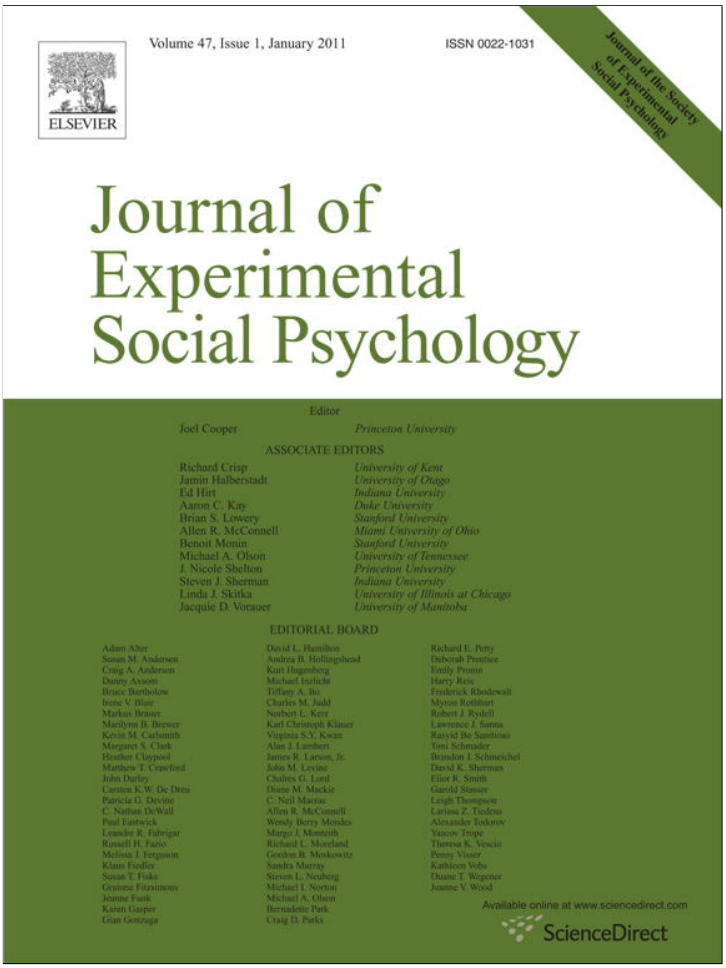

Volume 47, Issue 1, January 2011
ISSN 0022-1031
ELSEVIER
Journal of the Society of Experimental Social Psychology

Journal of Experimental Social Psychology

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

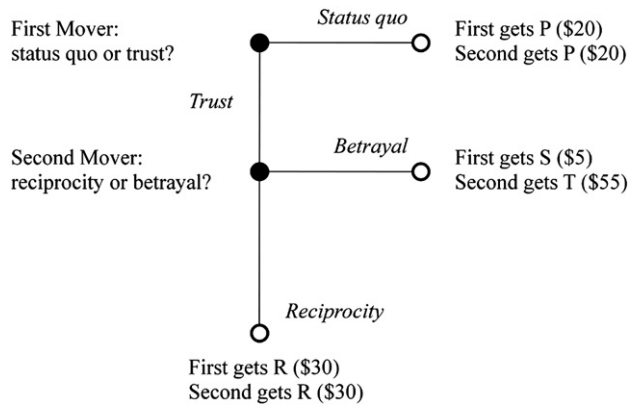

Fig. 1. An example of the sequential trust game. The first mover chooses between trust and status quo. If trust is chosen, then the second mover chooses between betrayal and reciprocity.

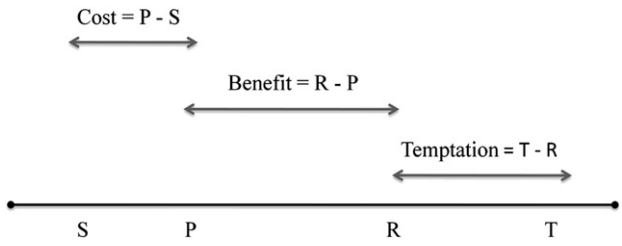

Fig. 2. Cost, benefit, and temptation in the trust game. By definition, S < P < R < T. Cost and benefit are factors related to the trustor's payoffs. Temptation describes the trustee's payoffs. Temptation is critical to the trustor because it implies the probability of reciprocity.

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:

$$\begin{aligned}\text{Expected value of trust} &= \begin{pmatrix}\text{Benefit of}\\ \text{reciprocity}\end{pmatrix} \times \begin{pmatrix}\text{Probability of}\\ \text{reciprocity}\end{pmatrix}\\ &+ \begin{pmatrix}\text{Cost of}\\ \text{betrayal}\end{pmatrix} \times \begin{pmatrix}\text{Probability of}\\ \text{betrayal}\end{pmatrix}\end{aligned}$$

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.

with preferences for an outcome with greater collective utility. In other words, the total amount of money shared by the players was fixed. By necessity then, a one-unit increase in cost (P–S) meant a one-unit decrease in S and a one-unit increase in T (the sum of S and T must stay constant; if S decreases, then T must increase to compensate). Mathematically, an increase in cost meant an increase in temptation ($[T+1]-R=T-R+1$). Equivalently, a one-unit increase in benefit (R–P) meant a one-unit increase in R. This increase in R also entailed (because $2R=S+T$) an increase in temptation ($[T+2]-[R+1]=T-R+1$).

We predicted a negative correlation between trust and cost, and a positive correlation between trust and benefit. The prediction that benefit and trust are positively associated challenges models of backward induction and expected value. Although increasing benefit improves the value of reciprocity, it simultaneously decreases the likelihood of reciprocity. Therefore, a positive correlation between benefit and trust would constitute strong evidence for egocentric decision-making.

Method

Participants

The experiment was administered online. Participants were recruited from university listservs and websites with listings of online experiments. Steps were taken to ensure the integrity of the data. Duplicate, multiple, and incomplete responders were identified and excluded (a total of 15 participants were removed). After these procedures, 240 participants remained. Fifty-six percent were women and the average age was 23, $SD=8.2$.

Procedure

Participants played 10 rounds of the game with changing payoffs. Each factor had four levels: 5, 10, 15, and 20. The values of R and S were adjusted while P was held constant at 20.[2] Each participant completed a subset of 10 games in random order without repetition. The games were played for hypothetical stakes (20 points represented $20). Participants were told that the study was designed to examine questions of personality in decision-making. They began by providing demographic information and completing an inventory of scales that were not analyzed for this report. Next, they received instructions for how to play the trust game (referred to as a decision-making task). Participants were told that they would play the role of the trustor with a simulated partner whose behavior was programmed to reflect human behavior in a laboratory version of the task performed for actual money (taken from Snijders & Keren, 1999). They would play the game with a different partner in each round; in other words, the trustee's behavior was independently determined for each game. To prevent feedback effects, players did not receive information on the other players' decisions until all 10 rounds of the game were completed. When the study was finished, participants received diagnostic feedback comparing their decisions and outcomes to other participants in the study.

Results

Descriptive statistics

Rates of trust ranged from 37% to 69% across the 16 payoff combinations. The overall rate was 52%. On average, participants trusted 5.09 times out of 10 ($SD=3.34$). Thirty-one participants (12.9%) never trusted and 33 (13%) always trusted. For most participants (74.1%), trust depended on the game's payoffs.

Determinants of trust

We used logistic regression to model the elements of trust with the participant's behavioral choice as the dependent variable in each round. Trust and status quo were respectively coded as 1 and 0. Benefit and cost were entered into the model as polynomial contrasts (linear, quadratic, and cubic). Each participant provided 10 data points. Because this method did not yield independent observations, we clustered the data by participant and calculated robust standard errors. Table 1 shows the regression coefficients of a full main-effects model with age, gender, cost, and benefit. Interaction terms were also tested, but none were found to be significant. Unstandardized beta weights and odds ratios are reported.

The overall model was significant and correctly classified 60.4% of observations. There were no meaningful effects of age or gender. As predicted, cost had a negative effect on trust, linearly ($b=-.14$, $p<.001$) and quadratically ($b=.10$, $p=.01$). Averaging across levels of benefit, increasing cost from the lowest (5) to the highest level (20) decreases trust from 62% to 44%. The decrease was most pronounced when cost increased from 5 to 10; trust declined from 62% to 50%. Also, as predicted, benefit had a positive linear effect, $b=.06$, $p<.01$. Averaging across levels of cost, increasing benefit from the lowest (5) to the highest level (20) increased trust from 46% to 56%. To test for possible shared variance among cost and benefit, we also performed regressions with single predictors. The results were consistent with the full model.

Discussion

The results of the Experiment 1 corroborate the idea that the decision to trust is boundedly rational. It is not surprising that greater cost increases the preference for the status quo, but the effect of benefit is not trivial. Recall that the experiment was designed so that any tendency to base trust on one's potential benefit would contradict backward induction and expected value maximization. Any increase in the trustor's benefit also increased the trustee's temptation to defect, thereby decreasing the probability of reciprocity. Previous research did not pit risk against temptation (Malhotra, 2004; Snijders & Keren, 1999). A limitation of this experiment is that risk and temptation were not manipulated independently. Although the results suggest that people reason egocentrically, the (comparatively larger) effect of risk may mask the effect of temptation. It is conceivable that people engage in perspective-taking when their own risk is low. We designed Experiment 2 to explore this possibility.

Table 1
Main effects of cost and benefit.

Variable	*b*	*b SE*	*log odds*	*p*
Age	.01	.01	1.00	.738
Gender	.09	.18	1.10	.583
cost (linear)* **	**−.14**	**.02**	**.87**	**<.001**
cost (quadratic)*	**.10**	**.04**	**1.10**	**.011**
cost (cubic)	−.01	.02	.99	.625
benefit (linear)* **	**.06**	**.02**	**1.06**	**.007**
benefit (quadratic)	.03	.04	1.01	.700
benefit (cubic)	.01	.01	1.01	.638
Wald Chi-Squared (8)	61.6			<.001

Significance tests were conducted using *z*-scores. Variables in bold have $p<.10$.
* indicates $p<.05$.
*** indicates $p<.001$.

[2] The possible values of S were 0, 5, 10, and 15. The possible values of R were 25, 30, 35, and 40.

Experiment 2

We hypothesized that people approach the decision to trust in a flexible manner. The risk of trust, the ratio of cost to benefit, should be of primary interest from a self-regarding point of view. When the risk is high, a decision not to trust may be reached relatively easily, regardless of the perceived probability of reciprocity. But when the risk is low, the decision might be influenced by the trustee's temptation to defect. In other words, we expected that a person's willingness to consider the other player's perspective depends on the judgment of whether the risk of trust is worth taking. If, as hypothesized, the decision-making process is more complex when personal risk is low than when it is high, such decisions would take longer. Arguably, evaluating both risk and temptation requires more time than evaluating risk alone.

We manipulated risk and temptation orthogonally. As the relative contributions of cost and benefit were no longer critical, we combined these two elements into a single index of risk with two levels: Low risk (cost 5/benefit 15) and high risk (cost 15/benefit 5). In Experiment 1, these payoff levels generated 61% (low risk) and 40% (high risk) rates of trust. We varied the second factor, temptation (T–R), over four levels (15, 25, 35, and 45) by changing the value of T.

Method

Participants

Participants were recruited from online listservs ($N=136$) and from an introductory psychology class ($N=47$). Five online participants were excluded because of missing data: The effective sample consisted of 178 participants; 68% women; average age = 25, $SD=10$. No significant differences were found between the two groups and they were combined for all subsequent analyses.

Procedure

As in the previous study, participants completed 10 rounds of the game against simulated partners. The eight payoff manipulations were randomized in rounds two through nine. Rounds 1 and 10 were held constant (cost = 10, benefit = 10, temptation = 50) to prevent any first or last round effects. We anticipated that the first and last rounds would have the longest and shortest response times, respectively. Furthermore, participants may be more or less likely to trust when it is the first or last round. For these reasons, we excluded these rounds from the following analyses.[3] Response time was recorded in seconds for each decision. The instructions and experimental conditions mirrored Study 1.

Results and discussion

Descriptive statistics

On average, participants showed medium levels of trust, $M=4.1$, $SD=3.1$. Only 34 of them (19.1%) never trusted and only 12 (6.7%) always trusted. Response time significantly decreased over the course of the study from an average of 11.1 s ($SD=8.1$) in the first round to an average of 5.4 s ($SD=4.1$) in the last. Response times were log-transformed to correct for skew.

Risk and temptation

As in the previous experiment, we used a logistic model with clustered observations to model the data, coding low and high risk respectively as 1 and 0; the variable "risk" indicates the main effect of risk. We transformed the variable of temptation into polynomial contrasts, and also entered the interactions between risk and temptation into the model. In Table 2, the terms linear, quadratic, and cubic describe the main effect of temptation. The interaction terms (i*linear, etc.) describe the interaction between risk and temptation. Fig. 3 shows the observed rates of trust across the eight payoff conditions.

Table 2
Incentive and temptation.

Variable	*b*	*b SE*	*log odds*	*p*
Age	.01	.01	1.01	.254
Gender	−.11	.23	.89	.619
risk* **	**.99**	**.14**	**2.69**	**<.001**
linear*	**−.02**	**.02**	**.98**	**.043**
quadratic	−.03	.03	.97	.594
Cubic	.01	.02	1.01	.818
i*linear	**−.06**	**.04**	**.94**	**.081**
i*quadratic*	**.15**	**.07**	**1.16**	**.050**
i*cubic	.01	.03	1.01	.849
Wald Chi-Squared (9)	73.14			<.001

Significance tests were conducted using *z*-scores.

The full model correctly classified 64% of observations. The pattern of results displayed in Fig. 3 was consistent with the central hypothesis. As before, risk strongly influenced trust. More importantly, however, the interaction effect of risk and temptation was fitted as a combination of linear and quadratic contrasts. Within the high-risk condition, the rate of trust was close to 30% across all levels of temptation. In contrast, trust in the low-risk condition varied with temptation; it decreased from 64% (temptation = 15) to 50% (temptation = 45). The linear main effect of temptation was also significant. As before, there were no meaningful age or gender effects.

Simple-effects analyses revealed no significant effect of temptation on trust when risk was high ($p>.4$) and the model was not significant, $\chi^2(5)=1.4$, $p=.52$. Conversely, temptation had a significant inhibitory effect when risk was low. The logistic model in the low-risk conditions was statistically significant, $\chi^2(5)=18.73$, $p=.001$, and the effect of temptation was fitted as a combination of linear ($b=-.08$, $p<.001$) and quadratic ($b=.11$, $p<.05$) contrasts. In the low-risk condition, increasing temptation caused trust to initially decline and then level off at 50%.

Next, we tested the prediction that decisions would take longer when risk was low. Hence, we submitted log-transformed response

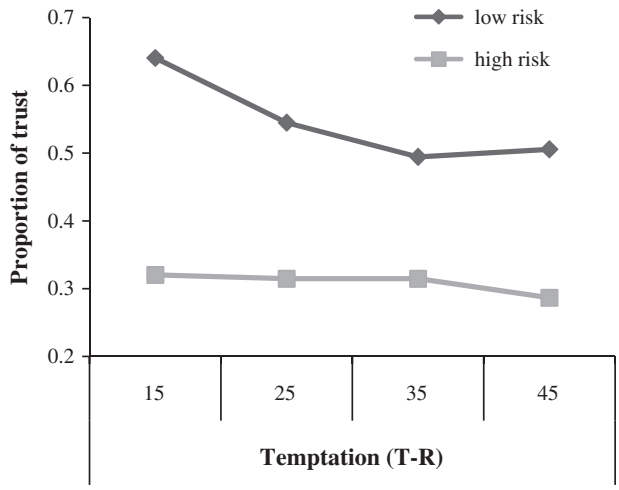

Fig. 3. Levels of trust by risk and temptation observed in Experiment 2. When the risk of trusting is high, temptation does not affect the level of trust; but when the risk is low, trust and temptation are negatively correlated.

[3] The rate of trust was 23% in the first round ($SE=3.1\%$) to 43% in the last round ($SE=3.7\%$).

times to a clustered linear regression with risk, temptation, round number, and decision outcome (trust versus distrust) as predictors. The overall model was significant, $F(4, 177) = 86$, $p < .001$, $R^2 = .12$, and so was the critical effect of risk, $b = -.16$ ($SE = .07$), $p = .01$.[4] Our full model predicts that participants made decisions more rapidly under high risk (untransformed $M = 6.4$ seconds, $SD = 2.2$) than under low risk ($M = 7.1$, $SD = 2.0$). As expected, latencies became shorter as participants gained practice, $b = -.11$ ($SE = .006$), $p < .001$. The outcome of the decision (trust versus distrust) did not significantly affect response time, $b = -.05$ ($SE = .06$), *ns*, nor did the level of temptation, $b = .006$ ($SE = .01$), *ns*. There were no significant interaction effects.

In summary, Experiment 2 supports the prediction that perspective-taking depends on the level of risk. When personal risk is high, people tend to move quickly and are more likely to distrust than to trust. Furthermore, the rate of trust is unaffected by changes in temptation, such that participants remain distrusting even when temptation is low. Conversely, when personal risk is low, people move more slowly and consider the other person's temptation to defect, leading them to be more likely to trust.

General discussion

Across two experiments, most participants were sensitive to the various payoffs of the trust game. Only a minority took a schematic, invariant response strategy. Consistent with previous research (Malhotra, 2004; Snijders & Keren, 1999)—and common sense—we found that low risk (i.e., a favorable ratio of cost to benefit) is conducive to trust. Experiment 1 went beyond prior work by showing that increasing benefit increases trust, even when this also means increasing the trustee's temptation to defect. This pattern suggests that players do not fully consider the trustee's perspective. Experiment 2 developed this finding by showing that perspective-taking is a conditional element of trust. Decisions were (negatively) correlated with the trustee's temptation, but only when the participants' own level of risk was low. When risk was high, most players quickly reached a decision not to trust.

These findings raise important theoretical issues. First, there was no support for the classic notion that no one will trust. This conclusion is in line with a body of empirical work and the everyday observation that many social and economic exchanges occur without contracts or enforceable promises. In other words, trust is pervasive in social life (Evans & Krueger, 2009). To do well, prosper economically, and be socially accepted, an individual must deploy trust wisely. This is a challenge that requires social intelligence and rationality.

Second, the canonical definition of trust (in terms of vulnerability and expectation) implies the estimation of expected values. A rational decision maker attends even-handedly to consequence (vulnerability) and probability (expectation). In the trust game, these elements are encoded in the payoffs. The person's vulnerability is given by cost and benefit (i.e., the risk), and the expectation of reciprocity is implied by the trustee's temptation to defect. Both experiments support the conclusion that the consequences of trust are viewed as more important than their respective probabilities. Thus, the expected value model of trust, which predicts that probabilities and consequences are equally important, fails at description.

Generally, expected value models are mute on the processes that give rise to decisions. In contrast, the perspective of bounded rationality suggests that rational decision-making is a joint function of the individual's mental capacity and the difficulty of the task (Simon, 1955). Trust is complex and cognitively demanding. We have suggested that trustors should derive an expectation of reciprocity from T–R, but this mapping is a shortcut; it ignores other possibilities. For example, trustees may be motivated by benevolence or competition (i.e., valuing respectively large or small benefits to the trustor). Needing to construct a probabilistic expectation, a trustor faces a situation of ambiguity, rather than simple uncertainty. Perhaps the former is more aversive than the latter (Rode, Cosmides, Hell, & Tooby, 1999).

The present work suggests that people reduce the complexity of trust by approaching the decision from an egocentric perspective. They consider the trustee's temptation, but only if their own risk is low. The possible reasons for this effect are varied. We suspect that people discount temptation because of the comparative processing advantage of self-related information (Jackson, Brunet, Meltzoff, & Decety, 2006; Krueger, 2003). Perspective-taking is more onerous and time-consuming (cf. Epley & Waytz, 2010). According to this account, people consider their own vulnerability before they form expectations about the behavior of others. If the risk is high, decisions are based solely on the consequences of trust (cost and benefit); but if the risk is low, people proceed to form and use expectations. This explanation predicts a two-stage decision process.

However, our experiments were not specifically designed to test the temporal aspect of decisions. An alternative description of our results is that people evaluate risk and temptation simultaneously, but then weight self-related information more strongly and selectively (cf. Alicke & Sedikides, 2010; Chambers & Windschitl, 2004). Likewise, responses may take longer in the low-risk condition because the optimal response (trust or distrust) is unclear. Within the high-risk condition, there was some consensus (70% distrust) about how to respond. In contrast, the average rate of trust in the low-risk condition (55%) was near the point of indifference. Arguably, it takes longer to make a final decision in the presence of a strong alternative. We find the sequential decision process to be the most likely explanation for the reported patterns of behavior and response latency, though the present evidence is not conclusive.

Additional methods are needed to support the proposed temporal sequence of processes and the effortful nature of perspective-taking. The claim that temptation is evaluated after risk could be tested using eye-tracking and information search methodologies. Furthermore, the assertion that perspective-taking is limited because it requires additional mental resources could be investigated using cognitive load, ego-depletion, or time-pressure manipulations.

The consequences of egocentrism

When the risk is high, most people heuristically choose distrust, neglecting the expectation of reciprocity. If the other player's temptation is high, this decision is correct and efficient; if, however the other player's temptation is low, distrust may be an error. Our research suggests that people would be more likely to trust if they always accounted for both consequences and probabilities. Overall, this pattern implies that people do not trust enough (Fetchenhauer & Dunning, in press).

If perspective-taking were to occur in high-risk games, there would be greater trust at low levels of temptation. To illustrate this point, we used the logistic model from Study 2 to estimate how perspective-taking would influence trust under high-risk conditions.[5] We calculated point estimates of trust using the slope coefficients from the simple effects test of temptation within the low risk condition. Fig. 4 compares the predicted effect of perspective-taking with the observed data. As the level of temptation decreases, there is an increasing disparity between observed and predicted rates of trust. People refrain from trusting because of their high vulnerability, but

[4] Unstandardized beta weights and robust standard errors are reported.

[5] We assumed that the effect of temptation would mirror results in the low-risk condition. In other words, we estimated a main effect of temptation with no interaction term.

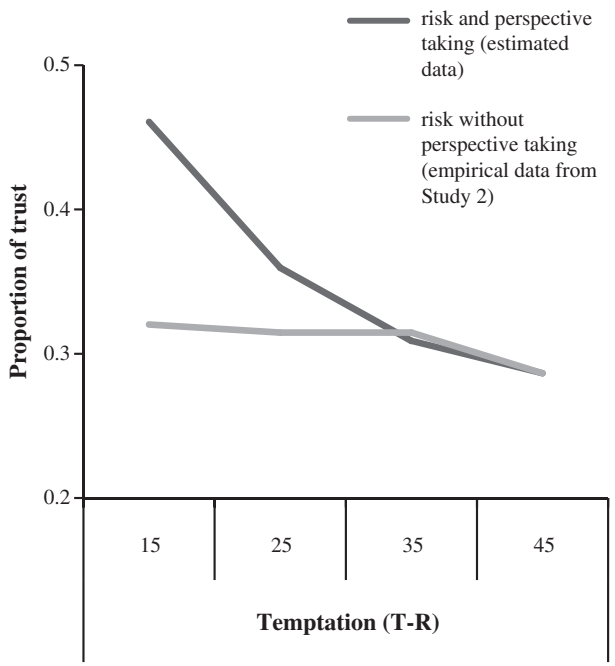

Fig. 4. In Experiment 2, trust and temptation were uncorrelated when the risk was high. Using data from Experiment 2, we estimated how trust decisions would change if participants considered the effects of both risk and temptation. Observed and predicted behaviors are compared, showing that participants are under-trusting when temptation is low.

they fail to account for the fact that the second mover has little incentive to defect. Trust would be evaluated more favorably if players accounted for both risk and temptation.

Two possible interventions could reduce this discrepancy: The first is to make the perspective of the trustee more salient. Trustors fail to engage in perspective-taking when the risk of trust is high, but they may become more willing to trust if they are directly led to consider the trustee's point of view. This solution assumes that egocentrism occurs because perspective-taking is difficult. Another possibility is to restructure the interaction to minimize risk. As Experiment 1 has demonstrated, reducing cost or increasing benefit can accomplish this. We found that reducing risk increases trust, even when the reduced risk is accompanied with increased temptation.

Applications

Understanding the process underlying trust has important applications for social exchanges. The proposed model is relevant when social information is minimal (and social distance is maximal). The development of trust can be facilitated by common group membership (Foddy, Platow, & Yamagishi, 2009) or a preexisting relationship (Glaeser, Laibson, Scheinkman, & Soutter, 2000), but trust dilemmas often occur in situations where people lack common ground or direct social contact. In these situations, decision processes may be more strategic than relational (Lewicki, 2006), making a process-oriented approach particularly useful.

Internet commerce is a timely example of trust among near-strangers. On the Internet, buyers and sellers interact with limited communication and apparent anonymity. The willingness to purchase a good, sight-unseen, requires the buyer to implicitly trust the seller. Our model suggests that the buyer's most important consideration is not the apparent trustworthiness of the seller, but the perceived cost and benefit of the purchase. Therefore, designing mechanisms that limit the apparent risk of trusting, such as a warrantee or buyer's insurance, would be effective in establishing trust between unfamiliar parties. Alternatively, egocentrism could be subverted by emphasizing the perspective of the seller, establishing that she has little temptation to betray the buyer's trust.

Concluding remarks

The finding that strategic decisions are motivated by egocentric concerns is not novel, but this tendency has new relevance when we evaluate it in the context of interpersonal trust. The trustee's temptation is the most important predictor of trustworthy behavior (Snijders & Keren, 1999). Therefore, the trustor's motive to consider the other player's perspective goes beyond altruism—perspective-taking is a matter of strategic self-interest. Consistent with past work, we found that trust is primarily motivated by egocentric risk taking. Still, we found a significant, if conditional, effect of perspective-taking. Players engage in perspective-taking only when risk is deemed sufficiently low. Under certain conditions, this failure to fully engage in perspective-taking is suboptimal and sometimes leads to inefficient outcomes. Rational trust, by definition, requires the evaluation of both vulnerability and expectation.

References

Alicke, M. D., Dunning, D., & Krueger, J. I. (2005). *The self in social judgment.* New York, NY: Psychology Press.

Alicke, M. D., & Sedikides, C. (2010). *Handbook of self-enhancement and self-protection.* New York, NY: Guilford.

Arrow, K. (1974). *The limits of organization.* New York, NY: Norton & Company Inc..

Camerer, C. F. (2003). *Behavioral game theory.* Princeton, NJ: Princeton University Press.

Chambers, J. R., & Windschitl, P. P. D. (2004). Biases in comparative judgments: The role of nonmotivated factors in above-average and comparative optimism effects. *Psychological Bulletin*, *130*, 813—838.

Costa-Gomes, M. A., & Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, *96*, 1737—1768.

Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology (Vol. 1)* (pp. 498—541)., 5th Edn. Hoboken, NJ: Wiley.

Evans, A. M., & Krueger, J. I. (2009). The psychology (and economics) of trust. *Social and Personality Psychology Compass*, *3*, 1003—1017.

Fehr, E., Naef, M., & Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American Economic Review*, *96*, 1912—1917.

Fetchenhauer, D., & Dunning, D. (in press). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science*. doi:10.1177/0956797609358586

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science*, *20*, 419—422.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, *115*, 811—846.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, *25*, 161—178.

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making*, 2nd Edn. Thousand Oaks, CA: Sage.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia*, *44*, 752—761.

Johnson, E. J., Camerer, C. F., Sen, S., & Rymon, T. (2002). Detecting failures of backward induction: Monitoring information search in ssequential bargaining. *Journal of Economic Theory*, *104*, 1—32.

Krueger, J. I. (2003). Return of the ego—Self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review*, *110*, 585—590.

Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence of social norms. *Negotiation & Conflict Management Research*, *1*, 31—52.

Lewicki, R. J. (2006). Trust, trust development, and trust repair. In M. Deutsch & P. Coleman (Eds.), *The handbook of conflict resolution* (pp. 86—107). San Francisco, CA: Jossey-Bass, Inc.

Luhmann, N. (2000). Familiarity, confidence, trust: Problems and alternatives. In D. Gambietta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 94—107). Oxford, UK: University of Oxford Press.

Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes*, *94*, 61—73.

Moshinsky, A., & Bar-Hillel, M. (2010). Loss aversion and status quo label bias. *Social Cognition*, *28*, 191—204.

Rode, C., Cosmides, L., Hell, W., & Tooby, J. (1999). When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition*, *72*, 269—304.

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science*, *12*, 185—190.
Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*, 393—404.
Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99—118.
Smith, V. L. (2003). Constructivist and ecological rationality in economics. *American Economic Review*, *93*, 465—508.
Snijders, C., & Keren, G. (1999). Determinants of trust. In D. V. Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior* (pp. 355—385). Mahwah, NJ: Lawrence Erlbaum.
Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*, 2nd Edn. Princeton, NJ: Princeton University Press.