

**John F. Kennedy School of Government  
Harvard University  
Faculty Research Working Papers Series**

**Repetition and Reputation:  
Implications for Trust and Trustworthiness in the Short and in the Long Run**

**Iris Bohnet and Steffen Huck**

**Nov 2003**

**RWP03-048**

The views expressed in the KSG Faculty Research Working Paper Series are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or Harvard University. All works posted here are owned and copyrighted by the author(s). Papers may be downloaded for personal use only.

**Repetition and Reputation:  
Implications for Trust and Trustworthiness in the Short and in the Long Run**

Iris Bohnet<sup>\*</sup> and Steffen Huck<sup>\*\*</sup>

Repeat transactions are not necessarily the rule in today's global economy. Indirect reputation systems, where buyers base their decisions on a seller's previous interactions with other buyers, are a potential substitute for personal interactions—provided such information is available. On eBay, for example, buyers are willing to pay a premium of 8.1% of the selling price to a seller with an established good reputation (Paul Resnick, Richard Zeckhauser, John Swanson and Kate Lockwood 2003). This paper examines experimentally to what degree indirect reputation building substitutes for direct reputation building in repeat interactions in the short run and analyzes the effects these environments have on behavior in the long run.

While a large number of earlier experimental studies compares one-shot with repeat interactions in the short run, indirect reputation building and long-term effects have received comparatively little attention so far.<sup>1</sup> Focusing on the long run we can analyze institutional change—how long it takes people to transition from one institutional environment to another, a question of paramount importance for the formerly Communist countries and many private and public organizations (see, e.g., Rose-Ackerman 2001), and whether trust and trustworthiness can be fostered by first exposing people to an environment conducive to trust. We are interested in whether (intrinsic) trust and trustworthiness can be induced in the long run by providing extrinsic incentives for trust and trustworthiness in the short run.

Reputation systems may provide incentives for trustworthiness and trust. Direct reputation building may occur in repeated games where pairs of subjects play the same stage

game for a finite number of times. Indirect reputation building may occur in population games where agents are randomly re-matched in every period but receive information about their opponents' past play. We compare the effects of direct and indirect reputation building in a binary-choice trust game where a buyer, the trustor, can either interact with the seller, the trustee, or exit. The trustee can either honor or exploit trust. The payoffs are such that a money-maximizing trustee prefers exploiting to honoring trust in a single-shot game—while a money-maximizing trustor prefers not offering trust to being exploited. The unique Nash equilibrium of the single-shot game predicts no trade. Figure 1 presents the game we implemented with the actual payoffs in cents used.

*Figure 1 about here*

In our experiment, subjects participate in the trust game in two blocks of 10 rounds each, which is common knowledge.<sup>2</sup> In phase 1, the first 10 rounds, they are confronted either with a standard, “one shot” random matching treatment (“stranger”); a fixed-pairs, “finitely repeated game” treatment (“partner”); or a random matching treatment (“reputation-stranger”). In the latter, trustors are informed about their trustees' past behavior in each round. In phase 2, rounds 11-20, all subjects interact in the stranger environment (without information about the past). 312 subjects participated in our experiment; 96 in the stranger treatment (4 sessions), 102 in the partner treatment (5 sessions) and 114 in the reputation-stranger treatment (4 sessions). Roles were randomly assigned and kept fix during the experiment.

For the short run (phase 1), standard theoretical models, where all agents are pure money maximizers and where this is commonly known, suggest that there should be no differences in behavior between one-shot games (with or without reputational information) and finitely repeated games. In contrast, models incorporating incomplete information about agents'

preferences and/or rationality allow for reputation building, directly and indirectly (David Kreps, Paul Milgrom, John Roberts and Robert Wilson 1982). If there are such reputation effects, there should be more trust and trustworthiness in the partner and the reputation-stranger treatments than in the stranger treatment in the short run and a decrease of trust and trustworthiness towards the end of the first 10 rounds. Our results for the first phase are in line with this prediction.

For the long run (phase 2), all orthodox models predict the same behavior. We should see low (or zero) levels of trust and trustworthiness in all three treatments since incentives for building a reputation have been removed. In that sense, orthodox models predict that history does not matter. If, on the other hand, reputation-based interactions “crowd in” trust and trustworthiness or evoke specific norms of behavior, differences between the treatments might be observed. Theoretically, such long-term effects require either changes in preferences (Bohnet, Frey and Huck 2001) or some inertia in adjustment and learning as, for example, in reinforcement learning models (Ido Erev and Alvin Roth 1998).

In this paper, we take an empirical approach and examine whether there are any history effects and, if so, whether they are systematic. In particular, we estimate subjects’ propensity to trust (or to be trustworthy) in the second phase of the experiment as a function of the institution they were exposed to in the first phase; their experiences in the first and second phase; their type (as measured by their initial propensity to trust and be trustworthy); and time. In our data, we find that subjects do understand changes in the incentive structure and fully discount previous experiences if they were not gained in the same environment. While this is in line with orthodox theory, our second main finding challenges it. We find that exposure to a partner treatment makes trustees more trustworthy in the long run. Partner and reputation-stranger treatments produce similar results in the short but not in the long run.

## I. Experimental results

Table 1 presents average trust and trustworthiness rates for each treatment and the two phases of the experiment (Figures 1 and 2 in the Appendix present the data by round for each of our treatments). Trust rates indicate the fraction of trustors offering trust in a given round; trustworthiness rates indicate the fraction of trustees honoring trust in a given round, conditional on having been offered trust.

*Table 1 about here*

In phase 1, trustors are significantly more likely to trust in the partner treatment than in the stranger treatment, and somewhat more likely to trust in the reputation-stranger treatment than in the stranger treatment. Trustees are equally likely to honor trust in the partner and the reputation-stranger treatments and significantly less likely to do so in the stranger treatment.<sup>3</sup> Our results show the existence of rather strong reputation effects, especially for trustees.

In phase 2, trust and trustworthiness rates in the partner treatment are slightly higher than in the two stranger treatments. The differences between the partner and the reputation-stranger treatments are significant,<sup>4</sup> i.e. there are history effects. To get a clearer picture of subjects' behavior in the second phase we estimate linear probability models for trustors' propensity to trust and trustees' propensity to be trustworthy.<sup>5</sup> We run two random-effects panel regressions:

$$\begin{aligned} \text{RATE}_{i,t} = & \delta C + \beta \text{RS}_i + \gamma \text{P}_i + \delta \text{EFP}_i + \epsilon \text{EFP}_i * \text{RS} + \zeta \text{EFP}_i * \text{P} + \eta \text{ESP}_{i,t} + \theta \text{ESP}_{i,t} * \text{RS} + \iota \text{ESP}_{i,t} * \text{P} \\ & + \kappa \text{TYPE}_i + \lambda \text{TYPE}_i * \text{RS} + \mu \text{TYPE}_i * \text{P} + \nu \text{RD}_t + \xi \text{RD}_t * \text{SR} + \omicron \text{RD}_t * \text{P} + \nu_i + e_{i,t} \end{aligned}$$

where  $\text{RATE}_{i,t}$  is subject  $i$ 's probability to trust (in the first regression) or be trustworthy (in the second regression).  $C$  is the constant;  $\text{RS}_i$  is a dummy variable equal to 1 if subject  $i$  is in the reputation-stranger treatment and zero otherwise;  $\text{P}_i$  is a dummy variable equal to 1 if subject  $i$  is in the partner treatment and zero otherwise;  $\text{EFP}_i$  is subject  $i$ 's experience during the first

phase—rounds 1-10 (i.e. trustor’s experienced trustworthiness rate or trustee’s experienced trust rate);  $ESP_{i,t}$  is subject  $i$ ’s experience up to period  $t-1$  in the second phase—rounds 11-20 (experienced trustworthiness rate for trustor and experienced trust rate for trustee up to  $t-1$ );  $TYPE_i$  captures subject  $i$ ’s initial propensity to trust or be trustworthy (a dummy variable for the trustor equal to 1 if the subject trusted in round 1 and 0 otherwise; and the average trustworthiness rate of subject  $i$  in the first 10 rounds for the trustee)<sup>6</sup>;  $RD_t$  is the round;  $v_i$  the idiosyncratic random-effect of subject  $i$ , and  $e_{i,t}$  the error term. Table 2 shows the results.

*Table 2 about here*

For trustors we find: The experience of trustworthiness in the first phase (EFP) and a subject’s initial propensity to trust (TYPE) only matter in the stranger treatment (without institutional change). In the other treatments where trustworthiness can be strategic in the first phase, neither others’ nor own actions in the first phase affect trust in the non-strategic second phase. In contrast, recent experiences in the previous rounds of phase 2 (ESP) are important in all treatments. While subjects in the partner treatment appear to be somewhat less sensitive to this, the more trustworthiness subjects have recently experienced, the more willing they are to trust. There is a significant time trend in all treatments: Subjects become less trusting as the end nears.

For trustees we find: The experience of trust does not matter for trustworthiness, independent of whether the experience was gained in the first or the second phase (EFP or ESP). Trust does not breed trustworthiness. A subject’s propensity to be trustworthy in the first phase (TYPE) is only relevant when it was non-strategic and, thus, a true matter of type. Finally, subjects in the partner treatment are substantially more trustworthy than subjects in the reputation-stranger treatment and a little more trustworthy than those in the stranger treatment. While Figure 2 suggests that trustworthiness decreases over time, the regressions show that this

is an artifact of the matching. There is no systematic decline in trustworthiness over time contrary to what trustors seem to expect.

## **II. Discussion and Conclusions**

Direct and indirect reputation systems increase trust and trustworthiness in the short run. Subjects strongly respond to the direct reputation building opportunities in a repeated game. With indirect reputation building, trustees appear to respond more strongly to the institutional environment than trustors. The benefits of this more complex and less familiar environment may not be as obvious as the advantages of repeat interactions, which may help explain why many consumers do not trust internet-based transactions using indirect reputation systems such as on eBay (Peter Kollock 1999). Our results suggest that they may be too pessimistic.

We also find that past experience is more relevant for trustors than for trustees. Experiences of trustworthiness increase the likelihood of trust in the same treatment; experiences of trust have no effect on the likelihood of trustworthiness, i.e. trust is not self-fulfilling as suggested, e.g., by Michael Bacharach and Diego Gambetta (2001).

Finally, and perhaps most importantly, we find that there are significant history effects. In particular, we find that trustees are more trustworthy after having been exposed to a partner treatment. The partner treatment is the most effective institutional arrangement to foster trust and trustworthiness in the short *and* in the long run. Indeed, it appears as if experiencing the intimate partner relationship (where trust and trustworthiness are perhaps more natural than in the reputation-stranger environment) breeds genuine trustworthiness. This might have important implications for issues in institutional design and education. Interactions in small closely-knit groups may have long-lasting beneficial consequences.

## References

- Andreoni, James. "Why Free Ride? Strategies and Learning in Public Goods Experiments." Journal of Public Economics, 1988, 37, pp. 291-304.
- Andreoni, James and Rachel Croson. "Partners versus Strangers: Random Rematching in Public Goods Experiments." In: Handbook of Experimental Economic Results. 2002. Forthcoming.
- Bacharach, Michael and Diego Gambetta. "Trust in Signs." In: K. Cook, ed., Trust in Society, New York: Russell Sage Foundation, 2001, pp. 148-184.
- Bohnet, Iris, Bruno S. Frey and Steffen Huck. "More order with less law: On Contract enforcement, trust and crowding." American Political Science Review, 2001, 95(1), pp. 131-144.
- Croson, Rachel. "Partners and strangers revisited," Economics Letters, 1996, 53, pp. 25-32.
- Erev, Ido and Alvin E. Roth. "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria." American Economic Review, 1998, 88(4), pp. 848-881.
- Fehr, Ernst and Simon Gächter. "Do Incentive Contracts Crowd Out Voluntary Cooperation?" Working paper, Institute for Empirical Research in Economics, University of Zurich, 2003.
- Fischbacher, Urs. "zTree: A Toolbox for Readymade Economic Experiments." Working paper, Institute for Empirical Research in Economics, University of Zurich, 1999.
- Kreps, David, Paul Milgrom, John Roberts and Robert Wilson. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." Journal of Economic Theory, 1982, 27, pp. 245-252.
- Kollock, Peter. "The production of trust in online markets," in: E.J. Lawler, M.W. Macy, S.R. Thye and H.A. Walker, eds., Advances in group processes, Greenwich, CT: JAI Press, 1999, pp. 99-123.
- Resnick, Paul, Richard Zeckhauser, John Swanson and Kate Lockwood. "The Value of Reputation on E-Bay: A Controlled Experiment." Working paper, Kennedy School of Government, Harvard University, 2003.
- Rose-Ackerman, Susan. "Trust, Honesty, and Corruption: Reflection on the State-Building Process." Working paper, Program for Studies in Law, Economics, and Public Policy, Yale Law School, 2001.

Figure 1: The trust game

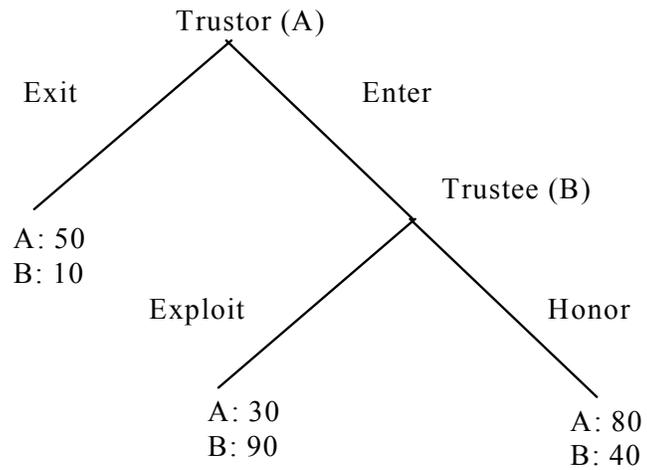


Table 1: Trust and trustworthiness rates in phase 1 (r. 1-10) and phase 2 (r. 11-20)

Treatment	Trust rate		Trustworthiness rate	
	Phase 1	Phase 2	Phase 1	Phase 2
S: Stranger (N=48)	0.32	0.23	0.30	0.18
R: Reputation-stranger (N=57)	0.43	0.19	0.55	0.18
P: Partner (N=51)	0.59	0.32	0.61	0.28

Table 2: Estimated trust and trustworthiness rates in phase 2—rounds 11-20

	Trust	Trustworthiness
Reputation-Stranger	-0.098 (0.161)	0.185 (0.389)
Partner	0.212 (0.168)	0.652 ^ (0.364)
EFP (Experience in first phase)	0.309 ^ (0.172)	-0.355 (0.317)
EFP * Reputation-Stranger	-0.226 (0.212)	0.019 (0.419)
EFP * Partner	-0.226 (0.211)	-0.151 (0.360)
ESP (Experience in second phase up to t-1)	0.444 ** (0.109)	-0.483 (0.682)
ESP * Reputation-Stranger	0.020 (0.144)	0.418 (0.935)
ESP * Partner	-0.208 (0.130)	0.984 (0.801)
Type	0.153 * (0.076)	0.492 ** (0.144)
Type * Reputation-Stranger	-0.142 (0.101)	-0.485 * (0.217)
Type * Partner	0.101 (0.103)	-0.522 * (0.230)
Round	-0.031 ** (0.006)	-0.007 (0.018)
Round * Reputation-Stranger	-0.008 (0.008)	-0.009 (0.025)
Round * Partner	-0.005 (0.008)	-0.034 (0.024)
Constant	0.564 ** (0.114)	0.029 (0.277)
#Observations	1380	289
#Subjects	138	119
R-square	0.171	0.130

Linear probability regressions, standard errors in parentheses.

^ for  $p < 0.1$ , \* for  $p < 0.05$ , \*\* for  $p < 0.01$ .

Figure 1: Trust rates

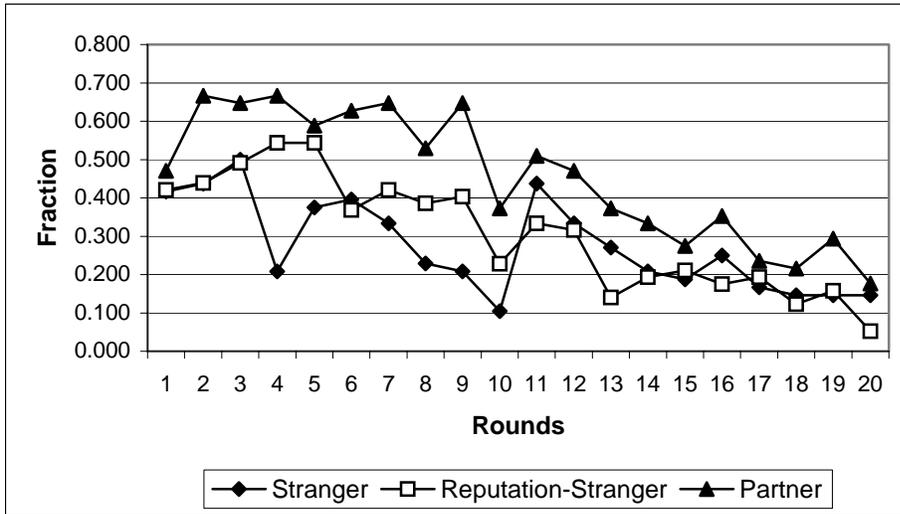
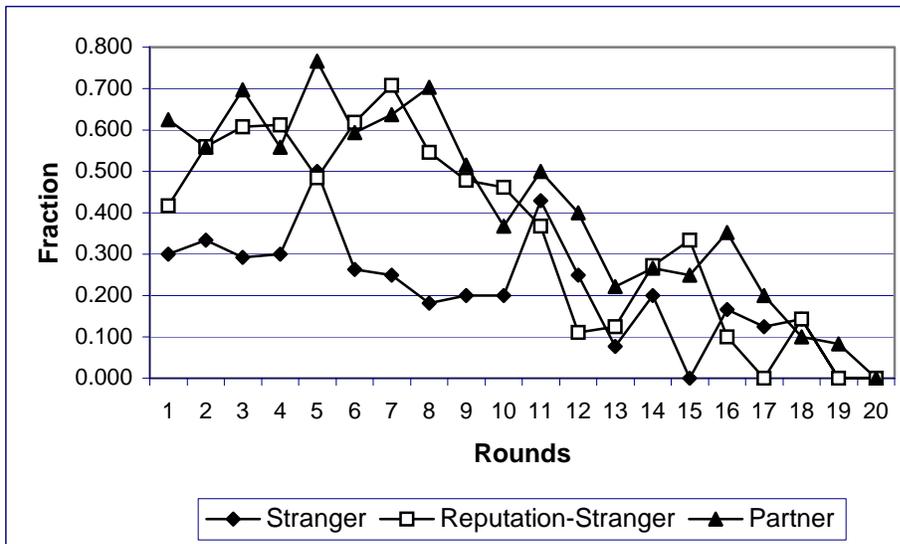


Figure 2: Trustworthiness rates



---

\* Corresponding author: Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA02139, email: Iris\_Bohnet@Harvard.edu, phone: (617) 495-5605, fax: (617) 496-5747.

\*\* Department of Economics and ELSE, University College London, Gower Street, London WC1E 6BT, UK, email: s.huck@ucl.ac.uk, phone: +44 207679 5895.

<sup>1</sup> For a recent survey, see James Andreoni and Rachel Croson (2002). Long-term effects have mainly been studied in experiments employing the same treatment condition in several blocks of rounds (“restart effects”, see, e.g., Andreoni 1988, Croson 1996). Exceptions comparing the effects of different institutional environments over time include Iris Bohnet, Bruno S. Frey and Steffen Huck (2001) and Ernst Fehr and Simon Gächter (2003) who both find evidence for institutional spill-over effects.

<sup>2</sup> Subjects are informed that there will be a “restart” in round 11 and that new information concerning rounds 11-20 will be provided in round 11. Subjects were recruited from universities in the greater Boston area and earned on average \$11 (plus a show-up fee of \$10). Sessions lasted approximately 1 hour. The experiments were computerized using Urs Fischbacher’s (1999) z-tree software. The instructions are available upon request.

<sup>3</sup> Mann-Whitney U-tests using session averages aggregated over the first 10 rounds as independent observations reveal the following p-values (two-tailed) for trust: partner—stranger:  $p=0.01$ ; reputation-stranger—stranger:  $p=0.08$ ; partner—reputation-stranger:  $p=0.14$ , and for trustworthiness: partner—stranger:  $p=0.01$ ; reputation-stranger—stranger:  $p=0.02$ ; partner—reputation-stranger:  $p=0.33$ .

<sup>4</sup> Mann-Whitney U-tests using session averages aggregated over the second 10 rounds as independent observations reveal the following p-values (two-tailed) for trust: partner—stranger:  $p=0.22$ ; reputation-stranger—stranger:  $p=0.56$ ; partner—reputation-stranger:  $p=0.03$ , and for trustworthiness: partner—stranger:  $p=0.12$ ; reputation-stranger—stranger:  $p=0.56$ ; partner—reputation-stranger:  $p=0.09$ .

<sup>5</sup> For samples of this size linear probability models are more robust than logit or probit models.

<sup>6</sup> Taking a similar average for trustors does not make sense since trust in later rounds may be driven by experience.