

SCIENTIFIC REPORTS



OPEN

Arbitrary Inequality in Reputation Systems

Vincenz Frey¹ & Arnout van de Rijt^{1,2,3}

Received: 21 September 2016

Accepted: 07 November 2016

Published: 20 December 2016

Trust is an essential condition for exchange. Large societies must substitute the trust traditionally provided through kinship and sanctions in small groups to make exchange possible. The rise of internet-supported reputation systems has been celebrated for providing trust at a global scale, enabling the massive volumes of transactions between distant strangers that are characteristic of modern human societies. Here we problematize an overlooked side-effect of reputation systems: Equally trustworthy individuals may realize highly unequal exchange volumes. We report the results of a laboratory experiment that shows emergent differentiation between ex ante equivalent individuals when information on performance in past exchanges is shared. This arbitrary inequality results from cumulative advantage in the reputation-building process: Random initial distinctions grow as parties of good repute are chosen over those lacking a reputation. We conjecture that reputation systems produce artificial concentration in a wide range of markets and leave superior but untried exchange alternatives unexploited.

Trust problems hamper mutually beneficial exchange across a broad swath of contexts. Trust is an issue whenever exchange requires that one party – the “trustor” – first expose herself to the risk of abuse by the other party – the “trustee”. Abuse may involve failed delivery, compromised quality, shirking, theft, physical violence, or disclosure of sensitive information^{1–3}. The threat of direct punishment by the trustor⁴ and the possibility that the trustor withdraws from future exchanges^{5–7} can mitigate the trust problem by providing incentives for trustworthiness. Legal institutions may also deter untrustworthy behavior and offer partial compensation for a trustor in the case of abuse⁸. However, in modern cross-border exchanges over the internet, these mechanisms often cannot warrant trust. Direct punishment is ineffective if the costs for the trustor are high⁹, the likelihood of repeat business is low¹⁰, and the cost of effective legal recourse is prohibitive so that trustees are not incentivized to honor trust.

Reputation systems can enable exchange when other mechanisms fall short^{5,11–17}. Reputation systems collate information on past exchanges voluntarily shared by trustors^{10,18,19}. This allows trustors to learn from the experiences of others and to selectively exchange with trustees of good repute^{5,20}. At the same time, reputation systems help incentivize trustees to act honorably, as a bad reputation prevents future exchange⁵. While prominent historical examples of reputation systems exist^{10,13}, recent technological advances have made the sharing of reputation information possible at minimal cost and unprecedented scale, enabling otherwise infeasible transactions across vast numbers of geographically dispersed parties.

Here we study an overlooked side-effect of reputation systems: Reputation building exhibits a form of cumulative advantage^{21–30}, resulting in arbitrary inequality in transaction volume among trustees. To minimize the risk of abuse, trustors may avoid many trustees who lack a transaction history in favor of a single trustee who was found trustworthy before. Random initial distinctions thus grow as parties of good repute are chosen over those lacking a reputation. The unintended consequence is a “reputation cascade” that keeps increasing the reputational advantage of one party while preventing others from building a reputation. (See model in Supplementary Information.) This inequality is arbitrary when the excluded parties are no less trustworthy than the trusted party.

Experiments

We tested the emergence of reputation cascades in a laboratory experiment. The experimental protocol was checked and approved by the IRB of Stony Brook University (CORIHS #2014-2787 F). The experiment was subsequently carried out in accordance with the approved protocol. 336 subjects played games in groups of four trustors and four trustees (Supplementary Information). A game consisted of one or more rounds and ended

¹Department of Sociology, Utrecht University, Padualaan 14, Utrecht, 3584CH, The Netherlands. ²Department of Sociology, Stony Brook University, 100 Nicolls Road, Stony Brook, NY 11794, USA. ³Institute for Advanced Computational Science, Stony Brook University, 100 Nicolls Road, Stony Brook, NY 11794, USA. Correspondence and requests for materials should be addressed to V.F. (email: v.c.frey@uu.nl)

after each round with probability 1/6 (Supplementary Information). In every round, one of the trustors chose whether to place trust in one of the trustees or to withhold trust. A selected trustee chose whether to honor or abuse trust. Games were played in turn-taking style, with trustor 1's turn in rounds 1, 5, etc., and trustor 2's turn in rounds 2, 6, etc.

Game play yielded subjects points that converted to 1.5 US dollar cents. In any round a trustor withholding trust, any trustor not in turn, as well as any unchosen trustee received 30 points. Honored trust paid the trustor and chosen trustee 50 points each. Abused trust resulted in 0 points for the victimized trustor and T points for the abusing trustee. Games were played in two trust conditions. In the condition "Trust Problem", T was 80 or 100 (Supplementary Information), so that the trustee earned a higher monetary payoff from abusing trust than from honoring trust. In the condition "No Trust Problem", T was 0 points, rendering the trustee's payoff of honoring trust higher than the payoff of trust abuse.

Games were played in three different reputation conditions. In the "Private" condition, the computer interface showed trustors only the results of their own turns (Supplementary Information), preventing cascading. In the "Partial" condition, the pair of even-numbered trustors (trustors 2 and 4) and the pair of odd-numbered trustors (trustors 1 and 3) could also see the results of each other's turns, allowing each pair to coordinate on a focal trustworthy trustee. In the "Full" condition, trustors could see the results of all turns, making it possible for all to rally around a single trustworthy trustee. Accordingly, we expect inequality in exchange volume between trustees in the Trust Problem condition to increase from Private to Partial to Full. Furthermore, inequality should be smaller in the No Trust Problem conditions where trustors lack an incentive to avoid untried trustees.

This experimental approach has three advantages over use of observational data on reputation systems. First, the distributional extremities others have observed in situations of reputation-enabled trust^{10,31,32}, while consistent with our argument, could also be caused, for instance, by trustee variability in quality, visibility, or price. In our design alternative sources of inequality are precluded through experimental control. Second, information sharing also enables forms of feedback other than reputation cascades such as information cascades³³, social influence^{20,27,29}, and increasing returns³⁴. While these forms of feedback are not driven by avoidance of trust abuse, they could also generate inequality in reputation systems. Our design allows isolating reputation cascades as the inequality-generating mechanism through a comparison of behavior in the Trust Problem condition and the No Trust Problem condition. Third, the mutual exclusivity of reputation information available to even- and odd-numbered trustors in the Partial condition allows a direct demonstration of arbitrariness in trustee selection as it makes it possible for two cascades to form involving two distinct trustees.

Results

We measured the prevalence of cascading as the proportion of times a trustor selected the trustee that had been selected on the most recent turn the trustor could observe, provided that trust had been honored (Fig. 1; Supplementary Information). Under the null-hypothesis that trustors randomly choose one of the four trustees every time they place trust, cascades should continue in only 25% of all cases. In the Trust Problem condition, cascades instead continued 59% of the time. This percentage increased monotonically as cascades grew in length, reaching 100% for cascades of length 5 or more (Supplementary Table 1). Remarkably, in the absence of a trust problem (Fig. 1: "No Trust Problem") the propensity for cascading completely vanished; with 17%, cascade continuation was even lower than expected under random trustee selection. These results show that subjects formed cascades not because of shared preferences for a particular trustee identity and not because of a general tendency to imitate the choices of others, but entirely out of fear of abuse.

In the Trust Problem condition greater degrees of information sharing produced higher levels of honored trust (Supplementary Table 2), confirming earlier studies that found that information sharing facilitates exchange^{9,10,35}. However, as a result of feedback in trustee selection enabled by information sharing, these gains in trust came with increased differentiation in exchange volume. Figure 2 shows that inequality in how often trustees were chosen, measured using the modified coefficient of variation³⁶ (Supplementary Information), increases monotonically from Private to Partial to Full information sharing.

To assess arbitrariness in trustee selection we exploited the feature of the Partial condition that the even- and odd-numbered pairs of trustors could not see one another's choices, by comparing how often a trustee was selected by either pair. If the inequalities produced under information sharing merely reflected differences in trustworthiness across trustees, a trustee who was often chosen by the odd-numbered trustors should also have been chosen often by the even-numbered trustors, and vice-versa. Instead Fig. 3 shows that in many cases, a trustee who was often chosen by one trustor pair was never chosen by the other pair. To statistically establish arbitrariness in trustee selection we compared the disagreement in trustee choices between pairs of information-sharing trustors and pairs of non-information-sharing trustors. We find that disagreement – the difference in the number of times two trustors trusted a given trustee, summed across the four trustees – to be significantly larger among non-information-sharing trustors than among information-sharing trustors (nested linear regression, $p = 0.007$; see Supplementary Information). This demonstrates that trustees were selected or excluded from exchange based on path-dependent histories of reputation-building.

Discussion

We conjecture that reputation cascades produce arbitrary inequality in a wide variety of everyday exchange settings that differ from the sterile conditions created in our laboratory. First, in unregulated economic exchange, established trustees can offer prices low enough to be preferable over the cheaper but riskier offers of newcomers. Laboratory experiments show that indeed trustors are willing to forgo more lucrative offers of untested parties in favor of the relative safety of exchanging in ongoing relationships with proven partners^{37–41}. Newcomers will face

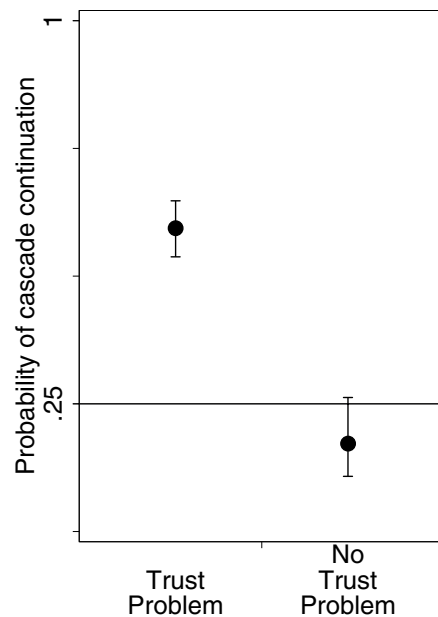


Figure 1. Cascading in situations with and without a threat of trust abuse. Shown is the proportion of times a trustor selected the trustee that had been selected on the preceding turn the trustor observed, given that trust had been honored. Confidence intervals are obtained from nested logistic regressions (see Supplementary Information).

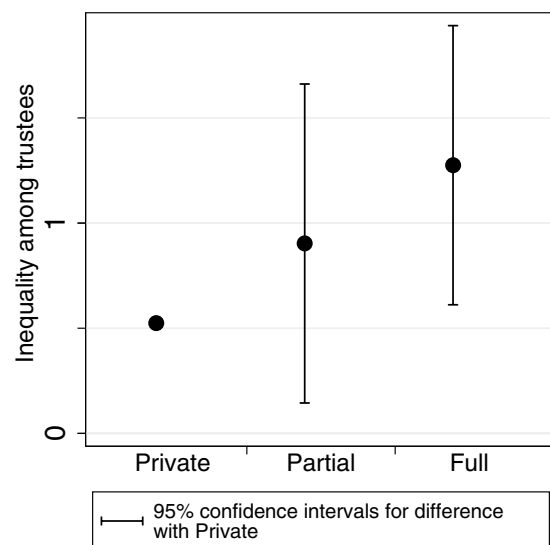


Figure 2. Effects of information sharing on inequality, when incentives for abuse are present (Trust Problem condition). Shown for each reputation condition is inequality among trustees in the number of times trusted in a game. 95% confidence intervals from nested linear regression models (Supplementary Information) indicate the significance of differences of Partial and Full with Private.

even greater barriers to entry in settings where market leaders can accumulate resources that grant them greater capacity to undercut prices.

Second, our finding that trustees with longer records of trustworthy behavior were more often chosen than those with shorter records suggests that reputation cascades are robust against random deviations. While the incidental choice of an untried trustee allows a demonstration of trustworthiness, that single act does not neutralize its disadvantage in attracting subsequent trustors vis-à-vis well-established competitors. Preferences for long-standing reputations over marginal ones will likely be even stronger in everyday settings where information is not always accurate and trustees can fabricate positive ratings (compare ref. 42).

While arbitrary inequality constitutes one undesired outcome^{43–45}, reputation cascades may have additional adverse effects. Sometimes market concentration will come with oligopolistic inefficiencies. In other instances,

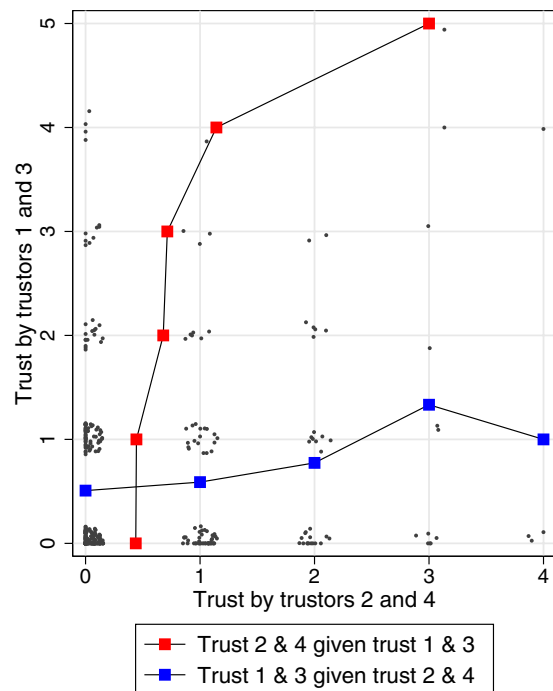


Figure 3. Arbitrariness of trustee selection in experimental reputation systems. Shown is the number of times a trustee was trusted by trustors 1 and 3 by the number of times that trustee was trusted by trustors 2 and 4. Data come from the Partial X Trust Problem condition, where pairs of trustors formed mutually exclusive information sharing groups and incentives for trust abuse were present.

exchange opportunities that provide a proximity advantage or superior value are foregone⁴⁶. Reputational feedback may also generate discrimination against groups marked by a systematic shortage of credit such as youth, innovators, and migrants. Public policy interventions – such as the EU’s prohibition of considering track records when comparing bids in public procurement or Germany’s ban on landlord requests for registries on timely payment from potential tenants – can aid in mitigating the negative consequences of reputation cascades.

References

1. Coleman, J. *Foundations of Social Theory* 91–116 (Belknap Press of Harvard Univ. Press, Cambridge, MA, 1990).
2. Dasgupta, P. In *Trust: Making and Breaking Cooperative Relations* (ed Gambetta, D.) 49–72 (Blackwell, Oxford, 1988).
3. Kreps, D. M. *Game Theory and Economic Modelling* (Clarendon, Oxford, 1990).
4. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
5. Buskens, V. & Raub, W. Embedded trust: control and learning. *Adv. Group Process.* **19**, 167–202 (2002).
6. Hirschman, A. O. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (Harvard Univ. Press, Cambridge, MA, 1970).
7. Lahno, B. Trust, reputation, and exit in exchange relationships. *J. Conf. Resolut.* **39**, 495–510 (1995).
8. Macaulay, S. Non-contractual relations in business: a preliminary study. *Am. Sociol. Rev.* **28**, 55–67 (1963).
9. Egas, E. & Riedl, A. The economics of altruistic punishment and the maintenance of cooperation. *Phil. Trans. R. Soc. B* **275**, 871–878 (2008).
10. Diekmann, A., Jann, B., Przepiorka, W. & Wehrli, S. Reputation formation and the evolution of cooperation in anonymous online markets. *Am. Sociol. Rev.* **79**, 65–85 (2014).
11. Bolton, G. E., Katok, E. & Ockenfels, A. How effective are electronic reputation mechanisms? An experimental investigation. *Manage. Sci.* **50**, 1587–1602 (2004).
12. Cook, K. S., Snijders, C., Buskens, V. & Cheshire, C. (eds) *eTrust: Forming Relationships in the Online World* (Russell Sage, New York, 2009).
13. Klein, D. B. *Reputation: Studies in the Voluntary Elicitation of Good Conduct* (University of Michigan Press, Ann Arbor, 1997).
14. Resnick, P., Zeckhauser, R., Friedman, E. & Kuwabara, K. Reputation systems. *Comm. ACM* **43**, 45–48 (2000).
15. Wu, J., Balliet, D. & Van Lange, P. A. M. Gossip versus punishment: the efficiency of reputation to promote and maintain cooperation. *Sci. Rep.* **6**, 23919 (2016).
16. Milinski, M. Reputation, a universal currency for human social interactions. *Phil. Trans. R. Soc. B* **371**, 20150100 (2016).
17. Kuwabara, K. Do reputation systems undermine trust? Divergent effects of enforcement type on generalized trust and trustworthiness. *Am. J. Sociol.* **120**, 1390–1428 (2015).
18. Heyes, A. & Kapur, S. Angry customers, e-word-of-mouth and incentives for quality provision. *J. Econ. Behav. Organ.* **84**, 813–828 (2012).
19. Feinberg, M., Willer, R., Stellar, J. & Keltner, D. The virtues of gossip: reputational information sharing as prosocial behavior. *J. Pers. Soc. Psychol.* **102**, 1015–1030 (2012).
20. Feinberg, M., Willer, R. & Schultz, M. Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* **25**, 656–664 (2014).
21. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
22. Denrell, J. & Le Mens, G. Interdependent sampling and social influence. *Psychol. Rev.* **114**, 398–422 (2007).
23. DiPrete, T. A. & Eirich, G. M. Cumulative advantage as a mechanism for inequality: a review of theoretical and empirical developments. *Annu. Rev. Sociol.* **32**, 271–297 (2006).

24. Lynn, F. B., Podolny, J. M. & Tao, L. A sociological (de) construction of the relationship between status and quality. *Am. J. Sociol.* **115**, 755–804 (2009).
25. Manzo, G. & Baldassarri, D. Heuristics, interactions, and status hierarchies: an agent-based model of deference exchange. *Sociol. Methods Res.* **44**, 329–387 (2015).
26. Merton, R. K. The Matthew effect in science: the reward and communication systems of science are considered. *Science* **159**, 56–63 (1968).
27. Muchnik, L., Aral, S. & Taylor, S. Social influence bias: a randomized experiment. *Science* **341**, 647–651 (2013).
28. Petersen, A. M., Jung, W.-S., Yang, J.-S. & Stanley, H. E. Quantitative and empirical demonstration of the Matthew Effect in a study of career longevity. *Proc. Natl. Acad. Sci. USA* **108**, 18–23 (2011).
29. Salganik, M. J., Dodds, P. S. & Watts, D. J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).
30. Van de Rijt, A., Kang, S., Restivo, M. & Patil, A. Field experiments of success-breeds-success dynamics. *Proc. Natl. Acad. Sci. USA* **111**, 6934–6939 (2014).
31. Barwick, P. J. & Pathak, P. A. The costs of free entry: an empirical study of real estate agents in Greater Boston. *RAND J. Econ.* **46**, 103–145 (2015).
32. Snijders, C. & Weesie, J. In *eTrust: Forming Relationships in the Online World* (eds Cook, K. S., Snijders, C., Buskens, V. & Cheshire, C.) 166–185 (Russell Sage, New York, 2009).
33. Bikhchandani, S., Hirshleifer, D. & Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **100**, 992–1026 (1992).
34. Arthur, W. B. Competing technologies, increasing returns, and lock-in by historical events. *Econ. J.* **99**, 116–131 (1989).
35. Cuesta, J. A., Gracia-Lázaro, C., Ferrer, A., Moreno, Y. & Sánchez, A. Reputation drives cooperative behaviour and network formation in human groups. *Sci. Rep.* **5**, 7843 (2015).
36. Allison, P. D. Estimation and testing for a Markov model of reinforcement. *Sociol. Methods Res.* **8**, 434–453 (1980).
37. Brown, M., Falk, A. & Fehr, E. Relational contracts and the nature of market interactions. *Econometrica* **72**, 747–780 (2004).
38. Cook, K. S., Rice, E. & Gerbasi, A. In *Creating Social Trust in Post-Socialist Transition* (eds Kornai, J., Rothstein, B. & Rose-Ackerman, S.) 193–212 (Palgrave Macmillan, New York, 2004).
39. Kollock, P. The emergence of exchange structures: an experimental study of uncertainty, commitment, and trust. *Am. J. Sociol.* **100**, 313–345 (1994).
40. Simpson, B. & McGrimmon, T. Trust and embedded markets: a multimethod investigation of consumer transactions. *Soc. Networks* **30**, 1–15 (2008).
41. Yamagishi, T., Cook, K. S. & Watabe, M. Uncertainty, trust, and commitment formation in the United States and Japan. *Am. J. Sociol.* **104**, 165–194 (1998).
42. Bozoyan, C. & Vogt, S. The impact of third-party information on trust: valence, source, and reliability. *PLoS ONE* **11**, e0149542 (2016).
43. Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. Egalitarian motives in humans. *Nature* **446**, 794–796 (2007).
44. Engelmann, D. & Strobel, M. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *Am. Econ. Rev.* **94**, 857–869 (2004).
45. Macro, D. & Weesie, J. Inequalities between others do matter: evidence from multiplayer dictator games. *Games* **7**, 11 (2016).
46. Nelson, P. Information and consumer behavior. *J. Polit. Econ.* **78**, 311–329 (1970).

Acknowledgements

We thank I. Akin, E. Eftekhari, U. Senn, and H.-G. Song for assistance in conducting the experiment and V. Buskens, R. Corten, J. Jones, M. Mäs, W. Przepiorka, and W. Raub for comments. This work was supported by the National Science Foundation, grant SES-1340122 to A.v.d.R.

Author Contributions

V.F. and A.v.d.R. designed the project and the experiment, performed the statistical analyses, and wrote the manuscript; V.F. developed the formal theory and programmed and executed the experiment.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Frey, V. and van de Rijt, A. Arbitrary Inequality in Reputation Systems. *Sci. Rep.* **6**, 38304; doi: 10.1038/srep38304 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016