Field experiments of success-breeds-success dynamics

Arnout van de Rijt[a,b,1], Soong Moon Kang[c], Michael Restivo[d], and Akshay Patil[e]

Departments of [a]Sociology and [e]Computer Science, State University of New York, Stony Brook, NY 11794; [b]Institute for Advanced Computational Science, Stony Brook, NY 11794; [c]Department of Management Science and Innovation, University College London, London WC1E 6BT, United Kingdom; and [d]Department of Sociology, State University of New York, Geneseo, NY 14454

Edited by Karen S. Cook, Stanford University, Stanford, CA, and approved March 28, 2014 (received for review September 10, 2013)

Seemingly similar individuals often experience drastically different success trajectories, with some repeatedly failing and others consistently succeeding. One explanation is preexisting variability along unobserved fitness dimensions that is revealed gradually through differential achievement. Alternatively, positive feedback operating on arbitrary initial advantages may increasingly set apart winners from losers, producing runaway inequality. To identify social feedback in human reward systems, we conducted randomized experiments by intervening in live social environments across the domains of funding, status, endorsement, and reputation. In each system we consistently found that early success bestowed upon arbitrarily selected recipients produced significant improvements in subsequent rates of success compared with the control group of nonrecipients. However, success exhibited decreasing marginal returns, with larger initial advantages failing to produce much further differentiation. These findings suggest a lesser degree of vulnerability of reward systems to incidental or fabricated advantages and a more modest role for cumulative advantage in the explanation of social inequality than previously thought.

Matthew effect | preferential attachment | scale-free networks | rich-get-richer effects | power law

Social scientists have long debated why we often see similar persons experience diverging trajectories of accomplishment, with some accumulating long strings of successes and others failing repeatedly. One explanation is that subtle variation along hard-to-observe dimensions of ability equips individuals with unequal a priori chances that gradually are revealed through differential achievement (1–5). A competing hypothesis states that "success breeds success" (4, 6–12). This hypothesis claims that the ultimate success of select persons may be born out of small, random initial advantages that grow ever larger through runaway positive feedback. Such cumulative advantage has been argued to produce significant, and arbitrary, inequality in many domains of human achievement (12–15). These two theoretical positions on the origins of societal inequities regularly meet in academic and public debate about whether great success is an accurate indicator of great talent (1–3, 5, 16–19).

Determining the origins of success in empirical studies is made difficult by the confounding of exogenous factors with endogenous processes. For instance, although some scholars have taken the extreme variance of success distributions as a tell-tale sign of cumulative advantage (8, 11, 16, 20), critics have pointed out that various other generative mechanisms, such as the existence of a convex correspondence between fitness and success (21, 22), can generate the same empirical regularities (17, 23–28). Further, in longitudinal records of success, unobserved dimensions of fitness generate apparent bias toward past winners (3, 4, 12, 15). In these cases, the higher success rates of talented, privileged, and well-connected individuals give rise to temporal correlations between successes, which may be erroneously interpreted as a causal effect of past on future success.

This problem of empirical confounding may be overcome through randomized experiments. Prior studies have used experimental methods to identify positive social feedback (13, 29–33). While these studies confirm the operation of reinforcement processes, they provide limited insight into the degree to which these processes distort the allocation of resources to individuals in various reward systems. First, the success-breeds-success hypothesis covers a much wider variety of types of success than previous experiments have investigated. In this paper we evaluate the presence of cumulative advantage by consistently applying the same experimental intervention across a diverse range of reward systems. The systems we study vary in the degree to which the rewards transferred carry immediate monetary value, affect the social status of recipients, or are of entirely ideological nature. Second, the degree to which cumulative advantage can disrupt meritocracies depends critically on whether greater initial advantages breed proportionately greater amounts of subsequent success. In our experiments we systematically vary the magnitude of the initial advantage to quantify the marginal effects on the size of the ultimate success gap.

We constructed an experimental design in which we explicitly control the allocation of success (*Materials and Methods*). In this setup, we bestow early successes upon randomly selected members of a population, thereby ensuring that the expectations of success before intervention are equal for recipients and nonrecipients. To allow a robust test of cumulative advantage in multiple contexts, we deployed this design in four naturally occurring systems, representing distinct forms of personal success—financial gain, endorsement, social status, and social support. First, in the financial domain, we applied the design to the crowdfunding website kickstarter.com, where creators of projects in the areas of technology, arts, and entertainment compete for donations from the general public. We sampled 200 new, unfunded projects and donated a percentage of the funding goal to 100

Significance

Social scientists have long debated why similar individuals often experience drastically different degrees of success. Some scholars have suggested such inequality merely reflects hard-to-observe personal differences in ability. Others have proposed that one fortunate success may trigger another, thus producing arbitrary differentiation. We conducted randomized experiments through intervention in live social systems to test for success-breeds-success dynamics. Results show that different kinds of success (money, quality ratings, awards, and endorsements) when bestowed upon arbitrarily selected recipients all produced significant improvements in subsequent rates of success as compared with the control group of nonrecipients. However, greater amounts of initial success failed to produce much greater subsequent success, suggesting limits to the distortionary effects of social feedback.

Author contributions: A.v.d.R., S.M.K., M.R., and A.P. designed research; A.v.d.R., S.M.K., M.R., and A.P. performed research; A.v.d.R., S.M.K., M.R., and A.P. analyzed data; and A.v.d.R. and S.M.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: arnout.vanderijt@stonybrook.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1316836111/-/DCSupplemental.

randomly chosen projects. Second, on the website epinions.com reviewers are paid for posting written evaluations of new products, and those evaluations subsequently are rated by website visitors as "very helpful," "helpful," "somewhat helpful," or "not helpful." Reviewers are paid more for evaluations that are considered more helpful. We sampled 305 new, unrated reviews that we evaluated as being very helpful and gave a random subset of these reviews a "very helpful" rating. Our third application involved the encyclopedia website wikipedia.org, where highly productive editors receive status awards from community members in recognition of their dedication (34). We sampled 521 editors who belonged to the top 1% of most productive editors and conferred an award to a randomly chosen subset of these editors. Fourth, on the petition website change.org individuals seek support from the general public for social and political goals through signature campaigns that can be signed electronically by any named or anonymous supporter. We sampled 200 early-stage campaigns and granted a dozen signatures to 100 randomly chosen petitions. In each experiment, we kept a daily record of subsequent donations, ratings, awards, and signatures given by third parties after the treatment in both the experimental and control condition. These four interventions thus represent a range of types of success, covering resource transfers in which both source and recipients are financially affected (kickstarter.com), transfers in which the recipient benefits materially without the source incurring a cost (epinions.com), conferrals of social status (wikipedia.com), and expressions of ideological support (change.org).

Results

To isolate the effect of our experimentally induced success on the rate of success accumulation in each study, net of any exacerbating or counteracting second-order effects that successive successes may have had on one another, we first calculated separately for both the experimental and control conditions the proportion of individuals who experienced at least one more success during the observation period. In all four domains, the experimental treatment produced significant increases in rewards for the treated individuals. Fig. 1 shows that in each study the artificial contribution of success had a positive effect on the rate of success. In the control condition of the crowd-funding study, 39% of project initiators received subsequent funding by one or more donors. In contrast, 70% of the individuals in the experimental condition received contributions from third parties, indicating that the mere presence of an initial donation made recipients about twice as likely to attract funding ($\chi^2 = 19.4$; $P = 0.000$). In the endorsement study, the baseline likelihood of success was much higher than in the other studies, with 77% of untreated product reviews receiving at least one "very helpful" rating during the 14 d immediately following the treatment. This percentage rose further to 90% in the treatment condition ($\chi^2 = 9.54$; $P = 0.002$), and this increase did not come at the expense of a parallel increase in less favorable ratings (*SI Results*). In the control condition of the Wikipedia study, 31% of the editors received a status award during the observation period. In comparison, 40% of editors who received their first award through our experiment received one or more other awards from fellow editors ($\chi^2 = 4.72$; $P = 0.030$). Finally, in the signature study, 52% of the individuals in the control condition received at least one more signature toward their petition goal during our observation period, whereas 66% of petitioners in the experimental condition subsequently accumulated additional signatures ($\chi^2 = 4.05$; $P = 0.044$). The effect of signatures contributed through our experiment on petitioners' yield of subsequent signatures suggests that social reinforcement effects are operative even for expressions of ideological support (35). Consistent with earlier results, we find a causal link between past and future rewards in four distinct substantive domains, providing robust evidence for positive feedback operating on arbitrary early advantages in the allocation of resources to individuals.

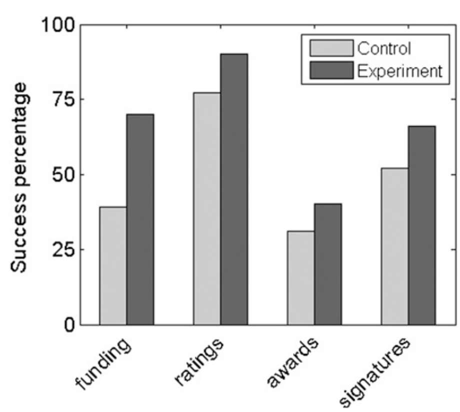

Fig. 1. Percentage of cases with posttreatment success. From left to right: percentage of crowd-funding project creators who collected subsequent funding; percentage of reviewers who subsequently received positive ratings; percentage of Wikipedia editors who subsequently received awards from third parties; and percentage of petitioners whose petitions were subsequently signed by others. The difference between conditions at the end of the observation period is statistically significant for funding ($\chi^2 = 19.4$; $P = 0.000$), ratings ($\chi^2 = 9.54$; $P = 0.002$), awards ($\chi^2 = 4.72$; $P = 0.030$), and signatures ($\chi^2 = 4.05$; $P = 0.044$).

Cumulative Advantage Dynamics. To assess whether the effect of our treatment was only transient or instead had an enduring impact on success accumulation, we calculated in each system the average number of posttreatment successes accumulated as a function of time (Fig. 2). All the posttreatment measures of success shown exclude the success applied through the treatment. In each study the arbitrary gap in subsequent success between the recipients and nonrecipients of early success persisted throughout the study. In the funding study, our donation increased the average number of subsequent donations from 1.11 in the control condition to 2.49 in the experimental condition. This difference between conditions is statistically highly significant (signed-rank test; $z = 3.95$; $P = 0.000$). In the endorsement study, 14 d after our ratings were applied, the number of subsequent positive ratings given by third parties still differed significantly, with a total of 11.4 in the control condition and 14.9 in the experimental condition (rank-sum test; $z = 3.213$; $P = 0.001$). In the awards study, 1 mo after our intervention, editors in the control condition had accumulated noticeably fewer awards on their user pages by fellow editors than editors in the experimental condition ($z = 2.635$; $P = 0.008$), and this difference still remained noticeable after 3 mo ($z = 1.982$; $P = 0.048$), at 0.17 and 0.28 awards per person, respectively. Finally, at 2 wk after intervention in the signature study, Fig. 2 shows a small gap remaining between the posttreatment signature yields of campaigns in the control condition, which had accumulated another 1.74 signatures on average, and those in our experimental condition, which had recruited an average of 2.32 additional signatures ($z = 1.759$; $P = 0.079$). In combination, these findings indicate that, despite qualitative differences in the nature of success across the four reward systems, an early advantage consistently drives a sustained difference between individuals with equal initial likelihood of success.

Marginal Returns of Success. Although the impact of the initial advantage in each study demonstrates the susceptibility of these reward systems to arbitrary and self-reinforcing differentiation between ex ante equivalent individuals, it tells us little about the extent to which inequalities can be affected. If more sizeable

SOCIAL SCIENCES

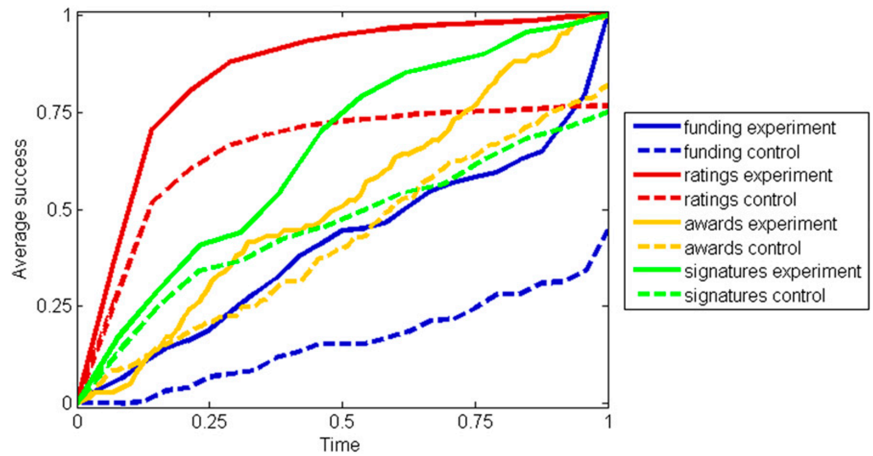

Fig. 2. The success-breeds-success effect over time. The curves represent running numbers of donations (blue), positive ratings (red), awards (yellow), and campaign signatures (green) in the experimental condition (solid lines) and the control condition (dashed lines). The horizontal axis is normalized so that 0 marks the time of experimental intervention, and 1 marks the end of the observation period. The vertical axis is normalized so that for each system a value of 1 equals the maximum across time and conditions.

initial differences were introduced between individuals, how much more severely would the subsequent allocation of resources be impacted?

To test the effects of larger initial endowments on cumulative returns, we subsequently varied the strength of the treatment in both the funding study and the rating study. In the funding study we included funding goals of up to \$5,000 and withheld a donation, donated 1% of the funding goal through one donor, or donated 4% of the funding goal through four separate donors. By holding the per-donor contribution level constant across treatment conditions, we neutralized any social influence effects that the size of the average prior contribution may exert on followers. In the rating study we again sampled previously unrated reviews, and when we found them to be of high quality, we left them unrated, rated them as "very helpful" once, or rated them as "very helpful" four times by four different raters.

Among subjects in the zero-donor condition, 32% attracted subsequent funding from one or more donors, whereas 74% of the subjects in the one-donor condition and 87% of the subjects in the four-donors condition collected subsequent funds. The difference between the one-donor condition and the control condition is statistically significant ($\chi^2 = 11.0$; $P = 0.001$), as is the difference between the four-donors condition and the control condition ($\chi^2 = 19.4$; $P = 0.000$). However, the increase in the size of the initial advantage as represented by the difference between the one-donor and four-donors conditions did not result in a significantly higher chance of one or more donations ($\chi^2 = 1.65$; $P = 0.199$). In the endorsements study, reviewers who wrote high-quality reviews but received no positive rating from us exhibited a 77% chance of receiving one or more positive ratings, compared with 90% of reviewers who received one positive review from us and 94% of reviewers who received four positive reviews from us. Again, the treatment effects are positive in both experimental conditions ($\chi^2 = 9.54$; $P = 0.002$ and $\chi^2 = 9.38$; $P = 0.002$) but do not differ from one another ($\chi^2 = 0.926$; $P = 0.336$). Together, these patterns of one-by-one comparisons between conditions suggest decreasing marginal returns: Each additional unit increase in input yields a progressively smaller increase in output. Indeed, in each experiment an increase in input from zero to one produces a significant increase in per-unit output, whereas the additional increase in input from one to four never yields a noticeable increase in per-unit output.

To quantify these marginal returns, we calculated average posttreatment success as a function of the number of successes applied through treatment, shown in Fig. 3. Fig. 3*A* displays the average total dollar amount raised by the number of donations bestowed. Fig. 3*B* displays the average number of donations accumulated by the number of donations made. Fig. 3*C* displays the number of positive ratings received by the number of positive ratings experimentally bestowed. The averages reported in each panel exclude the dollars, donations, and ratings applied through our experimental intervention. Consistently across all panels, the average marginal returns of an increase from zero to one exceed the average marginal returns of an increase from one to four. The average return on a single donation (on average \$24.52) is \$191.00, but the additional three donations are estimated to bring in only \$89.57 each (Fig. 3*A*). Accordingly, the difference in the amount of dollars raised between the zero-donations and one-donation conditions is significant (signed-rank test; $z = 3.02$; $P = 0.003$), and so is the difference between the zero-donations and the four-donations conditions ($z = 3.61$; $P = 0.000$), but the one-donation and four-donations conditions do not deviate significantly ($z = 1.70$; $P = 0.090$). A single donation raises the number of subsequent third-party donations by 4.3, whereas each of the additional three donations brings in only 1.7 subsequent third-party donations (Fig. 3*B*). Indeed, the difference in the number of donations elicited in the zero-donations and one-donation conditions is significant ($z = 3.20$; $P = 0.001$), as is the difference between the zero-donations and four-donations conditions ($z = 4.16$; $P = 0.000$), whereas the difference between the one-donation and four-donations conditions falls just short of statistical significance ($z = 1.95$; $P = 0.051$). Finally, a single "very helpful" rating given to a product reviewer increases the number of subsequent third-party "very helpful" ratings by 3.48, but awarding an additional three positive ratings does not appear to increase the expected number of "very helpful" ratings further, as indicated by a slightly negative marginal effect of −0.43 (Fig. 3*C*). The difference between the zero-ratings and one-rating conditions is significant (rank-sum test; $z = 3.21$; $P = 0.001$), but the four-ratings condition differs from neither the zero-ratings condition ($z = 1.83$; $P = 0.067$) nor the one-rating condition ($z = 1.07$; $P = 0.288$).

Discussion

Our findings reveal the presence of a noticeable feedback effect in each of the distinct settings that we investigated, in that initial arbitrary endowments create lasting disparities in individual success. These results suggest that the inadvertent magnification

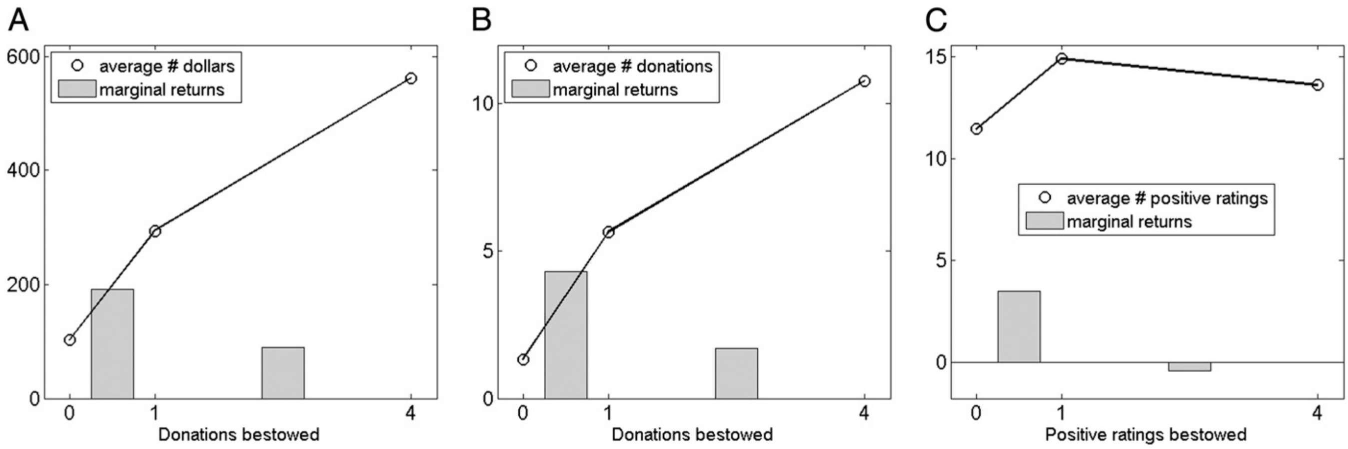

Fig. 3. Marginal returns of success. The horizontal axis measures the number of donations (in *A* and *B*) or ratings (in *C*) applied through experimental intervention, namely, none, one, or four. The circles measure the average dollar amount (*A*), number of donations (*B*), and positive ratings (*C*) obtained in each condition, excluding the treatment. Shaded bars measure the marginal returns, which are calculated as the slopes of the lines connecting the averages. In each panel, marginal returns decrease with the size of the treatment.

of arbitrary differences between individuals of comparable merit may be a common feature of many types of social reward systems. At the same time, our experimental demonstration of decreasing marginal returns to success may suggest bounds to the degree to which the natural allocation of resources can be disrupted by social feedback effects. Without a priori differentiation in quality or structural sources of advantage, cumulative advantage alone may not be able to generate the extreme kinds of runaway inequality that so commonly are attributed to it (4–16, 29, 30). The vulnerability of meritocracies to biases from success-breeds-success effects thus may be more limited than generally assumed.

The deliberate allocation of success in our experiments demonstrates that cascades of positive reinforcement can be initiated intentionally by a strategic actor. This form of purposive action presents the possibility of perverse effects, such as profit-seeking entities offering loans, positive reviews, and endorsements in exchange for the pecuniary equivalent of the anticipated ripple effect. It also raises the possibility of a philanthropic entity jumpstarting support through first-mover loans to underappreciated projects as a social policy instrument for counteracting nonmeritocratic disparities in populations. However, these opportunities for manipulation are offset by the decreasing marginal returns of success identified in our study, which suggests limits on the scale of such purposive intervention. Both the crowd-funding study and the rating study suggest that the reinforcement value of a single initial success is much larger than that of additional successes. The per-donor effect of a single donation by a single donor on fundraising success was greater than that of four donations by four separate donors. Similarly, a single positive rating of an unrated review increased the number of positive ratings by more than the increase resulting from each of the four positive ratings by four separate raters. Strategic contributions aimed at steering dynamics in a more positive direction thus may be less effective when made to campaigns that already have garnered some minimal degree of support. Hence, the susceptibility of reward systems to deliberate manipulation may be restricted mostly to interventions favoring those individuals who cannot muster any initial success otherwise.

Materials and Methods

The four studies (ID numbers 373335, 366647, 230771, and 442574) were approved by the Stony Brook University Human Subjects Committee and were conducted in compliance with the terms of use of kickstarter.com, epinions.com, wikipedia.org, and change.org.

Study 1. Kickstarter.com. ***Site.*** Kickstarter.com is a crowd-funding website launched in April 2009 for projects in diverse categories ranging from music, film, and video games to innovative products. The Kickstarter platform facilitates the gathering of monetary donations from the general public. Project creators choose a deadline and set a goal of raising a minimum amount of funds. If the chosen goal is not reached by the deadline, no funds are collected, and all donations are returned to the donors, or "backers." Kickstarter takes 5% of the funds raised. Payments are made through the online retailer Amazon.com, which charges an additional 3–5%. As of January 13, 2014, 128,887 projects had been launched, with $943 million raised (www.kickstarter.com/help/stats). For our experiments, we sampled from projects created by United States residents only.

Design and procedure. Fundraising goals in Kickstarter projects range from less than $100 to more than $1 million. To mitigate negative effects of this large variance on the statistical power of our study, we limited our sampling frame to projects with a low funding goal ($1,000 in round 1 and $5,000 in round 2) and matched projects across conditions according to their goal amount. To allow for comparative temporal analysis, we selected only projects that had to be funded within 28–30 d. Projects exhibit idiosyncratic funding behavior during their first 5–6 d, with some projects suddenly spiking in funding activity because of differentiated campaign efforts by project creators on social networking sites such as Facebook.com. To avoid this variability in the data, we selected only projects that had not yet been funded at 24 d before the end of the funding period. When a project fulfilling the selection criteria became available, we randomly assigned it to the experimental or the control condition. The next project with a similar goal amount that became available was assigned to the alternate condition. Because our donations were relatively small, the kickstarter.com website did not display the projects we invested in at a more prominent location than those in the control condition; thus we avoided a treatment effect caused by a difference in the visibility of the matched projects. We kept a daily record of donations until the funding deadline was reached. We collected a total of 293 projects in two rounds. Data collection took place between July 2012 and February 2014 for round 1 and between September 2013 and March 2014 for round 2.

Round 1. The treatment in round 1 of data collection involved the donation of 1% or 10% of a funding goal of up to $1,000. No donations were made to projects in the control condition.

Round 2. The treatment in round 2 involved the donation of either 1% by one donor or a total of 4% by four donors of a funding goal up to $5,000. No donations were made to projects in the control condition.

Study 2. Epinions.com. ***Site.*** Epinions.com is a popular general consumer review site that was established in 1999. In 2013, it had ~1 million visitors per month. The epinions platform gives users the ability to write reviews evaluating all types of products (www.epinions.com). Other users can rate reviews as being "very helpful," "helpful," "somewhat helpful," or "not helpful." The amount users are paid for writing reviews depends on the ratings their reviews receive from other users. [Epinions specifies that these earnings are determined by an undisclosed algorithm that provides

SOCIAL SCIENCES

greater financial compensation to authors of better-rated reviews (www.epinions.com/help/faq/show_~faq_earnings).]

Design and procedure. When an unrated review became available, we read the review to determine its quality, classifying it as either "high" or "low." We then randomly assigned high-quality reviews to either the experimental or control condition. High-quality reviews in the experimental condition were rated "very helpful," and high-quality reviews in the control condition were left unrated. After 14 d, we counted the number of ratings and the type of ratings that each of the selected reviews had accumulated. We collected a total of 481 cases in two rounds. Data collection took place between October 2012 and August 2013 for round 1 and between September 2013 and January 2014 for round 2.

Round 1. In round 1 of data collection, the treatment involved the application of a single rating. High-quality reviews received a single "very helpful" rating in the experimental condition and were left unrated in the control condition.

Round 2. In round 2 of data collection the treatment involved the application of either a single rating or four ratings on behalf of four members of the research team. High-quality reviews in the experimental condition received one or four "very helpful" ratings. High-quality reviews in the control condition were left unrated.

Study 3. Wikipedia.org. ***Site.*** Wikipedia is a collaboratively written encyclopedia, started in the United States in 2001, that as of January 13, 2014 encompassed 31 million articles in 285 languages (http://en.wikipedia.org/wiki/Wikipedia:About). Wikipedia is created through voluntary contributors, or "editors," who actively generate, update, and modify its content. The website poses very low barriers to participation by allowing any individual to edit its articles under a self-chosen pseudonym. For this research, we restricted our experiment to the English-language Wikipedia, which as of January 2014 had 4.4 million articles, 21 million registered accounts, and 118,082 active editors who had made at least one edit in the last month.

Design and procedure. On Wikipedia, editors can grant other editors a virtual award (http://en.wikipedia.org/wiki/Wikipedia:Awards) by posting such an award on the target editor's user page for public display. To ensure that we acted in accordance with informal Wikipedia rules to grant awards only to individuals who have contributed significantly to the project, we sampled only top contributors. We ranked the population of editors by their total number of edits in the preceding month and included only editors in the top 1% by edit count. We also eliminated from our target population anonymous editors and editors with special authority within the project.

We randomly placed 208 individuals in the experiment condition and anonymously gave them an award, which editors in the broader Wikipedia editing community could see when making their own decisions about whom to thank or reward. To do so, we posted a customized award on the focal editor's user page. In a prior experiment conducted 2 y before the present study, we used a similar study design but instead granted generic "barnstar" awards (http://en.wikipedia.org/wiki/Wikipedia:Barnstars) and focused on effects of these awards on productivity (34). The awards we gave out could be seen only on the editors' user pages, and nowhere did Wikipedia overview award recipients or sort them by popularity, so that recipients did not enjoy greater exposure than nonrecipients as a result of our treatment. After 90 d, we reviewed the editor pages for all 521 subjects in our sample and counted how many additional third-party awards an editor had received. Data collection spanned the period February 2012 to July 2012. Note that we collected data only on a sample of the target population, because the award data from digital historical records had to be collected manually (automatic data collection proved unreliable and error-prone).

Study 4. Change.org. ***Site.*** Change.org is a popular online petition website. Individuals who are passionate about a cause can initiate a signature campaign with only a few clicks after they have created an account with a functional e-mail and postal address. Visitors to the website can sign petitions either anonymously or by name. As of January 13, 2014 the website had hosted petitions initiated by more than 30 million people (www.change.org/). Petitions range in purpose from the mundane—"ban 'x' from our multiplayer gaming platform"—to topics that dominate the national news. These extreme discrepancies in the scope and importance of petitions create large natural variation in the potential for signatures. Indeed, most campaigns yield only a handful of sympathizers, but a few reach millions of supporters.

Design and procedure. We sampled from the population of new petitions whose creator had acquired at most 15 signatures. We could not sample petitions below the minimum of five signatures required for public posting on the website. The selection of small petitions prevented the petitions we studied from appearing on a list of popular petitions, thus precluding this sorting mechanism from mediating any treatment effect. We further selected petitions that had been initiated less than 14 d earlier, ensuring that the petitions had not lost their relevance. From among these recent petitions we selected those that had not been signed in the past 24 h, thus making sure that our treatment did not co-occur with an ongoing surge of petition signing. Two hundred petitions satisfied these criteria and passed a screening test against mal-intended campaigns that sought to do harm to an individual or group.

We randomly assigned 100 petitions to the experimental condition and 100 petitions to the control condition. We added 12 anonymous signatures to petitions in the experimental condition and withheld signatures from petitions in the control condition. We kept a daily record of the number of signatures the petitioners received for a period of 2 wk, after which interval most campaigns had stopped accumulating signatures. Data collection took place between July and August of 2012.

Further details on design and data analysis can be found in *SI Materials and Methods* and *SI Results*.

ACKNOWLEDGMENTS. We thank Damon Centola for helpful discussions, the editor and two anonymous reviewers for useful comments, and Idil Akin, Gabriella Gonzalez, Hua Mo, Fernanda Page, and Juhi Tyagi for assistance with data collection. This work was supported by National Science Foundation Grants SES-1340122 (to A.v.d.R.) and SES-1303522 (to M.R. and A.v.d.R.).

1. Herrnstein RJ, Murray C (1994) *The Bell Curve: Intelligence and Class Structure in American Life* (Free, New York).
2. Simonton D (1997) Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychol Rev* 104(1):66–89.
3. Huber J (1998) Cumulative advantage and success-breeds-success: The value of time pattern analysis. *J Am Soc Inf Sci* 49(5):471–476.
4. Allison P (1980) Estimation and testing for a Markov model of reinforcement. *Sociol Methods Res* 8(4):434–453.
5. Mankiw NG (2013) Defending the one percent. *J Econ Persp* 27(3):21–34.
6. Merton RK (1968) The Matthew effect in science. *Science* 159(3810):56–64.
7. Barabási AL (2012) Network science: Luck or reason. *Nature* 489(7417):507–508.
8. de Solla Price D (1976) A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inf Sci* 27(5):292–306.
9. Reskin B (1977) Scientific productivity and the reward structure of science. *Am Sociol Rev* 42(3):491–504.
10. Allison P, Long J, Krauze T (1982) Cumulative advantage and inequality in science. *Am Sociol Rev* 47(5):615–625.
11. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512.
12. DiPrete TA, Eirich GM (2006) Cumulative advantage as a mechanism for inequality. *Annu Rev Sociol* 32:271–297.
13. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.
14. Simcoe TS, Waguespack DM (2010) Status, quality and attention. *Manage Sci* 57(2):274–290.
15. Petersen AM, Jung WS, Yang JS, Stanley HE (2011) Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proc Natl Acad Sci USA* 108(1):18–23.
16. Simon H (1955) On a class of skew distribution functions. *Biometr* 42(3-4):425–440.
17. Mandelbrot B (1959) A note on a class of skew distribution functions: Analysis and critique of a paper by H. A. Simon. *Inform. & Contr* 2(1):90–99.
18. Murray C (2003) *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950* (HarperCollins, New York).
19. Denrell J, Liu C (2012) Top performers are not the most impressive when extreme performance indicates unreliability. *Proc Natl Acad Sci USA* 109(24):9331–9336.
20. Yule GU (1925) A mathematical theory of evolution. *Phil. Trans. B* 213:21–87.
21. Rosen S (1981) The economics of superstars. *Am Econ Rev* 71(5):845–858.
22. Frank RH, Cook PJ (1996) *The Winner-Take-All Society* (Free, New York).
23. Newman ME (2005) Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 46(5):323.
24. Huberman BA, Adamic LA (1999) Growth dynamics of the World-Wide-Web. *Nature* 401(6749):131.
25. Papadopoulos F, Kitsak M, Serrano MA, Boguñá M, Krioukov D (2012) Popularity versus similarity in growing networks. *Nature* 489(7417):537–540.
26. D'Souza RM, Borgs C, Chayes JT, Berger N, Kleinberg RD (2007) Emergence of tempered preferential attachment from optimization. *Proc Natl Acad Sci USA* 104(15):6112–6117.
27. Stumpf MPH, Porter MA (2012) Mathematics. Critical truths about power laws. *Science* 335(6069):665–666.
28. Adamic L (2011) Complex systems: Unzipping Zipf's law. *Nature* 474(7350):164–165.
29. Hanson WA, Putler DS (1996) Hits and misses: Herd behavior and online product popularity. *Mark Lett* 7(4):297–305.

30. Salganik MJ, Watts DJ (2008) Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Soc Psychol Q* 74(4):338–355.
31. Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Science* 341(6146):647–651.
32. Margetts H, John P, Escher T, Reissfelder S (2011) Social information and political participation on the internet: An experiment. *Eur. Pol. Sc. Rev* 3(3):321–344.
33. Ginsburgh VA, van Ours JC (2003) Expert opinion and compensation: Evidence from a music competition. *Am Econ Rev* 93(1):289–296.
34. Restivo M, van de Rijt A (2012) Experimental study of informal rewards in peer production. *PLoS ONE* 7(3):e34358.
35. Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *Am J Sociol* 113(3):702–734.

SOCIAL SCIENCES

Supporting Information

van de Rijt et al. 10.1073/pnas.1316836111

SI Materials and Methods

Overview. In this study we used a generic field experimental design which we deployed in four distinct, ongoing, real-world social settings online: a crowd-funding website (kickstarter.com), a product review website (epinions.com), an open-source encyclopedia (wikipedia.org), and an online petitioning website (change.org). The experimental intervention involved the random allocation of one or multiple successes to individuals. In each setting the success was in a different form (a dollar amount, a positive rating, an award, or a signature), but the communality of design permits a unique comparative analysis across social systems. The key advantage of field experiments is that they combine the potential for causal inference found in laboratory experiments with the external validity typical of naturalistic observation, by studying people inside the social systems of interest without having to remove them into an artificial environment. This feature is particularly important for the focal phenomenon in this study—success breeds success—for which problems of confounding are very difficult to address in observational studies, as we elaborate in the main text.

Ethical Considerations. Intervention in ongoing social systems naturally raises ethical concerns, and throughout our experiments we navigated these issues with the utmost care. All experiments were approved by Stony Brook University's Human Subjects Committee (ID nos. 373335, 366647, 230771, and 442574) and had the additional backing of the National Science Foundation (Award nos. 1303522 and 1340122). We were careful to abide by the terms of use of the internet sites. The guiding principles of minimal harm and minimal disruption led us to restrict our samples by scale. For example, in the crowd-funding study we did not select projects with a funding goal of more than \$5,000, and in the signatures study we selected only small petition campaigns with low signature goals. Additionally, as a research team we always acted in ways we could easily see ourselves act outside the experiments: We gave editing awards only to people who belonged to the 1% most prolific editors, we gave positive ratings only to good reviews and negative ratings only to bad reviews, and we sampled only petitions that sought no harm against a person or group. Finally, in the follow-up studies that increased the magnitude of the intervention, we refrained from intervening in two of the four settings (wikipedia.org and change.org) in which we thought that such intervention would lead to too severe a disruption. On wikipedia.org, we felt that we would not act within the spirit of the project if we gave editors more than one award within a short period, because receiving an award is a relatively rare event. On change.org a noticeable increase in our original intervention of 12 signatures by distinct signatories would have required many signatures per person and thereby would have violated the site's terms of use. For these reasons we increased the strength of the treatment in a second round of data collection only on kickstarter.com, where we increased the number of donors from one to four, and on epinions.com, where we increased the number of ratings from one to four.

SI Results

Site 1: Kickstarter.com. ***Descriptive statistics.*** There were two rounds of data collection. Round 1 focused on identifying the main effect of a donation on subsequent donations by third parties. The treatment involved a single donor donating a percentage of the funding goal to projects that had not raised any dollars 24 d before the funding deadline. Round 2 focused on identifying the relative effects of different numbers of donors. Here the treatment was either a donation of 1% by one donor or a total donation of 4% by four separate donors. For realism, in the case of four donors, we introduced a moderate variation in the size of donations across donors (e.g., one \$15 donation, two \$20 donations, and one \$25 donation for a funding goal of \$2,000). Table S1 shows the mean and median number of donations received and total dollar amounts raised during the 24-d observation period for both the experiment and the control conditions of both rounds, as well as the number of cases in each. The mean dollar amounts and mean number of donations in round 2 correspond to the averages reported in Fig. 3 *A* and *B*, respectively, in the main text.

Effect of treatment on subsequent fundraising (round 1). To produce the binary measure reported in Fig. 1 in the main text (which measures the immediate effect of the treatment on the rate of third-party giving, excluding effects between subsequent donations), we determined for the experimental and control condition of both rounds the number of cases in which no donations were made and the number of cases in which some additional donations were made by third parties. During the 24 d between the treatment and the deadline, 70% of subjects in the experimental condition and 39% of subjects in the control group received one or more third-party donations.. A χ^2 test shows a significant difference between conditions in the number of subjects receiving additional funding before the deadline ($\chi^2 = 19.4$; $P = 0.000$).

We also analyzed the effect of the treatment on cumulative measures of success that include possible second-order interdependencies among third-party donating events. Fig. S1*A* shows the distribution of the total number of dollars raised from third parties after the treatment, by condition, in round 1. A signed-rank test shows a significant difference between conditions in round 1 ($z = 4.13$; $P = 0.000$). Fig. S1*C* shows the distribution of the number of donations received from third parties after the treatment, by condition, in round 2. Again, a clear difference is visible between the experimental and control conditions ($z = 3.95$; $P = 0.000$).

Effect of treatment strength on subsequent fundraising (round 2). In the second round there also was a significant treatment effect on the percentage of cases experiencing at least one additional third-party donation during the 24-d observation period, but the strength of the treatment, which was varied in round 2, did not impact this percentage significantly. The percentages were 32%, 74%, and 87%, respectively, in the zero-donors, one-donor, and four-donors conditions. The difference between the zero-donors and one-donor conditions is significant ($\chi^2 = 11.0$; $P = 0.000$), as is the difference between the zero-donors and four-donors conditions ($\chi^2 = 19.4$; $P = 0.000$), but the difference between the one-donor and four-donors conditions is not significant ($\chi^2 = 1.65$; $P = 0.190$).

Fig. S1*B* shows the distribution of total numbers of dollars raised from third parties after the treatment, by condition, in round 2. The figure shows a large gap between the zero-donations and one-donation conditions, whereas the one-donation and four-donations conditions appear closer together. Signed-rank tests show that the difference in the total amount raised between the zero-donations and one-donation conditions in round 2 was statistically significant ($z = 3.02$; $P = 0.003$), as was the difference between the zero-donations and four-donations conditions ($z = 3.61$; $P = 0.000$), but the difference between the one-donation and four-donations conditions was not ($z = 1.70$; $P = 0.090$). Fig. S1*D* shows the distribution of the number of third-party donations after the treatment, by condition, in round 2. The number of donations clearly differs between the zero-donations and one-donation conditions, but the one-donation and four-donation conditions lie closer together. Indeed, the signed-rank tests find a difference between the zero-donors and one-donor conditions (z = 3.20; P = 0.001) and between the zero-donors and four-donors conditions (z = 4.16; P = 0.000) but not between the one-donor and four-donors conditions (z = 1.95; P = 0.051).

Goal amounts by round. Fig. S2*A* shows the distributions of funding goal amounts for rounds 1 and 2, respectively. In round 1 we set a maximum goal amount of $1,000, and in round 2 we increased the maximum goal amount to $5,000. Because in round 1 the pairs of treatment/control cases were matched in goal amount, the distribution of goal amounts is identical across the two conditions. Similarly, because in round 2 trios of cases were matched in goal amount, the distribution of goal amounts is identical across the three conditions.

Treatment effects on public enthusiasm. We measured expressions of enthusiasm through the number of Facebook likes accumulated at the end of each fundraising campaign. Fig. S2 *B* and *C* shows the distribution of these expressions in rounds 1 and 2 of the study, respectively, by condition. Large differences are visible in both graphs. In round 1 the difference between treatment and control is statistically significant (signed-rank test; z = 4.191; P = 0.000). In round 2, the difference between the one-donor condition and the zero-donors condition is significant (z = 3.35; P = 0.001), as is the difference between the four-donor condition and the zero-donors condition (z = 4.06; P = 0.000), but the difference between the one-donor and four-donors conditions is not (z = 1.59; P = 0.112). The significant treatment effects indicate that, despite the equal quality of projects across condition, random donations increased the level of enthusiasm for target projects, thus disconnecting public support from intrinsic merit.

Relationship between donations and public enthusiasm. We then predicted the incidence of donations from the treatment, controlling for the logarithm of the number of Facebook likes posted after the treatment, using negative binomial regression. Results are shown in Table S2. In models 2 and 4 the treatment effect loses a portion of the original effect size it had in models 1 and 3. In both round 1 and round 2, the treatment effect remains strong after controlling for the number of Facebook likes. These results suggest that a part of the treatment effect on subsequent donations was mediated by a greater level of public enthusiasm for treatment projects triggered through the intervention. We emphasize that the possibility of relevant unobservables prevents any hard conclusions about causal pathways and social mechanisms.

Effects of contribution percentage. In the first round, in which the treatment involved a single donor, we varied the size of this donor's contribution, donating either 1% of the funding goal (on average $6.77) or 10% of the funding goal (on average $66.76). Table S3 displays posttreatment third-party funding by the percentage donated. The percentage of projects with at least one posttreatment donation is significantly higher (χ^2 = 11.2; P = 0.001) when 1% is donated (68%) than when 0% is donated (39%) and also is significantly higher (χ^2 = 14.5; P = 0.000) when 10% is donated (72%) vis-à-vis the 0% condition but does not differ significantly between positive donations of varying magnitude (χ^2 = 0.191; P = 0.663). Similarly, the median number of third-party donations is significantly higher in projects that received donations of 1% than in projects that received no donation (rank-sum test; z = 3.42; P = 0.001) and in projects that received donations of 10% than in projects that received no donations (z = 4.02; P = 0.000) but does not differ between projects that receive donations of 1% and those that received a donation of 10% through the treatment (z = 0.583; P = 0.560). Also, the median number of dollars raised from third parties increases significantly in projects that received donations of 1% as compared with projects that received no donation (z = 3.30; P = 0.001) and in projects that received donations of 10% as compared with projects that received no donations (z = 3.74; P = 0.000) but again does not differ between projects that receive a 1% donation and those that receive a 10% donation through the treatment (z = 0.583; P = 0.672). These results indicate that potential third-party donors are sensitive to the presence of an initial donation but not to its size.

Site 2: Epinions.com. ***Descriptive statistics.*** We assessed the quality of new product reviews on epinions.com that as yet had received no ratings and categorized them as either "high quality" or "low quality." In case of a high-quality review, the treatment involved the application of one or more positive "very helpful" ratings to the review; in case of a low-quality review, the treatment involved the application of one or more negative "not helpful" ratings. Because only high-quality ratings are a form of success, the main text reports the results only for these cases. The data reported here come from two rounds of data collection. In the first round cases were assigned randomly to the zero-ratings or the one-rating condition; in the second round cases were assigned randomly to the zero-ratings, the one-rating, or the four-ratings condition. Below we combine these rounds of data collection for easier presentation, because we maintained the same rating procedure in the second round and find no differences between rounds. Table S4 shows the mean and median number of positive ratings received by high-quality reviews and the mean and median number of negative ratings received by low-quality reviews for the two experimental conditions and the control condition during the 14-d observation period as well as the number of cases in each condition.

Effect of treatment on subsequent ratings received. To produce the binary measure reported in Fig. 1 in the main text (which measures the immediate effect of the treatment on the rate of third-party ratings, excluding effects between subsequent ratings), we determined the number of cases in the one-rating and zero-ratings conditions in which no positive ratings were given after the treatment and the cases in which additional positive third-party ratings were given. During the 14 d immediately following the treatment, 90% of subjects in the one-rating condition and 77% of subjects in the zero-ratings condition received one or more positive third-party ratings (Fig. 1). A χ^2 test shows a significant difference between conditions in the percentage of subjects receiving one or more positive ratings 14 d after the treatment (χ^2 = 9.54; P = 0.002). In the negative-rating experiment, the percentage of cases with one or more negative third-party ratings was 50% in the experimental conditions and 16% in the control condition (χ^2 = 11.5; P = 0.001).

We also analyzed the effect of both treatment conditions on total positive and total negative ratings received (a measure that includes possible second-order interdependencies among third-party rating events). As we report in the main text, we find that the treatment had a significant effect on the number of positive ratings accumulated after 14 d by high-quality reviews (z = 3.21; P = 0.001), which were 11.4 in the zero-ratings condition and 14.9 in the one-rating condition. Analogously, we find that the treatment also had a significant effect on the number of negative ratings accumulated after 14 d by low-quality reviews (z = 3.44; P = 0.001), which were 0.581 in the zero-ratings condition and 2.40 in the one-rating condition.

Effect of treatment strength on subsequent ratings received. Fig. S3*A* combines these response percentages for high-quality reviews in the positive rating experiment with those found for low-quality reviews in the negative rating experiment. The difference in the percentage of high-quality reviews that received one or more positive ratings between the two treatment conditions was not significant (χ^2 = 0.304; P = 0.582), nor was there a significant difference in the percentage of low-quality reviews that received one or more negative ratings (χ^2 = 0.181; P = 0.670). Fig. S3*B* shows for each condition the change in the average number of positive ratings given to high-quality reviews in the positive ratings study and the average number of negative ratings given to low-quality reviews in the negative ratings study. Fig. S4 *A* and *B* shows the distribution of total numbers of positive and negative ratings received from third parties after the treatment, by condition. The high-quality reviews that received no positive ratings from us received significantly fewer third-party positive ratings during the observation period than did the reviews that received one positive rating during treatment ($z = 3.21$; $P = 0.001$) but not significantly fewer than those reviews that received four positive ratings during treatment ($z = 1.83$; $P = 0.067$). The reviews to which we gave four positive ratings and those to which we gave one positive rating do not differ significantly ($z = 1.07$; $P = 0.283$) in the number of subsequent positive third-party ratings. The low-quality reviews to which we gave no negative ratings received significantly fewer negative third-party ratings during the observation period than did the reviews to which we gave one negative rating ($z = 3.44$; $P = 0.001$) and also received fewer negative ratings than the reviews to which we gave four negative ratings ($z = 4.14$; $P = 0.000$). There is no difference in the number of negative third-party ratings between the reviews to which we gave four negative ratings and those to which we gave one negative rating ($z = 1.06$; $P = 0.288$).

Effect of treatment on positivity level of subsequent ratings. It is important to consider the possibility that, instead of encouraging third parties to give a review a rating similar to ours, our treatment simply increased the overall number of ratings given by third parties. In the latter case, the volume of ratings would have increased, but not necessarily how positive the ratings are. Accordingly, we calculated total number of ratings that were dissimilar to the rating applied through the treatment. For the high-quality reviews, where we applied no, one, or four "very helpful" ratings, we look at effects on the number of ratings worse than "very helpful" (i.e., the number of "helpful," "somewhat helpful," and "not helpful" ratings) received after the treatment. For the low-quality reviews, to which we applied no, one, or four "not helpful" ratings, we examine effects on the number of ratings better than "not helpful" (i.e., "very helpful,, "helpful," and "somewhat helpful" ratings) received after the treatment. If our treatment had merely increased the incidence of ratings of any kind without increasing the overall rating level, we would find increased frequencies of these other ratings in the experimental conditions vis-à-vis the control condition. Table S5 shows that, instead, the treatment consistently reduced the number of other ratings given by third parties. For high-quality reviews, the difference in the number of other ratings is significant between the zero-ratings and the one-rating conditions (rank-sum test; $z = 2.71$; $P = 0.007$) and between the zero-ratings and the four-ratings conditions ($z = 2.56$; $P = 0.009$) but is not significant between the one-rating and the four-ratings conditions ($z = 0.417$; $P = 0.677$). Similarly, for low-quality reviews, the difference in the number of other ratings is significant between the zero-ratings and the one-rating conditions ($z = 2.001$; $P = 0.045$) and between the zero-ratings and the four-ratings conditions ($z = 2.76$; $P = 0.006$) but is not significant between the one-rating and the four-ratings conditions ($z = 1.32$; $P = 0.186$). Taken together, Tables S4 and S5 thus show that one or more initial positive ratings increased the number of subsequent positive ratings and reduced the number of negative ratings given to high-quality reviews and, analogously, that one or more initial negative ratings increased the number of subsequent negative ratings and reduced the number of positive ratings given to low-quality reviews.

Site 3: Wikipedia.org. ***Descriptive statistics.*** Table S6 shows the mean and median number of awards received during the 90-d observation period in both the experiment and the control condition as well as the number of cases in each. We provide data for both 30 d and 90 d after the treatment to give the reader a sense of how effects changed with time over the course of the much longer observation period in this study.

Effect of treatment on awards received. To produce the binary measure reported in Fig. 1 in the main text (which measures the immediate effect of the treatment on the rate of success, excluding effects between subsequent awards), we determined the number of cases for each condition in which no additional awards were given and the remaining cases in which some additional awards were given by third parties. During the first 30 d after the treatment, 22% of editors in the experimental condition and 13% of editors in the control condition were subsequently given one or more awards from other users not involved in the experiment. A χ^2 test shows that this difference is significant ($\chi^2 = 7.18$; $P = 0.007$). After 90 d, these percentages had risen to 40% in the experimental condition and 31% in the control condition ($\chi^2 = 4.72$; $P = 0.030$); these are the percentages and significance levels reported in Fig. 1.

We also analyzed the effect of the treatment on the total number of awards received (a measure that includes possible second-order interdependencies among awarding events). Fig. S5*A* shows the distribution of awards received from third parties 90 d after the treatment, by condition. A rank-sum test shows a significant difference between the distributions ($z = 1.982$; $P = 0.048$).

Pretreatment awards by condition. Fig. S5*B* shows the cumulative distribution of awards received during the 30 d before the treatment. The minor differences between the curves indicate that randomization succeeded in balancing the two conditions reasonably in terms of pretreatment awarding. A rank-sum test shows no significant difference between the distributions of awards before treatment ($z = 0.991$; $P = 0.322$). Table S7 shows the results of logistic regression models predicting the likelihood of posttreatment awards, with and without controlling for pretreatment awards, 30 and 90 d after the treatment. The treatment effect is significantly positive throughout the four models shown, demonstrating that random differences in pretreatment awards across conditions do not impact the conclusion with respect to a cumulative advantage effect in award accumulation. Because 13% of editors in the control condition received more than one award after the treatment, Table S7 also shows the results of negative binomial regression models predicting the total number of posttreatment awards. The significant effect of the treatment in model 1 continues to be statistically significant in model 2 once pretreatment awards are controlled. In model 3 the treatment effect falls short of significance, but once pretreatment awards are controlled in model 4, the treatment effect becomes significant.

Effect of treatment on productivity. We measured posttreatment productivity as the number of edits to Wikipedia article pages made by editors during the 90-d observation period following the treatment and measured pretreatment productivity as the number of edits during the 30 d preceding the treatment. Fig. S5*C* shows the distribution of posttreatment productivity, by condition. The difference in posttreatment productivity is significant (rank-sum test; $z = 2.91$; $P = 0.004$), indicating that the treatment raised editors' productivity levels. Fig. S5*D* shows the distribution of pretreatment productivity, by condition. There is no significant difference in productivity before the treatment ($z = 0.516$; $P = 0.606$), indicating that randomization succeeded in balancing the two conditions in terms of productivity.

Relationship between awards and productivity. The positive effect of the treatment on productivity found in Fig. S5*C* raises the possibility that the mechanism driving the feedback effect in awarding was an intensification of editing behavior by award recipients, which in turn may have generated a merit-based response in awarding by third parties. To evaluate whether the treatment directly produced the observed increase in the likelihood of another award, rather than generating it indirectly through an increase in productivity, we predicted the probability of a posttreatment award from the treatment, controlling for posttreatment productivity, using logistic regression. Posttreatment productivity is measured in thousands of edits so that effect size and SE can be reported in regular decimal representation. Results are shown in Table S8. The significantly positive effect of productivity on awards in models 2 and 4 reflects the natural correlation between productivity and awarding that one would expect in a meritocratic system. The significantly positive effect of the treatment in models 1 and 3 reflects the main treatment effect identified earlier. In both models 2 and 4 the treatment effect maintains most of the original effect size it had in models 1 and 3, respectively, although in model 4 the significance level drops below the 95% confidence level. To evaluate if a treatment effect on the number of awards exists net of posttreatment productivity, we predicted the number of posttreatment awards 30 and 90 d after the treatment using negative binomial regression. The results are also shown in Table S8. We find that the treatment maintains most of its effect after controlling for productivity, both 30 d after the treatment (model 2) and 90 d after the treatment (model 4). Treatment effects after 30 d are significant (models 1 and 2), whereas after 90 d the effects fall short of significance, with or without the controlling effect of productivity. These results indicate that in large part the treatment directly affected award-giving behavior by third parties through the increased productivity of recipients. However, the possibility of relevant unobservables prevents any hard conclusions about causal pathways and social mechanisms.

Site 4: Change.org. ***Descriptive statistics.*** Table S9 shows the mean and median number of signatures received during the observation period of 2 wk in both the experiment and the control condition and the number of cases in each.

Effect of treatment on signatures received. The binary measure reported in Fig. 1 in the main text (which measures the immediate effect of the treatment on the rate of signatures, excluding effects between subsequent signatures) is based on a dichotomization of the posttreatment signature count contrasting cases in which some additional signatures were given and the remaining cases in which no additional signatures were given by third parties. During the 14 d after the treatment, 66% of subjects in the experimental condition subsequently received one or more signatures from third-party signatories, compared with 52% of subjects in the control group. A Pearson χ^2 test for independence shows a significant difference between conditions in the number of subjects receiving signatures 14 d after the treatment ($\chi^2 = 4.05$; $P = 0.044$).

We also analyzed the full effect of the treatment on total signatures received (a measure that includes possible second-order interdependencies among signatures). Fig. S6*A* shows the distribution of signatures solicited from third parties after the treatment, by condition. Fig. S6*A* indicates that signature totals in the experimental conditions were higher, but a rank-sum test shows that this difference falls short of statistical significance ($z = 1.76$; $P = 0.079$). Table S10 reports results from negative binomial regression models predicting total posttreatment signature counts from the treatment, controlling for pretreatment signature counts. The significantly positive effect of prior signatures on subsequent signatures in model 2 reflects the uncontrolled relationship between past and future support, which need not reflect a success-breeds-success effect because it may be produced spuriously by differences between campaigns in natural support base. The regression models find a positive treatment effect in model 1 which becomes statistically significant once orthogonal variance from pretreatment signatures is controlled in model 2, confirming the presence of a success-breeds-success effect on signature totals.

Goal amount. Fig. S6*B* shows the distribution of signature goal amounts by condition. Some imbalance is visible, with more petitions in the experiment condition than in the control condition having the typical, low goal amount of 100 signatures, although this difference is not statistically significant ($\chi^2 = 2.32$; $P = 0.128$). We found that higher goal amounts were associated significantly with greater numbers of posttreatment signatures (rank-sum test; $z = 2.00$; $P = 0.046$), suggesting that the slightly lower goal amounts in the treatment condition may have had a potential suppressing effect on signature totals. We explored the impact of this difference in goals on the treatment effect through negative binomial regression, reported in Table S10, model 3. A comparison with model 2 in Table S10 shows that the treatment effect indeed increases somewhat once the goal amount is controlled, thus confirming that the modest imbalance in the goal variable across conditions does not affect the conclusion about the presence of a success-breeds-success effect in signature accumulation.

Effect of treatment on public enthusiasm. To investigate whether the success effect in change.org produced differential expressions of enthusiasm about the campaigns across conditions, despite equivalence of expected project quality (because of randomization), we counted the number of supporting comments left on each campaign page. Fig. S6*C* shows the distribution of the number of supportive comments by condition. It shows that the number of comments in the treatment condition tended to be higher, but a rank-sum test identifies no significant difference ($z = 0.184$; $P = 0.854$).

Relationship between signatures and public enthusiasm. We then predicted the incidence of posttreatment signatures from the treatment, controlling for the number of supportive comments posted after the treatment, using negative binomial regression. The number of supportive comments was logged to correct for extreme variable skew. Results are shown in Table S10, model 4. The significantly positive effect of supportive comments on signatures in model 4 reflects the natural correlation between support in words and support in action. In model 4, the treatment effect maintains most of its original size in model 3 and remains significant. These results suggest that subsequent signatures were added in part because of increased public enthusiasm stemming from our intervention but that mostly it was the actual signatures we added that directly triggered subsequent signatures. We emphasize once more that the possibility of relevant unobservables prevents any hard conclusions about causal pathways and social mechanisms.

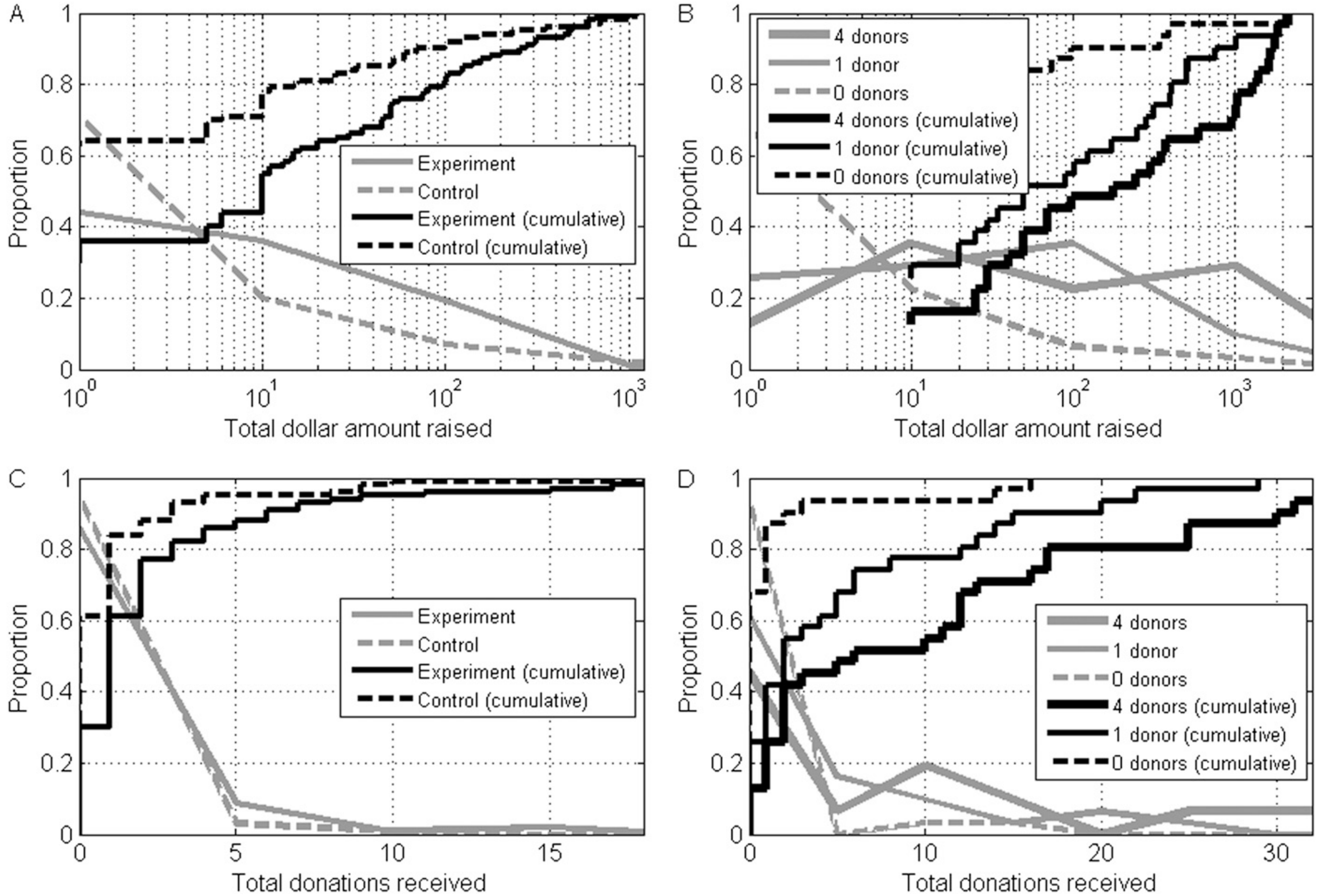

Fig. S1. Distributions of dollars and donations in the crowd-funding study. (*A*) Distribution of dollars raised in round 1, by condition. (*B*) Distribution of dollars raised in round 2, by condition. (*C*) Distribution of donations received in round 1, by condition. (*D*) Distribution of donations received in round 2, by condition.

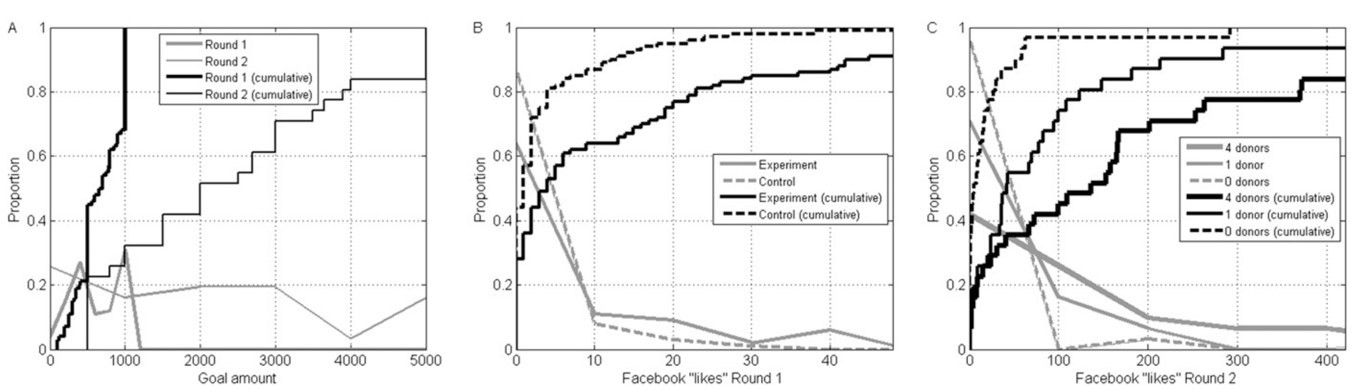

Fig. S2. Distribution of funding goal amount and Facebook likes in the crowd-funding study. (*A*) Distribution of funding goal amount in rounds 1 and 2. (*B*) Distribution of Facebook likes in round 1, by condition. (*C*) Distribution of Facebook likes in round 2, by condition.

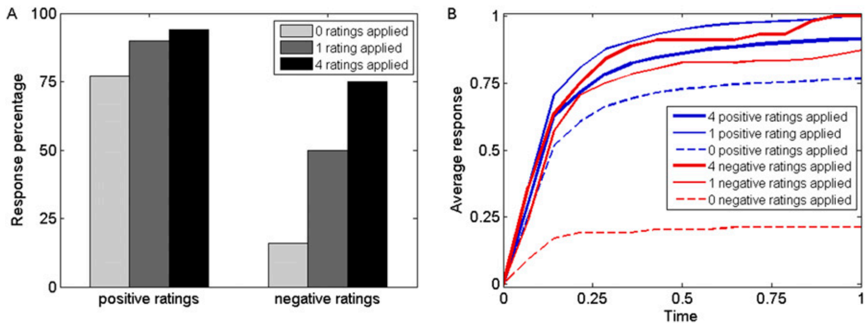

Fig. S3. Percentage and average running number of positive and negative ratings. (*A*) Percentage of high-quality reviews with one or more positive posttreatment ratings and percentage of low-quality reviews with one or more negative posttreatment ratings, by condition. (*B*) Average running number of positive ratings given to high-quality reviews and average running number of negative ratings given to low-quality reviews, over time and by condition. The horizontal axis is normalized so that 0 marks the time of experimental intervention, and 1 marks the end of the observation period. The vertical axis is normalized so that a value of 1 equals the maximum for each system across time and conditions.

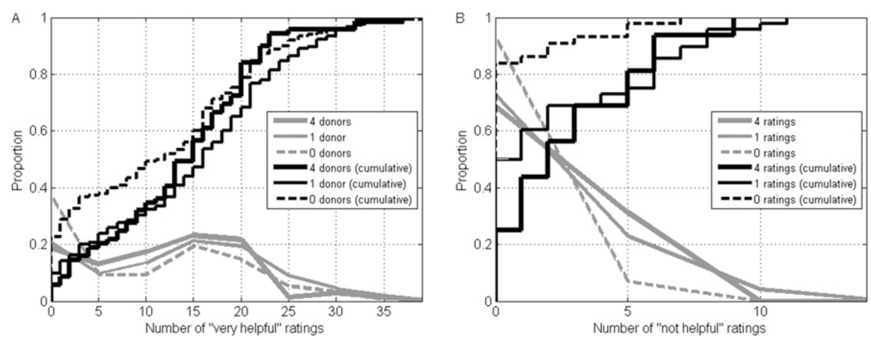

Fig. S4. Distribution of positive and negative ratings. (*A*) Distribution of positive ratings, by condition. (*B*) Distribution of negative ratings, by condition.

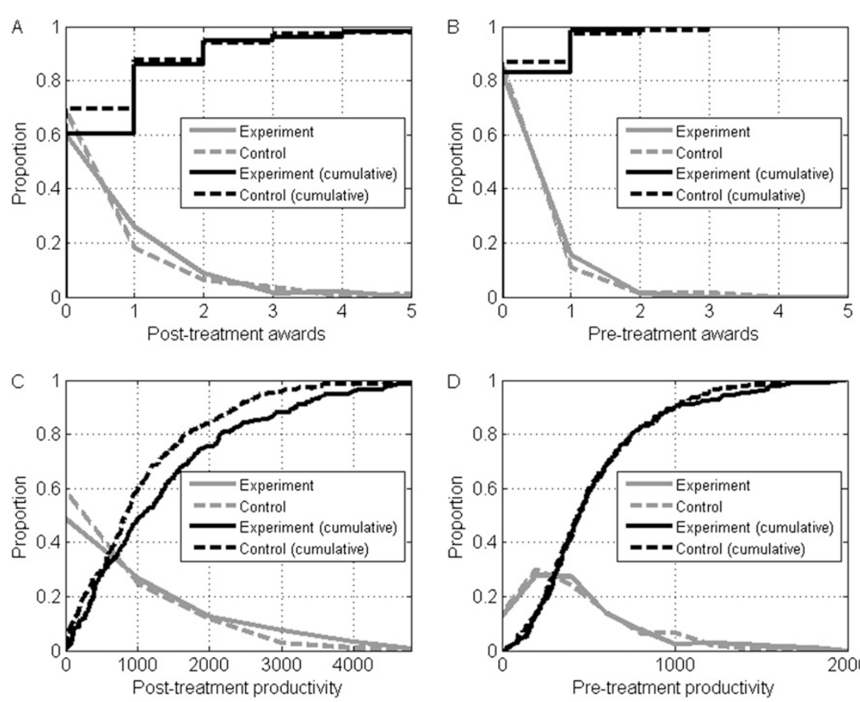

Fig. S5. Distribution of awards and productivity in the awards study. (*A*) Distribution of posttreatment awards received, by condition. (*B*) Distribution of pretreatment awards received, by condition. (*C*) Distribution of editors' posttreatment productivity, by condition. (*D*) Distribution of editors' pretreatment productivity, by condition.

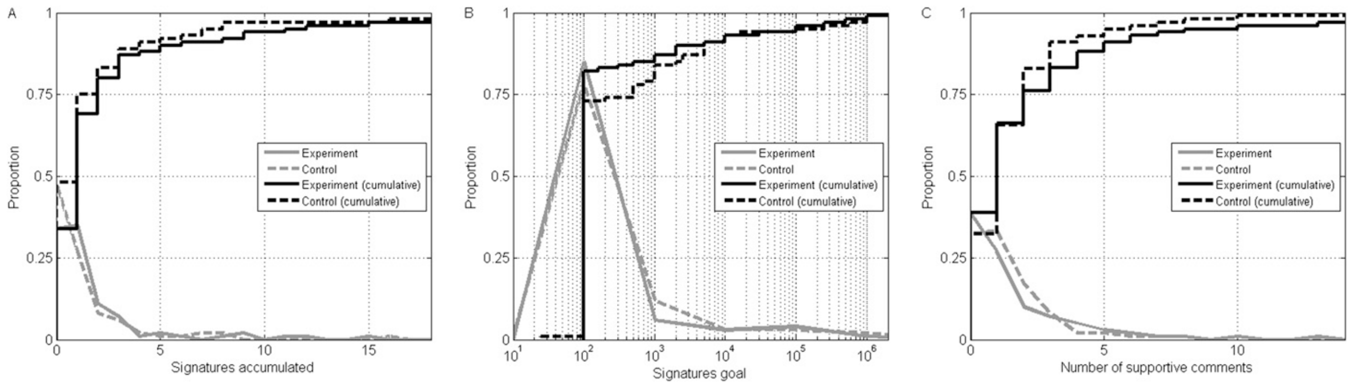

Fig. S6. Distribution of signatures and supportive comments in the signatures study. (*A*) Distribution of signatures solicited, by condition. (*B*) Distribution of signature goal amount, by condition. (*C*) Distribution of number of supportive comments on petitions, by condition.

Table S1. Crowd-funding study: Descriptive statistics of crowd-funding study, by condition

	Round 1		Round 2		
				Experiment	
Donations and dollars raised by day 24	Control 0 donations	Experiment 1 donation	Control 0 donations	1 donation	4 donations
Donations by day 24					
Mean	1.11	2.49	1.32	5.65	10.77
Median	0	1	0	2	6
SD	2.77	4.55	3.73	7.53	11.66
Dollars raised by day 24					
Mean	50.35	77.50	102.65	293.65	562.35
Median	0.00	10.00	0.00	50.00	180.00
SD	177.31	166.88	400.26	509.00	697.70
N	100	100	31	31	31

All statistics reported exclude the treatment.

Table S2. Crowd-funding study: Negative-binomial regression of posttreatment donations and dollars on treatment

	Effects on no. of donations				Effects on no. of dollars			
	Model 1		Model 2		Model 3		Model 4	
Round	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Round 1								
Treatment	0.785***	0.194	0.529**	0.203	0.859***	0.198	0.634**	0.208
Log no. of Facebook likes			0.313***	0.071			0.302***	0.074
N	200		200		200		200	
Round 2								
One donor	1.37***	0.376	0.924*	0.398	1.26**	0.372	0.776*	0.387
Four donors	1.85***	0.367	1.35***	0.386	1.74***	0.362	1.21**	0.375
Log no. of Facebook likes			0.257***	0.054			0.287***	0.056
N	93		93		93		93	

$*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

Table S3. Crowd-funding study: Donations and dollars raised in round 1 by percentage donated

Additional donations or dollars by day 24	0% donated	1% donated (mean, $6.77)	10% donated (mean, $66.76)
Additional donations, %	39	68	72
No. donations by day 24			
Mean	1.11	2.36	2.62
Median	0	1	1
SD	2.77	4.87	4.25
Dollars raised by day 24			
Mean	50.35	79.97	75.02
Median	0.00	10.00	10.50
SD	177.31	193.37	137.34
N	100	50	50

Table S4. Ratings study: Descriptive statistics of ratings study, by condition

	No. of "very helpful" or "not helpful" ratings applied*		
Type of review/rating	0	1	4
High-quality reviews: "very helpful" ratings by day 14			
Mean	11.4	14.9	13.6
Median	11.5	16	15
SD	10.1	9.68	8.26
N	150	155	69
Low-quality reviews: "not helpful" ratings by day14			
Mean	0.581	2.40	3.00
Median	0	0.5	2
SD	1.56	3.24	2.72
N	43	48	16

All statistics reported exclude the treatment.

*"Very helpful" ratings were applied to high-quality reviews; "not helpful" ratings were applied to low-quality reviews.

Table S5. Ratings study: Ratings other than treatment rating, by condition

	No. of "very helpful" or "not helpful" ratings applied*		
Type of review/rating	0	1	4
High-quality reviews: ratings worse than "very helpful" by day 14			
Mean	3.72	2.27	1.68
Median	0	0	0
SD	4.95	4.16	3.41
N	150	155	69
Low-quality reviews: ratings better than "not helpful" by day 14			
Mean	10.9	8.23	6.00
Median	9	8.5	5
SD	6.29	6.05	5.06
N	43	48	16

*"Very helpful" ratings were applied to high-quality reviews; "not helpful" ratings were applied to low-quality reviews.

Table S6. Awards study: Descriptive statistics of awards study, by condition

	After 30 d		After 90 d	
	Control	Experiment	Control	Experiment
Mean	0.169	0.284	0.559	0.683
Median	0	0	0	0
SD	0.519	0.689	1.13	1.21
N	313	208	313	208

All statistics reported exclude the treatment.

Table S7. Awards study: Regression of posttreatment awards on treatment and pretreatment awards

	30 d after treatment				90 d after treatment			
	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Logistic								
Effect of treatment	0.634**	0.239	0.656**	0.243	0.406*	0.187	0.408*	0.194
Effect of pretreatment awards			0.797***	0.209			1.18***	0.227
Negative binomial								
Effect of treatment	0.544*	0.215	0.550**	0.212	0.272	0.146	0.293*	0.143
Effect of pretreatment awards			0.751***	0.140			0.743***	0.090
N	521		521		521		521	

$*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

Table S8. Awards study: Regression of posttreatment awards on treatment and productivity

	30 d after treatment				90 d after treatment			
	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Logistic								
Effect of treatment	0.634**	0.239	0.583*	0.241	0.406*	0.187	0.343	0.190
Effect of productivity			0.641*	0.262			0.205*	0.086
Negative binomial								
Effect of treatment	0.544*	0.215	0.489*	0.216	0.272	0.146	0.205	0.147
Effect of productivity			0.624***	0.214			0.178**	0.061
N	521		521		521		521	

$*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

Table S9. Signatures study: Descriptive statistics of signatures study, by condition

	After 14 d	
	Control	Experiment
Mean	1.74	2.32
Median	1	1
SD	3.96	4.59
N	100	100

Note: All statistics reported exclude the treatment.

Table S10. Signatures study: Negative binomial regression of posttreatment signatures

Predictor	Model 1 Coefficient	Model 1 SE	Model 2 Coefficient	Model 2 SE	Model 3 Coefficient	Model 3 SE	Model 4 Coefficient	Model 4 SE
Treatment	0.288	0.231	0.525*	0.241	0.600**	0.220	0.478*	0.212
Pretreatment signatures			0.164***	0.037	0.114***	0.037	0.091**	0.091
Signature goal					1.043***	0.249	0.659**	0.252
Supportive comments							0.382**	0.117
N	200		200		200		199	

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.