EXTENDED PDF FORMAT
SPONSORED BY
Stem Cell Marker Antibodies
>Learn More
R&D SYSTEMS
www.rndsystems.com

Science
AAAS

Social Influence Bias: A Randomized Experiment

Lev Muchnik *et al.*

Science **341**, 647 (2013);
DOI: 10.1126/science.1240466

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 8, 2013):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/341/6146/647.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2013/08/07/341.6146.647.DC1.html

This article **cites 41 articles**, 20 of which can be accessed free:
http://www.sciencemag.org/content/341/6146/647.full.html#ref-list-1

Downloaded from www.sciencemag.org on August 8, 2013

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2013 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

BVOCs. The contribution from hundreds of individually minor compounds is important. Future research is required to determine if this is also true for ecosystems that emit larger quantities of isoprene and/or monoterpenes.

The minor species with lower fluxes and concentrations are not in current BVOC emission models, but their sizes, chemical formulae, and sum suggest that they should be important for SOA formation and regional photochemistry. Their presence in the atmosphere may also account for a sizable amount of the missing OH chemical reactivity and O_3 chemical loss observed in plant environments.

References and Notes

1. A. Guenther *et al.*, *J. Geophys. Res. Atmos.* **100**, 8873 (1995).
2. W. L. Chameides, R. W. Lindsay, J. Richardson, C. S. Kiang, *Science* **241**, 1473–1475 (1988).
3. M. O. Andreae, P. J. Crutzen, *Science* **276**, 1052–1058 (1997).
4. M. S. Jang, N. M. Czoschke, S. Lee, R. M. Kamens, *Science* **298**, 814–817 (2002).
5. A. H. Goldstein, I. E. Galbally, *Environ. Sci. Technol.* **41**, 1514–1521 (2007).
6. M. R. Kurpius, A. H. Goldstein, *Geophys. Res. Lett.* **30**, 1371 (2003).
7. P. Di Carlo *et al.*, *Science* **304**, 722–725 (2004).
8. R. Holzinger, A. Lee, K. T. Paw, A. H. Goldstein, *Atmos. Chem. Phys.* **5**, 67–75 (2005).
9. T. M. Ruuskanen *et al.*, *Atmos. Chem. Phys.* **11**, 611–625 (2011).
10. M. Hallquist *et al.*, *Atmos. Chem. Phys.* **9**, 5155–5236 (2009).
11. T. Karl *et al.*, *Science* **330**, 816–819 (2010).
12. J.-H. Park *et al.*, *Atmos. Chem. Phys.* **13**, 1439–1456 (2013).
13. A. Jordan *et al.*, *Int. J. Mass Spectrom.* **286**, 122–128 (2009).
14. M. Graus, M. Müller, A. Hansel, *J. Am. Soc. Mass Spectrom.* **21**, 1037–1044 (2010).
15. R. Holzinger, A. Kasper-Giebl, M. Staudinger, G. Schauer, T. Rockmann, *Atmos. Chem. Phys.* **10**, 10111–10128 (2010).
16. See supplementary materials on *Science* Online.
17. A. Lee *et al.*, *J. Geophys. Res. Atmos.* **111**, 21 (2006).

Acknowledgments: This work was supported by the Netherlands Organization for Scientific Research (NWO) under the ALW-Middelgroot program (grant 834.08.002), the California Air Resources Board (Award 06-329), the Citrus Research Board, and the University of Utrecht short stay fellowship program. We thank the Gorden Ranch for hosting this study and for the assistance provided by the University of California Lindcove Research and Extension station.

Supplementary Materials

www.sciencemag.org/cgi/content/full/341/6146/643/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S3
References (*18–28*)

11 January 2013; accepted 2 July 2013
10.1126/science.1235053

Social Influence Bias: A Randomized Experiment

Lev Muchnik,[1] Sinan Aral,[2]* Sean J. Taylor[3]

Our society is increasingly relying on the digitized, aggregated opinions of others to make decisions. We therefore designed and analyzed a large-scale randomized experiment on a social news aggregation Web site to investigate whether knowledge of such aggregates distorts decision-making. Prior ratings created significant bias in individual rating behavior, and positive and negative social influences created asymmetric herding effects. Whereas negative social influence inspired users to correct manipulated ratings, positive social influence increased the likelihood of positive ratings by 32% and created accumulating positive herding that increased final ratings by 25% on average. This positive herding was topic-dependent and affected by whether individuals were viewing the opinions of friends or enemies. A mixture of changing opinion and greater turnout under both manipulations together with a natural tendency to up-vote on the site combined to create the herding effects. Such findings will help interpret collective judgment accurately and avoid social influence bias in collective intelligence in the future.

We rely on ratings contributed by others to make decisions about which hotels, books, movies, political candidates, news, comments, and stories are worth our time and money (*1*). Given the widespread use and economic value of rating systems (*2–4*), it is important to consider whether they can successfully harness the wisdom of crowds to accurately aggregate individual information. Do they produce useful, unbiased, aggregate information about the quality of the item being rated? Or, as suggested by the experiments of Salganik *et al.* (*5*), are outcomes path dependent, yielding different aggregate ratings for items of equivalent quality?

Collective intelligence has recently been heralded as a harbinger of accelerated human potential (*6*). But, social influence on individuals' perceptions of quality and value could create herding effects that lead to suboptimal market outcomes (*7*, *8*); rich-get-richer dynamics that exaggerate inequality (*9–12*); a group think mentality that distorts the truth (*13*); and measurable disruptions in the wisdom of crowds (*14*). If perceptions of quality are biased by social influence, attempts to aggregate collective judgment and socialize choice could be easily manipulated, with dramatic consequences for our markets, our politics, and our health.

The recent availability of population-scale data sets on rating behavior and social communication enable novel investigations of social influence (*1*, *15–20*). Unfortunately, our understanding of the impact of social influence on collective judgment is limited because empirically distinguishing influence from uninfluenced agreement on true quality is nearly impossible in observational data (*21–27*). For example, popular products may be popular because of the irrational effect of past positive ratings, or alternatively, the best products may become popular because they are of the highest quality. We must distinguish these explanations to determine the extent to which social influence creates irrational herding.

We therefore designed and analyzed a large-scale randomized experiment to quantify the effects of social influence on users' ratings and discourse on a social news aggregation Web site, where users contribute news articles and discuss them. Users of the site that we studied write comments in response to posted articles, and other users can then "up-vote" or "down-vote" these comments, yielding an aggregate current rating for each posted comment equal to the number of up-votes minus the number of down-votes. Users do not observe the comment scores before clicking through to comments—each impression of a comment is always accompanied by that comment's current score, tying the comment to the score during users' evaluation—and comments are not ordered by their popularity, mitigating selection bias on high (or low) rated comments. Similar scoring mechanisms are widely used on the Web to reward users for supplying insightful or interesting analysis, while penalizing those posting irrelevant, redundant, or low-quality comments. The vast majority of interuser relations occur on the Web site, in contrast to Web sites whose members also interact offline. The data therefore provide a unique opportunity to comprehensively study social influence bias in rating behavior.

Over 5 months, 101,281 comments submitted on the site were randomly assigned to one of three treatment groups: up-treated, down-treated, or control. Up-treated comments were artificially given an up-vote (a +1 rating) upon the comment's creation, whereas down-treated comments were given a down-vote (a –1 rating) upon the comment's creation. Users were unaware of the

[1]School of Business Administration, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, 91905 Israel. [2]MIT Sloan School of Management, 100 Main Street, Cambridge, MA 20142, USA. [3]NYU Stern School of Business, 44 West 4th Street, New York, NY 10012, USA.

*Corresponding author. E-mail: sinana@mit.edu

Downloaded from www.sciencemag.org on August 8, 2013

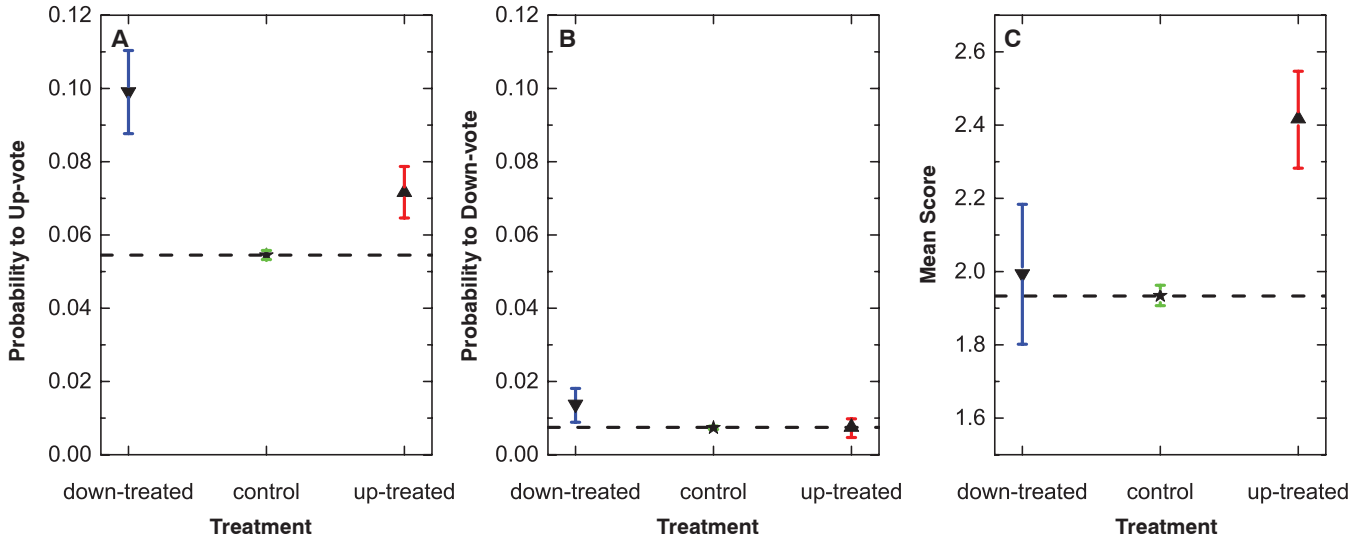

Fig. 1. Effect of manipulation on voting behavior. The positively manipulated treatment group (up-treated), the negatively manipulated treatment group (down-treated), and the control group (dotted line) are shown. The probabilities to up-vote (**A**) and down-vote (**B**) positively manipulated, negatively manipulated, and control group comments are shown by the first unique viewer; 95% confidence intervals are inferred from Bayesian logistic regression with commenter, rater, and commenter-rater random effects. (**C**) The mean final scores of positively manipulated, negatively manipulated, and control group comments are shown with 95% confidence intervals inferred from Bayesian linear regression of the final comment score with commenter random effects. Final mean scores on this Web site are measured as the number of up-votes minus the number of down-votes. We discuss the implications of this measurement in greater detail in the supplementary materials.

Downloaded from www.sciencemag.org on August 8, 2013

manipulation and unable to trace votes to any particular user. As a result of the randomization, comments in the control and treatment groups were identical in expectation along all dimensions that could affect users' rating behavior except for the current rating. This manipulation created a small random signal of positive or negative judgment by prior raters for randomly selected comments that have the same quality in expectation, enabling estimates of the effects of social influence holding comment quality and all other factors constant. The 101,281 experimental comments (of which 4049 were positively treated and 1942 were negatively treated to reflect the natural proportions of up- and down-votes on the site) were viewed more than 10 million times and rated 308,515 times by subsequent users.

We may sample users multiple times, and nonrandom heterogeneity may exist in users' comment-generating process, their selection of comments to vote on, and in relationships between commenters and raters. We therefore estimated hierarchical Bayesian models of voting behavior and specified commenter, rater, and commenter-rater pair random effects; i.e., the confidence intervals are based on repeated resampling, creating a distribution of parameter estimates from which the 95% confidence bands are derived (see materials and methods in the supplementary materials).

We first compared the probabilities that comments in each group would be up-voted or down-voted by the first viewer after the manipulation. These probabilities measure the immediate effect of current ratings on users' rating behavior. We then analyzed comments' long-run ratings distributions and final mean scores by aggregating all users' ratings for comments in the three groups over time.

Fig. 2. Effects of topic on herding. Mean final scores of positively manipulated and control group comments are shown with 95% confidence intervals inferred from Bayesian linear regression of the final comment score with commenter random effects across the seven most active topic categories on the site, ordered by the magnitude of the difference between the mean final score of positively manipulated comments and the mean final score of control comments in each category.

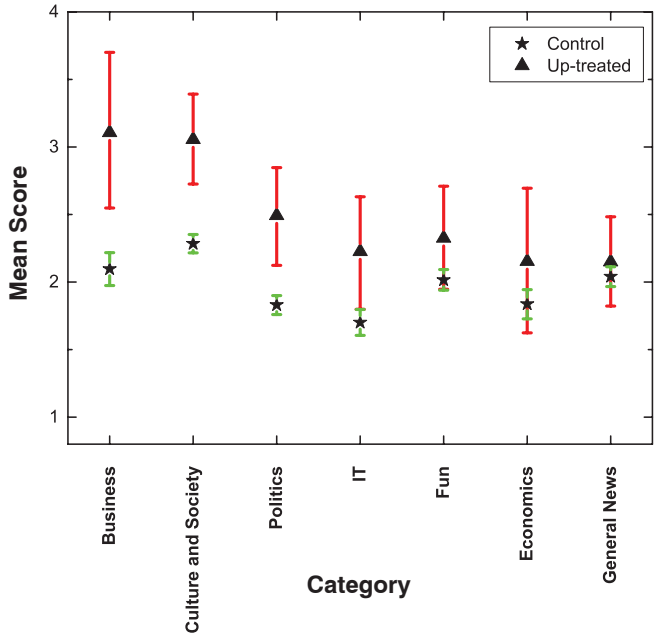

Figure 1A shows the immediate up-vote and down-vote probabilities for the first viewer of comments in each of the three categories. Up-votes were 4.6 times as common as down-votes on this site, with 5.13% of all comments receiving an up-vote by the first viewer of the comment and only 0.82% of comments receiving a down-vote by the first viewer. The up-vote treatment significantly increased the probability of up-voting by the first viewer by 32% over the control group ($P = 1.0 \times 10^{-6}$) (Fig. 1A). Up-treated comments were not down-voted significantly more or less frequently than the control group, so users did not tend to correct the upward manipulation. In the absence of a correction, positive herding accumulated over time.

The positive manipulation created a positive social influence bias that persisted over our 5-month observation window, generating accumulating herding effects that increased comments' final mean ratings by 25% relative to the final mean ratings of control group comments (χ^2 test;

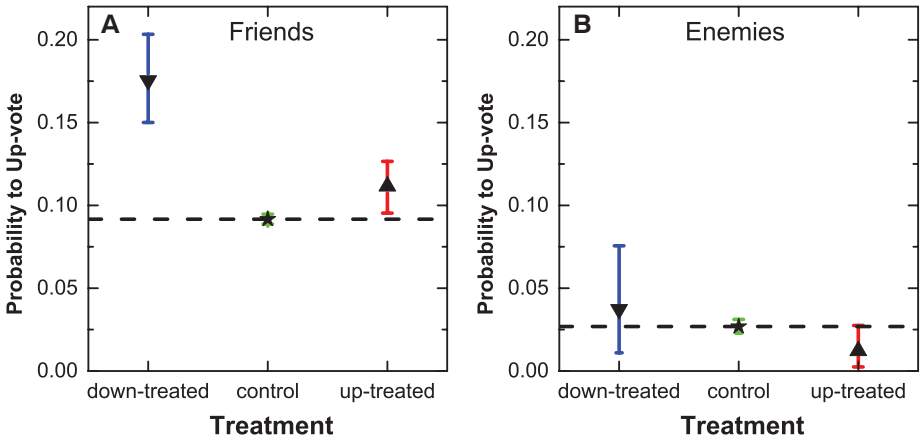

Fig. 3. Effects of friendship on rating behavior. The figure shows the probability of a friend (**A**) and enemy (**B**) of the commenter to up-vote positively manipulated, negatively manipulated, and control group comments. Friends and enemies are defined as users who had previously clicked a button on the Web site labeling the commenter as someone they "liked" or "disliked," respectively.

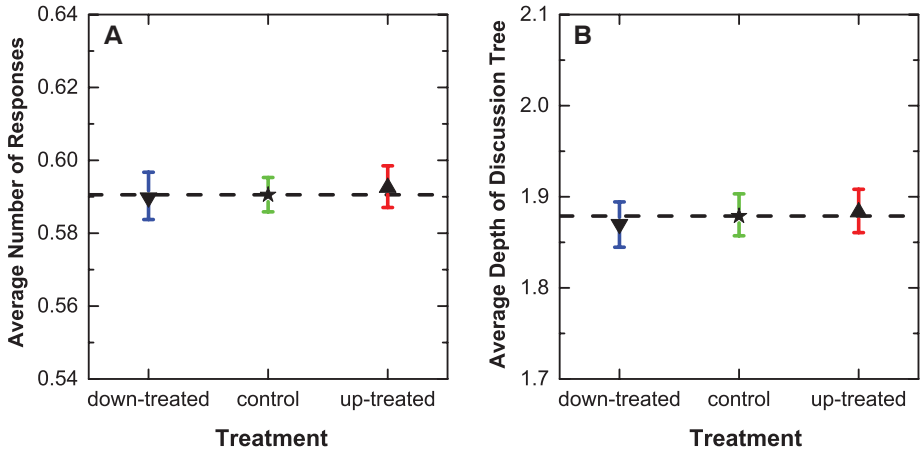

Fig. 4. Effects on subsequent discourse. The figure displays the average number of responses (**A**) and the average depth of the discussion tree (**B**) that developed in response to positively manipulated, negatively manipulated, and control group comments; 95% confidence intervals are inferred from Bayesian linear regressions with author random effects.

$P = 2.3 \times 10^{-11}$) (Fig. 1C), and Kolmogorov-Smirnov (K-S) tests showed that the final score distribution of up-treated comments was significantly shifted toward higher scores (K-S test statistic: 0.083; $P = 1.2 \times 10^{-23}$). Comments in the up-treated group were also significantly more likely than those in the control group to accumulate exceptionally high scores. Up-treated comments were 30% more likely to reach or exceed a score of 10 (6.4% versus 4.9% in the control group, χ^2 test; $P = 2.0 \times 10^{-5}$). The small manipulation of a single random up-vote when the comment was created resulted in significantly higher accumulated ratings due to social influence.

Positive and negative social influence created asymmetric herding effects. The probability of down-treated comments receiving subsequent down-votes was 0.014, whereas the probability of control comments receiving a down-vote was 0.007; a significant difference (χ^2 test; $P = 1.1 \times 10^{-3}$) (Fig. 1B). However, this effect was offset by a larger correction effect. The probability that a down-treated comment would subsequently be up-voted was 0.099, whereas the probability that a control comment would be up-voted was significantly different at 0.054 (χ^2 test; $P = 1.0 \times 10^{-30}$) (Fig. 1A). This correction neutralized social influence in the ratings of negatively manipulated comments, and their final mean ratings were not statistically different from the control group's final mean ratings (Fig. 1C).

We next estimated changes in the final mean score for up-treated comments compared to control comments in the seven most active topic categories on the Web site. We found significant positive herding effects for comment ratings in "politics," "culture and society," and "business," but no detectable herding behavior for comments in "economics," "IT," "fun," and "general news" (Fig. 2). These differences are not due to the frequency of commenting in these categories, as categories with significant differences in control and treatment ratings and those with no significant differences had similar levels of activity. There was no significant negative herding in any category.

Friendship also moderated the impact of social influence on rating behavior (Fig. 3, A and B). The Web site has a feature whereby users can indicate that they "like" or "dislike" other users, forming "friends" and "enemies" social preference graphs. Unsurprisingly, friends of the commenter were more likely to up-vote a comment than those who disliked him or her (9.2% versus 2.7%, χ^2 test; $P = 2.7 \times 10^{-49}$) [compare the average (dotted line) in Fig. 3A to the average (dotted line) in Fig. 3B]. Friends also tended to herd on current positive ratings (friends' probability to up-vote a positively manipulated comment: 0.122 versus friends' probability to up-vote a control comment: 0.092; χ^2 test; $P = 1.4 \times 10^{-2}$) and to correct comments with negatively manipulated ratings (friends' probability to up-vote a negatively manipulated comment: 0.176 versus friends' probability to up-vote a control comment: 0.092; χ^2 test; $P = 4.0 \times 10^{-12}$) (Fig. 3A), mirroring the cooperation found in human social networks (*28*). By contrast, enemies of the commenter were not susceptible to social influence. Enemies' ratings were unaffected by our treatments, possibly because of the small sample of potential first ratings by enemies (though there are a substantial number of enemies in the community, they are less active) (Fig. 3B).

Finally, social influence in ratings behavior did not affect discourse in our setting during the 5-month observation period. Neither the positive nor the negative manipulation affected the average number of replies (Fig. 4A) or the average depth of the discussion tree created in response to a comment (Fig. 4B).

Several data-generating processes could explain our findings. A selection effect could inspire different populations of voters to turn out to rate the item (selective turnout)—for example, if the negative manipulation inspired voters who tend to down-vote (negative voters) to vote in higher proportion. Alternatively, prior ratings could bias users' voting behavior by changing their opinions about comment quality and therefore their votes (opinion change). We analyzed changes in turnout (the likelihood of rating rather than just viewing comments) and changes in positivity (the proportion of positive ratings) across subgroups in our study population to identify variance in our results explained by turnout effects and opinion change, respectively. We divided raters and commenters, on the basis of their rating history, along four subgroup dimensions by their positivity (proportion of positive votes), the commenters' quality (prior scores of their comments), the frequency with which a rater rated a particular commenter, and whether raters were friends or enemies with commenters. We then compared treatment effects with expected voting behavior in these subgroups. Our analysis revealed several

Downloaded from www.sciencemag.org on August 8, 2013

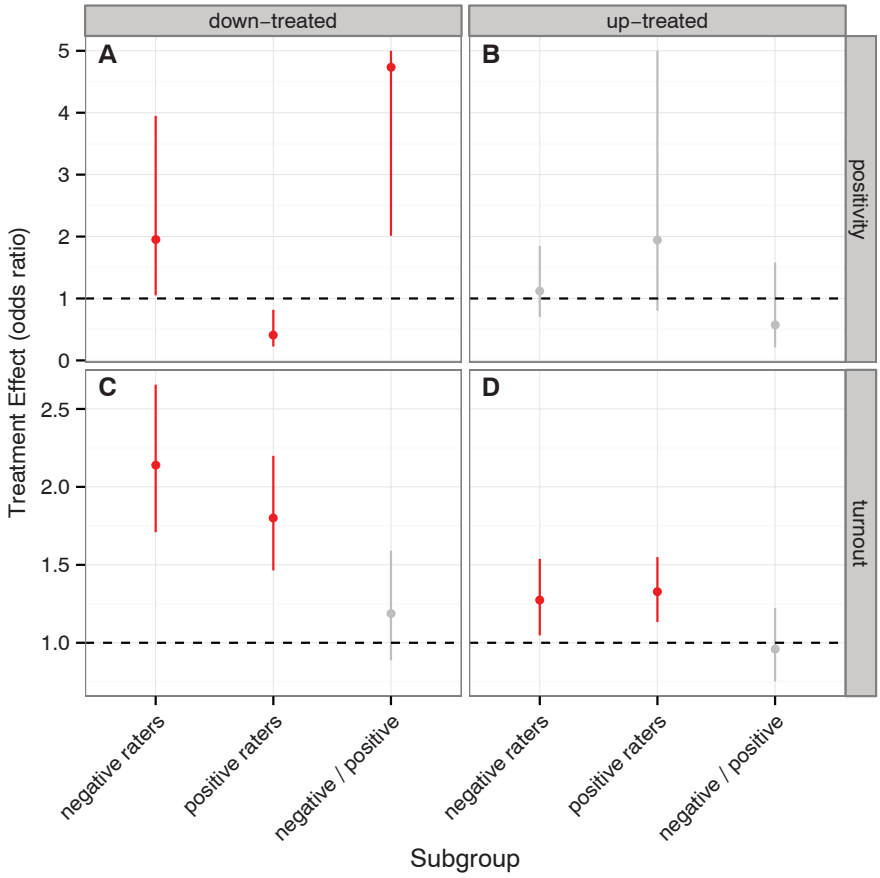

Fig. 5. Effects on turnout versus positivity. The figure displays treatment effects of the negative and positive manipulation on turnout (the likelihood of rating) and positivity (the proportion of positive votes) for negative (positive) raters, defined as active raters (with at least 100 votes on control comments) who display less than (greater than) the median proportion of positive votes on control comments. Results displayed in red are statistically significant at the 95% level, whereas results displayed in gray are not. (**A**) Treatment effect of the negative manipulation on positivity; (**B**) treatment effect of the positive manipulation on positivity; (**C**) treatment effect of the negative manipulation on turnout; and (**D**) treatment effect of the positive manipulation on turnout. In each panel, the first estimate is the treatment effect on negative raters, the second estimate is the treatment effect on positive raters, and the third estimate is the ratio of the treatment effects on negative/positive raters. The treatment effects displayed are odds ratios of the treatment effect on the treated compared to the control for negative and positive raters. The negative manipulation created countervailing opinion change for positive and negative raters, increasing the positivity of negative raters and decreasing the positivity of positive raters, whereas both treatments increased turnout uniformly across both subgroups.

behavioral mechanisms that together explain the experimental results (see "Estimating Behavioral Mechanisms" in the supplementary text).

First, both treatments increased turnout (by 82 and 28% for first viewers of down- and up-treated comments, respectively), but neither created differential turnout for different types of voters. This suggests that selection, or differential turnout by voter type (e.g., selecting different proportions of positive or negative voters, or frequent or infrequent voters), cannot explain our results. Second, we found statistically significant opinion change in two of the four subgroup dimensions. The positive treatment created a systematic increase in the proportion of positive ratings for raters with little prior experience rating the particular commenter whose comment was manipulated (a 7% increase in the ratio of positive to negative votes and a 50% increase in the probability of up-votes for these raters) and no decrease in positivity in any subgroup. This implies that positive opinion change explains part of the variation in positive herding. The negative treatment, by contrast, created countervailing opinion change for positive and negative raters, cancelling out any aggregate-level evidence of opinion change (Fig. 5). This in part explains why we find no aggregate trend in either direction for the negative treatment. Third, both treatments created a uniform increase in turnout across voter types. This overall increase in turnout, combined with a general preference for positivity on the site, created a tendency toward positive ratings under both treatments. Prior work has shown a preference for positivity in ratings in other contexts (*1*), which suggests that these results will generalize. Together, these findings suggest that a mixture of changing opinion and the natural tendency to up-vote together with greater turnout under both manipulations, combine to create the herding effects we see (see "Estimating Behavioral Mechanisms" in the supplementary text).

Our results demonstrate that whereas positive social influence accumulates, creating a tendency toward ratings bubbles, negative social influence is neutralized by crowd correction (*29*). Our findings suggest that social influence substantially biases rating dynamics in systems designed to harness collective intelligence. Future research that further explores the mechanisms driving individual and aggregate ratings—especially in in vivo social environments—will be essential to our ability to interpret collective judgment accurately and to avoid social influence bias in collective intelligence. We anticipate that our experiment will inspire more sophisticated analyses of social influence bias in electoral polling, stock market prediction, and product recommendation and that these results will be used to adapt rating and review technologies to account for social influence bias in their outputs.

References and Notes

1. J. A. Chevalier, D. Mayzlin, *J. Mark. Res.* **43**, 345–354 (2006).
2. D. G. Pope, *J. Health Econ.* **28**, 1154–1165 (2009).
3. C. Dellarocas, *Manage. Sci.* **49**, 1407–1424 (2003).
4. L. Cabral, A. Hortacsu, "The dynamics of seller reputation: Theory and evidence from eBay." National Bureau of Economic Research, working paper w10363 (2004).
5. M. J. Salganik, P. S. Dodds, D. J. Watts, *Science* **311**, 854–856 (2006).
6. A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, T. W. Malone, *Science* **330**, 686–688 (2010).
7. S. Bikhchandani, D. Hirshleifer, I. Welch, *J. Polit. Econ.* **100**, 992 (1992).
8. D. Hirshleifer, in *The New Economics of Human Behaviour*, M. Tommasi, K. Ierulli, Eds. (Cambridge Univ. Press, Cambridge, UK, 1995), pp. 188–215.
9. R. K. Merton, *Social Theory and Social Structure* (Free Press, New York, 1968).
10. H. A. Simon, *Biometrika* **42**, 425 (1955).
11. A.-L. Barabási, R. Albert, *Science* **286**, 509–512 (1999).
12. J.-P. Onnela, F. Reed-Tsochas, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18375–18380 (2010).
13. I. L. Janis, *Groupthink: Psychological Studies of Policy Decisions and Fiascoes* (Cengage Learning, Boston, MA, ed. 2, 1982).
14. J. Lorenz, H. Rauhut, F. Schweitzer, D. Helbing, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9020–9025 (2011).
15. C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, L. Lee, How opinions are received by online communities: A case study on amazon.com helpfulness votes, in *Proceedings of the 18th International Conference on World Wide Web* (Association for Computing Machinery, New York, 2009), pp. 141–150.

Downloaded from www.sciencemag.org on August 8, 2013

16. C. A. Hidalgo, R. A. Castro, C. Rodriguez-Sickert, *New J. Phys.* **8**, 52 (2006).
17. E. Bakshy, B. Karrer, L. A. Adamic, Social influence and the diffusion of user created content, in *Proceedings of the 10th ACM Conference on Electronic Commerce* (Association for Computing Machinery, New York, 2009), pp. 325–334.
18. J. Leskovec, A. Singh, J. Kleinberg, Patterns of influence in a recommendation network, in *Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science* (Springer, Berlin, Heidelberg, 2006), vol. 3918, pp. 380–389.
19. F. Wu, B. A. Huberman, *Internet and Network Economics* **5385**, 334–341 (2008).
20. N. Archak, A. Ghose, P. G. Ipeirotis, *Manage. Sci.* **57**, 1485–1509 (2011).
21. C. F. Manski, *Rev. Econ. Stud.* **60**, 531–542 (1993).
22. S. Aral, L. Muchnik, A. Sundararajan, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21544–21549 (2009).
23. S. Aral, *Mark. Sci.* **30**, 217–223 (2011).
24. S. Aral, D. Walker, *IEEE Intell. Syst.* **26**, 91–96 (2011).
25. S. Aral, D. Walker, *Manage. Sci.* **57**, 1623–1639 (2011).
26. S. Aral, D. Walker, *Science* **337**, 337–341 (2012).
27. C. R. Shalizi, A. C. Thomas, *Sociol. Methods Res.* **40**, 211–239 (2011).
28. C. L. Apicella, F. W. Marlowe, J. H. Fowler, N. A. Christakis, *Nature* **481**, 497–501 (2012).
29. D. Centola, R. Willer, M. Macy, *Am. J. Sociol.* **110**, 1009–1040 (2005).

Acknowledgments: We thank D. Eckles and D. Watts for invaluable discussions. This work was supported by a Microsoft Research faculty fellowship (S.A.) and by NSF Career Award 0953832 (S.A.). The research was approved by the New York University institutional review board. There are legal obstacles to making the data available and revealing the name of the Web site, but code is available upon request. All of the user data that we analyzed is publically available, except the treatment assignment and the random identifier that links a deidentified user to a vote, and therefore does not contain any other information that cannot be obtained by crawling the Web site. The randomized testing that the Web site performed is covered by the Web site's terms of service. Opt-in permissions were granted by the users when they registered for the Web site. No data on nonregistered users was collected.

Supplementary Materials
www.sciencemag.org/cgi/content/full/341/6146/647/DC1
Materials and Methods
Supplementary Text
Tables S1 to S21
Figs. S1 to S16
References (*30–57*)

14 May 2013; accepted 16 July 2013
10.1126/science.1240466

Pluripotent Stem Cells Induced from Mouse Somatic Cells by Small-Molecule Compounds

Pingping Hou,[1]* Yanqin Li,[1]* Xu Zhang,[1,2]* Chun Liu,[1,2]* Jingyang Guan,[1]* Honggang Li,[1]* Ting Zhao,[1]† Junqing Ye,[1,2]† Weifeng Yang,[3]† Kang Liu,[1]† Jian Ge,[1,2]† Jun Xu,[1]† Qiang Zhang,[1,2]† Yang Zhao,[1]‡ Hongkui Deng[1,2]‡

Pluripotent stem cells can be induced from somatic cells, providing an unlimited cell resource, with potential for studying disease and use in regenerative medicine. However, genetic manipulation and technically challenging strategies such as nuclear transfer used in reprogramming limit their clinical applications. Here, we show that pluripotent stem cells can be generated from mouse somatic cells at a frequency up to 0.2% using a combination of seven small-molecule compounds. The chemically induced pluripotent stem cells resemble embryonic stem cells in terms of their gene expression profiles, epigenetic status, and potential for differentiation and germline transmission. By using small molecules, exogenous "master genes" are dispensable for cell fate reprogramming. This chemical reprogramming strategy has potential use in generating functional desirable cell types for clinical applications.

Pluripotent stem cells, such as embryonic stem cells (ESCs), can self-renew and differentiate into all somatic cell types. Somatic cells can be reprogrammed to become pluripotent via nuclear transfer into oocytes or through the ectopic expression of defined factors (*1–4*). However, exogenous pluripotency-associated factors, especially *Oct4*, are indispensable for establishing pluripotency (*5–7*), and previous reprogramming strategies have raised concerns regarding the clinical applications (*8*, *9*). Small molecules have advantages because they can be cell permeable, nonimmunogenic, more cost-effective, and more easily synthesized, preserved, and standardized. Moreover, their effects on inhibiting and activating the function of specific proteins are often reversible and can be finely tuned by varying the concentrations. Here, we identified small-molecule combinations that were able to drive the reprogramming of mouse somatic cells toward pluripotent cells.

To identify small molecules that facilitate cell reprogramming, we searched for small molecules that enable reprogramming in the absence of *Oct4* using *Oct4* promoter-driven green fluorescent protein (GFP) expression (OG) mouse embryonic fibroblasts (MEFs), with viral expression of *Sox2*, *Klf4*, and *c-Myc*. After screening up to 10,000 small molecules (table S1A), we identified Forskolin (FSK), 2-methyl-5-hydroxytryptamine (2-Me-5HT), and D4476 (table S1B) as chemical "substitutes" for *Oct4* (Fig. 1, A and B, and figs. S1 and S2). Previously, we had developed a small-molecule combination "VC6T" [VPA, CHIR99021 (CHIR), 616452, Tranylcypromine], that enables reprogramming with a single gene, *Oct4* (*6*). We next treated OG-MEFs with VC6T plus the chemical substitutes of *Oct4* in the absence of transgenes. We found that VC6T plus FSK (VC6TF) induced some GFP-positive clusters expressing E-cadherin, a mesenchyme-to-epithelium transition marker, reminiscent of early reprogramming by transcription factors (*10*, *11*) (Fig. 1C and fig. S3). However, the expression of *Oct4* and *Nanog* was not detectable, and their promoters remained hypermethylated, suggesting a repressed epigenetic state (fig. S3).

To identify small molecules that facilitate late reprogramming, we used a doxycycline (DOX)–inducible *Oct4* expression screening system, adding DOX only in the first 4 to 8 days (*6*). Small-molecule hits, including several cAMP agonists (FSK, Prostaglandin E2, and Rolipram) and epigenetic modulators [3-deazaneplanocin A (DZNep), 5-Azacytidine, sodium butyrate, and RG108], were identified in this screen (fig. S4 and table S1B).

To achieve complete chemical reprogramming without the *Oct4*-inducible system, these small molecules were further tested in the chemical reprogramming of OG-MEFs without transgenes. When DZNep was added 16 days after treatment with VC6TF (VC6TFZ), GFP-positive cells were obtained more frequently by a factor of up to 65 than those treated with VC6TF, forming compact, epithelioid, GFP-positive colonies without clear-cut edges (Fig. 1, D and E, and fig. S5). In these cells, the expression levels of most pluripotency marker genes were elevated but were still lower than in ESCs, suggesting an incomplete reprogramming state (fig. S6). After switching to 2i-medium with dual inhibition (2i) of glycogen synthase kinase-3 and mitogen-activated protein kinase signaling after day 28 posttreatment, certain GFP-positive colonies developed an ESC-like morphology (domed, phase-bright, homogeneous with clear-cut edges) (Fig. 1F) (*12*, *13*). These colonies could be further cultured for more than 30 passages, maintaining an ESC-like morphology (Fig. 1, G and H). We refer to these 2i-competent, ESC-like, and GFP-positive cells as chemically induced pluripotent stem cells (CiPSCs).

Next, we optimized the dosages and treatment duration of the small molecules and were able to generate 1 to 20 CiPSC colonies from 50,000

[1]College of Life Sciences and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China. [2]School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen 518055, China. [3]Beijing Vitalstar Biotechnology Co., Ltd., Beijing 100012, China.

*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. E-mail: hongkui_deng@pku.edu.cn (H.D.); yangzhao@pku.edu.cn (Y.Z.)

Downloaded from www.sciencemag.org on August 8, 2013