

# The Psychology (and Economics) of Trust

Anthony M. Evans\* and Joachim I. Krueger

*Brown University*

---

## Abstract

Interpersonal trust is a mental construct with implications for social functioning and economic behavior. We review contemporary theories of trust from behavioral economics and social psychology. Neoclassical economic theory considers trust in strangers to be irrational, but observed behavior reveals widespread trust and trustworthiness. Theories of social preferences and adherence to social norms have been proposed to rationalize trust. Psychological approaches investigate trusting behavior in terms of an underlying disposition, intergroup processes, and cognitive expectations. The breadth of these approaches illustrates the multi-faceted nature of trusting behavior. The determinants of trust are related to the relevant characteristics of the individual, the situation, and their interaction.

---

‘The entire fabric of our day-to-day living, of our social world, rests on trust – buying gasoline, paying taxes, going to the dentist, flying to a convention – almost all our decisions involve trusting someone else’. (Rotter, 1970, p. 443)

Psychologists from diverse backgrounds have identified interpersonal trust as a social construct of far-reaching significance. The noted developmentalist Erik Erikson (1950), called trust ‘the first task of the ego’ (p. 221). He argued that the ability to trust others is necessary to form relationships and function in the social world. There are inherent risks in any social situation and we often trust that others will not attack us, tell our secrets, or break our hearts. Without the capacity to trust, we would be unable to accept these ordinary uncertainties. Contemporary research examines trust in many domains; such as close relationships (Simpson, 2007), political beliefs (Brewer, 2004), community involvement (Rahn & Transue, 1998), and organizational behavior (Kramer, 2007). The consequences of trust are observed in the workplace, in intimate relationships, and in society at large.

The question of why we trust also has implications in economic exchanges. Nobel laureate Kenneth Arrow (1974) called trust ‘a lubricant for social systems’ (p. 23). A consumer trusts that the car he buys is not a lemon; an employer trusts that his new assistant will be dependable; and an investor trusts that corporate accountants are not cooking the books. The micro-level trust that exists between individuals contributes to the economic development of nations (Fukuyama, 1995; Zak & Knack, 2000). Trust allows a society to invest more capital and conduct transactions with greater efficiency. In recent years, many behavioral (or psychological) economists have begun to examine trust and its consequences. Their work extends traditional theories of economic behavior with theoretical constructs and experimental methods adopted from psychology.

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). There are numerous variations on this definition, but the critical elements of any definition of trust

are vulnerability and expectation. Trust presupposes risk because there must be a possibility for loss and regret. Without personal vulnerability, trust devolves into confidence – a belief without consequences (Luhmann, 2000). Similarly, trust cannot exist without a positive expectation. When people trust, they accept risk because they believe they can avoid a negative outcome. Trust without a positive expectation is self-destructive.

Because of its far-reaching implications, interpersonal trust is a topic with great potential for interdisciplinary collaboration. Our present focus is on trust in social exchanges; in particular, we review trust in situations involving individuals with undeveloped or short-term relationships. Other areas, such as trust in a close relationship or trust in an institution, are beyond the scope of our review. We focus on research conducted by psychologists and economics that illuminate some of the determinants of trusting behavior.

### **Economic Dilemmas of Trust and Reciprocity**

Neoclassical economic theory assumes that individuals are purely self-interested (Von Neumann & Morgenstern, 1944). In other words, people always seek to maximize their own monetary payoffs, without regard for how their choices might affect others. According to this view of rationality, trustworthy behavior is irrational because it works against the person's own material interests. On this assumption, there is no reason to take the risk of trusting someone else. In a world of pure self-interest, every exchange must be enforced with brute force or legal contract.

Fortunately, human behavior does not strictly conform to the pessimistic expectations of classic game theorists. Even behavior in experiments (in which participants remain anonymous) yields strong evidence for other-regarding behavior (Smith, 2003). Although self-interest matters, few people are motivated by nothing else (Krueger, Massey, & DiDonato, 2008). Trust is an important example of a generalized human tendency towards unselfish behavior (see Camerer, 2003 for an overview). In the following section, we describe findings from selected experimental games to show that trust occurs even when it opposes self-interest. These examples demonstrate that the classical model of rationality is insufficient to describe behavior. We then compare preference and norm-based theories of trust.

#### *Empirical paradigms*

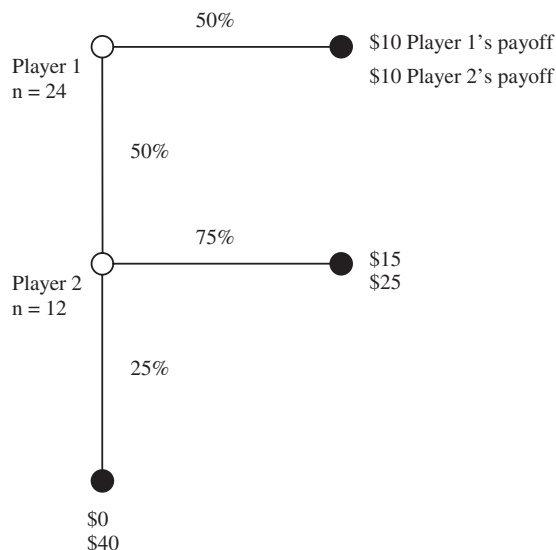
Economists measure trust by studying behavior in laboratory experiments called games. Like many experiments in psychology, games are designed to represent certain critical features of real-life situations. Although many economists view certain experimental practices in psychology with suspicion, they accept the validity of games because players in these games receive motivating financial incentives and they are not deceived (Camerer & Fehr, 2004; Levitt & List, 2007). There are several common features of the games used to measure interpersonal trust: The first player (the trustor) has a choice between trust and mistrust. There is a potential benefit for choosing trust, but the decision to trust requires the acceptance of risk and vulnerability. Ultimately, the outcome of trust depends upon the behavior of the second player (the trustee). Experimental games are useful for studying trust because they provide an external, quantifiable measure of the underlying psychological state of trust.

*The investment game.* The investment game is the standard dilemma of trust and reciprocity in behavioral economics (Berg, Dickhaut, & McCabe, 1995). There are two

players: a sender and a responder. The sender receives a small amount of money (\$10) and an opportunity to invest it. However, much money she invests is tripled and given to the responder, the rest she keeps. The responder then decides how much of the investment to give back to the sender (and how much to keep for herself). The rational equilibrium is clear: If the responder is self-interested, she will keep the sender's entire investment. Knowing this, the sender will not invest anything. By this reasoning, self-interest undercuts all trust and thereby vitiates reciprocity. Yet, Berg et al. found varying degrees of trust. On average, senders invested \$5.16 and received \$4.66 in payback. Of the 28 senders who invested, 11 profited. In a study with a similar design, Pillutla, Malhotra, and Murnighan (2003) found that payback increased as the investment size increased. Most senders who invested the full amount broke even and some made a profit.

*The trust game.* Figure 1 is an example of the trust game, which is a simplified version of the investment game. Decisions in the trust game are discrete choices (McCabe & Smith, 2000). The sender has two options: trust (down) and mistrust (right). If the sender chooses trust, then the returner chooses reciprocity (right) or betrayal (down). Again, the self-interest hypothesis predicts betrayal and mistrust, but as in the investment game, many participants are both trusting and trustworthy. In the McCabe and Smith study, 50% of senders chose trust, and 75% of responders who were trusted chose reciprocity. Rates of trust and trustworthiness vary with experimental conditions and payoffs (Evans & Krueger, 2008; Snijders & Keren, 1999). The likelihood of trust decreases as the risk increases. People become less likely to trust when there is little to gain and a lot to lose. Similarly, reciprocity decreases as the temptation to violate trust increases.

*Trust and the prisoner's dilemma.* The prisoner's dilemma is a canonical example from game theory where two players simultaneously choose whether to cooperate or defect. The players can potentially benefit if they both cooperate. However, each player could



**Figure 1** An example of the extensive-form trust game played by university undergraduates.

earn more (at the other player's expense) by defecting. Again, neoclassical rationality predicts an inefficient equilibrium: mutual defection. Figure 2 depicts a standard form of the game. As in the trust game, players can profit from friendly behavior; the socially optimal outcome of the game is mutual cooperation. In this situation, the risk of defection is symmetric. This is different from the trust game, where the risk is assumed by the sender. The prisoner's dilemma can be thought of as a two-sided trust game – each player is both a sender and a returner.

The prisoner's dilemma can be modified to simultaneously measure mutual cooperation and one-sided trust (Van Lange & Visser, 1999; Yamagishi, Kanazawa, Mashima, & Terai, 2005). As in the standard dilemma, players choose to cooperate or defect. But players also choose whether to increase or decrease their payoff dependence on the other player. Before the game begins, one player chooses between two possible payoff matrices that differ in dependence. Figure 3 illustrates how the game's payoffs change with dependence. As dependence increases, the potential benefits of mutual cooperation (and the potential cost of mutual defection) increase. Thus, participants who wish to increase dependence will choose the top payoff matrix over the bottom. Choosing to increase dependence is a sign of trust because it signifies the expectation that the players will achieve mutual cooperation. Behavior in the modified dilemma reveals that there is a cyclical relationship between trust and cooperation. Individuals increase their trust when the other player has cooperated in the past. Similarly, trust is interpreted as a promise to cooperate in the future.

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	18, 18	-50, 50
	Defect	50, -50	-18, -18

**Figure 2** The normal-form prisoner's dilemma, where players choose between cooperation and defection.

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	34, 34	-50, 50
	Defect	50, -50	-34, -34

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	2, 2	-50, 50
	Defect	50, -50	-2, -2

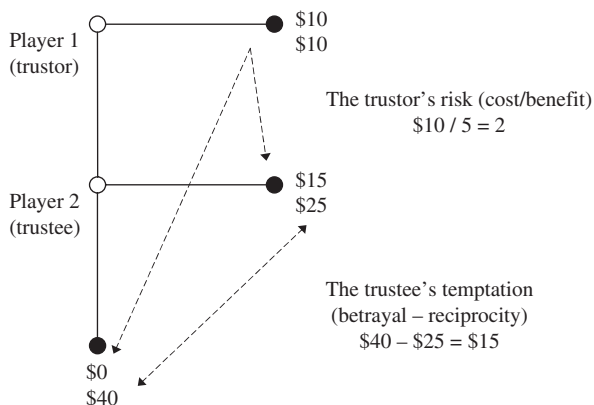
**Figure 3** Players choose between possible payoff matrices of high (top) and low (bottom) interdependence.

*The public-goods game.* The public-goods game measures trust in a group or society. Players are assigned to groups (of two or more) and each person receives a small amount of money. They can invest the money in a fund that benefits everyone in the group or keep it for themselves. The decision of whether to contribute is a question of how much the individual trusts the other group members. If everyone in the group contributes, then the players will collectively profit. But, as in the previous examples, self-interest tempts the players to keep their money (and free-ride on others' contributions). By now, it should come as no surprise that players often contribute to the public good (Iishi & Kurzban, 2008; Yamagishi, 1988). The public goods dilemma can be thought of as a multiple person prisoner's dilemma, it conflates trust with cooperation.

### *Theories of social preferences and norm adherence*

The preceding examples illustrate that the classical view of rationality fails to predict how people act. Behavioral economists have proposed theories of rationality that are more consistent with observed behavior. Social preference theories explain trust without abandoning the framework of utility maximization (Bolton & Ockenfels, 2000; Fehr & Schmitt, 1999). These theories assume that at least some people treat the payoffs of others as positive utilities. Individuals with benevolent preferences trust, cooperate, and reciprocate because it makes them feel good. Their economic decisions are difficult in situations in which social preferences are directly opposed to self-interest. When cooperation is costly, rational agents need to weigh self-interest against the satisfaction of costly other-regarding motives.

There are two major types of social preference models. According to models of inequality-aversion, people prefer their own payoffs to be neither higher nor lower than the payoffs of others (Bolton & Ockenfels, 2000; Fehr & Schmitt, 1999). Like Robin Hood – they want to establish equality by transferring resources from the rich to the poor. Other models stress fairness and reciprocity (Rabin, 1993). Reciprocity means that people want to help others who treat them kindly and to punish those who hurt them. The two models make distinct predictions. Inequality-aversion predicts that we help those less fortunate, whereas reciprocity predicts we help those who help us. There is some evidence supporting both models, but recent work suggests that the reciprocity motive is particularly powerful (Falk, Fehr, & Fischbacher, 2008; Güth, Kliemt,



**Figure 4** Trustee's temptation and trustor's risk in the trust game.

& Ockenfels, 2003). The recent economic models are supported by decades of social-psychological research on social values (e.g., Messick & Sentis, 1985; Van Lange, 1999). There is consistent empirical evidence that people often engage in prosocial behavior and prefer equality.

A criticism of the early social preference models is that they are consequentialist. Social preferences assume that decisions depend on the anticipated final consequences of behavior without regard to how these consequences come about (Smith, 2003). As the following examples demonstrate, intentions may have a significant effect on trust and reciprocity (Cox, 2004; McCabe, Rigdon, & Smith, 2003). Consider the following situations:

- (1) *The Investment Game*: Player A is given \$10 to invest. She decides to invest all of it. Player B is given \$30 and decides how much to return to A.
- (2) *The Dictator Game*: A has no decision to make. B is given \$30 and decides how much to share with A.

In terms of consequences, these two situations are identical to Player B and social preference theories predict that Player B will send back the same amount of money in each situation. However, Cox (2004) found that Player B returns more money in the first situation. McCabe et al. (2003) obtained the same result with the trust game. The decision to trust creates goodwill between the players, which leads to reciprocity.

In an interesting twist, Kiyonari, Yamagishi, Cook, and Cheshire (2006) reported a contrary effect that they labeled 'negative reciprocity'. Kiyonari and colleagues compared the traditional trust game with the faith game. The faith game is similar to the trust game, except that the responder does not know that the sender was given a choice about whether to trust.<sup>1</sup> As in the study conducted by Cox (2004), the responder believes she is playing a dictator game - the consequences are the same. But unlike the previous studies, responders in the faith game returned more money than those in the trust game. The effect of negative reciprocity was minor (average amounts returned were 39 and 33% of the amount sent) and only marginally significant, but this finding suggests that additional studies are needed to qualify the conditions that affect the emergence of reciprocity.

Another experiment was conducted to test whether the decision of Player A, the sender, was motivated by a preference to help Player B or the expectation that Player B would return money (Cox, 2004). Consider the following games:

- (1) *The Investment Game*: Player A is given \$10 to invest, however much she invests will be tripled and given to Player B. Player B then decides how much to return.
- (2) *The Dictator Game*: A is given \$10 to invest, however much she invests will be tripled and given to B. In this game, B has no opportunity to return money to A.

If Player A invests because of a desire to help Player B, then these two situations should be identical. However, Player A tends to send more money in the first case. Player A's investing is motivated in large part by an expectation that Player B will return money (in other words, trust), instead of a benevolent concern for Player B's well-being. Cox (2004) argues that the decisions of both the sender and the responder are motivated by intentions and expectations, rather than outcomes. The ongoing debate over consequentialism has motivated recent preference models to acknowledge the potential role of intentions (Cox, Sadiraj, & Sadiraj, 2008; Falk et al., 2008).

Norm-based theories of trust and reciprocity are alternative approaches to preference models (Burnham, McCabe, & Smith, 2000). Rather than assuming that social choices are made with the goal of maximizing outcomes (either for the self or others), these theories assume that decisions are affected by the context in which they occur. Norms (such as fairness or revenge) are activated by cues from the situation and knowledge of others' past actions (Bicchieri, 2008). According to this approach, trust and trustworthiness occur because individuals are motivated by the desire to follow social norms (Krueger et al., 2008).

Burnham et al. (2000) found that referring to players as 'partners' rather than 'opponents' increased both trust and trustworthiness. In particular, rates of trustworthiness doubled, increasing from 33 to 68 percent. Social preference theories cannot explain context effects, such as changing the instructions of games (Smith, 2003). The norm-adherence theory of trust, in contrast, suggest that what matters most in economic games is how the individual perceives the situation.

Context shapes behavior in many ways. For example, framing a situation as 'economic' has detrimental effects on cooperative behavior. Vohs, Mead, and Goode (2006) found that people become more self-focused and less likely to help others when they are experimentally induced to think about money. Heymann and Ariely (2006) noted that markets are either monetary or social. Whereas monetary markets encourage quid-pro-quo behavior, social markets inspire less calculated acts of reciprocity. Although it is acceptable to reward a hard-working employee with a \$10,000 bonus, it is socially inappropriate to give a bonus to a spouse as an anniversary present. In social markets, it's the thought that counts. The distinction of calculus and identity-based trust is a useful framework for interpreting these differences in behavior (Lewicki, 2006): Calculus-based trust exists in relationships where prosocial behavior is maintained through the threat of punishment and the potential benefits of reciprocity. Trust is calculated by weighing risks and benefits. Identification-based trust, in contrast, emerges through empathy and identification with another person's intentions and desires. The characteristics and strength of trust depend on whether the relationship is primarily economic (calculus-based) or intimate (identification-based).

Economic interactions occur in a domain where selfishness is not only lucrative, but also considered rational. It is not surprising that among college students, economics majors are comparatively less likely to cooperate in social dilemmas (Frank, Gilovich, & Regan, 1993). Economic framing has an antisocial influence, but self-report data suggest that people perceive games of trust and reciprocity as moral, rather than strategic, dilemmas. Krueger et al. (2008) asked participants to rate the perceived morality and rationality of different decisions in the investment game. They found that decisions of investing and reciprocating were only correlated with perceptions of morality. Malhotra (2004) also found that players interpreted the behavior of others in moral terms. Yet, they perceived their own behavior in terms of rationality. These results suggest that rational behavior is not limited to utility maximization. What is rational depends on the characteristics of the environment and the individual. According to norm-based theories of trust, cooperation can be both rational and profitable.

The work stimulated by social preference and norm-based theories shows that behavioral economics has overcome the limitations of the traditional self-interest hypothesis. Trust and reciprocity are not necessarily irrational when they are understood in these terms. As the experimental data demonstrate, other-regarding behavior often leads to socially desirable results. In the next section, we review social-psychological perspectives on trust and consider their implications for economic research.

## Social-Psychological Determinants of Trust

Although many behavioral economists utilize laboratory studies in their research, there is still a salient divide between psychology and economics. Many of the differences between the fields are methodological in nature. Economists tend to distrust the validity of self-report data. They insist that behavior must be financially motivated and cannot occur under false pretenses (deception is not permitted). There are also deep theoretical differences between the approaches of the two fields. A traditional framework for social psychology is to interpret behavior as the product of factors from the person, the situation, and their specific interaction (Krueger, 2009; Lewin, 1936). It is accepted that social behavior is the result of numerous interacting determinants. This approach is distinct from the economic tradition, which emphasizes the need for a unified, rational explanation of behavior. Furthermore, economic approaches are supported by the use of formal mathematical models. The traditional economic method is to describe behavior as the maximization of utilities (either social or selfish). These differences (both methodological and theoretical) inhibit cross talk between the two fields.

In this section, we present three avenues of trust research from social psychology. There is evidence that the decision to trust is related to personality differences, social identity, and expectations. Although these areas are not a comprehensive representation of all psychological approaches, they provide insight into the intrapersonal processes that motivate trust in economic exchanges.

### *Individual differences in trust*

Experimental games illustrate the simple point that some people are more trusting than others. Recall the investment game: Berg et al. (1995) reported that players invested and returned variable amounts of money. Out of 32 senders, five invested the full amount, five invested half, and two invested none, with the rest sending amounts in between. There was no single tendency to describe how people acted. When trusting behavior is measured, there are significant and consistent individual differences.

These differences are in part predicted (and perhaps caused) by an underlying propensity to trust (Evans & Revelle, 2008; Rotter, 1967). This propensity is most relevant in ambiguous situations; disposition matters most when there is little first-hand knowledge or past experience. Rotter (1967) defined dispositional trust as the general expectation that others will behave fairly and responsibly. Individuals with a high propensity to trust are generally optimistic, especially in novel or unscripted situations. Other dispositional theories are similar to Rotter's as they emphasize the importance of situational uncertainty (Yamagishi & Yamagishi, 1994). Hence, dispositional theories assume that the choice between trust and distrust is most keenly experienced in interactions with strangers. Trust in an established relationship is more constrained by context and experience. No matter how strong their propensity is, people hesitate to trust salespersons with dubious reputations or friends who have disappointed them in the past. Yet, even the most hardened cynic has someone that she can tell her secrets to.

Economists report that personality differences play a weak or unobservable role in the decision to trust (Glaeser et al., 2000; Holm & Nystedt, 2008). However, these results were obtained with measures of dubious psychometric quality. When reliable and validated instruments are used, behavior is consistently correlated with disposition. Survey measurements of propensity correlate with decisions in games of trust.



Some evidence suggests that a trusting disposition has a genetic component. In a study of monozygotic and dizygotic twins, Cesarin et al. (2006) detected a genetic influence on decisions in the investment game. Common genes had a larger effect than shared environment (growing up in the same household). This finding is consistent with evidence that facet-level personality traits, including the propensity to trust, have significant genetic components (Jang, McCrae, Angleitner, Rieman, & Livesly, 1998).

Individual differences add psychological depth to the analysis of economic behavior, but the propensity to trust alone does not offer a complete explanation for trusting behavior. The situation and context matter. Consider the small-group situation characterized by the public goods game. Yamagishi (1988) found that disposition predicted the size of contributions. When players had opportunities to punish free-riders at some cost to themselves, the trait of trust was negatively correlated with own contributions to the public good. Low-trust individuals sacrificed more of their own money to punish defectors, but they also invested more. Such a result highlights the need to search for person-by-situation interaction effects in addition to main effects of person and situation.

### *Depersonalized trust and ingroup cooperation*

When the decision to trust occurs in an intergroup context, the general tendency to favor the ingroup becomes relevant. For example, you may be more likely to lend a stranger money if you know that she is a member of your local church. Or you may be more likely to invest in a start-up company if the CEO was a member of your college fraternity. In these situations, you are engaging in depersonalized trust – trusting someone because of a shared social identity (Brewer, 1986, 2008). This form of trust is motivated by the expectations that members of the same group will reciprocate and cooperate with one another, and that those who do not meet these expectations will be punished by the group. Depersonalized trust is an implicit contract shared by ingroup members (James, 2002).

When social identity is manipulated in the investment game, ingroup favoritism increases the frequency of trust. Tanis and Postmes (2005) conducted an experiment where participants played the role of the sender and learned about the social identity of the responder. Social identity was manipulated such that members of the ingroup (outgroup) were students at the same (rival) university. When participants were only shown social identity information, they were more likely to send money to someone from their ingroup than their outgroup. Expectations of reciprocity mediated this effect; senders trusted members of their ingroup more because they believed that their partners were more likely to send money back. In another experimental condition, participants learned about both the social and personal identities of the other player. Personal identity was controlled by showing a photo of the other player. When the picture was added, group identity no longer predicted investing. The decision to trust was based on personal (not social) identity. These results illustrate the difference between person-based and depersonalized trust. The mental processes underlying trust depend on whether we perceive the trustee as an individual or as a member of a group.

Two recent studies by Foddy, Platow, and Yamagishi (2009) provide further evidence (and qualification) of depersonalized trust. In a first study, participants played a game where they were given the choice to receive unknown amounts of money from either an ingroup or outgroup member (the allocator game). In their design, participants are required to trust one of the potential allocators, but they are free to choose who they trust. The experimenters manipulated whether the potential allocators knew about

the participant's group membership: When group membership was common knowledge, participants unanimously preferred the ingroup member. But when group membership was concealed, only 53% chose to trust the ingroup. This finding suggests that participants show ingroup trust because they expect reciprocity based on group membership. In the private knowledge condition, stereotypes about the characteristics of the ingroup (instead of expectations) predicted ingroup trust.

In a second study, Foddy et al. examined members of two different outgroups: economics and nursing majors (the participants were psychology majors). These two outgroups have contrasting stereotypes; nursing (economics) majors are perceived positively (negatively). As in the first study, ingroup favoritism was strong in the common knowledge condition. But a strong difference emerged in the private knowledge condition: 80% still chose the ingroup allocator when the outgroup member was an economics major, whereas only 40% chose the ingroup over a nursing major. These studies demonstrate the strong effects of depersonalized trust, but show that it is less prevalent when ingroup membership is not common knowledge. In the private knowledge condition, ingroup trust was contingent on stereotypes of the ingroup and outgroups.

Generally, trust decreases when the social distance between the trustor and the trustee increases. People are more likely to trust others when they perceive them as familiar and socially similar. Glaeser et al. (2000) found that trust was positively correlated with the amount of time that the players knew each other and the number of common friends. Participants were also less trustworthy when interacting with someone from another country. Greiner, Ockenfels, and Werner (2007) conducted trust games among players with equal and unequal endowments. The equally endowed players were initially more trusting. Evans and Revelle (2008) compared one-sided and symmetric investment games. In the symmetric games, both players in the game simultaneously played the role of the sender. Investing increased when participants knew that the other player was currently facing the same situation.

The impact of depersonalized trust depends on the type of social group involved and on local cultural norms. Güth, Levati, and Ploner (2008) studied behavior among minimal groups (a canonical psychological paradigm), that is, participants were randomly assigned to groups using arbitrary symbols. Although participants liked and identified with ingroup members, there was no evidence of depersonalized trust. Johansson-Sterman, Mahmud, and Martinsson (2005) measured trusting behavior among Hindu and Muslim participants from Bangladesh. The authors found no ingroup preference based on religious orientation. Yuki, Maddox, Brewer, and Takemura (2005) compared depersonalized trust among American and Japanese participants. American participants showed evidence of depersonalized trust, but the Japanese were more likely to base their trust on their individual relationships with members of the ingroup and outgroup. These diverse results show that although there is some evidence of depersonalized trust, the phenomenon is far from universal.

Depersonalized trust is one of several cognitive explanation of why individuals are more likely to cooperate with ingroup members. Brewer (2008) identified three additional mechanisms that could facilitate trust between ingroup members: heuristic thinking, social projection, and social identification and goal transformation. The heuristic approach suggests that people follow a simple mental rule to trust ingroup members (and mistrust outgroup members). Ingroup cooperation is a mental shortcut that simplifies the decision of whom and when to trust. The social projection account suggests that individuals project their own characteristics onto members of the ingroup (Krueger & DiDonato, 2008).

Believing that others share our social values, we trust ingroup members inasmuch as we regard ourselves as trustworthy. In contrast, the social identity approach suggests that individuals form emotional attachments to their groups and become motivated to pursue the best interests of the group (Tajfel & Turner, 1979). Depersonalized trust and these alternative mechanisms are not mutually exclusive. The processes that influence the decision to trust depend upon the characteristics of the group and the salient information presented in the situation.

### *Expectations of reciprocity and betrayal*

By definition, trust presupposes the expectation of how others will act when they are trusted (Rousseau et al., 1998). Trust without the expectation of reciprocity is self-destructive. Dispositional and depersonalization theories of trust promote different approaches to the idea of expectation: The dispositional approach posits that trust is motivated by individual differences in a general expectation. The depersonalization account proposes that the expectation of reciprocity depends on shared social identity. Recent studies have examined the relationship between expectations of reciprocity and trusting behavior. Rationality predicts that there should be a one-to-one correlation between expectations and decisions, but empirical results show that this relationship is complex.

In trust games, economic rationality predicts that the trustor should form expectations that are based on the incentives of the trustee. A rational person should be less likely to trust if she knows that the trustee has a lot to gain from betrayal. Snijders and Keren (1999) manipulated the payoffs of the trust game to test this prediction. The sender's risk (a ratio of potential costs and benefits) and the responder's temptation (the benefit for betraying the sender) were examined as predictors of behavior. Figure 4 illustrates these two constructs in a basic trust game. Responders were more likely to act selfishly when temptation was high. Every trustee, it seems, has a price. Surprisingly, the senders were barely influenced by the responder's temptation. They paid more attention to their own costs and benefits. Evans and Krueger (2008) explored this result in greater detail and found that senders rated the probability of betrayal as being higher when the potential costs of being betrayed increased. Players paid more attention to their own payoffs, in spite of the prediction that they should base their expectations on the incentives of the responder.

Expectations depend on how players think about the situation. Kugler, Connolly, and Kausel (2009) primed players to think of trust in terms of its possible outcomes, activating what the authors called 'consequentialist thinking'. In one condition, participants were asked to predict how much money the other player would return if they invested different amounts. Behavior was radically changed by this manipulation; players invested less after thinking about the possible consequences. This change supports the hypothesis that people do not normally think about the responder's incentives. It also suggests that people are normally over-trusting, they invest without thinking about what the responder will do with their investment. Fetchenhauer and Dunning (2009) reached a similar conclusion. After measuring investing decisions and expectations, they found that given their own level of cynicism about others people trusted too much. According to this logic, trust is a lack of foresight. An alternative interpretation is that trust is associated with a mental state where we give others the benefit of the doubt by default (Schul, Mayo, & Burnstein, 2004, 2008). When we put trust in another person, we do not need to think about the possible consequences.

Expectations may cause decisions, as any rational actor theory assumes, but there may also be a reciprocal relationship between beliefs and choices. Dawes, McTavish, and Shacklee (1977) found that expectations are influenced by decision making. People who choose to defect expect a greater percentage of others to defect as well, either to justify their own decision or because they assume that others share their values. Messé and Sivacek (1979) reported evidence supporting both explanations. People assume that strangers will make decisions that mirror their own, but this effect is stronger when they make predictions about people with whom they interact directly. When we trust a stranger, we need to justify accepting the risk that our trust could be violated.

## Conclusion

Psychologists and economists have proposed numerous explanations for why we trust. The construct of trust can be viewed in terms of economic preferences, social norms, personality traits, group processes, or expectations. The breadth of these approaches illustrates the multi-faceted nature of trust. There is no single force that explains why we accept or reject the vulnerability and uncertainty in dilemmas of trust. To understand why one person trusts another, we must consider the characteristics of the specific individuals and the situation. What is the social relationship between the trustor and the trustee? How does this situation relate to their past experiences and social expectations? What are the potential benefits and costs? The factors that determine behavior vary from situation to situation and from person to person. As Lewin (1936) wrote, 'Every psychological event depends upon the state of the person and at the same time on the environment' (p. 12).

## Short Biography

Anthony Evans' research examines the psychological and economic aspects of interpersonal trust and reciprocity; his work integrates methods and theoretical perspectives from social and personality psychology and behavioral economics. His current work is investigating how interpersonal exchanges are affected by self-regulatory processes; this research project is funded by a grant from the Russell Sage Behavioral Economics Roundtable. Other ongoing projects study the dynamics of role choice in strategic games and the role of individual differences in economic decisions. Evans holds a BA in Mathematical Methods in the Social Sciences and Psychology from Northwestern University and a MS in Experimental Psychology from Brown University.

Joachim I. Krueger is Professor of Psychology at Brown University. His research interests include self-perception, social stereotyping, and strategic reasoning in interpersonal and intergroup contexts. He has recently edited volumes on 'The self in social judgement' (with M. Alicke & D. Dunning, Psychology Press, 2005) and 'Rationality and Social responsibility' (Psychology Press, 2008). More information can be found at [http://research.brown.edu/research\\_profile.php?id=10378](http://research.brown.edu/research_profile.php?id=10378).

## Endnotes

\* Correspondence address: Department of Psychology, Brown University, Box 1853, 89 Waterman St., Providence, RI 02912, USA. Email: [Anthony\\_Evans@Brown.edu](mailto:Anthony_Evans@Brown.edu)

<sup>1</sup> Another difference is that the trustor's choice is dichotomous (all or nothing), while the trustee's choice is continuous.

## References

- Arrow, K. (1974). *The Limits of Organization*. New York, NY: Norton & Company, Inc.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social-history. *Games and Economic Behavior*, **10**, 122–142.
- Bicchieri, C. (2008). How expectations affect behavior. In J. I. Krueger (ed.), *Rationality and Social Responsibility: Essays in Honor of Robyn Mason Dawes* (pp. 187–214). New York, NY: Taylor & Francis Psychology Press.
- Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and cooperation. *American Economic Review*, **90**, 166–193.
- Brewer, M. B. (1986). Ethnocentrism and its role in interpersonal trust. In M. Brewer & B. Collins (Eds.), *Scientific Inquiry and the Social Sciences: A Volume in Honor of Donald T. Campbell* (pp. 345–360). San Francisco, CA: Jossey-Bass.
- Brewer, P. (2004). Public trust in (or cynicism about) other nations across time. *Political Behavior*, **26**, 317–341.
- Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In J. I. Krueger (Ed.), *Rationality and Social Responsibility: Essays in Honor of Robyn Mason Dawes* (pp. 215–232). New York, NY: Taylor & Francis Psychology Press.
- Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior and Organization*, **43**, 57–73.
- Camerer, C. F. (2003). *Behavioral Game Theory*. Princeton, NJ: Princeton University Press.
- Camerer, C. F., & Fehr, E. F. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr & F. Gintis (Eds.) *Foundations of Human Sociality: Experimental and Ethnographic Evidence From Fifteen Small Scale Societies* (pp. 55–95). Oxford, UK: Oxford University Press.
- Cesarin, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2006). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of Sciences*, **105**, 3721–3726.
- Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, **46**, 260–281.
- Cox, J. C., Sadiraj, K., & Sadiraj, V. (2008). Implications of trust, fear, and reciprocity for modeling economic behavior. *Experimental Economics*, **11**, 1–24.
- Dawes, R., McTavish, J., & Shacklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology*, **35**, 1–11.
- Erikson, E. (1950). *Childhood and Society*. New York, NY, Norton & Company, Inc.
- Evans, A. M., & Krueger, J. I. (2008). *Expectation, Risk, and Disposition in the Trust Game*. Poster presented at the Society for Judgment and Decision Making Annual Meeting, Chicago, IL: ?????.
- Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, **42**, 1585–1593.
- Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness: Intentions matter. *Games and Economic Behavior*, **62**, 287–303.
- Fehr, E., & Schmitt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, **114**, 817–867.
- Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science*, **20**, 419–422.
- Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *The Journal of Economic Perspectives*, **7**, 159–171.
- Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. New York, NY: Free Press.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, **115**, 811–846.
- Greiner, B., Ockenfels, A., & Werner, P. (2007). *The Dynamic Interplay of Inequality and Trust: An Experimental Study. Working Paper*. Cambridge, MA: Harvard University Business School.
- Güth, W., Kliemt, H., & Ockenfels, A. (2003). Fairness versus efficiency: An experimental study of (mutual) gift giving. *Journal of Economic Behavior & Organization*, **50**, 465–475.
- Güth, W., Levati, V., & Ploner, M. (2008). Social identity and trust: An experimental investigation. *The Journal of Socio-Economics*, **37**, 1293–1308.
- Heymann, J., & Ariely, D. (2006). Effort for payment: A tale of two markets. *Psychological Science*, **15**, 787–793.
- Holm, H., & Nystedt, P. (2008). Trust in surveys and games: A methodological contribution on the influence of money and location. *Journal of Economic Psychology*, **29**, 522–542.
- Iishi, K., & Kurzban, R. (2008). Public goods games in Japan: Cultural and individual differences in reciprocity. *Human Nature*, **19**, 138–156.
- James, H. S. (2002). The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior & Organization*, **47**, 291–307.

- Jang, K. L., McCrae, R. R., Angleitner, A., Rieman, R., & Livesly, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample : Support for a hierarchical model of Personality. *Journal of Personality and Social Psychology*, **74**, 1556–1565.
- Johansson-Sterman, O., Mahmud, M., & Martinsson, P. (2005) *Trust and Religion: Experimental Evidence From Bangladesh. Working Paper*. Göteborg, Sweden: Göteborg University.
- Kiyonari, T., Yamagishi, T., Cook, K. S., & Cheshire, C. (2006). Does trust beget trustworthiness? Trust and trustworthiness in two games and two cultures. *Social Psychology Quarterly*, **69**, 270–283.
- Kramer, R. M. (2007). *Organizational Trust: A Reader*. Oxford, UK: Oxford University Press.
- Krueger, J. I. (2009). A componential model of person, situation, and person by situation effects on social behavior. *Journal of Research in Personality*, **43**, 127–136.
- Krueger, J. I., & DiDonato, T. E. (2008). Social Categorization and the Perception of Groups and Group Differences. *Social and Personality Psychology Compass*, **2**, 733–750.
- Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence to social norms. *Negotiation and Conflict Management Research*, **1**, 31–52.
- Kugler, T., Connolly, T., & Kausel, E. E. (2009). The effect of consequentialist thinking on trust game behavior. *Journal of Behavioral Decision Making*, **22** (2), 101–119.
- Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, **21**, 153–174.
- 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.
- Lewin, K. (1936). *Principles of Topological Psychology*. New York, NY: McGraw-Hill.
- 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.
- McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*, **52**, 267–275.
- McCabe, K. A., & Smith, V. L. (2000). A comparison of naive and sophisticated subject behavior with game theoretic predictions. *Proceedings of the National Academy of Sciences*, **97**, 3777–3781.
- Messé, L. A., & Sivacek, J. M. (1979). Prediction of others' responses in a mixed-motive game: self-justification or false consensus? *Journal of Personality and Social Psychology*, **37**, 602–607.
- Messick, D. M., & Sentis, K. P. (1985). Estimating social and nonsocial utility functions from ordinal data. *European Journal of Social Psychology*, **15**, 389–399.
- Pillutla, M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, **39**, 448–455.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, **83**, 1281–1301.
- Rahn, W. M., & Transue, J. E. (1998). Social trust and value change: The decline of social capital in American youth, 1976–1995. *Political Psychology*, **19**, 545–565.
- Rotter, J. B. (1967). A new scale for measurement of interpersonal trust. *Journal of Personality*, **35**, 651–655.
- Rotter, J. B. (1970). Generalized expectancies for interpersonal trust. *American Psychologist*, **26**, 443–452.
- 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.
- Schul, Y., Mayo, R., & Burnstein, E. (2004). The value of distrust. *Journal of Experimental Social Psychology*, **44**, 1293–1302.
- Schul, Y., Mayo, R., & Burnstein, E. (2008). Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology*, **86**, 668–679.
- Simpson, J. (2007). Psychological foundations of trust. *Current Directions in Psychological Science*, **16**, 264–268.
- 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.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (eds.), *Psychology of Intergroup Relations* (pp. 33–47). Chicago, IL: Nelson-Hall.
- Tanis, M., & Postmes, T. (2005). A social identity approach to trust: Interpersonal perception, group membership, and trusting behavior. *European Journal of Social Psychology*, **35**, 413–424.
- Van Lange, P. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, **77**, 337–349.
- Van Lange, P., & Visser, K. (1999). Locomotion in social dilemmas: How people adapt to cooperative, tit-for-tat, and noncooperative partners. *Journal of Personality and Social Psychology*, **77**, 762–773.
- Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science*, **314**, 1154–1156.

- Von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ, Princeton University Press.
- Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly*, **51**, 265–271.
- Yamagishi, T., Kanazawa, S., Mashima, R., & Terai, S. (2005). Separating trust from cooperation in a dynamic relationship. *Rationality and Society*, **17**, 275–308.
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, **18**, 129–166.
- Yuki, M., Maddox, W. M., Brewer, M. B., & Takemura, K. (2005). Cross-Cultural differences in relationship- and group-based trust. *Personality Social Psychology Bulletin*, **31**, 48–62.
- Zak, P. J., & Knack, S. (2000). Trust and growth. *The Economic Journal*, **111**, 295–321.