trust in experiments

Iris Bohnet
From *The New Palgrave Dictionary of Economics*, Second Edition, 2008
Edited by Steven N. Durlauf and Lawrence E. Blume

Abstract

Trust is the willingness to make oneself vulnerable to another person's actions, based on beliefs about that person's trustworthiness. This article focuses on interpersonal trust and trustworthiness between two people, a trustor and a trustee, as measured in laboratory experiments. A trustee behaves trustworthily if he voluntarily refrains from taking advantage of the trustor's vulnerability. Trust applies to all transactions where the outcome is partly under the control of another person and not fully contractible. The article discusses measurement issues, the motives for and influences on trust and trustworthiness (incentives, repetition and demographic variables) and questions of external validity.

Keywords

cooperation; fairness; folk theorem; investment game; public goods games; reciprocity; repeated games; social preferences; trust; trust in experiments; trust games; ultimatum games

JEL classifications

C9

Article

Trust is the willingness to make oneself vulnerable to another person's actions, based on beliefs about that person's trustworthiness.

This article focuses on interpersonal trust between two people, a trustor and a trustee. A trustee behaves trustworthily if he voluntarily refrains from taking advantage of the trustor's vulnerability. Trust applies to all transactions where the outcome is partly under the control of another person and not fully contractible, for example, between employers and employees or patients and doctors. Trust and trustworthiness are typically measured in surveys or laboratory experiments. We shortly discuss some survey evidence but focus on behavioural measures of trust.

Measurement

Trust attitudes have typically been measured by the following survey question (for example, used in the General Social Survey and the World Values Survey): 'Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?' Based on this question, trust has declined dramatically across the world since the 1960s. There are noticeable cross-country differences, with Scandinavians most and Latin Americans least likely to

trust others. An empirical literature building on Knack and Keefer (1997) shows that trust attitudes are positively correlated with various measures of a country's economic performance.

Around 1990, two seminal papers started a new wave in the economic research on trust. In 1988, Camerer and Weigelt employed a binary-choice trust game, and in 1995 Berg, Dickhaut and McCabe the 'investment game' to study trust. In the binary-choice trust game, the trustor decides between a sure outcome and trust. If she chooses the sure thing, she and her trustee both receive (S, S). If she is willing to trust, both either end up with a moderate payoff exceeding S (M, M), or the trustor receives a lower payoff than if she had not trusted, and the trustee the highest possibly payoff, (L, H). Thus, for the trustor, $M > S > L$, and for the trustee, $H > M > S$. In the investment game, a trustor and a trustee are endowed with a certain amount of money, A (in some experiments, only trustors are endowed). The trustor can send any amount, $X \leq A$, to the trustee. X is multiplied by $k > 1$ by the experimenter. In most experiments, $k = 3$. Trustees receive kX and then decide how much of it, $Y \leq A + kX$, to return to their trustor. The final payoffs are $A - X + Y$ for the trustor and $A + kX - Y$ for the trustee. X is commonly referred to as trust and Y, or more precisely, Y/X, measures trustworthiness for $X > 0$. In both games, the equilibrium prediction based on selfish money-maximization and rationality is zero trustworthiness and zero trust.

The relationship between trust attitudes, as measured in surveys, and trust behaviour, as measured in experiments, is not clear. Some have found that they are related (for example, Fehr and Schmidt, 2002), others that they are not (for example, Glaeser et al., 2000). While the investment game and the binary-choice trust game have turned out to be the most widely used games to study trust experimentally, related games include the 'gift exchange game', the 'moonlighting game,' and standard public goods games (for a review, see Camerer, 2003).

What motivates trust and trustworthiness?

Trust is based on preferences, namely, the willingness to be vulnerable to someone else, and on expectations, namely, the belief about someone else's trustworthiness. A person's willingness to be vulnerable may be related to her attitudes to risk (for example, Eckel and Wilson, 2004), her social preferences (for example, Cox, 2004), and her willingness to accept the risk of betrayal (Bohnet and Zeckhauser, 2004). Bohnet and Zeckhauser introduced an analytical framework to disentangle the various motives, and show that people dislike making themselves vulnerable to the actions of another person more than to natural circumstances. This suggests betrayal aversion: people care not only about outcomes but also about how outcomes come to be. This finding was supported by neuroscientific evidence (Kosfeld et al. 2005).

The relevance of expectations of trustworthiness for trust has typically been measured by including a question about trustors' beliefs. While this measure is not perfect, generally the relationship between expectations of trustworthiness and trust is very strong. For example, using a within-subject design with behavioural controls for risk and social preferences, Ashraf, Bohnet and Piankov (2006) found that expectations of trustworthiness explain most of the variance in trust in an investment game but that social preferences also matter.

Trustworthiness is based on trustees' social preferences, which may be related either to outcomes (for a survey, see Fehr and Schmidt, 2002) or to what the trustors'

actions reveal about their intentions. In a seminal paper, Rabin (1993) introduced a theoretical model of intention-based preferences, reciprocity, into the literature. A large number of empirical studies suggests the importance of reciprocity in trust interactions (for example, Fehr, Gächter and Kirchsteiger, 1997) although outcome-based social preferences also play an important role for trustworthiness (for example, Cox, 2004; Ashraf, Bohnet and Piankov, 2006).

What influences trust and trustworthiness?

Incentives

According to most models, trustors should be more likely to trust the higher the expected returns are from trusting. Bohnet, Herrmann and Zeckhauser (2006) measured the elasticity of trust and found that trust is responsive both to changes in the likelihood and to the cost of betrayal in Western countries. However, this does not necessarily apply in other parts of the world. For example, in Persian Gulf countries people hardly responded to such changes. Instead, many basically demanded a guarantee of trustworthiness before trusting, suggesting substantial aversion to betrayal. In addition, incentives may also not work as predicted by theory if they not only affect behaviour directly but also exhibit an influence on preferences, thus either fostering or undermining people's willingness to accept vulnerability and be trustworthy voluntarily (Bohnet, Frey and Huck, 2001).

Repetition

Generally, people are more likely to trust and be trustworthy in repeated than in one-shot interactions. Theoretically, this result is expected in a traditional model when interactions are indefinitely repeated (folk theorem) but not in finitely repeated games. In support of the theory, experimental evidence suggests that trust and trustworthiness rates are generally higher in indefinitely than in finitely repeated games but they are also higher in the latter than in one-shot interactions. The equilibrium prediction of no trust and trustworthiness is generally refuted, although trust and trustworthiness rates typically drop substantially as the end of the game draws nearer (for example, Gächter and Falk, 2002).

Demographic variables

Generally, the evidence is not as conclusive as we might expect or wish. While in theory variables such as gender, race or country of origin should be easy to control for, experiments produce different results precisely because of the different sets of control variables and the different subject pools used. The most promising approaches include those identifying overarching frameworks able to account for a variety of studies. We discuss three such frameworks here: history of discrimination, societal organization and market integration.

Groups that historically have been discriminated against, such as women and minorities, are generally less likely to trust. At the same time, often these groups are more trustworthy (for example, Alesina and LaFerrara 2002; Buchan, Croson and Solnick, 2003; Eckel and Wilson, 2003).

Group-based societal organization based on long-standing relationships and repeated interactions within groups can substantially reduce the social uncertainty involved in trust. It is often referred to as 'collectivist' in contrast to the Western 'individualist' model of organization, which produces trust through more anonymous, institutional arrangements such as contracts and insurance. Trust in strangers has often been found to be higher in individualist (for example, the United States or Switzerland) than in collectivist countries (for example, Japan or the Persian Gulf countries), although the rather small number of studies and sample sizes does not allow any definite conclusions at this point (for example, Bohnet, Herrmann and Zeckhauser, 2006; but see also Croson and Buchan, 1999).

The degree of market integration is related to norms of cooperation and fairness in public goods and ultimatum games. Similarly, the norms of reciprocity typically found in trust experiments in developed countries seem to apply more strongly in societies in which goods and services are exchanged in the market rather than in informal reciprocal-exchange arrangements. Greig and Bohnet's survey of the evidence (2006) suggested that the positive relationship between trust and trustworthiness, normally taken to indicate reciprocity, is more pronounced in developed than in developing countries.

External validity

Experiments allow for maximum internal control. Concerns typically arising in field settings such as lack of randomization, selection and endogeneity can easily be addressed by experimental design. To address concerns about the subject pools experimentalists typically use, that is, North American or European students, experiments are now run with representative samples (for example, Fehr et al., 2002 in Germany) and with student and non-student subjects in other parts of the world (for example, Cardenas and Carpenter, 2005, for a survey). To directly test the external validity of trust experiments, Karlan (2005) ran investment games with members of a group lending association in Peru, and compared trustworthiness in the experiment with repayment rates. The more trustworthy subjects indeed were significantly more likely to repay their loans a year later.

See Also

- altruism in experiments
- behavioural game theory
- experimental economics
- public goods experiments
- reciprocity and collective action
- risk aversion
- social capital

Bibliography

Alesina, A. and LaFerrara, E. 2002. Who trusts others? *Journal of Public Economics* 85, 207–34.

Ashraf, N., Bohnet, I. and Piankov, N. 2006. Decomposing trust and trustworthiness. *Experimental Economics* 9, 193–208.

Berg, J., Dickhaut, J. and McCabe, K.A. 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10, 290–307.

Bohnet, I., Frey, B.S. and Huck, S. 2001. More order with less law: on contract enforcement, trust and crowding. *American Political Science Review* 95, 131–44.

Bohnet, I., Herrmann, B. and Zeckhauser, R. 2006. The requirements for trust in Gulf and Western countries. *Working paper.*

Bohnet, I. and Zeckhauser, R. 2004. Trust, risk and betrayal. *Journal of Economic Behavior and Organization* 55, 467–84.

Buchan, N., Croson, R. and Solnick, S. 2003. Trust and gender: an examination of behavior, biases, and beliefs in the investment game. Working paper, Wharton School, University of Pennsylvania.

Camerer, C.F. 2003. *Behavioral Game Theory.* Princeton: Princeton University Press.

Camerer, C.F. and Weigelt, K. 1988. Experimental tests of a sequential equilibrium reputation model. *Econometrica* 56, 1–36.

Cardenas, J.C. and Carpenter, J. 2005. Experiments and economic development: lessons from field labs in the developing world. Working paper, Middlebury College.

Cox, J.C. 2004. How to identify trust and reciprocity. *Games and Economic Behavior* 46, 260–81.

Croson, R. and Buchan, N. 1999. Gender and culture: international experimental evidence from trust games. *American Economic Review* 89, 386–91.

Eckel, C.C. and Wilson, R.K. 2003. Conditional trust: sex, race and facial expressions in a trust game. Working paper, Rice University.

Eckel, C.C. and Wilson, R.K. 2004. Is trust a risky decision? *Journal of Economic Behavior and Organization* 55, 447–66.

Fehr, E. and Schmidt, K. 2002. Theories of fairness and reciprocity – evidence and economic applications. In *Advances in Economics and Econometrics*, ed. M. Dewatripont, L. Hansen and S. Turnovsky. Cambridge: Cambridge University Press.

Fehr, E., Gächter, S. and Kirchsteiger, G. 1997. Reciprocity as a contract enforcement device: experimental evidence. *Econometrica* 64, 833–60.

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J. and Wagner, G. 2002. A nation-wide laboratory-examining trust and trustworthiness by integrating behavioral experiments into representative surveys. *Schmollers Jahrbuch* 122, 519–42.

Gächter, S. and Falk, A. 2002. Reputation and reciprocity: consequences for the labour relation. *Scandinavian Journal of Economics* 104, 1–26.

Glaeser, E.L., Laibson, D.I., Scheinkman, J.A. and Soutter, C.L. 2000. Measuring trust. *Quarterly Journal of Economics* 115, 811–46.

Greig, F. and Bohnet, I. 2006. Is there reciprocity in a reciprocal-exchange economy? Evidence of gendered norms from a slum in Nairobi, Kenya. Working paper, Kennedy School of Government, Harvard University.

Karlan, D. 2005. Using experimental economics to measure social capital and predict financial decisions. *American Economic Review* 95, 1688–99.

Knack, S. and Keefer, P. 1997. Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics* 112, 1251–88.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U. and Fehr, E. 2005. Oxytocin increases trust in humans. *Nature* 435, 673–6.

Rabin, M. 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83, 1281–302.