Journal of Experimental Social Psychology 47 (2011) 171–177

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Report

Elements of trust: Risk and perspective-taking

Anthony M. Evans *, Joachim I. Krueger

Brown University

ARTICLE INFO

Article history:
Received 10 August 2009
Revised 27 July 2010
Available online 14 August 2010

Keywords:
Interpersonal trust
Behavioral economics
Rationality

ABSTRACT

Trust is essential to personal well-being and economic success, but it cannot occur without accepting the possibility of betrayal. In the experimental trust game, game-theoretic rationality prescribes that trust decisions should depend on the potential risk (egocentric costs and benefits) and the probability of reciprocity (derived from the trustee's temptation to defect). The current work tests the relative weights of these elements. Experiment 1 shows that trust increases when costs decrease and benefits increase. The latter finding is critical because increasing the trustor's benefit also means increasing the trustee's temptation to defect. Hence, this finding suggests that egocentrism prevails over perspective-taking. Experiment 2 shows that the trustee's temptation to defect (negatively) affects trust, but only when the trustor's cost and benefit are favorable. Results are interpreted as reflecting a boundedly rational decision process.

Why and when do people trust one another? The answers to these questions have important implications in social exchanges. Nobel laureate Arrow (1974) called trust "a lubricant for social systems" (p. 23). A consumer trusts that purchased goods will work as promised; a manager trusts that a new employee will be dependable; and an investor trusts that corporate accountants will report honest figures. But what is trust? According to a widely accepted definition, trust is "a psychological state comprising the intention to accept vulnerability based upon the positive expectations of the intentions or behavior of another" (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 395).

The trust game has become an influential experimental paradigm in psychology and behavioral economics (Evans & Krueger, 2009). Fig. 1 displays the game in its extensive form. The first mover, or trustor, chooses between the status quo and trust. With the status quo, the game ends and both players receive the outcome labeled P. With trust, the game enters a second stage in which the second mover, or trustee, chooses between reciprocity and betrayal. With reciprocity, both players receive outcome R; with betrayal, the trustor receives S and the trustee receives T. In the game, these payoffs are ordered T>R>P>S. The psychological situations are notably different for the two players. The trustee faces a choice between fairness and selfishness: she can reciprocate and distribute the money equally (each player receives R) or betray trust by taking more money for herself (the trustor gets S and the trustee gets T, S<T). On the other hand, the trustor has a strategic dilemma that is not solvable without making assumptions about the trustee. The trustor knows that reciprocated trust yields a better outcome than the status quo (R>P), but there is no guarantee that reciprocity will occur. Betrayed trust leads to an outcome that makes the trustor worse off than the status quo (S<P). The trustor must decide if this risk is worth taking.

A theory seeking to explain decisions in the trust game must recognize that the payoff structure comprises three distinct elements. The first two elements of trust are *cost* and *benefit*, which jointly correspond to the concept of personal vulnerability or risk. The trustor's potential *cost* is the difference between the status quo and betrayal (P–S), and the potential *benefit* is the difference between reciprocity and the status quo (R–P). The third element, *temptation*, is the difference in the trustee's payoffs between betrayal and reciprocity (T–R). This element is relevant to the concept of expectation because it captures the trustee's temptation to defect if trusted. In other words, this difference is a cue to the probability that the trustee will reciprocate trust. Empirically, temptation is the best predictor of reciprocity among strangers (Snijders & Keren, 1999). Fig. 2 illustrates the numerical relationships among the three factors. These elements capture the idea that trust presupposes both vulnerability and expectation. Without personal vulnerability, the expectation that others will act benevolently would be an ordinary social prediction (Luhmann, 2000). Conversely, trust without the expectation of reciprocity is self-defeating.

Having identified the elements of trust in the payoff structure, we can predict how they influence the process of decision-making. First, consider the perspective of classic game theory (von Neumann & Morgenstern, 1947). This approach assumes that both players are strictly self-interested. If so, a player reasoning by backward induction must conclude that trust is irrational. Assuming that decisions depend only on the ordering of the payoffs, it is clear that a trustee will choose betrayal if trusted (T>R). Knowing this, the trustor will choose the status quo, noting that P>S. Though analytically elegant, classic game

* Corresponding author. Department of Psychology, Brown University, Box 1853, 89 Waterman St., Providence, RI 02912, USA. Fax: +1 401 863 1300.
E-mail address: Anthony_Evans@Brown.edu (A.M. Evans).

0022-1031/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2010.08.007

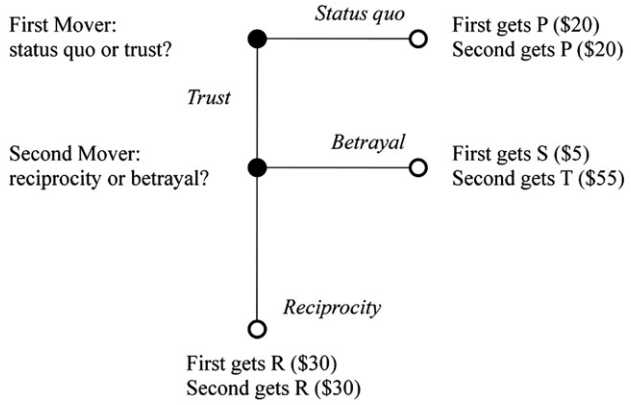

Fig. 1. An example of the sequential trust game. The first mover chooses between trust and status quo. If trust is chosen, then the second mover chooses between betrayal and reciprocity.

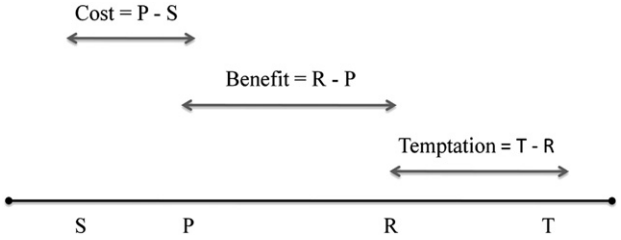

Fig. 2. Cost, benefit, and temptation in the trust game. By definition, S<P<R<T. Cost and benefit are factors related to the trustor's payoffs. Temptation describes the trustee's payoffs. Temptation is critical to the trustor because it implies the probability of reciprocity.

theory does not fare well as a descriptive theory. Scores of experiments show that trust is a common choice, even under conditions of complete anonymity (Camerer, 2003; Smith, 2003).

Why does classic game theory fail at description? Process data suggest that reasoning by backward induction is counterintuitive. Using a sequential bargaining task, Johnson, Camerer, Sen, and Rymon (2002) assessed cognitive processes by tracking visual attention to payoffs.[1] Like the trust game, the bargaining game can be rationally solved by viewing the payoffs in later stages of the game and then to reason backward to the game's beginning. Few participants were able (or willing) to do this; instead they tended to focus on the payoffs in the first round and reasoned *forward*. Players followed this strategy even after they gained experience with the game. In a second study, another group of participants received explicit descriptions of backward induction. Here, visual attention and behavior quickly converged on the predicted rational equilibrium (see Costa-Gomes & Crawford, 2006, for similar results in a two-stage "beauty contest" guessing game). Reasoning by backward induction appears to occur rarely on its own, but is an acquirable skill.

A revised rational model predicts that people estimate the expected value of trust. Unlike backward induction, this model allows for the subjective probability of reciprocity to be greater than zero. According to this view, people are sensitive to their own costs and benefits, and to the trustee's temptation to defect. These variables should not interact with one another; instead, a player integrates consequences (cost/benefit) with respective probabilities (Hastie & Dawes, 2010). The first mover calculates the expected value of trust and compares this amount to the status quo:

$$\text{Expected value of trust} = \left(\begin{matrix}\text{Benefit of}\\ \text{reciprocity}\end{matrix}\right) \times \left(\begin{matrix}\text{Probability of}\\ \text{reciprocity}\end{matrix}\right) + \left(\begin{matrix}\text{Cost of}\\ \text{betrayal}\end{matrix}\right) \times \left(\begin{matrix}\text{Probability of}\\ \text{betrayal}\end{matrix}\right)$$

Distrust is not a foregone conclusion if expected values are estimated. People will choose trust when its expected value is greater than the status quo payoff. The model grants that a trustor may have a reasonable expectation that the trustee will reciprocate. The expectation of reciprocity may be justified, for example, by appeals to the trustee's aversion to inequality (Fehr, Naef, & Schmidt, 2006) or adherence to norms of reciprocity (Gouldner, 1960; Krueger, Massey, & DiDonato, 2008). As noted above, a trustor who does not know the trustee's values can estimate the probability of reciprocity from temptation (T–R).

Although the expected value model is more plausible than backward induction, it also has drawbacks. One challenging finding is that people are often biased in assessing consequences and probabilities. For example, people underweight very low or very high probabilities when an event is emotionally arousing (Rottenstreich & Hsee, 2001). Other findings point to biases of loss aversion and status quo respect (Moshinsky & Bar-Hillel, 2010). Such concerns lead to a third decision model, which assumes that people focus on those elements of trust that are salient or easy to compute (cf. Hastie & Dawes, 2010). Typically, information is more salient and accessible when it is relevant to the self (Alicke, Dunning, & Krueger, 2005).

In the trust game, it follows that egocentric costs and benefits loom larger than the trustee's temptation to defect. In short, this model suggests that the decision to (dis)trust is egocentric by default and only mindful of others by elaboration. Corroborating this model, Snijders and Keren (1999) found that the rate of trust was strongly associated with personal risk (i.e., cost/benefit), but only weakly associated with temptation (see Malhotra, 2004 for similar results). Previous experiments support this explanation, but none have directly tested the predictions of an egocentric model. In the present research, we sought to provide direct evidence for this process-oriented description of trust.

Our central hypothesis is that people are focused on their own potential costs and benefits. By comparison, perspective-taking plays a limited role, in part because it requires additional mental work. In this way, we view trust as boundedly rational. We hypothesize that trustors engage in perspective-taking (by evaluating the trustee's temptation) only when they deem the personal risk to be sufficiently low. The trustee's temptation does not factor into the decision process if the personal risk of trust is sufficiently high. We conducted two experiments to test our central hypothesis: Our first experiment investigated the frequencies of trust under varying levels of cost and benefit, replicating and extending previous results. The purpose of this experiment was to test if cost and benefit are indeed interpreted egocentrically. In Experiment 2, risk (the ratio of cost and benefit) and temptation were manipulated orthogonally. Using this design, we tested the egocentric model's critical prediction that perspective-taking (evaluating temptation) only occurs when the risk of trust is low.

Experiment 1

The objective of Experiment 1 was to replicate and extend the findings of Snijders and Keren (1999). The critical innovation was that the first mover's risk was decomposed, that is, costs and benefits were manipulated orthogonally, thereby permitting estimates of their unique contributions. Trust decisions were measured as outcomes across different payoff levels. The temptation of the second player (T–R) was not directly manipulated. We imposed an efficiency constraint (2R = S + T) to ensure that trust and reciprocity were not confounded

[1] The game's parameters were concealed from participants, but were temporarily revealed by moving the mouse over targets. Thus, the experimenters observed the process players used to reason about the game's payoffs.