

Bounded Propection in Dilemmas of Trust and Reciprocity

Anthony M. Evans
Tilburg University

Joachim I. Krueger
Brown University

Knowing when to trust others is an important social skill, but recent findings suggest that humans struggle with this dilemma—trusting strangers more than they should. Although trust decisions often do not meet the standards of rationality, they appear to be boundedly rational. We present a model of heuristic trust, according to which people focus on their own potential outcomes (what may be gained or lost from trusting), but neglect the probabilities of those outcomes occurring. We examine how trustors form expectations of reciprocity, and how those expectations relate to optimal trust decisions: some previous research suggests that people underestimate the probability of reciprocity and, relative to their subjective expectations, trust strangers too much. In contrast, our heuristic model allows for fine-grained predictions of when people trust too much and when they trust too little. The accuracy of trust depends on the selection and use of available cues; errors occur when trustors neglect valid, but difficult to process, cues and overemphasize salient cues lacking validity.

Keywords: trust, social decision making, behavioral economics, bounded rationality, social cognition

Do people know when to trust others? And if they do not, are there strategies they can use to improve their decision-making? Dilemmas of interpersonal trust pervade social and professional life. They affect the shape and quality of personal relationships (Simpson, 2007) and economic exchanges (Dasgupta, 2007). The ability to judge whom and when to trust lies at the core of psychological and material well-being (Arrow, 1974; DeSteno, 2014; Evans & Krueger, 2015; Rotter, 1971; Zak & Knack, 2001). Given the centrality of trust to human flourishing, it is alarming that a growing body of research suggests that most people do not meet this challenge well. People are biased and ill-informed in how they form expectations of reciprocity (Kausel & Connolly, 2014; Olivola, Funk, & Todorov, 2014), and many trust decisions fail to correspond with the decision-maker's subjective preferences and expectations (Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014). These findings suggest that trust is in need of repair, lest people remain in economically self-defeating patterns.

While acknowledging that many trust decisions fall short of the demanding standards of full rationality, we show that such decisions are *boundedly rational* (Simon, 1955). Trustors use a heuristic strategy, focusing on cues that are easily processed and sufficiently valid (Goldstein & Gigerenzer, 1999). Heuristic decision-making can produce satisfactory results while conserving psychological and material resources (Evans, Dillon, Goldin, & Krueger, 2011; Mascampo & Baumeister, 2008). In our research, we have found that trustors focus

on their own potential gains and losses while neglecting (though not completely ignoring) expectations of the trustee's behavior. In particular, trustors fail to appreciate the trustee's financial temptation to act selfishly (Evans, Athenstaedt, & Krueger, 2013; Evans & Krueger, 2011, 2014). This egocentric focus on one's own potential outcomes may be subsumed under a generalized tendency to focus on salient, easy-to-process cues. Our model of heuristic decision-making has implications for accuracy. In dilemmas of trust, decision quality depends on cue selection: errors occur when trustors neglect valid, but difficult to process, cues and overemphasize salient cues lacking validity.

Defining and Measuring Trust

Trust is important in psychology (Simpson, 2007) and economics (Dasgupta, 2007). Recently, Thielmann and Hilbig (2015a) defined trust as “a risky choice of making oneself dependent on the actions of another in a situation of uncertainty, based upon some expectation of whether the other will act in a benevolent fashion despite an opportunity to betray” (p. 251). This definition emphasizes three core components of trusting behavior: First, trust is a risky *choice* and it involves accepting an uncertain or ambiguous outcome. Second, trust requires *vulnerability*. When trust is betrayed, the victim suffers a financial or nonmaterial loss. Third, trust is based on *expectations* of reciprocity, that is, prospective beliefs that the trusted party will choose not to betray the initial act of trust. The social nature of these expectations differentiates trust decisions from individual decisions under risk. Expectations regarding the behavior of other people require an element of mind-reading; the trustor must assess the trusted party's preferences and intentions (Mayer, Davis, & Schoorman, 1995).

To measure trust and reciprocity, many researchers use variants of the trust game (Figure 1), an economic interaction involving two agents, the trustor and the trustee (Camerer, 2003; Evans &

This article was published Online First January 18, 2016.

Anthony M. Evans, Department of Social Psychology, Tilburg University; Joachim I. Krueger, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University.

Correspondence concerning this article should be addressed to Anthony M. Evans, Department of Social Psychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: a.m.evans@uvt.nl

Krueger, 2009).¹ The game begins with the trustor choosing between the status quo and trust. If trust occurs, then the trustee chooses between reciprocity and betrayal. If the trustee chooses reciprocation, both players receive outcomes better than the status quo. If the trustee chooses betrayal, the trustee does even better, whereas the trustor ends up with an outcome worse than the status quo. To capture the idea that trust-based exchanges tend to create economic wealth (Arrow, 1974), the initial act of trust increases the total amount of wealth available to both players.²

Research has shown that there is an imperfect correspondence between feelings of trust—the belief that a person is generally trustworthy—and behavior in the trust game (Glaeser, Laibson, Scheinkman, & Soutter, 2000; Johnson & Mislin, 2012). Behavior is also shaped by the degree of situational vulnerability, the potential consequences of choosing trust in a specific situation (Snijders & Keren, 1999). Not all dilemmas are created equally. In some situations, showing trust involves minimal vulnerability; the potential cost of betrayal is small and the potential benefit of reciprocity is large. If there is little to lose, people may be willing to trust even if there is a minimal expectation of reciprocity. On the contrary, sometimes trusting behavior involves accepting extreme vulnerability. Trust can mean risking one's life or livelihood (Rand & Epstein, 2014). When the stakes are high, people may be unlikely to trust even if they hold near-certain expectations of reciprocity. The distinction between low- and high-vulnerability situations means that trustors must meet two related challenges. First, he must judge if the other party is likely to reciprocate; second, he must reach a decision based on his expectations and the potential consequences. Fully rational decision-making requires the trustor to succeed in both tasks.

Rational and Expressive Models of Trust

In long-term relationships, trust decisions depend on the outcomes of previous interactions (Anderhub, Engelmann, & Güth, 2002; Charness, Du, & Yang, 2011) and successful decision-making is a matter of learning to identify trustworthy individuals (Basu, Dickhaut, Hecht, Towry, & Waymire, 2009; King-Casas et al., 2005). Yet, many important decisions involve strangers and unfamiliar situations, and these decisions must be made on the basis of little or no information. Although reasoning under conditions of uncertainty is difficult (Hastie & Dawes, 2009; Taleb, 2010), solutions that attempt to minimize uncertainty often result

in less efficient interactions. For example, uncertainty can be reduced through the use of formal contracts (Sitkin & Roth, 1993), but contracts are costly to enforce and undermine the parties' intrinsic motives to cooperate after the contracts have concluded (Malhotra & Murnighan, 2002). Uncertainty is necessary to reap the full benefits of trust. Psychologists and economists have proposed normative standards for how people should reason under the conditions of zero-acquaintance interactions, and they have tested whether actual behavior lives up to these standards.

Rational Models

Beginning with Berg, Dickhaut, and McCabe's (1995) pioneering research, scores of studies have shown that rational game-theoretic models do a poor job predicting human behavior. A strict model of self-interested rationality predicts that trust among strangers cannot occur. A self-interested trustee will not reciprocate trust, a simple deduction available to the equally self-interested trustor. Hence, the inefficient status quo will prevail. Trust should only occur in situations in which the trustee has financial incentives to reciprocate—for example, when there is the possibility of a repeated interaction (B6, 2005) or when there is a sanctioning system in place to punish unfair behavior (Fehr & Gächter, 2002). Refuting the dire predictions of conventional game theory, trust and reciprocity are consistently observed in laboratory studies, even when anonymous strangers make one-shot, high-stakes decisions (Johnson & Mislin, 2011).

To better explain trust and reciprocity among strangers, the game-theoretic approach can be expanded by including social preferences or values, such as the desire to maximize joint outcomes or to make outcomes equal for both agents (Fehr & Schmidt, 1999; Van Lange, 1999). These models are outcome-based or consequentialist in nature: assuming that people are utility maximizers, but that the expected utility of a decision depends, in part, on the outcomes others receive. Social preference models raise the possibility that trust may be understood as a form of rational prosociality. Because trust creates wealth (Arrow, 1974; Zak & Knack, 2001), acts of trust are potentially beneficial for both parties. There is empirical evidence supporting such models in that trusting behavior is correlated with prosocial preferences (Kanagaretnam, Mestelman, Nainar, & Shehata, 2009) and with cooperative behavior in other social dilemmas, such as the prisoner's dilemma (Peysakhovich, Nowak, & Rand, 2014; Yamagishi et al., 2013, 2015).

However, trust cannot be fully explained as an expression of altruism. After all, trustors stand to gain from their investments if the trustees reciprocate. The potential to gain from reciprocity is, in fact, important to the trustor, since the willingness to trust decreases sharply when the trustee is unable to reciprocate (Cox, 2004; Smith, 2003) and when the outcome of trust (reciprocation vs. betrayal) is determined by chance (Bohnet & Zeckhauser, 2004). Further limiting the potential role of social preferences, process tracing data reveal that trustors often decide without even

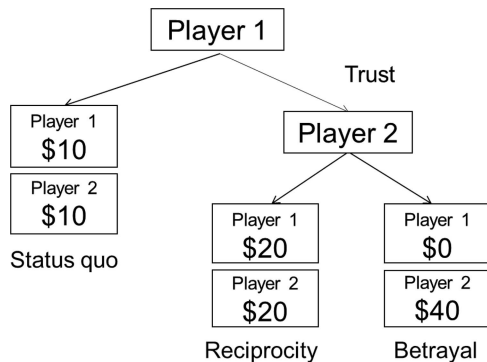


Figure 1. An example of the economic trust game.

¹ Many studies use a variant called "the investment game," which follows the same sequential structure and represents the same basic dilemma. In the investment game, the trustor and trustee make incremental, as opposed to all-or-nothing, decisions.

² In this article, we refer to the trustor as "he" and the trustee as "she."

searching for information about the trustee's potential payoffs (Evans & Krueger, 2014). In short, trust cannot be explained purely in terms of the desire for mutually beneficial outcomes or altruistic concern for the trustee's well-being.

Social preferences may predict trusting behavior, in part, because prosocial trustors are more likely to believe that trustees are also prosocial (Krueger, Massey, & DiDonato, 2008; Thielmann & Hilbig, 2014), but they do not illuminate the process of how trust decisions are made. Social preference models do a better job accounting for the trustee's behavior, since the dilemma of reciprocity deals directly with resource distribution, rather than strategic uncertainty (Thielmann & Hilbig, 2015b). If social preferences cannot explain trust, then we must move beyond game-theoretic models and the benchmark of fully rational decision-making (Basu, 1994; Krueger, Evans, & Heck, in press).

Expressive Trust

Rational models assume that decisions arise from prospective reasoning. People decide by evaluating all possible outcomes and their respective probabilities of occurring (Seligman, Railton, Baumeister, & Sripada, 2013). When outcomes and probabilities are fixed, a generalized model of rationality does not distinguish between dilemmas of interpersonal trust and nonsocial problems of risky choice. In both cases it is assumed that decisions are fully determined by the calculation of expected value or expected utility. There is now clear evidence that social trust and nonsocial risk-taking are shaped by different mental processes. First, there is only a weak correlation between trust and risk-seeking behavior (Ashraf, Bohnet, & Piankov, 2006; Eckel & Wilson, 2004; Houser, Schunk, & Winter, 2010). People who are more willing to accept vulnerability in the context of individual risk-taking are only slightly more likely to trust others. Second, expectations of reciprocity are only weakly correlated with trusting behavior (Evans & Krueger, 2014; Malhotra, 2004; Snijders & Keren, 1999). Increasing the likelihood of reciprocity only has a small effect on the rate of trust. Taken together, these findings suggest that trust decisions have little to do with risk, vulnerability, and expectations, the specific features that are supposed to define trusting behavior (Dunning & Fetchenhauer, 2013; Thielmann & Hilbig, 2015a).

Arguing that social trust defies the logic of prospection, Dunning and Fetchenhauer (2013) focused on the expressive function of trust. Unlike individual risk-taking, where decisions are strongly influenced by anticipated outcomes and probabilities (Kahneman & Tversky, 1979; Mellers & McGraw, 2001), trusting behavior may serve the pursuit of immediate, nonmaterial goals. For example, people may trust strangers because they construe trust as an injunctive norm (Dunning et al., 2014; but see Bicchieri, Xiao, & Muldoon, 2011). People feel they should trust others even when they do not want to and even when they believe that reciprocity is unlikely to occur. According to this account, trust among strangers is blind to consequences and probabilities. The act of trust is rewarding for its own sake: it is not based on the characteristics of the trustee or the financial costs and benefits.

The theory of expressive trust offers a counterpoint to rational models of game-theoretic reasoning, but fails to account for all the evidence suggesting at least a bounded role of prospection. Contrary to the idea that trust is driven primarily by a sense of social obligation, ratings of what people *want* to do are a stronger

predictor of behavior than ratings of what people feel they *should* do (Dunning et al., 2014). Moreover, although the effects of expectations on decisions tend to be weaker than normative models of rationality would demand, these effects do emerge reliably over studies: positive expectations consistently predict trust (Evans & Krueger, 2014; Snijders & Keren, 1999). As expected from rational individuals, people use an array of informational cues to form expectations of reciprocity (Thielmann & Hilbig, 2015a). Some cues, such as the trustee's physical appearance, have large and robust effects on trusting behavior (Bonnenfon, Hopfensitz, & De Neys, 2013; Eckel & Petrie, 2011; Wilson & Eckel, 2006). These findings demonstrate that prospection plays some role in the development of trust among strangers, though previous theories have not explained how reasoning in dilemmas of trust differs from the processes suggested by game-theoretic reasoning. To make progress in this direction, we now outline a model of boundedly rational trust.

Bounded Rationality and Trust

We propose that trust decisions are neither fully rational nor blindly expressive. Instead, trust can be understood as boundedly rational (Simon, 1955). People lack the time and cognitive capacity to approach trust with a fully calculative mindset, and instead they rely on a heuristic strategy to reach a decision (Gigerenzer, Todd, & the ABC Research Group, 1999). Heuristic processes are fast and can be executed with little effort (Masicampo & Baumeister, 2008; Rand, Greene, & Nowak, 2012); in some environments, heuristics perform as well as (or even better than) fully rational analyses (Todd & Gigerenzer, 2007). The framework of bounded rationality suggests that seemingly irrational behavior observed in the context of anonymous, one-shot laboratory experiments can be understood in broader terms as a cognitively efficient and robust strategy to resolve the dilemma of trust (Hertwig & Herzog, 2009).

Heuristics in Individual Decision-Making

Over the past 40 years, researchers in psychology and behavioral economics have documented examples suggesting that individual decision-making routinely violates the laws of statistical probability and expected utility theory (Hastie & Dawes, 2009; Kahneman, 2011). Prospect theory (Kahneman & Tversky, 1979) and models of hyperbolic discounting (Laibson, 1997), attempt to amend or adapt normative models to better reflect human behavior. However, these violations of rationality can also be understood in terms of simple heuristic processes that are fast and simple to execute (Gigerenzer et al., 1999). Heuristic models have been proposed to describe decisions involving risky outcomes (Brandstätter, Gigerenzer, & Hertwig, 2006; Hilbig, 2008), multiattribute choices (Hogarth & Karelaia, 2005), forecasts (Åstebro & Elhedhli, 2006; Goldstein & Gigerenzer, 2009), and intertemporal tradeoffs (Marzilli Ericson, White, Laibson, & Cohen, 2015).

Many heuristic models assume that people use lexicographic strategies; they reduce the complexity of decisions by comparing relevant cues in a sequential order until a satisfactory decision can be reached (Goldstein & Gigerenzer, 1999). For example, the priority heuristic states that when people choose between two risky gambles, they sequentially compare (a) the worst possible outcome, (b) the probabilities of receiving the worst possible out-

comes, and (c) the best possible outcomes (Brandstätter et al., 2006). Importantly, lexicographic strategies predict that people stop evaluating cues as soon as the evidence suggests that one alternative is “good enough.” As a result, final choices are often made without considering all available information. Other models assume that people use tallying strategies, which posit that people use all available information, but simplify the process of calculation by assigning each attribute an equal decision weight (Dawes, 1979). Tallying rules can perform almost as well as (and sometimes better than) a multiple regression approach that assigns an optimal decision weight to each cue (Czerlinski, Gigerenzer, & Goldstein, 1999).

Egocentric Outcomes and Expectations

Recent work has begun to extend the framework of bounded rationality to investigate how heuristics shape social decision-making (Hertwig & Herzog, 2009), but this research has not yet addressed the role that heuristic thinking plays in shaping trusting behavior. We argue that bounded rationality is a lens to understand dilemmas of trust (Dunning et al., 2014). Trust decisions may follow from a simplified or reduced form of prospective reasoning, where decisions are based on a single, highly salient cue (Goldstein & Gigerenzer, 1999), instead of a comprehensive strategy where informational cues are fully integrated (Payne, Bettman, & Johnson, 1993).

When faced with uncertainty, people tend to focus on accessible information and are less likely to focus on information that is ambiguous or difficult to evaluate (Alter & Oppenheimer, 2009; Morewedge & Kahneman, 2010; Slovic, Finucane, Peters, & MacGregor, 2007). In social situations, this tendency to prioritize accessible information leads people to emphasize their own experiences and preferences, while giving less thought to other points of view (Krueger, 2003). Overriding this egocentric tendency to consider other perspectives requires time and cognitive effort (Lin, Keysar, & Epley, 2010; Tamir & Mitchell, 2013). Hence, the central premise of our model is that the tendency to focus on easily processed information leads trustors to prioritize their own potential outcomes (what they will gain or lose from trusting) and ignore expectations of reciprocity. This egocentrism arises because outcomes can be assessed directly, whereas expectations must be formed through the integration of base-rate information and informational cues.

Risk and temptation. To investigate the heuristic model of trust, we first consider the question of how trust is influenced by changes in the structure of the decision environment; Snijders and Keren (1999) proposed that the trust game is defined by two structural factors: risk and temptation (Figure 2). Risk is the ratio of the trustor’s potential cost of betrayal and the trustor’s potential gain from reciprocity. Risk is great when there is much to lose and little to gain from trusting. The trustee’s temptation is the financial incentive for choosing betrayal instead of reciprocity. The trustee’s temptation is large when there is a large bonus for acting selfishly.

Rational and expressive theories of trust make distinct predictions about how behavior should be related to changes in the two payoff factors: theories of full rationality predict that trust should closely track any changes in both risk and temptation. When risk increases, the outcomes of trust become less favorable, whereas when temptation increases, the probability of reciprocity decreases

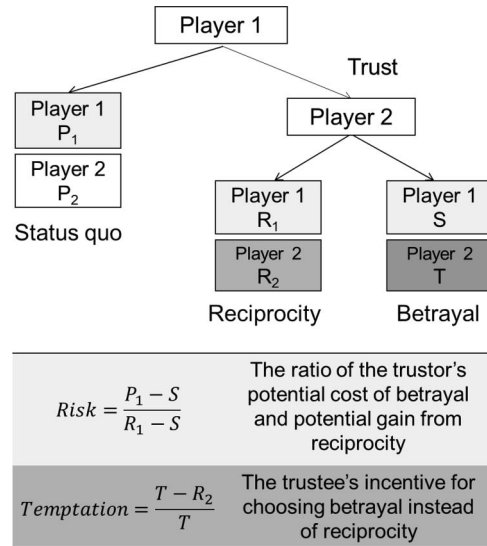


Figure 2. Risk and temptation in the trust game.

(Malhotra, 2004; Snijders & Keren, 1999). Hence, a rational decision-maker should use the trustee’s temptation to estimate the probability of reciprocity and then weight the potential outcomes of trust by their probabilities of occurring. Conversely, the theory of purely expressive trust suggests that changes in risk and temptation have little effect on trusting behavior. Expressive reasoning predicts that trust occurs irrespective of the potential outcomes and the probability of reciprocity (Dunning et al., 2014).

In contrast, our heuristic approach predicts that trust should be strongly related to the trustor’s risk and weakly related to the trustee’s temptation. This prediction stems from the idea that risk is egocentrically relevant and is therefore easier to evaluate than the trustee’s temptation. Indeed, the rate of trust is highly sensitive to changes in risk, but only weakly influenced by changes in the trustee’s temptation (Evans et al., 2013; Evans & Krueger, 2011, 2014; Malhotra, 2004; Snijders & Keren, 1999). One possible explanation for the weak effect of temptation on trust is that trustors focus on other sources of information to form expectations of reciprocity. For example, trustors may ignore temptation and instead rely on their general expectations of whether people are trustworthy (Rotter, 1967). To test this possibility, Evans and Krueger (2014) measured participants’ expectations of reciprocity and the effects of these expectations on trust decisions. The level of temptation was strongly correlated with trustors’ expectations of reciprocity, but these expectations were only weakly correlated with their ultimate trust decisions. Trust decisions were more sensitive to changes in outcomes than expectations, even after accounting for their differing effects on the expected value of trust (Evans & Krueger, 2014). In other words, trustors understood that the trustee’s temptation was important, but did not fully apply this knowledge when making decisions. In other words, trustors are capable, but unwilling, to engage in the computations of fully rational decision-making.

Process-tracing trust. An important aspect of our heuristic model is that trustors may not use all available information at the moment of decision-making. Heuristic models emphasize that the

process of decision-making terminates once an option is deemed “good enough” (Brandstätter et al., 2006). In particular, trustors are likely to reach a decision without considering the trustee’s payoffs. To test this prediction, Evans and Krueger (2014) studied decisions in the trust game using process-tracing methods (Schulte-Mecklenbeck, Kühberger, & Ranyard, 2011). In one experiment, participants were presented with variants of the trust game where the potential outcomes of the game were hidden in boxes. To learn about the different outcomes, participants needed to drag the mouse over each box to reveal its contents. Trustors often decided without gathering information about the trustee’s outcomes (R_2 and T , the payoffs associated with the trustee’s temptation), even though searching for information was costless and trustors explicitly understood that temptation was strongly related to the probability of reciprocity. In contrast, trustors almost always searched for the payoffs associated with risk (P_1 , R_1 , and S). Trustees were similarly egocentric, consistently searching for information related to temptation, but ignoring the trustor’s level of risk.

Lexicographic trust. Evans and Krueger (2011) raised the possibility that trust is based on a lexicographic process, with relevant cues evaluated sequentially until a decision can be reached. Reaction time and behavioral data suggest that decisions occur in two stages (Figure 3): first, trustors evaluate the egocentric risk of trusting. If the potential outcomes of trust are acceptable, they engage in a limited, secondary stage where they consider the probability that the trustee will reciprocate (Evans & Krueger, 2011). When the outcomes of trust are unacceptable (the level of risk is high), trustors decide quickly and do not consider the trustee’s temptation. However, if the potential outcomes are acceptable (the level of risk is low), trustors decide more slowly and are more likely to be influenced by the level of temptation.

Summary. Models of full rationality predict that trust should be influenced by both outcomes and expectations, whereas models of expressive trust predict that trust should be influenced by neither. Our recent work finds that potential outcomes influence decision-making in dilemmas of trust, but expectations have a weak or secondary effect on final decisions (Evans & Krueger, 2011, 2014). Trustors often decide without searching for information related to the trustee’s potential outcomes (Evans & Krueger, 2014); and the trustee’s temptation is only evaluated when personal outcomes have been evaluated favorability (Evans & Krueger, 2011). Taken together, these findings point to the conclusion that trust, being neither fully rational nor blindly expressive, is based off of bounded prospection.

Implications of Bounded Rationality

We now consider the broader relevance of our heuristic model, and suggest that it may provide new answers to outstanding questions in trust research. First, we examine whether trustors give equal consideration to prospective gains and losses, or if trustors are loss averse, prioritizing losses over gains (Kahneman & Tversky, 1979; Thielmann & Hilbig, 2015a). Then, we consider the role of effort in decision-making: does trusting behavior require self-control, or does it occur effortlessly and automatically? Our model suggests that perspective-taking, rather than trust itself, requires cognitive resources. The depletion of self-control can potentially increase or decrease trusting behavior (Evans et al.,

2011). Finally, we address the developmental trajectory of trust. Our work suggests that older children are more likely to trust, but they are also more discerning in when to trust others (Evans et al., 2013).

Trust and loss aversion. In our previous discussion of the trustor’s decision-making, we focused on the distinction between outcomes and expectations, but did not address the possibility that trustors may be differentially sensitive to potential gains and losses (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Kahneman & Tversky, 1979). Indeed, Evans and Krueger (2011) found that trustors were more sensitive to changes in the cost of betrayal ($P_1 - S$) compared to changes in the benefit of reciprocity ($R_1 - P_1$), consistent with the idea that loss aversion influences trusting behavior (Thielmann & Hilbig, 2015a).

This raises the question of whether trustors integrate potential outcomes, forming an overall assessment of their desirability, or if they focus primarily on the cost of betrayal and give less attention to the benefit of reciprocity.³ Two findings suggest that trustors simultaneously evaluate and integrate the potential costs and benefits of trust: first, information search data suggest that trustors are equally likely to access information about the cost of betrayal and the benefit of reciprocity (Evans & Krueger, 2014). Note that a lexicographic search process predicts that trustors are more likely to search for payoffs related to losses compared to payoffs related to gains. Second, manipulations of the cost of betrayal and the benefit of reciprocity have independent, rather than interactive, effects on the rate of trust (Evans & Krueger, 2011). Changes in the benefit of reciprocity influence trust even when there is an extremely high cost of betrayal, suggesting that positive and negative outcomes are evaluated simultaneously, rather than lexicographically.

Trust and self-control. Recently, researchers have asked whether trusting behavior is effortless and automatic, or if acts of trust require reflection and effortful self-control (Ainsworth, Baumeister, Ariely, & Vohs, 2014; Evans et al., 2011). Our heuristic model suggests that the ability to form and use accurate expectations of reciprocity, rather than trust itself, requires self-control. Our model predicts that trustors are less likely to carefully process relevant informational cues in the environment when self-control resources are depleted. In some cases, this will result in a decrease in trusting behavior (Ainsworth et al., 2014). The depletion of self-control is likely to reduce trusting behavior in situations where trustors may miss subtle cues that the trustee is, in fact, trustworthy. Yet, in other situations, ego-depletion may increase trusting behavior. For example, Evans and colleagues (2011) found that the depletion of self-control decreases trust, but only when distrust is the preselected default response, the option that requires the least amount of physical effort. When trust is the preselected response, the depletion of self-control increases trusting behavior. Ego-depletion causes individuals to focus on salient, easily processed information (such as the default response) and to neglect information that requires reflective processing (such as expectations of reciprocity). Therefore, the depletion of self-

³ A model of trust based on the priority heuristic (Brandstätter et al., 2006) would predict that trustors sequentially evaluate (a) the cost of betrayal, (b) the probability of betrayal, and (c) the benefit of reciprocity. In contrast, our model predicts that trustors simultaneously evaluate the cost of betrayal and the benefit of reciprocity.

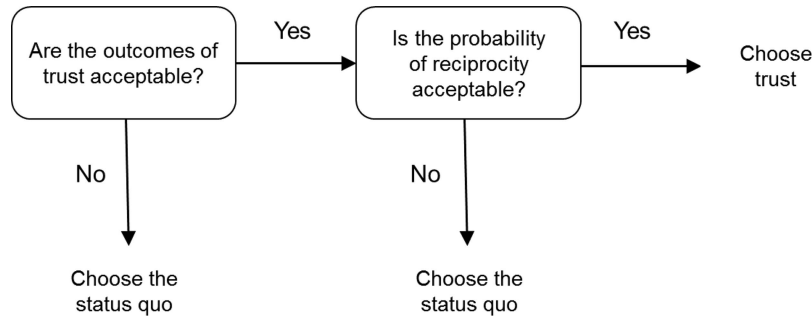


Figure 3. Reaction time and behavioral data suggest that trust decisions occur in two stages (Evans & Krueger, 2011): First, people evaluate the potential outcomes of trust. If the outcomes are acceptable, then trustors engage in a secondary stage where they consider the probability of reciprocity. According to this model, trust only occurs when both outcomes and expectations of reciprocity are acceptable.

control can make people more or less trusting, depending on the structure of the decision environment.

The development of trust. Do humans learn when to trust others at an early age (Zaki & Mitchell, 2013), or does trust require cognitive sophistication? Previous research has argued that trust increases linearly with age (Sutter & Kocher, 2007). A series of developmental studies suggests that age-related changes in trust are related to changes in the ability to account for expectations of reciprocity (Evans et al., 2013). Elementary schoolchildren (10- and 11-year-olds) were more trusting than kindergarteners (5- to 6-year-olds), but they were also more careful about when to trust. Consistent with our heuristic model, younger children were not sensitive to situational cues, such as the level of the trustee's temptation, and instead focused on their own outcomes when making decisions. Older children were more trusting and more altruistic, but age-related increases in altruism did not predict increased trust. Accurate trust decisions require cognitive skills that develop with age, though even then adults underemphasize the role of expectations (Evans & Krueger, 2011, 2014).

The Limits of Egocentric Trust

The framework of bounded rationality posits that trustors underemphasize expectations because they must be derived from base-rate information and the trustee's situational temptation. This suggests that trustors will be more likely to focus on expectations when they are based on salient, easy-to-process sources of information, such as past behavior (Bohnet & Huck, 2004; King-Casas et al., 2005) compared to when they are based on the trustors' subjective beliefs or difficult to evaluate cues (Evans & Krueger, 2014). Indeed, Bonnefon and colleagues (2013) found that when participants were shown full-color photos of trustees' faces, there was a near-perfect correlation between expectations of trustworthiness and trustworthy behavior; however, when participants were shown cropped, black and white photos of the trustees, the correlation between expectations and trustworthy behavior was significantly attenuated. This difference could be explained by the difficulty of forming a clear impression from a cropped photograph. Similarly, trustors may be less likely to focus on potential outcomes (i.e., potential gains and losses) when the outcomes of trusting are ambiguous or difficult to evaluate (Ellsberg, 1961; Hogarth, 1989).

Our work also raises questions about whether the primacy of outcomes over expectations extends to other game-theoretic dilemmas related to cooperation and altruism. In contrast to models of social preferences, which assume that people apply strategies consistently across situations (Peysakhovich et al., 2014; Yamagishi et al., 2013), the bounded rationality approach assumes that the heuristics people use are game-specific. In particular, there may be important differences between reasoning in asymmetric dilemmas (e.g., the trust game and the ultimatum game) and symmetric dilemmas (e.g., the prisoner's dilemma, the game of chicken, and the public goods game). In symmetric dilemmas, where the players decide simultaneously and face the same strategic dilemma, it may be easier to use social projection to predict the behavior of other players (Krueger, 2014; Krueger, DiDonato, & Freestone, 2012; Thielmann & Hilbig 2015a), resulting in a stronger link between expectations and cooperative behavior (Balliet & Van Lange, 2013).

Bounded Rationality and Accuracy

In addition to investigating how trust decisions are made (Evans & Krueger, 2009; Thielmann & Hilbig, 2015a), researchers have also questioned whether trust decisions are made accurately (Bonnefon et al., 2013; DeSteno et al., 2012). Dilemmas of trust involve two related challenges: (a) forming expectations of reciprocity, and (b) using outcomes and expectations to reach a decision. Recent work on accuracy has focused on directional errors (Krueger & Funder, 2004), asking if trustors' expectations of reciprocity are overly optimistic or overly cynical and if people trust too much or too little (Dunning et al., 2014; Dunning & Fetchenhauer, 2013; Fetchenhauer & Dunning, 2009; Schlösser, Mensching, Dunning, & Fetchenhauer, 2015; Van Lange, 2015). Directional errors suggest that the accuracy of trust can be improved through a process of deliberate self-correction (Herzog & Hertwig, 2009; Krueger & Chen, 2014).

Alternatively, our heuristic model suggests that people err in how they select and prioritize cues. Trustors neglect valid but difficult-to-evaluate information (Evans et al., 2013; Evans & Krueger, 2011, 2014) and overweight easy-to-evaluate cues that lack validity (Kausel & Connolly, 2014; Rule, Krendl, Ivcevic, & Ambady, 2013). This cue-focused view of accuracy suggests that directional errors (e.g., having overly cynical expectations of rec-

iprocity or trusting too much) are inconsistent. Sometimes people trust too much, other times they trust too little—the direction of error depends on how trustors weight the available information. Thus, improvements in accuracy depend on the ability to identify valid cues and to ignore irrelevant ones.

Forming Expectations of Reciprocity

Fully successful trust decisions depend, in part, on the ability to form accurate expectations of reciprocity. Expectations are influenced by generalized beliefs about whether others are trustworthy (Evans & Revelle, 2008; Rotter, 1967; Yamagishi & Yamagishi, 1994), as well as inferences based on personal and situational cues (Thielmann & Hilbig, 2015a): Personal cues are the characteristics of the trustee's appearance and behavior that may reveal her psychological state (Steyer, Ferring, & Schmitt, 1992) and stable dispositions (Gosling, 2009). Situational cues are characteristics of the specific interaction that may have a main effect on trustworthiness (Snijders & Keren, 1999) or interact with personal cues to predict behavior (Snyder, 1983).

Cynicism and the norm of self-interest. When interacting with a stranger, do people tend to over- or underestimate the likelihood of reciprocity? A number of studies suggest that people underestimate the overall level of trustworthiness (Dunning et al., 2014; Fetchenhauer & Dunning, 2012; Schlösser et al., 2015; Vuolevi & Van Lange, 2010). People believe that much of human behavior is determined by self-interest, and perceive overtly prosocial behaviors in terms of hidden selfish motives (Critcher & Dunning, 2011; De Vito & Bonnefon, 2014; Miller, 1999). These cynical beliefs are perpetuated by lack of feedback; distrustful individuals never have the opportunity to learn that others are, in fact, more trustworthy than expected (Carter & Weber, 2010; Denrell, 2005; Fetchenhauer & Dunning, 2010; Sturgis, Read, & Allum, 2010).

This tendency to underestimate trustworthiness may be explained in terms of the minimization of costly errors (Haselton & Buss, 2000; Haselton, 2007). Error management theory proposes that biased judgments may be influenced by the asymmetric costs of different outcomes. There are two possible errors in a dilemma of trust: you may trust someone who betrays you (Type I) or fail to trust someone who is actually trustworthy (Type II). If we assume that people are more sensitive to losses than gains (Kahneman & Tversky, 1979) and are also particularly sensitive to negative social experiences (Eisenberger & Lieberman, 2004), then it may be advantageous to minimize the possibility of betrayal by systematically underestimating the likelihood of reciprocity.

Informational cues and accuracy. Rather than focusing on directional errors, our boundedly rational approach suggests that it is more important to consider the types of information trustors use to form expectations: Trustors overemphasize easily accessed information (Schwarz, 1998) and ignore cues that are ambiguous or difficult to evaluate (Einhorn & Hogarth, 1985). As the following examples illustrate, this tendency can result in overly cynical or overly optimistic expectations of reciprocity.

Because trustors are motivated to conserve cognitive resources, they may underemphasize difficult-to-process information that strongly predicts trustworthy behavior: for example, the trustee's temptation is difficult to evaluate, but has a large effect on trustworthy behavior (Evans & Krueger, 2014; Malhotra, 2004; Sni-

iders & Keren, 1999). Trustors understand that temptation has a large effect on trustworthy behavior, but they underestimate the size of this effect (Figure 4). Evans and Krueger (2014) found that trustors underestimate the probability of reciprocity when temptation is low, but they overestimate the probability of reciprocity when temptation is high. In this case, trustors are neither overly cynical nor overly optimistic, the direction of error depends on the level of the trustee's temptation.

In other cases, people rely on easy-to-process information that fails to reliably predict trustworthy behavior. For example, expectations of reciprocity are strongly influenced by the physical appearance of the trustee. When judging a face for the first time, people quickly and automatically form an impression about whether that person is trustworthy (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Although judgments based on appearances are easily formed and significantly influence behavior, they often lack validity: for example, Wilson and Eckel (2006) found that trustors believe that attractive individuals are more likely to be trustworthy, although there was no actual correlation between beauty and trustworthiness (see Andreoni & Petrie, 2008 for similar results in the public goods game). Looking at a range of real-life domains, Rule and colleagues (2013) found that people are not able to identify trustworthiness from physical appearance, even though subjective impressions of trustworthiness are highly consistent across individuals. Critically, judgments based on physical appearances can lead to expectations of reciprocity that are overly cynical (Kausel & Connolly, 2014) or overly optimistic (Wilson & Eckel, 2006).

Research on the effects of the trustee's temptation (Evans & Krueger, 2014) and physical appearance (Wilson & Eckel, 2006) on expectations of reciprocity reveal that errors in the formation of expectation arise from how trustors use (or neglect to use) specific informational cues. Hence, accuracy in the formation of expectations depends on the process of cue selection, rather than a general tendency to over- or underestimate the likelihood of reciprocity.

The Optimal Level of Trust

In addition to forming expectations of reciprocity, trustors must use this information to reach a final decision. To ask if people make optimal trust decisions, researchers have compared trust

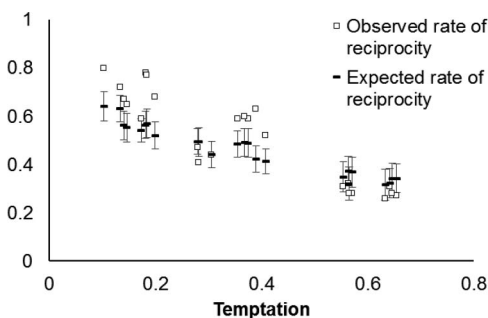


Figure 4. Study 2 of Evans and Krueger (2014) compared the effects of the trustee's temptation on the observed and expected rates of trustee. When the trustee's temptation was low, trustors underestimated the likelihood of reciprocity, but when temptation was high they overestimated the probability of reciprocity. Error bars denote 95% CI.

decisions with comparable dilemmas involving individual risk-taking. Different studies have concluded that people are too fearful (Bohnet & Zeckhauser, 2004) or too trusting (Fetchenhauer & Dunning, 2009) when they interact with strangers. In contrast, our heuristic model reconciles these findings by assuming that trustors differentially weight expectations in dilemmas of trust and individual risk-taking (Evans & Krueger, 2014). Expectations play a larger role in individual risk-taking decisions than trust decisions. This leads to the prediction that people trust too much when the probability of reciprocity is low, and trust too little when the probability of reciprocity is high.

Betrayal aversion and excessive trust. Assuming identical consequences and probabilities, are people more or less willing to trust than take individual risks? Bohnet and Zeckhauser (2004) compared decisions in a trust game and a risky gamble. In the trust game, participants chose between the guaranteed status quo and an uncertain outcome determined by another person. In the individual gamble, the participant's payoffs were identical but the uncertain outcome was determined by chance instead of another person.⁴ Participants were asked to state the minimum acceptable probability (MAP) of reciprocity in the trust game (or the analogous probability of winning in the individual gamble) necessary for the player to choose the uncertain outcome instead of the guaranteed status quo. Decision-makers stated higher MAPs in the trust game compared to the individual gamble (Bohnet & Zeckhauser, 2004), suggesting that participants perceived an additional, nonmonetary cost associated with betrayal. Losing money due to betrayal was worse than losing the same amount due to chance. The finding of betrayal aversion has been replicated in diverse cultural contexts (Aimone & Houser, 2012; Bohnet, Greig, Herrmann, & Zeckhauser, 2008), and the effect is stronger among high-status groups (Hong & Bohnet, 2007).

Betrayal aversion is consistent with psychological findings that people are highly sensitive to negative social experiences, such as interpersonal rejection and ostracism (Williams, 2007). Human beings have a fundamental need to belong (Baumeister & Leary, 1995) and negative social experiences that threaten this need activate several of the brain regions associated with physical pain (Eisenberger & Lieberman, 2004; MacDonald & Leary, 2005). People prefer not to be socially excluded, even when being included in the group means earning less money (Van Beest & Williams, 2006). The experience of betrayal is particularly strong when there is a preexisting relationship (Shackelford & Buss, 1996), but ostracism still has a strong negative effect even when people are excluded by strangers or the members of a disliked outgroup (Gonsalkorale & Williams, 2007). The fear of betrayal may be a persistent obstacle to the creation of trust among strangers.

Yet, other experiments show instances of excessive trust, situations where players are more willing to trust others than take individual risks. Fetchenhauer and Dunning (2009) conducted a series of studies comparing decisions in the trust game with MAPs in an individual gamble. The authors used MAPs in the individual gamble to predict whether trustors, given their subjective expectations of reciprocity, should choose trust or the status quo. Trustors were consistently more willing to trust others than to accept risky gambles. Fetchenhauer and Dunning (2012) obtained similar results comparing trust games and individual gambles where the

probability of reciprocity (or winning the gamble) was explicitly stated in the experiment (see also Schlösser et al., 2015).

Excessive trust can be explained by respect for social norms (Dunning et al., 2014). People trust, in part, because it is something they feel they should do even if it is not in their best material interests. In fact, it may be socially risky not to trust. The recipients of trust react negatively when they feel they have not been trusted completely (Pillutla, Malhotra, & Murnighan, 2003). Individuals who refuse to trust strangers are seen as less moral (Krueger et al., 2008), though people do not believe that others should be punished for choosing not to trust (Bicchieri et al., 2011).

Heuristic trust and expectations of reciprocity. Our heuristic model suggests that the primary difference between decisions involving social trust and individual risk-taking is in how people weight expectations. In the domain of individual risk-taking, outcomes and probabilities both have large effects on final decisions (Slovic & Lichtenstein, 1968); and when risk-takers search for information, they are equally likely to search for outcome- and probability-relevant information (Johnson, Schulte-Mecklenbeck, & Willemsen, 2008; Michalaszek & Sokolowska, 2010). In contrast, trustors focus on outcomes and routinely ignore probabilities (Evans & Krueger, 2014). Changes in probabilities have larger effects on individual risk-taking decisions than trust decisions (Fetchenhauer & Dunning, 2012). Hence, our model predicts that people trust too much when the probability of reciprocity is low and trust too little when the probability is high, as illustrated in Figure 5. This pattern can potentially explain why previous studies found that people trust too much in some situations and too little in others.

Conclusion

The ability to make accurate decisions about when to trust another person in a context of social or economic exchange is essential for individual well-being and societal growth (Arrow, 1974; Erikson, 1955; Harris, 2012; Hosking, 2014; Kohn, 2008; Van Lange, 2015). Given the complexity and difficulty of real-life trust decisions, it is not surprising that wildly divergent theoretical perspectives have been proposed. At one end of the spectrum, rationalist theories assume that people should perform an exhaustive evaluation and integration of all relevant information. Specifically, people should be willing and able to assess the costs and benefits of trust and distrust, and weigh them against the probability that their trust will be betrayed. At the other end of the spectrum, expressive theories assume that people trust inasmuch as they perceive the presence of a social norm to trust. These latter theories suggest that people tend to trust more than is in their own best interest.

In this article, we staked a middle ground between these two extremes. Rather than mixing or averaging elements of full rationalism with expressive decision-making, we have sketched a

⁴ Bohnet and Zeckhauser (2004) also presented participants with a risky dictator game, where the participant's decision also affected a second player who received the same payoffs as the second player in the trust game. This second party made no active choice and the consequences of choosing the uncertain option were again determined by chance. This condition controlled for the fact that trust also conferred benefits to a second party, but there were no significant differences between the risky gamble and the risky dictator game.

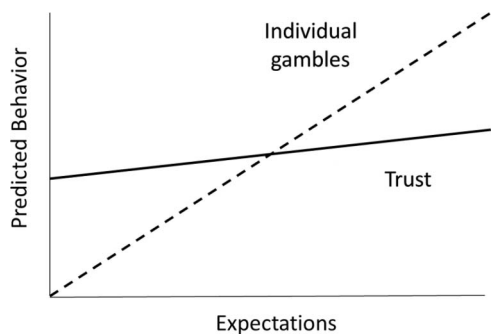


Figure 5. The heuristic model of trust predicts that people do not sufficiently account for expectations when making trust decisions.

model of bounded rationality in the tradition of Simon (1955) and Gigerenzer (e.g., Gigerenzer & Gaissmaier, 2011). Our model assumes that people consider relevant cues in a simple and lawful sequence, and that they reach a decision regarding trust when they feel that the evidence (in either direction) has reached a satisfactory threshold. This model is supported by empirical evidence, and it is consistent with a broad range of findings in the areas of social cognition and judgment and decision-making. One particularly important achievement of this model is that it can make predictions about when people trust too much (a Type I error) and when people trust too little (a Type II error). Moving the study of trust beyond simple (or simplistic) inferences about human nature, the model of boundedly rational trust can help set the stage for further research that recognizes the fact that the dilemma of trust is among the most intricate challenges members of a self-aware social species face.

References

- Aimone, J. A., & Houser, D. (2012). What you don't know won't hurt you: A laboratory analysis of betrayal aversion. *Experimental Economics*, 15, 571–588. <http://dx.doi.org/10.1007/s10683-012-9314-z>
- Ainsworth, S. E., Baumeister, R. F., Ariely, D., & Vohs, K. D. (2014). Ego depletion decreases trust in economic decision making. *Journal of Experimental Social Psychology*, 54, 40–49. <http://dx.doi.org/10.1016/j.jesp.2014.04.004>
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13, 219–235. <http://dx.doi.org/10.1177/1088868309341564>
- Anderhub, V., Engelmann, D., & Güth, W. (2002). An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior & Organization*, 48, 197–216. [http://dx.doi.org/10.1016/S0167-2681\(01\)00216-5](http://dx.doi.org/10.1016/S0167-2681(01)00216-5)
- Andreoni, J., & Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology*, 29, 73–93. <http://dx.doi.org/10.1016/j.joep.2007.07.008>
- Arrow, K. J. (1974). *The limits of organization*. New York, NY: Norton.
- Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9, 193–208. <http://dx.doi.org/10.1007/s10683-006-9122-4>
- Åstebro, T., & Elhedhli, S. (2006). The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science*, 52, 395–409. <http://dx.doi.org/10.1287/mnsc.1050.0468>
- Balliet, D., & Van Lange, P. A. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, 139, 1090–1112. <http://dx.doi.org/10.1037/a0030939>
- Basu, K. (1994). The traveler's dilemma: Paradoxes of rationality in game theory. *The American Economic Review*, 84, 391–395.
- Basu, S., Dickhaut, J., Hecht, G., Towry, K., & Waymire, G. (2009). Recordkeeping alters economic history by promoting reciprocity. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 106, 1009–1014. <http://dx.doi.org/10.1073/pnas.0811967106>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323–370. <http://dx.doi.org/10.1037/1089-2680.5.4.323>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497–529. <http://dx.doi.org/10.1037/0033-2909.117.3.497>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10, 122–142. <http://dx.doi.org/10.1006/game.1995.1027>
- Bicchieri, C., Xiao, E., & Muldoon, R. (2011). Trustworthiness is a social norm, but trusting is not. *Politics, Philosophy & Economics*, 10, 170–187. <http://dx.doi.org/10.1177/1470594X10387260>
- Bó, P. D. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *The American Economic Review*, 95, 1591–1604. <http://dx.doi.org/10.1257/000282805775014434>
- Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *The American Economic Review*, 98, 294–310. <http://dx.doi.org/10.1257/aer.98.1.294>
- Bohnet, I., & Huck, S. (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *The American Economic Review*, 94, 362–366. <http://dx.doi.org/10.1257/0002828041301506>
- Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55, 467–484. <http://dx.doi.org/10.1016/j.jebo.2003.11.004>
- Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General*, 142, 143–150. <http://dx.doi.org/10.1037/a0028930>
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113, 409–432. <http://dx.doi.org/10.1037/0033-295X.113.2.409>
- Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Carter, N. L., & Weber, J. M. (2010). Not Pollyannas: Higher generalized trust predicts lie detection ability. *Social Psychological and Personality Science*, 1, 274–279. <http://dx.doi.org/10.1177/1948550609360261>
- Charness, G., Du, N., & Yang, C. L. (2011). Trust and trustworthiness reputations in an investment game. *Games and Economic Behavior*, 72, 361–375. <http://dx.doi.org/10.1016/j.geb.2010.09.002>
- Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, 46, 260–281. [http://dx.doi.org/10.1016/S0899-8256\(03\)00119-2](http://dx.doi.org/10.1016/S0899-8256(03)00119-2)
- Critcher, C. R., & Dunning, D. (2011). No good deed goes unquestioned: Cynical reconstructions maintain belief in the power of self-interest. *Journal of Experimental Social Psychology*, 47, 1207–1213. <http://dx.doi.org/10.1016/j.jesp.2011.05.001>
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & the ABC Research Group. (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York, NY: Oxford University Press.
- Dasgupta, P. (2007). *Economics: A very short introduction*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/actrade/9780192853455.001.0001>

- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582. <http://dx.doi.org/10.1037/0003-066X.34.7.571>
- Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, *112*, 951–978. <http://dx.doi.org/10.1037/0033-295X.112.4.951>
- DeSteno, D. (2014). *The truth about trust: How it determines success in life, love, learning, and more*. New York, NY: Penguin.
- DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science*, *23*, 1549–1556. <http://dx.doi.org/10.1177/0956797612448793>
- De Vito, S., & Bonnefon, J. F. (2014). People believe each other to be selfish hedonic maximizers. *Psychonomic Bulletin & Review*, *21*, 1331–1338. <http://dx.doi.org/10.3758/s13423-014-0592-7>
- 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. <http://dx.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. <http://dx.doi.org/10.1177/1745691612474319>
- Eckel, C. C., & Petrie, R. (2011). Face value. *The American Economic Review*, *101*, 1497–1513. <http://dx.doi.org/10.1257/aer.101.4.1497>
- Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization*, *55*, 447–465.
- Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, *92*, 433–461. <http://dx.doi.org/10.1037/0033-295X.92.4.433>
- Eisenberger, N. I., & Lieberman, M. D. (2004). Why rejection hurts: A common neural alarm system for physical and social pain. *Trends in Cognitive Sciences*, *8*, 294–300. <http://dx.doi.org/10.1016/j.tics.2004.05.010>
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, *75*, 643–669. <http://dx.doi.org/10.2307/1884324>
- Erikson, E. H. (1955). *Childhood and society*. New York, NY: Norton.
- Evans, A. M., Athenstaedt, U., & Krueger, J. I. (2013). The development of trust and altruism during childhood. *Journal of Economic Psychology*, *36*, 82–95. <http://dx.doi.org/10.1016/j.joep.2013.02.010>
- Evans, A. M., Dillon, K. D., Goldin, G., & Krueger, J. I. (2011). Trust and self-control: The moderating role of the default. *Judgment and Decision Making*, *6*, 697–705.
- Evans, A. M., & Krueger, J. I. (2009). The psychology (and economics) of trust. *Social and Personality Psychology Compass*, *3*, 1003–1017. <http://dx.doi.org/10.1111/j.1751-9004.2009.00232.x>
- Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk and perspective-taking. *Journal of Experimental Social Psychology*, *47*, 171–177. <http://dx.doi.org/10.1016/j.jesp.2010.08.007>
- Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust. *Judgment and Decision Making*, *9*, 90–103.
- Evans, A. M., & Krueger, J. I. (2015). The edge of trust: An introduction. *Social Cognition*, *33*, 359–367. <http://dx.doi.org/10.1521/soco.2015.33.5.359>
- Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*, 1585–1593. <http://dx.doi.org/10.1016/j.jrp.2008.07.011>
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140. <http://dx.doi.org/10.1038/415137a>
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*, 817–868. <http://dx.doi.org/10.1162/003355399556151>
- Fetchenhauer, D., & Dunning, D. (2009). Do people trust too much or too little? *Journal of Economic Psychology*, *30*, 263–276. <http://dx.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*, 189–193. <http://dx.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 & Organization*, *81*, 534–541. <http://dx.doi.org/10.1016/j.jebo.2011.07.017>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. <http://dx.doi.org/10.1146/annurev-psych-120709-145346>
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (Eds.). (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, *115*, 811–846. <http://dx.doi.org/10.1162/003355300554926>
- Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & the ABC Research Group. (Eds.), *Simple heuristics that make us smart* (pp. 37–58). New York, NY: Oxford University Press.
- Goldstein, D. G., & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, *25*, 760–772. <http://dx.doi.org/10.1016/j.ijforecast.2009.05.010>
- Gonsalkorale, K., & Williams, K. D. (2007). The KKK won't let me play: Ostracism even by a despised outgroup hurts. *European Journal of Social Psychology*, *37*, 1176–1186. <http://dx.doi.org/10.1002/ejsp.392>
- Gosling, S. (2009). *Snoop: What your stuff says about you*. New York NY: Basic Books.
- Harris, P. H. (2012). *Trusting what you're told: How children learn from others*. New York NY: Belknap Press. <http://dx.doi.org/10.4159/harvard.9780674065192>
- Haselton, M. G. (2007). Error management theory. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (Vol. 1, pp. 311–312). Thousand Oaks, CA: Sage.
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*, 81–91. <http://dx.doi.org/10.1037/0022-3514.78.1.81>
- Hastie, R., & Dawes, R. M. (2009). *Rational choice in an uncertain world: The psychology of judgment and decision making*. New York, NY: Sage.
- Hertwig, R., & Herzog, S. M. (2009). Fast and frugal heuristics: Tools of social rationality. *Social Cognition*, *27*, 661–698. <http://dx.doi.org/10.1521/soco.2009.27.5.661>
- Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*, 231–237. <http://dx.doi.org/10.1111/j.1467-9280.2009.02271.x>
- Hilbig, B. E. (2008). One-reason decision making in risky choice? A closer look at the priority heuristic. *Judgment and Decision Making*, *3*, 457–462.
- Hogarth, R. M. (1989). Ambiguity and competitive decision making: Some implications and tests. *Annals of Operations Research*, *19*, 29–50. <http://dx.doi.org/10.1007/BF02283513>
- Hogarth, R. M., & Karelaia, N. (2005). Simple models for multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Science*, *51*, 1860–1872. <http://dx.doi.org/10.1287/mnsc.1050.0448>
- Hong, K., & Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology*, *28*, 197–213. <http://dx.doi.org/10.1016/j.joep.2006.06.003>

- Hosking, G. (2014). *Trust: A history*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780198712381.001.0001>
- Houser, D., Schunk, D., & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization*, 74, 72–81. <http://dx.doi.org/10.1016/j.jebo.2010.01.002>
- Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, 115, 263–272. <http://dx.doi.org/10.1037/0033-295X.115.1.263>
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32, 865–889. <http://dx.doi.org/10.1016/j.joep.2011.05.007>
- Johnson, N. D., & Mislin, A. A. (2012). How much should we trust the World Values Survey trust question? *Economics Letters*, 116, 210–212. <http://dx.doi.org/10.1016/j.econlet.2012.02.010>
- Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Straus & Giroux.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291. <http://dx.doi.org/10.2307/1914185>
- Kanagaretnam, K., Mestelman, S., Nainar, K., & Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. *Journal of Economic Psychology*, 30, 368–380. <http://dx.doi.org/10.1016/j.joep.2008.12.003>
- Kausel, E. E., & Connolly, T. (2014). Do people have accurate beliefs about the behavioral consequences of incidental emotions? Evidence from trust games. *Journal of Economic Psychology*, 42, 96–111. <http://dx.doi.org/10.1016/j.joep.2014.02.002>
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308, 78–83. <http://dx.doi.org/10.1126/science.1108062>
- Kohn, M. (2008). *Trust: Self-interest and the common good*. New York, NY: Oxford University Press.
- Krueger, J. I. (2003). Return of the ego—Self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review*, 110, 585–590. <http://dx.doi.org/10.1037/0033-295X.110.3.585>
- Krueger, J. I. (2014). Heuristic game theory. *Decision*, 1, 59–61. <http://dx.doi.org/10.1037/dec0000002>
- Krueger, J. I., & Chen, L. J. (2014). The first cut is the deepest: Effects of social projection and dialectical bootstrapping on judgmental accuracy. *Social Cognition*, 32, 315–336. <http://dx.doi.org/10.1521/soco.2014.32.4.315>
- Krueger, J. I., DiDonato, T. E., & Freestone, D. (2012). Social projection can solve social dilemmas. *Psychological Inquiry*, 23, 1–27. <http://dx.doi.org/10.1080/1047840X.2012.641167>
- Krueger, J. I., Evans, A. M., & Heck, P. R. (in press). Let me help you help me: Trust between profit and prosociality. In P. A. M. Van Lange, B. Rockenbach, & T. Yamagishi (Eds.), *Social dilemmas: New perspectives on trust*. New York, NY: Oxford University Press.
- Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, 27, 313–327. <http://dx.doi.org/10.1017/S0140525X04000081>
- 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. <http://dx.doi.org/10.1111/j.1750-4716.2007.00003.x>
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112, 443–478. <http://dx.doi.org/10.1162/003355397555253>
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46, 551–556. <http://dx.doi.org/10.1016/j.jesp.2009.12.019>
- MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, 131, 202–223. <http://dx.doi.org/10.1037/0033-2909.131.2.202>
- Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes*, 94, 61–73. <http://dx.doi.org/10.1016/j.obhdp.2004.03.001>
- Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*, 47, 534–559. <http://dx.doi.org/10.2307/3094850>
- Marzilli Ericson, K. M., White, J. M., Laibson, D., & Cohen, J. D. (2015). Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does. *Psychological Science*, 26, 826–833. <http://dx.doi.org/10.1177/0956797615572232>
- Masicampo, E. J., & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychological Science*, 19, 255–260. <http://dx.doi.org/10.1111/j.1467-9280.2008.02077.x>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20, 709–734.
- Mellers, B. A., & McGraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science*, 10, 210–214. <http://dx.doi.org/10.1111/1467-8721.00151>
- Michalaszek, A., & Sokolowska, J. (2010). The relative input of payoffs and probabilities into risk judgment. *Polish Psychological Bulletin*, 41, 46–51. <http://dx.doi.org/10.2478/v10059-010-0006-2>
- Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, 54, 1053–1060. <http://dx.doi.org/10.1037/0003-066X.54.12.1053>
- Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, 14, 435–440. <http://dx.doi.org/10.1016/j.tics.2010.07.004>
- Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, 18, 566–570. <http://dx.doi.org/10.1016/j.tics.2014.09.007>
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139173933>
- Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a ‘cooperative phenotype’ that is domain general and temporally stable. *Nature Communications*, 5, 4939. <http://dx.doi.org/10.1038/ncomms5939>
- Pillutla, M. M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39, 448–455. [http://dx.doi.org/10.1016/S0022-1031\(03\)00015-5](http://dx.doi.org/10.1016/S0022-1031(03)00015-5)
- Rand, D. G., & Epstein, Z. G. (2014). Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PLoS ONE*, 9, e109687–e109687. <http://dx.doi.org/10.1371/journal.pone.0109687>
- Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489, 427–430. <http://dx.doi.org/10.1038/nature11467>
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651–665. <http://dx.doi.org/10.1111/j.1467-6494.1967.tb01454.x>
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, 26, 443–452. <http://dx.doi.org/10.1037/h0031464>
- Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology*, 104, 409–426. <http://dx.doi.org/10.1037/a0031050>

- 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.
- Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making, 6*, 733–739.
- Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review, 2*, 87–99. http://dx.doi.org/10.1207/s15327957pspr0202_2
- Seligman, M. E., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science, 8*, 119–141. <http://dx.doi.org/10.1177/1745691612474317>
- Shackelford, T. K., & Buss, D. M. (1996). Betrayal in mateships, friendships, and coalitions. *Personality and Social Psychology Bulletin, 22*, 1151–1164. <http://dx.doi.org/10.1177/01461672962211006>
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118. <http://dx.doi.org/10.2307/1884852>
- Simpson, J. A. (2007). Psychological foundations of trust. *Current Directions in Psychological Science, 16*, 264–268. <http://dx.doi.org/10.1111/j.1467-8721.2007.00517.x>
- Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organization Science, 4*, 367–392. <http://dx.doi.org/10.1287/orsc.4.3.367>
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research, 177*, 1333–1352. <http://dx.doi.org/10.1016/j.ejor.2005.04.006>
- Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology, 78*, 1–18. <http://dx.doi.org/10.1037/h0026468>
- Smith, V. L. (2003). Constructivist and ecological rationality in economics. *The American Economic Review, 93*, 465–508. <http://dx.doi.org/10.1257/000282803322156954>
- 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: Erlbaum.
- Snyder, M. (1983). The influence of individuals on situations: Implications for understanding the links between personality and social behavior. *Journal of Personality, 51*, 497–516. <http://dx.doi.org/10.1111/j.1467-6494.1983.tb00342.x>
- Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8*, 79–98.
- Sturgis, P., Read, S., & Allum, N. (2010). Does intelligence foster generalized trust? An empirical test using the UK birth cohort studies. *Intelligence, 38*, 45–54. <http://dx.doi.org/10.1016/j.intell.2009.11.006>
- Sutter, M., & Kocher, M. G. (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior, 59*, 364–382. <http://dx.doi.org/10.1016/j.geb.2006.07.006>
- Taleb, N. N. (2010). *The black swan: The impact of the highly improbable*. New York, NY: Random House.
- Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General, 142*, 151–162. <http://dx.doi.org/10.1037/a0028232>
- Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality, 50*, 61–65. <http://dx.doi.org/10.1016/j.jrp.2014.03.006>
- Thielmann, I., & Hilbig, B. E. (2015a). Trust: An integrative review from a person–situation perspective. *Review of General Psychology, 19*, 249–277. <http://dx.doi.org/10.1037/gpr0000046>
- Thielmann, I., & Hilbig, B. E. (2015b). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin, 41*, 1523–1536. <http://dx.doi.org/10.1177/0146167215600530>
- Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science, 16*, 167–171. <http://dx.doi.org/10.1111/j.1467-8721.2007.00497.x>
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*, 813–833. <http://dx.doi.org/10.1521/soco.2009.27.6.813>
- van Beest, I., & Williams, K. D. (2006). When inclusion costs and ostracism pays, ostracism still hurts. *Journal of Personality and Social Psychology, 91*, 918–928. <http://dx.doi.org/10.1037/0022-3514.91.5.918>
- Van Lange, P. A. (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. <http://dx.doi.org/10.1037/0022-3514.77.2.337>
- Van Lange, P. A. (2015). Generalized trust: Four lessons from genetics and culture. *Current Directions in Psychological Science, 24*, 71–76. <http://dx.doi.org/10.1177/0963721414552473>
- Vuolevi, J. H., & Van Lange, P. A. (2010). Beyond the information given: The power of a belief in self-interest. *European Journal of Social Psychology, 40*, 26–34.
- Williams, K. D. (2007). Ostracism. *Annual Review of Psychology, 58*, 425–452. <http://dx.doi.org/10.1146/annurev.psych.58.110405.085641>
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592–598. <http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59*, 189–202. <http://dx.doi.org/10.1177/106591290605900202>
- Yamagishi, T., Akutsu, S., Cho, K., Inoue, Y., Li, Y., & Matsumoto, Y. (2015). Two-component model of general trust: Predicting behavioral trust from attitudinal trust. *Social Cognition, 33*, 436–458. <http://dx.doi.org/10.1521/soco.2015.33.5.436>
- Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., . . . Simunovic, D. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes, 120*, 260–271. <http://dx.doi.org/10.1016/j.obhdp.2012.06.002>
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 129–166. <http://dx.doi.org/10.1007/BF02249397>
- Zak, P. J., & Knack, S. (2001). Trust and growth. *The Economic Journal, 111*, 295–321. <http://dx.doi.org/10.1111/1468-0297.00609>
- Zaki, J., & Mitchell, J. P. (2013). Intuitive prosociality. *Current Directions in Psychological Science, 22*, 466–470. <http://dx.doi.org/10.1177/0963721413492764>

Received June 18, 2015

Revision received November 21, 2015

Accepted November 23, 2015 ■