Trust versus Rationality

David Dunning
University of Michigan

Thomas Schlösser
South Westphalia University of Applied Sciences

Detlef Fetchenhauer
University of Cologne

Abstract

Trust between people is essential for individuals and societies to flourish, yet making oneself vulnerable to another person is unreasonable under a rational actor model—a theoretical perspective that has dominated many behavioral and biological sciences over the last century. We discuss this paradox and what it means for our understanding of trust behavior. We suggest that although people may trust others in part out of rational self-interest, work on economic games suggests that trust is driven by a normative component. People trust because they believe it is what they should do. Thus, people do not interpret trust decisions along economic principles of gain and loss as much as they along matters of morality, etiquette, and social obligation. We demonstrate most directly by showing how differently people act in the trust game versus another classic economic game, the prisoner's dilemma. This perspective suggests that people may act more because of emotion rather than calculation, are focused more on their actions rather than the outcomes those actions might bring, may not choose out of preference, are not guided by altruism, and may not consciously understand what leads them to trust. Trust has cognitive, behavioral, and emotional components, and those components may not cohere.

One aspect of trust has received far too little attention, just how much of a scandal the existence of human trust represents. Much like other forms of other-regarding behavior, it rather outrageous that people choose to act as though they can rely on the integrity of people they know, along with the kindness of strangers. As science and technology theorists (Haraway, 1991), biologists (Williams, 166), and philosophers (Hull, 2001) have pointed out, with the possible exceptions of close family members or tight circles of well-established friends, there is little rationale for trust among disposable acquaintances, and none whatsoever between strangers who will never see one another again.

To be sure, like other forms of human altruism and cooperation, interpersonal trust undeniably confers benefits to individuals, organizations, and to societies that practice it (see Fetchenhauer, this volume). These benefits are considerable. People who trust gain more income over time (Stavrova & Ehlebracht, 2016), ascend to positions of leadership (Stavrova et al., 2024), are treated better by others (Stavrova et al., 2020), experience healthier intimate relationships (e.g., (Murray et al., 2025), and attain better overall well-being (Zhang, 2020). Trust in organizations facilitates positive performance, citizenship behavior, and reduces conflict and its associated stress (Dirks & Ferrin, 2001). Nations high in trust experience higher economic standards of living (Fetchenhauer & Van der Vegt, 2001; Knack & Keefer, 1997) and greater economic growth (Zak & Knack, 2001). Trust in civic institutions lies at the heart of thriving democracies (Rohrschneider & Schmitt-Beck, 2002; Warren, 1999).

Thus, trust is adaptive, bringing untold benefits to human life at both the individual and societal level. Philosophers (e.g., Aristotle, 1876; Hobbes, 1651; Kant, 1909) and behavioral scientists (e.g., Rotter, 1971; Simpson, 2007) have gone to great lengths to extol just how essential trust is for any beneficent social interaction to develop, thrive, and persist.

The scandal emerges, despite its adaptive benefits, because trust contradicts another core element of human flourishing, one that stands also at the heart of successful scholarly work striving for understanding of the human condition. One predominant theoretical framework underpinning much of the biological and behavioral sciences focuses on the crucial need for human rationality. People think, calculate, apply logic, and they do so primarily toward a singular purpose. As many scholars have noted for over a century, people follow a rational actor model, individuals act overwhelmingly out their own material self-interest. First and foremost, they wish to survive, procreate, maintain their health, grow their wealth, and make life easier for themselves. They make careful analyses of what is in their best interests and then act accordingly. Thus, the rational actor model ensures that they gain all those things.

The rational actor, or rational choice, model has been a firm foundation for many insights in a myriad of intellectual disciplines for a very long time. It guides those academic pursuits just as well as, it can be argued, it has guided human (and nonhuman) behavior through the millennia. It has been influential in economic theory for over a century. Adam Smith, for example, cited the awesome power of self-interest in his doctrine of the invisible hand (Smith, 1853), but the rational choice model only reached a zenith of dominance in the mid- to late-twentieth century following World War II (Arrow, 1959; Becker, 1962; Von Neumann & Morgenstern, 1953). Its intellectual reach extended far beyond matters of money. It was deployed to explain voting behavior

in political science (Downs, 1957), choices that seem more intimate such as whether to stay in married (Becker et al., 1977) and applied to behaviors that seem downright irrational or pathological like committing crime (Becker et al., 1968) or taking drugs (Becker & Murphy, 1968).

How Trust Challenges Rationality

Interpersonal trust, though, is difficult to reconcile with rationality. The contradiction is straightforward. As one popular academic definition puts it, trust comprises "the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau et al., 1998, p. 395). Rationally, however, no one should ever hold positive expectations about the intentions or behavior about another person except under very constrained circumstances.

For example, in behavioral economics, consider the laboratory game known as the trust or the investment game (as originated by Berg et al., 1993), which lays out the bare bones of trust in a visible way. In the game, displayed schematically in Figure 1, a person is given an amount of money—in our American version of the game it is US$5—and they are told that they can either keep the money or give it to some partner we have assigned to them. If they give it to the partner, the amount is inflated to US$20, and the partner is given a decision to make: They can give US$10 back to the original person or they can keep the entire amount for themselves. According to the rational actor model, that partner, responding to the dictates of their material self-interest, should always keep the money, giving nothing back. Thus, the original person should never trust. The rational analysis is rather simple, and the decision to avoid trusting another person compellingly clear.

To be sure, if the person doing the trusting and the partner to be trusted are in a permanent relationship where reciprocal transactions are common and goodwill must be maintained, then one can possibly have positive expectations that one's trust will be honored. Thus, it may be rational to give the US$5, but that does not extend to many daily interactions arising in everyday life, especially when people leave their house to shop at a new store, or leave town, talk to a new salesperson, or shop at a website. Many transactions with others are one-offs, and there is no reason to guarantee the other person will harbor a mutually self-interested reason to reciprocate trust. Perhaps one could be trust that another person if one had a readout of their past behavior establishing that they tend to honor trust, but that is often unavailable. Or, instead, one could trust if one had some leverage over that person's behavior that forced that person to be trustworthy, but then again that would hardly qualify as a trusting situation and another word should be used altogether.

Those safer circumstances do not usually apply, yet people go ahead and extend trust, both in the real world and in the laboratory (Johnson & Mislin, 2011, 2012). More confounding, that trust is often reciprocated when the other person is under no formal obligation or pressure to do so (Berg et al., 1993; Johnson & Mislin, 2011, 2012). More tellingly, in behavioral economic games like the laboratory trust game, people trust complete strangers, including anonymous ones whose identities they do not know, whom they will never meet face-to-face, and in one-time transactions containing no further consequences that could motivate trust, such as establishing one's reputation or goodwill for a more permanent relationship (Berg et al., 1993). Thus, trust behavior fails

central tenets of the rational choice model, as does the reciprocal act of honoring that trust.

That said, one could still argue that trust still fits a second, looser definition of rationality, based on the notion that people know that they live in communities that strike mutually beneficial social routines, and so they are safe to make themselves vulnerable to other people who live in those communities because others can be expected to honor their trusting actions. That arrangement is a bargain that the community—whether it be a tribe, a town, or a society—have struck among its members and does not have to be renegotiated with each new pairing of members within it. Much like their peers in economics, post-war sociologists wrote much about this type of rationality in social folkways and contracts that made living less nasty, brutish, and short (Blau, 1964/2017; Coleman, 1994; Homans, 1958). Thus, the act of trust could be rational because of an unwritten but quite understood social arrangement between participants and partners.

However, trust behavior, at least in the laboratory, fails that test of rationality. If there is a social contract, people do not seem to be consciously aware of it. In our studies using the $5/$20 format, or its variants, we have found little evidence that participant sare acting rationally because they know of a social arrangement that ensures their self-interest if they choose to trust. In our original studies (Fetchenhauer & Dunning, 2009), we asked participants the likelihood that their partner would return money to them. In Study 1, participants said the chance on average was only 45%; in Study 2 the overage estimate was 59%. These are estimates that are far lower than ones that would ensure trust. In a second scenario, participants were asked what minimum probability of winning they would require to gamble US$5 to win $US10. In Study 1, that probability was 62%; in Study 2, 75%.

Given these estimates, participants should have been reticent to trust their assigned partner. Indeed, only 30% in both studies should have been willing, in that for them the odds they thought that their partner would reciprocate their trust was equal or greater than the minimum odds they would accept to gamble US$5 in the lottery. However, a full 64% of Study 1 and 56% of Study 2 participants chose to trust their partner. Thus, rationality at this second circle of defense, again, failed—a finding we have repeatedly obtained (Dunning et al., 2014; Fetchenhauer & Dunning, 2012). To be sure, if people make their trust decisions in a group, discussing their options, they do end up trusting less than if they make decisions individually, showing perhaps more rationality (Kugler et al., 2007). The irony, however, is that this rationality leads to decisions that deprive people of the benefits that would come by trusting people more (Fetchenhauer & Dunning, 2010).

Beyond Rationality in Trust Behavior

Thus, trust offers a conundrum, a mystery—for lack of a better term, a scandal. It does not follow a completely rational economic calculus. So what underlies it? What psychological, sociological, economic, biological, or cultural factors are responsible for its emergence?

Caveats

Before addressing that question, three caveats are in order. First, one must not dismiss rational or economic considerations completely. People do trust more when they

believe that their trust will be rewarded (e.g., Dunning et al., 2014; Fetchenhauer & Dunning, 2009). Expectation, however, is not the only input that governs trust, and often it is not a primary one, although many accounts of trust virtually equate the concept of trust with expectation that it will be reciprocated, such as the Rotter Scale for Interpersonal Trust (Rotter, 1971) and the World Value Survey (Haerpfer et al., 2022).

Second, although trust may not be rational, it is adaptive, in the sense that it leads to beneficial consequences. In laboratory settings, it does so visibly. People are somewhat pessimistic that their trust will be rewarded and have been so since the earliest studies on the game (see Berg et al., 1993, in which participants expected a negative return for the actions they took). However, in reality, people prove to be quite trustworthy. In our original studies, 79% of Study 1 and 90% of Study 2 participants decided to reward trust and give money back—a rate far greater than what their peers expected (Fetchenhauer & Dunning, 2009; see also Dunning et al., 2014). Raising people's expectations that their trust will be honored does cause them to trust more, and to profit from it (Fetchenhauer & Dunning, 2010).

Third, trust is not motivated by some other material or social gain that lies outside of the monetary values involved. One might think that perhaps people are willing to take a potential loss in the trust game because they want to look like a good person in the eyes of the experimenter or anyone else privy to their action, including their partner. That is, they give the money to buy a good reputation, at least for the moment. Making the trust decision completely private, however, does nothing to diminish rates of trust (Dunning et al., 2014, Study 3), although it does interestingly lower rates at which that trust is honored. In addition, making the trust decision merely hypothetical, which means the decision carries no consequence except as a statement of reputation, prompts trust rates to go down, not up (Fetchenhauer & Dunning, 2009; Holm & Nystedt, 2008)—against an explanation centered on reputation.

Further, people do not trust because they care about the outcomes of their partner. More specifically, one could argue that people may decide to trust the other partner because doing so creates US$20 of worth to the world instead of merely US$5, even if they fail to share in any of it, and they wish to create that larger value. We have tested this explanation by comparing behavior in the trust game with behavior in an "extended lottery game" in which a person can gamble their US$5 on a coin flip. If they win, not only do they receive US$10 but another person receives the same amount. This added benefit to another individual does nothing to make people more willing to bet than a simple coin flip involving only benefits to the self (Schlösser et al., 2015). Further, knowing that the other person starts with an equal amount of money on their own (i.e. US$5 in the example above), rather than nothing, fails to reduce rates of trust (Fetchenhauer et al., 2025; Fetchenhauer & Dunning, 2012).

Trust as Noneconomic: An Example and Comparison

So, what is the "else" that governs trust behavior? Consider the following set of two studies, in which we compared behavior in the trust game with a different, and classic, economic game (Dunning et al., 2025). The payoffs of the trust game were modified somewhat, in that both the participant and their partner received US$5 if the participant decided not to trust. If the participant decided to trust, they both received US$10 if the partner honored that trust. If the partner violated that trust, that partner

received US$15 and the original participant nothing. Figure 2 (left panel) depicts the flow of choices available to participant and partner in the trust game.

The classic game used as a comparison was the prisoner's dilemma (Axelrod, 2006). In this game, participants were assigned a partner and asked to execute a one-shot transaction with them. The set of choices contained the same payoffs but were described differently. In essence, instead of the sequential order inherent in the trust game, in which there was a primary mover (the participant) and a secondary responder (the partner), in the prisoner's dilemma both participants moved simultaneously. Each participant each chose between two options, as outlined in the right panel of Figure 2, with the participant choosing between Up and Down and the partner choosing between Left and Right. If the participant chose Down, they received US$5, regardless of what their partner selected. Choosing Up, however, was more interesting. In that case, if the partner chose Left, they each received US$10. If the partner chose Right instead, then the partner received US$15 and the participant nothing.[1]

In terms of economic elements, these two games were equivalent. They shared the same payoff structure, and it emerged that participants held pessimistic expectations their partners would act in a prosocial way roughly at the same rate across the two games. To be sure, across the two studies, , participants expected their partners to be more likely to trust (44% and 49% in Studies 1 and 2, respectively) than they expected them to cooperate in reaction to the prisoner's dilemma (39% and 34% in the two studies, respectively), but those expectations accounted for some but hardly all of the differences in behavior between the two games.

Across the two studies, 121 college students in Study 1 and 97 in Study 2 were presented with both the trust and prisoner dilemma games, with the two decisions separated by ten to fifteen minutes in a single laboratory session, filling out related and unrelated questionnaires in-between. The order of the games was counterbalanced, but order of presentation did not have a consistent effect across the two studies.

As seen in Figure 3, students in both studies did not respond to the two games in the same way. Students chose to trust much more in the trust game (63% in Study 1 and 56% in Study 2) than they chose to cooperate in the prisoner's dilemma (31% and 36% across the two studies, respectively). The rate of trust participants showed across the two studies far outstripped the rate they should have trusted rationally given their expectations about their partner's trustworthiness and their level of risk tolerance (see Figure 3). The level of cooperation seen in the prisoner's dilemma, however, did not.

Even more intriguing, and telling, was a common reaction from participants during debriefing sessions. When we stated that two of the decision scenarios they had confronted were largely equivalent, essentially the same situation, we began noticing that participants often looked puzzled and would go back through their questionnaires. Many of them did not perceive the trust and prisoner dilemma scenarios to hold any

[1] Aficionados will recognize that the prisoner's dilemma game does not technically meet all the criteria of the game. In the formal definition of the game, the selfish choice must be more rewarding for both participant and partner than cooperating no matter what the other person does. For example, if the partner chooses Left, the participant should be tempted by more than US$10 for choosing Down. The intention in this study, however, was to clothe the trust game in the procedure of the prisoner's dilemma to see if it changed people's tendency to trust others, but we note that in equating the playoff structure of the two games, we did not create a classic prisoner's dilemma game.

resemblance whatsoever. To them, the situations were quite distinct. However they interpreted those situations, it was not due to their shared economic features.

Perhaps we should have anticipated that reaction, for those differing reactions were the point of the study. The trust situation may wear the clothing of an economic transaction, but it is not how it is typically conceived. In pilot studies that led up to the laboratory study, we had presented online respondents with trust and prisoner dilemma game instructions and asked them to list everyday situations that these games reminded them of. We then asked a second set of roughly 100 respondents to rate thematically common examples on how much they resembled the two games.

Reading instructions for the prisoner's dilemma, respondents tended to rate it as similar to situations involving gambling or competition, such as *playing chess*, or *poker*, *gambling at a casino*, and *betting on a roulette wheel*. Respondents saw the trust game as less like those situations but more like situations involving informal social exchange or goodwill. These situations were also viewed as serving more social or normative goals, defined as acting in the appropriate or right way, as defined by Lindenberg (2013, 2015), rather than gain goals, which are more aligned with economic aims regarding self-interest. The examples *included helping a colleague at work, donating to a charity, giving someone a birthday present, doing someone a favor*, or *helping a friend*. Trust, it appeared, has elements of risk-taking to it, but it also brought to mind a good dose of social concern for others and acting the way that one ought to.

Trust as Normative Behavior

In short, a good deal of trust, at least as it is embodied in the laboratory trust game, involves norm-driven elements. To be sure, participants are concerned in part about gain and loss, but they are also concerned about what they should be doing irrespective of the economics (Dunning et al., 2014; Schlösser et al., 2015). More important, we have shown that "excessive" trust, or the high level of trust beyond that suggested by high levels of social cynicism plus a heavy tilt toward risk aversion, is explained statistically by the fact that people report that trust is what they ought to do.

Elsewhere, we have also provided data specifying more precisely the norm that is in play. People appear to be reluctant to call the character of their partner into question, to potentially insult that person's integrity by withholding trust (Dunning et al., 2014, 2016). This is a norm seen in other corners of social psychology. Financial advisors who disclose their conflicts of interest, a common practice designed to give clients more leeway to reject advice, paradoxically make it more difficult for those clients to reject it, for that rejection would impugn the sincerity of the advisor, a phenomenon known as *insinuation anxiety* (Sah et al., 2019). People accede to requests for their help because they do not wish to indicate that the request is inappropriate (Flynn & Lake, 2008). In linguistics, this norm has a name. It is called maintaining *positive face* (Brown & Levinson, 1987), and is why people remain polite despite negative internal opinions.

We have shown the operation of this norm by making it irrelevant in the trust game (Dunning et al, 2016, Study 6). For example, in one study, we varied whether the partner had the power to make the decision whether to honor the participants trust or instead had to flip a coin to make the decision, thus removing issues of their character from consideration. Forcing the partner to decide based on a coin flip, even though it slightly improved the odds the odds that participant would receive a reward, caused

trust rates to collapse from 67% to 44%. Participants also reported it was less rude, impolite, and disrespectful to just keep their US$5 when a coin flip was involved, and these perceptions statistically mediated the appearance of the collapse.

The normative dictates inherent in the trust game, and potentially in a broad range of trust decisions as well, are thus different from other situations that seem more economic in nature. For example, take the trust game and compare its normative directives to a purely economic decision, whether to gamble on a flip of a coin. Each decision has a normative choice: One should trust the other person in the trust game, and one should refuse to gamble on the coin unless the odds of winning are quite good, according to participants. Each decision scenario not only has an economic analysis that can be applied, but each also has an "ought," and participants in studies are quite able to report which is the choice they should be making (Schlösser et al., 2015).

However, how participants define "should" differs between the two decision scenarios. When asked what "should" means for the coin flip, participants respond like logicians. They state they should make a decision that is the most *rational, logical, intelligent*, and *objective* to choose. When weighing the trust game, their interpretation of "should" shifts to feature more whether the decision is *socially proper, polite, considerate*, and *respectful*. Defining the "should" in social terms mediated greater risks participants took in trusting other people than they did in gambling on a coin flip (Schlösser et al., 2015).

Implications for Understanding Trust

If trust has significant normative components which lessen its calculative, rational, and economic facets, then there are several ways in which scholars (and laypeople) should change their understanding of trust behavior.

The Role of Emotions

First, a normative perspective allows for trust to be more of an emotional act than a cold, calculative one, as suggested by economics. Indeed, one of the best ways to reveal that trust is driven by normative pressures is to track people's emotions as they consider it. In our studies, we have asked participants how they feel about trusting the other person versus keeping their original money. They report, for example, that giving the money will provide them with the "warm glow" often associated with altruism, such as feeling *happy, pleased,* and *content*, but the association with trust is weak.

The emotions that correlate most with trust behavior is agitation, or rather feeling *tense, anxious, guilty,* and *remorseful*. Those who report that withholding trust would make them feel more agitated than trusting the other person are the ones who give the US$5 to their partner. Across the studies where we have data in Dunning et al. (2014), that difference explains more variance in trust behavior (13%) than does expectation of reward (9%), risk tolerance (0.3%), and warm glow emotions (6%). It also explains high levels of trust. People on average feel more agitated about withholding trust than they do about extending it (see also Schlösser et al., 2016). This pattern stands in stark contrast to other risky decisions, such as gambling on coin flips, lotteries, or the prisoner's dilemma, where people report feeling more agitated about taking the gamble than standing pat. See the left panel of Figure 4, which displays the amount of agitation people reported as they considered keeping versus risking money in the trust game

versus coin flip gambles (with only private benefits and ones involving possible benefits to others (Schlösser et al., 2016). The left panel of the figure shows how agitation shifts as the game moves from trust to prisoner's dilemma (Dunning et al., 2025).

Agitation emotions are the signature of normative pressures, and that people are acting to be the person they ought to be (Higgins, 1987). However, that should not be taken to mean that such emotions have a causal influence on trust behavior. They signal that a norm is in play, and it is the norm that is the causal agent producing both the agitation and the behavior. That said, the feelings associated with norms may be an accelerant. Kugler et al. (2012) placed people in situations where they experienced anger or fear and then had them consider a gamble or a game in which they had to choose whether to coordinate with another person (i.e., a stag hunt game). Fear, an emotion close to the agitation emotions associated with norms, made participants less likely to gamble on their own but more likely to risk by coordinating with another person.

Action, Not Outcome, Matters

The salience of emotion also suggests another fundamental way that trust behavior differs from economic choice. In economics, decisions are consequentialist. They rest first and foremost on the considerations of the outcomes of one's decisions. In the case of our lab's trust game, those outcomes would be keeping the US$5 versus increasing it to US$10 or ending up empty-handed.

Our emotion measures suggest that people make their decisions based not on outcomes but on the actions themselves—that is, whether to say "yes" or "no" to the option to trust. Their decisions, thus, are non-consequentialist (Dunning & Fetchenhauer, 2013). They may be considering whether there is a rule to follow, or they may be considering the utility they derive from taking the action itself (e.g., avoiding the guilt of not trusting the other person) rather than from any downstream consequences.

We have found that emotions associated with actions more strongly associate with what people choose to do than any emotions connected to downstream consequences (Schlösser et al., 2013, 2016). Not surprisingly, when asked, people report that they will feel betrayed if they trust their partner and that person leaves with all the money. However, the depth of that anticipated feeling does not correlate their decision to trust (Fetchenhauer et al., 2020; Schlösser et al., 2016). The degree of agitation they feel about withholding trust--the action, not the outcome—does (Schlösser et al., 2016).

This finding has good company. In the moral psychology literature, a growing body of evidence suggests that emotions attached to actions are more closely tied to moral behavior than any tied to outcomes. For example, college students are not all that disturbed to watch another person point a toy gun at a laboratory research assistant's face and pull the trigger. However, ask them to be the one pointing the gun the gun and pulling the trigger and they become quite upset and refuse to do it. The two situations share the same outcome; it is the person doing the action who is different. People appear not averse to the consequence, but they do appear quite averse to the performing the action (Cushman et a., 2012).

Awareness of Cause

The normative nature of trust decisions may also provide an explanation for

another phenomenon that has vexed trust researchers for a long time (e.g., Zak, 2008). People do not seem to have much awareness about why they choose to trust the other person. In debriefings in our own labs (Dunning et al., 2012), we ask participants what led them to send the US$5 to their partner, particularly when they had little expectation of receiving any money back, and the most common answer is “It’s only $5.” When we point out, on occasion, that they turned down a chance to gamble the same amount on a lottery, and so what explains the difference if it is only $5, they sit in unclear silence.

Normative dictates are often so well-learned that people know enough to follow them but do not know or have forgotten the reasons why they do so, much like expert typists cannot report verbally where the letters of the alphabet are on the keyboard without silently moving their fingers. Or, the fact that it is perfectly acceptable to talk about one’s “lovely little old rectangular green French silver whittling knife,” but not one’s “old green French silver whittling rectangular little lovely knife” (Forsyth, 2013). Underneath consciousness, we know that adjectives in English must follow the ordering of *opinion-size-age-shape-color-origin-material-purpose*, and use that order when speaking, but cannot consciously describe that linguistic norm if it is violated.

A classic scenario from behavioral economics shows that for social grammar people also know the right conclusion but not the logic, either cognitive or social, that gets them there. Consider this scenario (Kahneman et al., 1986):

> A small photocopying shop has one employee who has worked in the shop for six months and earns $9 per hour. Business continues to be satisfactory, but a factory in the area has closed and unemployment has increased. Other small shops have now hired reliable workers at $7 an hour to perform jobs similar to those done by the photocopy shop employee. The current employee leaves, and the owner decided to pay a replacement $7 an hour.

In the original study, 73% of respondents thought it appropriate to start the new employee at the lower wage, but only 17% thought it would be appropriate lower the current employee’s wage to $7 if they stayed on the job, even though the two scenarios are economically equivalent. Beyond muttering about “fairness,” or “that’s not right,” people struggle to articulate the underlying rationale that makes lowering the wage acceptable in the first scenario but not in the second, a form of *moral dumbfounding* (Haidt, 2001).

Trust Is Not a Preference

The normative nature of trust also suggests that it is not a preference, in the sense that it is something that people aspire to do, but rather an obligation, something they feel a duty or obligation to perform. The fact that predicting trust hinges on agitation emotions rather than on warm glow emotions provides the clearest evidence for this assertion. Happiness and pleasure occur after aspirations are achieved; the avoidance of tension and guilt occur after obligations have been discharged (Higgins, 1987).

Towards that end, it might be fruitful to consider other behavioral economic games. In those games, when expectations fail to predict behavior, theorists argue that people have formed preferences for the options they have chosen. Those choices, however, may not be preferences in the way that people typically define it.

For example, though plagued by inconsistent results, people change their choices in prisoner’s dilemma games depending on the game’s description. For

example, when the game is described as “The Community Game,” they are more likely to cooperate than when it is labeled “The Wall Street Game” (Kay & Ross, 2003; Liberman et al., 2004). Theorists have argued whether these “framing” effects are due to changes in belief or preference (Ellingsen et al., 2012). By belief, theorists point to expectations about the choices each game player will favor, akin to a demand characteristic, or perhaps a focal point in economic terms. By preference, theorists are more unclear, meaning that players may want to make a choice to establish a public reputation, fit better their personal identities, or adhere to a social norm.

Our work suggests a more specific explanation for framing effects, although empirical data is needed to validate it—and one that we would not label a preference. We would not assert that our participants prefer trust. Instead, they feel a duty or obligation to trust the other person. They may want to do it, but it is the injunctive norm that drives that appears to drive the excess trust beyond any that would be motivated by their expectations and appetite for risk. Indeed, in Dunning et al. (2014), a full 41% of participants who chose trust reported not wanting to or being ambivalent about it; only 18% said the same about whether they “should” do it. Thus, we would replace the label of “preference” for “obligation,” (Dunning et al., 2020).

Trust is Not Altruism

This perspective of trust as obligation leads to another observation, that trust is not altruism, at least as traditionally defined. It does not necessarily correlate with it. Degree of trust in the trust game does not correlate with generosity in the dictator game, where people can just share all or a portion of a monetary endowment bestowed on them to another person (Brülhart & Usunier, 2012; Yamagishi et al., 2013).

Further, trust behavior among children develops earlier than selfless action. Schoolchildren in Austria were given a chance to give a small bag of toys to another child, who would then receive four bags of toys with an option to return two of those bags to the first child. Older children, roughly ten years old trusted at a high rate, roughly 70%, relative to kindergarteners, of whom only 27% trusted. Altruism, revealed in a separate decision to forgo a bag of toys so that another child could have four bags for themselves, remained low for both groups (27% vs. 17% of the older and younger children, respectively, selflessly chose to give up their bag) (Evans et al., 2013).

However, if one reconceptualizes altruism as normatively driven, more giving out of an obligation than from a freely chosen preference, perhaps trust is a form of altruism. Evidence suggests that much altruism is not so much giving but giving in to social pressure (Cain et al., 2014; Flynn & Lake, 2008). In one famous example, having people costumed as Santa Claus outside a Boston department store expressly asking customers for donations increased contributions to the Salvation Army by 70% from those walking out the door. However, it also increased the number of customers walking out a side door where there was no Santa by 30%, suggesting not everyone held a preference to give if they had foresight to avoid it (Andreoni et al., 2017).

Trust Is Not a Unitary Phenomenon

Our studies also suggest that trust is much like the concept of attitudes in having three different components: cognitive, affective, and behavioral (Dunning et al., 2019). However, unlike attitudes, where those three components are presumed to cohere to

some degree, our research suggests that these components of trust fracture and dissociate in important ways. At the cognitive level, people are distrustful. They hold cynical views about whether others honor trust (Fetchenhauer & Dunning, 2009, 2010). Yet, at the behavioral level, they trust at high levels (Fetchenhauer & Dunning, 2009, 2012; Dunning et al., 2014). Affectively, it appears that trust is associated with an avoidance of guilt and anxiety centered on not trusting rather than a positive feeling about trust itself (Dunning et al., 2014). To understand trust, one must study all three components separately and not assume that data about one component reveals the nature of the other components.

Conclusion

However, all things considered, there still may be a way to consider trust decisions to be largely rational. The theoretical move would be to abandon the way that sociologists and economists have traditionally defined rationality in social exchange. Those theorists that defined rationality at the level of the individual, and if any norm promoting trust developed, it developed in particular social relationships or over a repeated history of positive experiences of trustworthiness (Blau, 1964/2017; Coleman, 1994; Homans, 1958). Trust was built from the ground up, one interaction at a time. Ultimately, everyone acted in a way to promote their own self-interest, and in doing so built a community in which all acted in a trustful and trustworthy way.

The move would abandon this perspective that focused on the individual to center on the community instead. It would place the community as the unit of analysis and make it the actor of interest that either promotes or undermines trust. In sociology, such perspectives are not uncommon, such as the structural functionalist approach, associated with such thinkers as Auguste Comte, Émile Durkheim, Talcott Parsons, and Robert Merton, which examined how societies formed and then maintained shared values and norms that promoted orderly and beneficial social relations as a system or unit (Kingsbury & Scanzoni, 1993). Such communities might produce norms for trust that are already in place for young children to learn and follow from a young age which do not have to be built up one relationship at a time, without understanding exactly why the norm is there but following it nonetheless and enjoying its advantages.

This move brings us back to the issue of rationality. Moving the responsibility for trust from the individual to the community is not necessarily such a novel approach. Recently, in psychology, there is a growing recognition that attributes often thought to dwell in individuals might instead be more properly placed in the community they inhabit (Hofstede, 2001). Implicit attitudes seem not to reside in single people but rather in the college campuses and metropolitan areas in which they live (Payne et al., 2017). Practices of intellectual humility may not characterize individuals as much as they do organizations and professions (Dunning, 2023).

Similarly, the trait of rationality that promotes positive behavior in social interaction, and in matters such as trust, cooperation, and altruism, may dwell not in individuals but in the communities or cultures in which they are embedded. Under that perspective, trust and rationality may not oppose one another. And trust may pose no scandal at all.

References

Andreoni, J., Rao, J. M., & Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of political Economy*, *125*(3), 625-653. https://doi.org/10.1086/691703

Aristotle. (1876). *The Nicomachean ethics.* London, UK: Longmans, Green.

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, *26*(102), 121-127. https://doi.org/10.2307/2550390

Axelrod, Robert (2006). *The evolution of cooperation.* (revised ed.). New York: Basic Books.

Becker, G. S. (1962). Irrational behavior and economic theory. *Journal of political economy*, *70*(1), 1-13. https://doi.org/10.1086/258584

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of political economy*, *76*(2), 169-217. https://doi.org/10.1086/259394

Becker, G. S., Landes, E. M., & Michael, R. T. (1977). An economic analysis of marital instability. *Journal of political Economy*, *85*(6), 1141-1187. https://doi.org/10.1086/260631

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *Journal of political Economy*, *96*(4), 675-700. https://doi.org/10.1086/261558

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Blau, P. (1964/2017). *Exchange and power in social life*. New York: Routledge.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* Cambridge, UK: Cambridge University Press.

Brülhart, M., & Usunier, J. C. (2012). Does the trust game measure trust? *Economics Letters, 115, 20–23. doi:10.1016/j.econlet.2011.11.039*

Cain, D. M., Dana, J., & Newman, G. E. (2014). Giving versus giving in. *Academy of Management Annals*, *8*(1), 505-533. https://doi.org/10.5465/19416520.2014.911576

Coleman, J. S. (1994). *Foundations of social theory*. Harvard University Press.

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion, 12*(1), 2-7. https://doi.org/10.1037/a0025071

Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87*(4), 611–628. https://doi.org/10.1037/0021-9010.87.4.611

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, *65*(2), 135-150. https://doi.org/10.1086/257897

Dunning, D. (2023). Where does intellectual humility reside? *The Journal of Positive Psychology, 17*(2), 264-266. https://doi.org/10.1080/17439760.2022.2155220

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107,* 122-141. https://doi.org/10.1037/a0036673

Dunning, D., & Fetchenhauer, D. (2013). Behavioral influences in the present tense: On expressive versus instrumental action. *Perspectives on Psychological Science, 8*, 142-145. https://doi.org/10.1177/1745691612474319

Dunning, D., Fetchenhauer, D., & Schlösser, T. (2012). Trust as a social and emotional

act: Noneconomic considerations in trust behavior. *Journal of Economic Psychology*, *33*(3), 686–694. https://doi.org/10.1016/j.joep.2011.09.005

Dunning, D., Fetchenhauer, D., & Schlösser, T. (2016). The psychology of respect: A case study of how behavioral norms regulate human action. In A. Elliot (Ed.), *Advances in motivation science* (vol. 3; pp. 1-34). New York: Elsevier. https://doi.org/10.1016/bs.adms.2015.12.003

Dunning, D., Fetchenhauer, D., & Schlösser, T. (2019). Why people trust: Solved puzzles and open mysteries. *Current Directions in Psychological Science, 28,* 366-371. https://doi.org/10.1177/0963721419838255

Dunning, D., Fetchenhauer, D., & Schlösser, T. (2020). Obligation at zero acquaintance. *Behavioral and Brain Sciences. 43,* E69. Doi: 10.1017/S0140525X19902498.

Dunning, D., Fetchenhauer, D., & Schlösser, T. (2025). *Divergent construals and behavior in the trust game and the prisoner's dilemma: On the cogntive and emotive bases of social choice.* Unpublished manuscript, University of Michigan and the University at Cologne.

Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, *76*(1), 117-130. https://doi.org/10.1016/j.geb.2012.05.007

Evans, A. M., Athenstaedt, U., & Krueger, J. I. (2013). The development of trust and altruism during childhood. *Journal of Economic Psychology*, *36*, 82-95. https://doi.org/10.1016/j.joep.2013.02.010

Fetchenhauer, D., & Dunning, D. (2009). Do people trust too much or too little? *Journal of Economic Psychology, 30*, 263-276. https://doi.org/10.1016/j.joep.2008.04.006

Fetchenhauer, D., & Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science, 21*(2), 189–193. https://doi.org/10.1177/0956797609358586

Fetchenhauer, D., & Dunning, D. (2012). Betrayal aversion versus principled trustfulness: How to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior and Organization, 81*, 534-541. https://doi.org/10.1016/j.jebo.2011.07.017

Fetchenhauer, D., Dunning, D., Ehlebracht, D., Gracyk, T., & Schlösser, T. (2025). *Trust in competence rather than character: A matter of respect or investment?* Manuscript under review, University of Cologne.

Fetchenhauer, D., Ehlebracht, D., Lang, A.-S., Schlösser, T., & Dunning, D. (2020). Does betrayal aversion really guide trust decisions. *Journal of Behavioral Decision Making, 33*, 556-566. https://10/1002/bdm.2166

Fetchenhauer, D., & van der Vegt, G. (2001). Honesty, trust and economic growth. *Zeitschrift für Sozialpsychologie, 32*(3), 189-200. https://doi.org/10.1024/0044-3514.32.3.189

Flynn F. J., Lake V. K. B. (2008). If you need help, just ask: Underestimating compliance with direct requests for help. *Journal of Personality and Social Psychology*, 95, 128-143. https://doi.org/10.1037/0022-3514.95.1.128

Forsyth, M. (2013). *The elements of eloquence.* New York: Berkley Books.

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano J., M. Lagos, P. Norris, E. Ponarin & B. Puranen (eds.). 2022. World Values Survey:

Round Seven – Country-Pooled Datafile Version 6.0. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. https://doi.org/10.14281/18241.24

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haraway, D. J. (1991). *Simians, cyborgs and women: The reinvention of nature*, New York: Routledge.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*, 319–340.

Hobbes T. (1651) *Leviathan*. Cambridge, UK: Cambridge University Press

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: SAGE Publications.

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.

Homans, G. C. (1958). Social behavior as exchange. *American Journal of Sociology*, *63*(6), 597-606. https://doi.org/10.1086/222355

Hull, D. L. (2001). *Science and selection: Essays on biological evolution and the philosophy of science*. Cambridge University Press.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *35*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Johnson, N. D., & Mislin, A. (2012). How much should we trust the World Values Survey trust question? *Economic Letters, 116*, 210-212. https://doi.org/10.1016/j.econlet.2012.02.010

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, *76*(4), 728-741. https://doi.org/10.1515/9781400829118-011

Kant, I. (1909). *Critique of practical reason.* London, UK: Huge Print Press.

Kay, A. C., & Ross, L. (2003). The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner's Dilemma. *Journal of Experimental Social Psychology*, *39*(6), 634-643. https://doi.org/10.1016/S0022-1031(03)00057-X

Kingsbury, N., & Scanzoni, J. (1993). Structural-functionalism. In P.. Boss, W. Doherty, R. LaRossa, W. Schumm, & S. Steinmetz (Eds.) *Sourcebook of family theories and methods: A contextual approach* (pp. 195-221). Boston, MA: Springer US.

Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics*, *112*(4), 1251-1288. https://doi.org/10.1162/003355300555475

Kugler, T., Bornstein, G., Kocher, M. G., & Sutter, M. (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic psychology*, *28*(6), 646-657. https://doi.org/10.1016/j.joep.2006.12.003

Kugler, T., Connolly, T., & Ordóñez, L. D. (2012). Emotion, decision, and risk: Betting on gambles versus betting on people. *Journal of Behavioral Decision Making, 25*(2),

123–134. https://doi.org/10.1002/bdm.724
Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin, 30*(9), 1175–1185. https://doi.org/10.1177/0146167204264004
Lindenberg, S. (2013). Social rationality, self-regulation and well-being: The regulatory significance of needs, goals, and the self. In R. Wittek, T.A.B. Snijders, & V. Nee (Eds.) *Handbook of rational choice social research* (pp. 72-112). Stanford: Stanford University Press.
Lindenberg, S. (2015). The third speed: Flexible activation and its link to self-regulation. *Review of Behavioral Economics, 2,* 147–160. https://doi.org/10.1561/105.00000024
Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233-248. https://doi.org/10.1080/1047840X.2017.1335568
Rohrschneider, R., & Schmitt-Beck, R. (2002). Trust in democratic institutions in Germany: Theory and evidence ten years after unification. *German Politics*, *11*(3), 35-58. https://doi.org/10.1080/714001314
Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(5), 443–452. https://doi.org/10.1037/h0031464
Rousseau, D. M., Sitkin, S. I. M. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617
Sah, S., Loewenstein, G., & Cain, D. (2019). Insinuation anxiety: Concern that advice rejection will signal distrust after conflict of interest disclosures. *Personality and Social Psychology Bulletin, 45*(7), 1099-1112. https://doi.org/10.1177/0146167218805991
Schlösser, T., Dunning, D., & Fetchenhauer, D. (2013). What a feeling: the role of immediate and anticipated emotions in risky decisions. *Journal of Behavioral Decision Making*, *26*(1), 13-30. https://doi.org/10.1002/bdm.757
Schlösser, T., Fetchenhauer, D., & Dunning, D. (2016). Against all odds? The emotional dynamics underlying trust. *Decision, 3*, 216-230. https://doi.org/10.1037/dec0000048
Schlösser, T., Mensching, O., Dunning, D., & Fetchenhauer, D. (2015). Trust and rationality: Shifting normative analyses in risks involving other people versus nature. *Social Cognition, 33,* 459-482. https://doi.org/10.1521/soco.2015.33.5.459
Smith, A. (1853). *The theory of moral sentiments*. London, UK: HG Bohn.
Simpson, J. A. (2007). Foundations of interpersonal trust. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 587–607). The Guilford Press.
Stavrova, O., & Ehlebracht, D. (2016). Cynical beliefs about human nature and income: Longitudinal and cross-cultural analyses. *Journal of Personality and Social Psychology, 110*(1), 116–132. https://doi.org/10.1037/pspp0000050
Stavrova, O., Ehlebracht, D., & Ren, D. (2024). Cynical people desire power but rarely acquire it: Exploring the role of cynicism in leadership attainment. *British Journal*

of Psychology, 115(2), 226–252. https://doi.org/10.1111/bjop.12685
Stavrova, O., Ehlebracht, D., & Vohs, K. D. (2020). Victims, perpetrators, or both? The vicious cycle of disrespect and cynical beliefs about human nature. *Journal of Experimental Psychology: General, 149*(9), 1736–1754. https://doi.org/10.1037/xge0000738
Von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
Warren, M. E. (Ed.). (1999). *Democracy and trust*. Cambridge University Press.
Williams, G. C. (1966), *Adaptation and natural selection*, Princeton: Princeton University Press.
Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., Miura, A., Inukai, K., Tanida, S., Kiyonari, T., Takagishi, H., & Simunovic, D. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes, 120*(2), 260-271. https://doi.org/10.1016/j.obhdp.2012.06.002
Zak, P. J. (2008). The neurobiology of trust. *Scientific American*, *298*(6), 88-95. https://www.jstor.org/stable/26000645
Zak, P. J., & Knack, S. (2001). Trust and growth. *The Economic Journal*, *111*(470), 295–321. https://doi.org/10.1111/1468-0297.00609
Zhang, R. J. (2020). Social trust and satisfaction with life: A cross-lagged panel analysis based on representative samples from 18 societies. *Social Science & Medicine*, *251*, 112901. https://doi.org/10.1016/j.socscimed.2020.112901

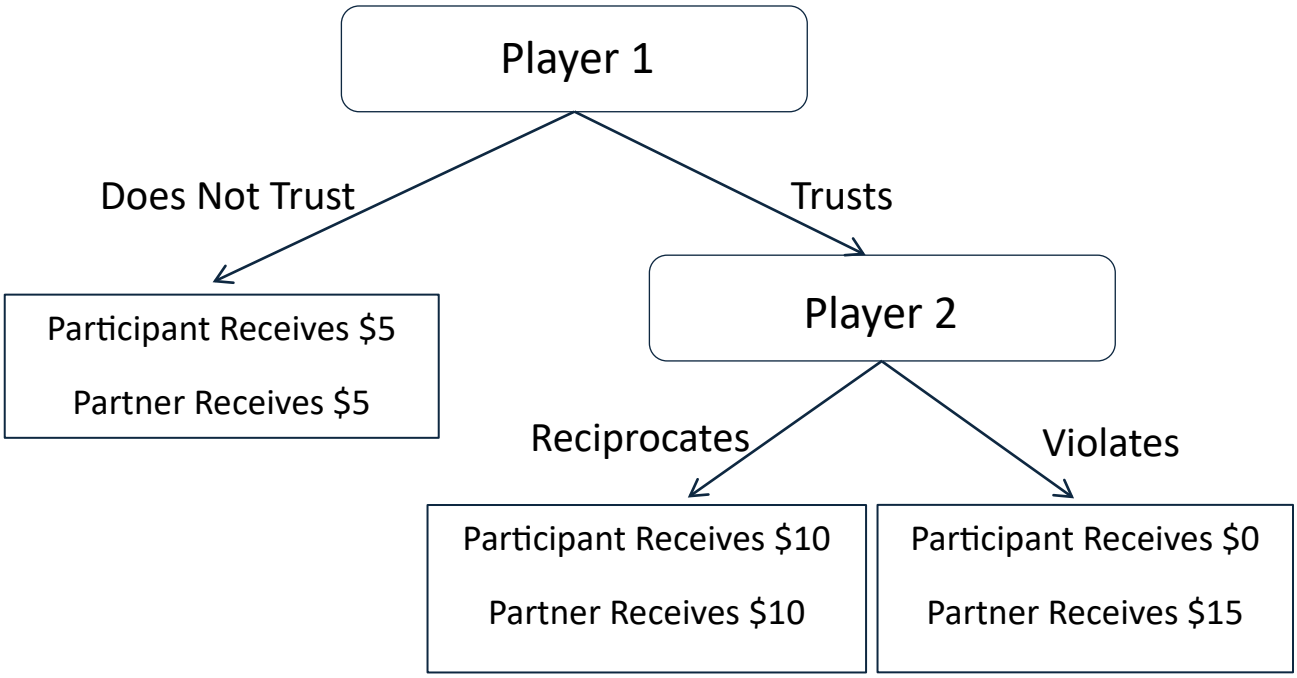
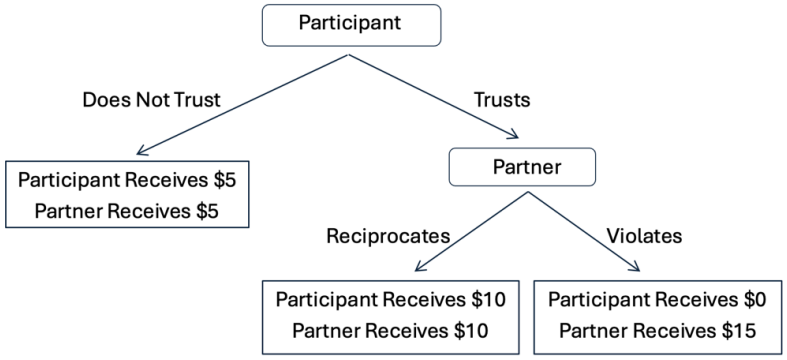


Figure 1. Schematic representation of a trust game.

A. Trust Game



B. Prisoner's Dilemma Game

	Partner Choice	
Participant Choice	Left	Right
Up	Participant: $10 Partner: $10	Participant: $ 0 Partner: $15
Down	Participant: $ 5 Partner: $ 5	Participant: $ 5 Partner: $ 5

Figure 2. Schematic representation of trust game and prisoner's dilemma presented in Dunning et al. (2025). *Divergent construals and behavior in the trust game and the prisoner's dilemma: On the cogntive and emotive bases of social choice.* Unpublished manuscript, University of Michigan and the University at Cologne.

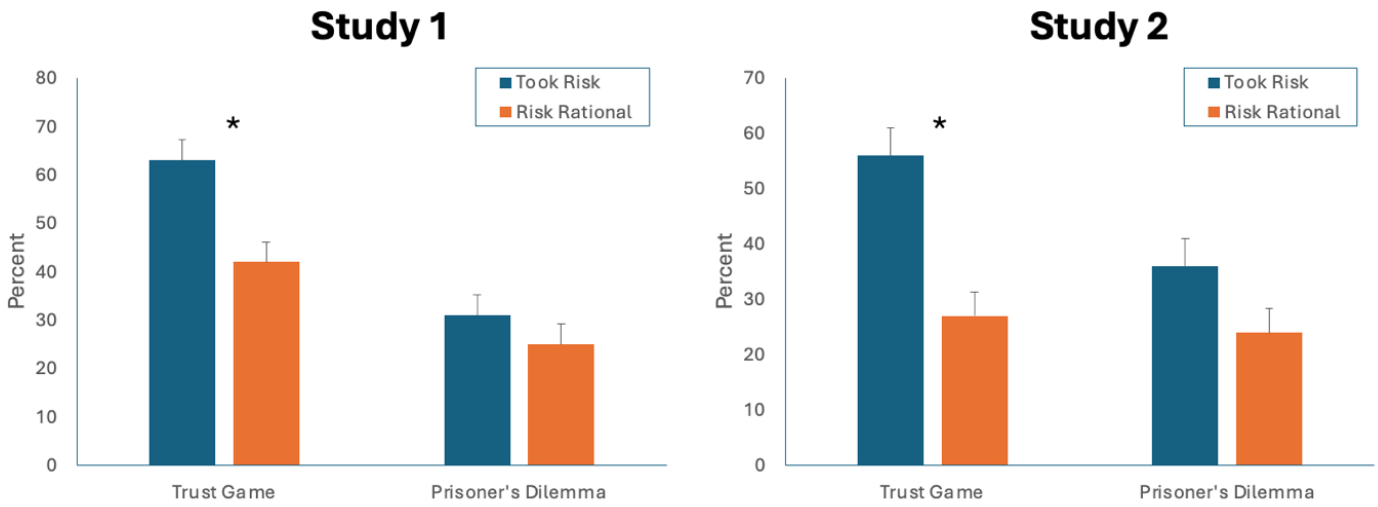


Figure 3. Percentage of participants risking money in trust and prisoner dilemma games (blue) versus percentage for whom it is rational to do so (orange). Asterisks indicate that percentage risking is significantly greater than anticipated by rationality rate ($P < .005$). Error bars refer to standard errors. Adapted from Dunning et al. (2025). *Divergent construals and behavior in the trust game and the prisoner's dilemma: On the cogntive and emotive bases of social choice.* Unpublished manuscript, University of Michigan and the University at Cologne.

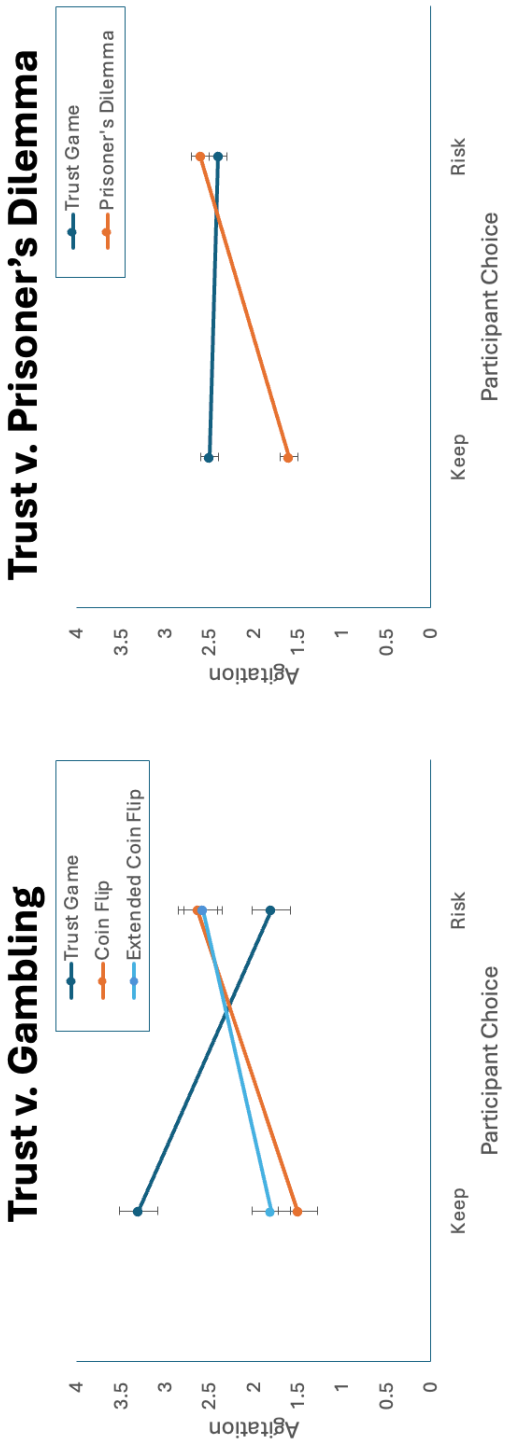


Figure 4. Degree of agitation reported when contemplating keeping versus risking money. Left Panel: Trust game versus coin flip and extended coin flip gambles. (Schlosser et al., 2016). Right Panel: Trust game versus Prisoner's dilemma (Dunning et al., 2025). Error bars refer to standard errors.