) shares by increasing (decreasing) an existing YES position, or decreasing (increasing) an existing NO position. The market maker ensures that the value of a YES share is \$1 minus the value of a NO share. Parameter  $b$  determines the liquidity and the maximal subsidies provided by the market maker and controls how strongly the market price is affected by a trade. We set the liquidity parameter to  $b = 100$  (points). This means that, by investing 1,000 points (i.e., 1/10 of the initial endowment), traders can move the price of a single market from 50% to about 55%; and investing the entire initial endowment into a single market moves the price from 50% to 82%.

For the first set of prediction markets, investments were settled 5 mo after the market had closed according to actual results of the replications in the cases where the outcome was available and to market value in the cases where the replications were not yet finished. At the time of the close of the market, only eight results were known by the replicating researcher, where all replicating researchers had agreed to not share the results with anyone until after the market closed. For the second set of prediction markets, investments were similarly settled 4.5 mo after the markets had closed. At the time of the close of the second market, one result was known by the replicating researcher; all replicating researchers agreed here too to not share their results with anyone until the market had closed.

**ACKNOWLEDGMENTS.** We thank Agneta Berge for research assistance; and Juergen Huber, Willemien Kets, and Pranjal Mehta for comments on a previous version of the manuscript. We thank the Jan Wallander and Tom Hedelius Foundation (P2012-0002:1, P2013-0156:1, and P2015-0001:1), the Knut and Alice Wallenberg Foundation [Wallenberg Academy Fellows Grant (to A.D.)], the Swedish Foundation for Humanities and Social Sciences (NHS 14-1719:1), and the National Science Foundation (Grant CCF-0953516) for financial support.

- Prinz F, Schlange T, Asadullah K (2011) Believe it or not: How much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 10(9):712.
- Begley CG, Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. *Nature* 483(7391):531–533.
- Freedman LP, Cockburn IM, Simcoe TS (2015) The economics of reproducibility in preclinical research. *PLoS Biol* 13(6):e1002165.
- Button KS, et al. (2013) Power failure: Why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14(5):365–376.
- Hewitt JK (2012) Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behav Genet* 42(1):1–2.
- Lawrence MS, et al. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–218.
- Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 22(11):1359–1366.
- Carpenter S (2012) Psychology research. Psychology's bold initiative. *Science* 335(6076):1558–1561.
- Open Science Collaboration (2012) An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspect Psychol Sci* 7(6):657–660.
- Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science* 349(6251):aac4716.
- Bohannon J (2014) Psychology. Replication effort provokes praise—and "bullying" charges. *Science* 344(6186):788–789.
- Ioannidis J, Doucouliagos CJ (2013) What's to know about the credibility of empirical economics? *J Econ Surv* 27(5):997–1004.
- Maniatis 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.
- Ioannidis JPA (2005) Why most published research findings are false. *PLoS Med* 2(8):e124.
- Nuzzo R (2014) Scientific method: Statistical errors. *Nature* 506(7487):150–152.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR (1991) Publication bias in clinical research. *Lancet* 337(8746):867–872.
- Stern JM, Simes RI (1997) Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 315(7109):640–645.
- Ioannidis JPA, Munafò MR, Fusar-Poli P, Nosek BA, David SP (2014) Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends Cogn Sci* 18(5):235–241.
- Miguel E, et al. (2014) Social science. Promoting transparency in social science research. *Science* 343(6166):30–31.
- Hanson RD (1995) Could gambling save science? Encouraging an honest consensus. *Soc Epistemology* 9(1):3–33.
- Almenberg J, Kittlitz K, Pfeiffer T (2009) An experiment on prediction markets in science. *PLoS One* 4(12):e8500.
- Park I-U, Peacey MW, Munafò MR (2014) Modelling the effects of subjective and objective decision making in scientific peer review. *Nature* 506(7486):93–96.
- Wolfers J, Zitzewitz E (2004) Prediction markets. *J Econ Perspect* 18(2):107–126.
- Tziralis G, Tsioupolous I (2007) Prediction markets: An extended literature review. *J Pred Markets* 1(1):75–91.
- Arrow KJ, et al. (2008) Economics. The promise of prediction markets. *Science* 320(5878):877–878.
- Berg J, Forsythe R, Nelson F, Rietz T (2008) Results from a dozen years of election futures markets research. *Handbook of Experimental Economics Results*, eds Plott CR, Smith VL (North-Holland, Amsterdam), pp 742–751.
- Manski CF (2006) Interpreting the predictions of prediction markets. *Econ Lett* 91(3):425–429.
- Wolfers J, Zitzewitz E (2006) *Interpreting Prediction Market Prices as Probabilities* (National Bureau of Economic Research, Cambridge, MA), NBER Working Paper No. 12200.
- Manning WG, et al. (1987) Health insurance and the demand for medical care: Evidence from a randomized experiment. *Am Econ Rev* 77(3):251–277.
- Hanson R (2006) Decision markets for policy advice. *Promoting the General Welfare: American Democracy and the Political Economy of Government Performance*, eds Patashnik EM, Gerber AS (Brookings Institution Press, Washington, DC), pp 151–173.
- Chen Y, Kash IA, Ruberry M, Shnayder V (2014) Eliciting predictions and recommendations for decision making. *ACM Trans Econ Comput* 2(2):6:1–6:27.
- Hanson R (2007) Logarithmic market scoring rules for modular combinatorial information aggregation. *J Pred Markets* 1(1):3–15.
- Richeson JA, Trawalter S (2008) The threat of appearing prejudiced and race-based attentional biases. *Psychol Sci* 19(2):98–102.

34. Reynolds M, Besner D (2008) Contextual effects on reading aloud: Evidence for pathway control. *J Exp Psychol Learn Mem Cogn* 34(1):50–64.
35. Rule NO, Ambady N (2008) The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychol Sci* 19(2):109–111.
36. Morris AL, Still ML (2008) Now you see it, now you don't: Repetition blindness for nonwords. *J Exp Psychol Learn Mem Cogn* 34(1):146–166.
37. Shnabel N, Nadler A (2008) A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *J Pers Soc Psychol* 94(1):116–132.
38. Correll J (2008) If noise and effort on implicit measures of bias. *J Pers Soc Psychol* 94(1):48–59.
39. Fischer P, Greitemeyer T, Frey D (2008) Self-regulation and selective exposure: The impact of depleted self-regulation resources on confirmatory information processing. *J Pers Soc Psychol* 94(3):382–395.
40. Alter AL, Oppenheimer DM (2008) Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle). *Psychol Sci* 19(2):161–167.
41. Mirman D, Magnuson JS (2008) Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *J Exp Psychol Learn Mem Cogn* 34(1):65–79.
42. Estes Z, Verges M, Barsalou LW (2008) Head up, foot down: Object words orient attention to the objects' typical location. *Psychol Sci* 19(2):93–97.
43. Nairne JS, Pandeirada JNS, Thompson SR (2008) Adaptive memory: The comparative value of survival processing. *Psychol Sci* 19(2):176–180.
44. White PA (2008) Accounting for occurrences: A new view of the use of contingency information in causal judgment. *J Exp Psychol Learn Mem Cogn* 34(1):204–218.
45. Pacton S, Perruchet P (2008) An attention-based associative account of adjacent and nonadjacent dependency learning. *J Exp Psychol Learn Mem Cogn* 34(1):80–96.
46. Pleskac TJ (2008) Decision making and learning while taking sequential risks. *J Exp Psychol Learn Mem Cogn* 34(1):167–185.
47. Masicampo EJ, Baumeister RF (2008) Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychol Sci* 19(3):255–260.
48. Janssen N, Schirm W, Mahon BZ, Caramazza A (2008) Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *J Exp Psychol Learn Mem Cogn* 34(1):249–256.
49. Henderson MD, de Liver Y, Gollwitzer PM (2008) The effects of an implemental mindset on attitude strength. *J Pers Soc Psychol* 94(3):396–411.
50. Turk-Browne NB, Isola PJ, Scholl BJ, Treat TA (2008) Multidimensional visual statistical learning. *J Exp Psychol Learn Mem Cogn* 34(2):399–407.
51. Vul E, Nieuwenstein M, Kanwisher N (2008) Temporal selection is suppressed, delayed, and diffused during the attentional blink. *Psychol Sci* 19(1):55–61.
52. Vohs KD, Schooler JW (2008) The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychol Sci* 19(1):49–54.
53. Fischer P, Schulz-Hardt S, Frey D (2008) Selective exposure and information quantity: How different information quantities moderate decision makers' preference for consistent and inconsistent information. *J Pers Soc Psychol* 94(2):231–244.
54. Bressan P, Stranieri D (2008) The best men are (not always) already taken: Female preference for single versus attached males depends on conception risk. *Psychol Sci* 19(2):145–151.
55. Sahakyan L, Delaney PF, Waldum ER (2008) Intentional forgetting is easier after two "shots" than one. *J Exp Psychol Learn Mem Cogn* 34(2):408–414.
56. Lobue V, DeLoache JS (2008) Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychol Sci* 19(3):284–289.
57. Nurmsoo E, Bloom P (2008) Preschoolers' perspective taking in word learning: Do they blindly follow eye gaze? *Psychol Sci* 19(3):211–215.
58. Wolf ST, Insko CA, Kirchner JL, Wildschut T (2008) Interindividual-intergroup discontinuity in the domain of correspondent outcomes: The roles of relativistic concern, perceived categorization, and the doctrine of mutual assured destruction. *J Pers Soc Psychol* 94(3):479–494.
59. Förster J, Liberman N, Kuschel S (2008) The effect of global versus local processing styles on assimilation versus contrast in social judgment. *J Pers Soc Psychol* 94(4):579–599.
60. Lau GP, Kay AC, Spencer SJ (2008) Loving those who justify inequality: The effects of system threat on attraction to women who embody benevolent sexist ideals. *Psychol Sci* 19(1):20–21.
61. Stanovich KE, West RF (2008) On the relative independence of thinking biases and cognitive ability. *J Pers Soc Psychol* 94(4):672–695.
62. Lemay EP, Jr, Clark MS (2008) "Walking on eggshells": How expressing relationship insecurities perpetuates them. *J Pers Soc Psychol* 95(2):420–441.
63. Stinson DA, et al. (2008) The cost of lower self-esteem: Testing a self- and social-bonds model of health. *J Pers Soc Psychol* 94(3):412–428.
64. McCreary SM (2008) Self-handicapping, excuse making, and counterfactual thinking: Consequences for self-esteem and future motivation. *J Pers Soc Psychol* 95(2):274–292.
65. Halevy N, Bornstein G, Sagiv L (2008) "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm. *Psychol Sci* 19(4):405–411.
66. Yap MJ, Balota DA, Tse CS, Besner D (2008) On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *J Exp Psychol Learn Mem Cogn* 34(3):495–513.
67. Koo M, Fishbach A (2008) Dynamics of self-regulation: How (un)accomplished goal actions affect motivation. *J Pers Soc Psychol* 94(2):183–195.
68. Schmidt JR, Besner D (2008) The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *J Exp Psychol Learn Mem Cogn* 34(3):514–523.
69. Dodson CS, Darragh J, Williams A (2008) Stereotypes and retrieval-provoked illusory source recollections. *J Exp Psychol Learn Mem Cogn* 34(3):460–477.
70. Fiedler K (2008) The ultimate sampling dilemma in experience-based decision making. *J Exp Psychol Learn Mem Cogn* 34(1):186–203.
71. Berry CJ, Shanks DR, Henson RNA (2008) A single-system account of the relationship between priming, recognition, and fluency. *J Exp Psychol Learn Mem Cogn* 34(1):97–111.
72. Tamir M, Mitchell C, Gross JJ (2008) Hedonic and instrumental motives in anger regulation. *Psychol Sci* 19(4):324–328.
73. Liefvooghe B, Barrouillet P, Vandierendonck A, Camos V (2008) Working memory costs of task switching. *J Exp Psychol Learn Mem Cogn* 34(3):478–494.
74. Farris C, Treat TA, Viken RJ, McFall RM (2008) Perceptual mechanisms that characterize gender differences in decoding women's sexual intent. *Psychol Sci* 19(4):348–354.
75. Purdie-Vaughns V, Steele CM, Davies PG, Dittmann R, Crosby JR (2008) Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *J Pers Soc Psychol* 94(4):615–630.
76. Bassok M, Pedigo SF, Oskarsson AT (2008) Priming addition facts with semantic relations. *J Exp Psychol Learn Mem Cogn* 34(2):343–352.