WHO IS TRUSTWORTHY?

PREDICTING TRUSTWORTHY INTENTIONS AND BEHAVIOR

Emma E. Levine[1], T. Bradford Bitterly[2], Taya R. Cohen[3], Maurice E. Schweitzer[2]

[1]The University of Chicago Booth School of Business

[2]The Wharton School, The University of Pennsylvania

[3]Tepper School of Business, Carnegie Mellon University

*Forthcoming in the *Journal of Personality and Social Psychology**

The authors are grateful for the financial support of the Wharton Behavioral Lab and the Russell Ackoff Fellowship of the Wharton Risk Management and Decision Processes Center, as well as the Center for Decision Research and the Charles E. Merrill Faculty Research Fund at the University of Chicago. We are also grateful for research assistance from Soaham Bharti, Ilan Wolff, Sophia Yang, and Catherine O'Donnell.

Address correspondence to Emma E. Levine, The University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60647. E-mail: Emma.Levine@chicagobooth.edu.

WHO IS TRUSTWORTHY?

PREDICTING TRUSTWORTHY INTENTIONS AND BEHAVIOR

ABSTRACT

Existing trust research has disproportionately focused on what makes people more or less trusting, and has largely ignored the question of what makes people more or less *trustworthy.* In this investigation, we deepen our understanding of trustworthiness. Across six studies using economic games that measure trustworthy behavior and survey items that measure trustworthy intentions, we explore the personality traits that predict trustworthiness. We demonstrate that guilt-proneness predicts trustworthiness better than a variety of other personality measures, and we identify sense of interpersonal responsibility as the underlying mechanism by both measuring it and manipulating it directly. People who are high in guilt-proneness are more likely to be trustworthy than are individuals who are low in guilt-proneness, but they are not universally more generous. We demonstrate that people high in guilt-proneness are more likely to behave in interpersonally sensitive ways when they are more responsible for others' outcomes. We also explore potential interventions to increase trustworthiness. Our findings fill a significant gap in the trust literature by building a foundation for investigating trustworthiness, by identifying a trait predictor of trustworthy intentions and behavior, and by providing practical advice for deciding in whom we should place our trust.

Abstract word count: 195

Key words: trust, trustworthiness, guilt-proneness, personality, responsibility, Rely-or-Verify game, trust game

Trust is critical for effective organizational performance (Dirks & Ferrin, 2001; Jones & George, 1998; Kramer, 1999; Salamon & Robinson, 2008) and interpersonal functioning (Rempel, Holmes, & Zanna, 1985; Simpson, 2007). Trust increases leadership efficacy and organizational commitment, improves job satisfaction, and decreases conflict (Dirks & Ferrin, 2002; Zaheer, McEvily, Perrone, 1998). Trust also promotes positive perceptions of one's relationships (Luchies, et al., 2013; Rempel, Ross, & Holmes, 2001) and increases forgiveness after interpersonal transgressions (Molden & Finkel, 2010). Accordingly, scholars have argued that trust "may be the single most important ingredient for the development and maintenance of happy, well-functioning relationships" (Simpson, 2007, p. 264). Because of trust's central role in social life, a substantial body of research has investigated when and why individuals decide to trust others (e.g., Dunn, Ruedy, & Schweitzer, 2012; Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014; Kim, Ferrin, Cooper, & Dirks, 2004; Lewicki, & Bunker, 1995; Lewicki, McAllister, & Bies, 1998; Lount, 2010; Lount & Pettit, 2012; Lount, Zhong, Sivanathan, & Murnighan, 2008; Pillutla, Malhotra, & Murnighan, 2003; Tomlinson & Mayer, 2009).

In addition to promoting beneficial outcomes, however, trust can facilitate exploitation (Yip & Schweitzer, 2015). It is only when trust is well placed, in targets who are trustworthy, that trust yields substantial benefits. Surprisingly, existing trust research provides surprisingly little insight into *whom* to trust. Rather than examining the actual qualities that make a person *trustworthy*, prior investigations in economics, organizational behavior, and social psychology have focused largely on what makes people more or less *trusting*. In other words, much existing trust literature has deeply explored only one side of an inherently dyadic relationship.

In this paper, we advance our understanding of trust and trustworthiness in several ways. First, we shift the focus of trust research to *trustworthiness.* We draw on research from

criminology, clinical psychology, and personality psychology to identify a robust individual-level predictor of trustworthy intentions and behavior: guilt-proneness. Importantly, we find that guilt-proneness predicts trustworthiness better than more commonly examined personality traits (i.e., the Big Five). We also document interpersonal responsibility as the underlying mechanism. By measuring interpersonal responsibility and manipulating it directly, we find that interpersonal responsibility mediates the relationship between guilt-proneness and trustworthiness. Across our studies, we introduce two new measures to assess trustworthiness: a self-report measure of trustworthy intentions and a behavioral measure of integrity-based trustworthiness. By developing our understanding of the qualities that predict trustworthiness, we make an important theoretical contribution to our understanding of trust and offer practical insights into how to curtail the risk of misplacing trust.

Trust and Trustworthiness

We conceptualize trust as the willingness to be vulnerable to exploitation within a social interaction (e.g. Lewicki & Bunker, 1995; Rousseau et al., 1998). Decades of research within economics, behavioral decision theory, organizational behavior, and social psychology have examined factors that make trusters more or less trusting. Prior research has conceptualized the decision to trust others as a function of characteristics of the truster, situational factors, and perceptions of the trustee. For example, a truster's likelihood of trusting others is influenced by attributes such as their status (Lount & Pettit, 2012), their fear of social exclusion (Derfler-Rozin, Pillutla, & Thau, 2010), their nationality (Özer, Zheng, & Ren, 2014; Özer & Zheng, 2017), their gender (Buchan, Croson, & Solnick, 2008; Larrick, 2016), and their dispositional propensity to trust (Mayer, Davis, & Schoorman, 1995). Similarly, situational factors, such as the presence of sanctions and monitoring systems (Malhotra & Murnighan, 2002; Mulder, Van Dijk,

De Cremer, & Wilke, 2006; Schweitzer & Ho, 2005; Schweitzer, Ho, & Zhang, 2016) and a truster's incidental emotions (Dunn & Schweitzer, 2005; Lount, 2010), influence trusting decisions. Recent research also demonstrates that individuals' trusting decisions are often motivated by a sense of social duty (Dunning et al., 2014; Dunning, Fetchenhauer, & Schlösser, 2012).

A substantial literature has also explored perceptions of trustworthiness. The dominant paradigm for understanding perceptions of trustworthiness is Mayer, Davis and Schoorman's (1995) ability, benevolence, and integrity (ABI) model. According to the ABI model, individuals are most likely to trust people whom they perceive as having high ability (intelligent, competent, capable), high benevolence (kind, caring, empathic), and high integrity (consistent, principled, and ethical). Trusters make judgments about ability, benevolence, and integrity by drawing on a variety of personal, social, and situational cues, such as whether a trustee has previously broken a promise (Schweitzer, Hershey, & Bradlow, 2006), whether a trustee has an ulterior motive or a conflict of interest (Sah, Loewentstein, & Cain, 2011, 2013), and whether a trustee apologizes for or denies a potential transgression (De Cremer, van Dijk, & Pillutla, 2010; Kim et al., 2006; Kim et al., 2004; Brooks, Dai, & Schweitzer, 2014; Schweitzer, Brooks & Galinsky, 2015).

A substantial trust literature examines trustworthiness from the perspective of the truster and conceptualizes trustworthiness as a *perception* that inspires trust (Mayer et al., 1995; Tomlinson & Mayer, 2009; Whitener, Brodt, Koresgaard, & Werner, 1998). This research offers fundamental insights into the cues and behaviors that engender trust. However, trust scholars have largely overlooked trustworthiness and important questions remain with respect to the traits and qualities that predict *actual* trustworthiness. That is, prior work in the trust literature offers little guidance with respect to the likelihood that trust will be honored or exploited once it has

been conferred. In Figure 1, we propose a more complete theoretical model for the study of interpersonal trust, which highlights the importance of studying trust from both the truster's and trustee's perspective.

--Figure 1 here--

Although prior work has largely characterized trust as a beneficial force that is essential for establishing a variety of positive outcomes ranging from stable political exchange (Hosmer, 1995) to effective leadership (Dirks & Ferrin, 2002) and flourishing marriages (Finkel, 2017; Luchies, et al., 2013; Miller & Rempel, 2004), trust is not intrinsically helpful. By trusting the wrong people (those who in reality are untrustworthy), individuals *misplace* their trust and risk exploitation (Yip & Schweitzer, 2015). Conversely, by failing to trust people who are actually trustworthy, individuals fail to realize the joint gains of mutual trust and trustworthiness. In the present research, we shift focus from the perception of trustworthiness to the actual quality of trustworthiness. In doing so, we build a more complete understanding of the dyadic nature of trusting interactions.

We define trustworthiness as *the propensity to fulfill another's positive implicit or explicit expectations regarding a particular action.* Trust and trustworthiness are related, but distinct, constructs (Ashraf, Bohnet, & Piankov, 2006; Ben-Ner & Halldorsson, 2010; Colquitt, Scott, & LePine, 2007; Glaeser, Laibson, Scheinkman, & Soutter, 2000; Hardin, 2004). Trust reflects the truster's willingness to be vulnerable based on positive expectations of the trustee, whereas trustworthiness reflects the trustee's propensity to fulfill those positive expectations. Therefore, being trustworthy requires recognizing that another party has expectations, and feeling responsible for fulfilling those expectations (Salamon & Robinson, 2008). We conceptualize positive expectations as desirable from the perspective of the truster, but note that

trusters may expect others to engage in trustworthy but unethical behavior. For example, a boss may trust their employee to keep a secret about misconduct that endangers an organization.

Trustworthiness is both a trait and a state construct. We define trait level trustworthiness as *the general propensity to fulfill others' positive expectations, across time and circumstances*. We define state level trustworthiness as *the fulfillment of a specific person's (the truster's) positive implicit or explicit expectations regarding a particular action*. State-level trustworthiness is a behavior that emerges as a response to a specific act of trust and can be elicited by anyone facing the temptation to violate trust. Just as past research distinguishes between an individual's general propensity to trust others (trait-level trust) and an individual's willingness to be vulnerable to a specific party based on expectations of trustworthiness (state-level trust; Mayer et al., 1995), we distinguish between trait and state trustworthiness.

Importantly, our definition of trustworthiness also differs from prior work that has conceptualized trustworthiness as a calculative reaction to trusting behavior (e.g., Ashraf et al., 2006; Buchan et al., 2008; Glaeser et al., 2000; Malhotra, 2004; Pillutla et al., 2003). Trustworthy actions are not always characterized by calculated reciprocity. Many acts of trust, such as trusting someone to mail an important letter, keep a secret, or give sound advice, fulfill a truster's positive expectations without requiring an initial generous act that might trigger reciprocity. In addition, different individuals may be far more or less trustworthy in different contexts even when they are similarly trusted.

Predictors of Trustworthiness

Relatively few scholars have investigated the correlates of trustworthiness, and those who have, have primarily examined the relationship between the HEXACO and Big Five personality traits (extraversion, openness, agreeableness, neuroticism, conscientiousness, and honesty-

humility; Costa & McCrae, 1992; John, Donahue, & Kentle, 1991; Thielmann & Hilbig, 2015) and return behavior in the trust game (see Zhao & Smillie, 2015, for review). Although honesty-humility has been linked with return behavior in the trust game, recent research demonstrates that this relationship is driven by the association between honesty-humility and unconditional kindness (Thielmann & Hilbig, 2015). Individuals who are high in unconditional kindness give and return more money in economic games, such as the dictator game and trust game, regardless of the behavior of their counterpart (Ashraf, Bohnet, & Piankov, 2006; Thielmann & Hilbig, 2015). This result suggests that honesty-humility predicts generosity, rather than trustworthiness per se. In addition, agreeableness, which captures one's tendency towards cooperation and concern for social relationships, is the dimension of the Big Five that has most frequently been linked to trustworthiness in the trust game (Becker, Deckers, Dohmen, Falk, & Kosse, 2012; Ben-Ner & Halldorsson, 2010; Evans & Revelle, 2008; Lönnqvist, Verkasalo, Wichardt, & Walkowitz, 2012). However, in several studies, the relationship between agreeableness and trustworthiness has been weak (e.g., Evans & Revelle, 2008) or was only significant when combined with other traits (e.g., low neuroticism; Lönnqvist, Verkasalo, Wichardt, & Walkowitz, 2012). Thus, it remains unclear whether agreeableness is truly a robust predictor of trustworthy behavior in the trust game.

Although trust and trustworthiness have primarily been studied by economists, social psychologists, and organizational scholars, it is important to note that research in criminology, personality psychology, and clinical psychology has also explored links between individual differences and behaviors that are likely to be related to trustworthiness. In criminology, scholars have linked Criminogenic Cognitions (notions of entitlement, failure to accept responsibility, short-term orientation, insensitivity to impact of crime, and negative attitudes toward authority)

with criminal activity and the ability to rationalize deviant and unethical behavior (Tangney, Mashek, & Stuewig, 2007; Tangney, Stuewig, Furukawa, Kopelovich, Meyer, & Cosby, 2012). In personality psychology, scholars have linked the Dark Triad – Machiavellianism, subclinical narcissism, and subclinical psychopathy – with selfishness (Hodson et al., 2018; Muris, Merckelbach, Otgaar, & Meijer, 2017), manipulation and lying (Jakobwitz & Egan, 2006; Jones & Paulhus, 2017; Kashy & DePaulo, 1996), and low remorse (Paulhus, & Williams, 2002). In clinical psychology, scholars have investigated the extreme and clinical forms of the Dark Triad, finding that these traits predict violence, crime, and even murder (Barkataki, Kumari, Das, Taylor, & Sharma 2006; Eronen, Angermeyer, & Schulze, 1998; Muris, Merckelbach, Otgaar, & Meijer, 2017). These individual differences and behaviors are likely to be correlated with trustworthiness, but they are conceptually distinct from trustworthiness. Trustworthiness requires recognizing another individual's expectations and feeling a sense of responsibility to fulfill those expectations. In contrast, the Dark Triad reflects self-interest, insensitivity to punishment, and the propensity to exploit others. These personality traits neither reflect the ability to recognize others' expectations or feel a sense of responsibility to fulfill those expectations. For example, people who are high in Machiavellianism may choose untrustworthy actions, but this is likely because they are motivated to pursue their self-interest, rather than their preference for undermining or fulfilling others' expectations (Gunnthorsdottir, McCabe, & Smith, 2002). In fact, individuals high in Machiavellianism may choose trustworthy actions when demonstrating trustworthiness advances their self-interest. That is, the Dark Triad traits predict antisocial and selfish behavior broadly, rather than untrustworthiness specifically. In the present work, we examine a facet of personality that is likely to be uniquely related to trustworthy intentions and behavior: guilt-proneness.

Guilt-Proneness

Guilt-proneness is a facet of personality that directly relates to one's sense of interpersonal responsibility (Schaumberg & Flynn, 2012). Thus, we hypothesize that guilt-proneness is particularly well-suited for predicting trustworthy intentions and behavior. Guilt-proneness is correlated with, but is distinct from honesty-humility and the Dark Triad (Cohen, Panter, Turan, Morse, & Kim, 2014). In addition, guilt-proneness is related to, but distinct from the emotion of guilt. Guilt is a negative, self-conscious emotion that is evoked in response to wrongdoing (Bohns & Flynn, 2012; Cohen et al., 2011; De Hooge, Zeelenberg, Breugelmans, 2007; De Hooge, Nelissen, Breugelmans, Zeelenberg, 2011; Grant & Wrzesniewski, 2010; Tangney, 1996; Tangney & Dearing, 2002; Tangney et al., 2007). When people experience guilt, they focus on the transgression and become motivated to repair the perceived harm caused by their transgression (De Hooge et al., 2011). Guilt is functional insofar as it protects and restores social relationships (Baumeister, Stillwell, & Heatherton, 1994; De Hooge et al., 2011).

The experience of guilt—the emotion—elicits reparative behavior following a transgression. In contrast, guilt-proneness—the individual difference that captures the *anticipation* of guilt over wrongdoing—causes people to avoid transgressing in the first place (Cohen et al., 2012; Tangney et al., 2007). Specifically, individuals who anticipate feeling guilty over wrongdoing (i.e., those with high levels of guilt-proneness) avoid norm violations, such as taking credit for a colleague's work, that would cause them to feel guilt.

We expect guilt-proneness to predict trustworthiness across time and contexts because highly guilt-prone individuals generally have a strong sense of interpersonal responsibility (Cohen et al., 2012; Schaumberg & Flynn, 2012; Tangney et al., 2007; Wiltermuth & Cohen, 2014). As a result, they tend to work harder in organizations (Flynn & Schaumberg, 2012) and

emerge as effective leaders (Schaumberg & Flynn, 2012). In contrast, individuals with low levels of guilt-proneness are more likely than others to engage in unethical and even criminal behaviors (Cohen, Panter et al., 2014; Stuewig et al., 2015; Tangney, Stuewig, & Martinez, 2014). In light of these prior findings, we expect individuals high in guilt-proneness to be particularly trustworthy.

Importantly, we expect that highly guilt-prone individuals will not be unconditionally prosocial. Rather, we expect them to be particularly prosocial when others are relying on them. Individuals who are high in guilt-proneness seek to avoid disappointing others (Wiltermuth & Cohen, 2014), and are particularly sensitive to others' expectations (Pinter et al., 2007; Wildschut & Insko, 2006). This helps to explain why guilt-proneness may be more related to trustworthiness than other traits. By trusting, trusters make themselves vulnerable to a trustee and signal their expectations. Someone who is high in guilt-proneness is likely to feel particularly responsible for meeting those expectations, for example, by returning money in a trust game. However, absent expectations to return money, we do not expect highly guilt-prone individuals to be more likely to return money than individuals who are low in guilt-proneness.

We expect the relationship between guilt-proneness and trustworthiness to operate through both trait and state processes. At the trait-level, we expect guilt-proneness to elicit trustworthy intentions across time and circumstances because guilt-prone individuals generally feel responsible towards others and are averse to letting others down. At the state-level, we expect guilt-proneness to cause individuals to feel a greater sense of interpersonal responsibility when specific opportunities to violate trust arise, and thus be more likely to engage in trustworthy behavior. Notably, no prior research has explored the potentially important link between guilt-proneness and trustworthiness.

Overview of Research

We establish the relationship between guilt-proneness and trustworthiness across six studies. In Study 1, we demonstrate that guilt-proneness is associated with trustworthy intentions across time and situations. In Studies 2 and 3, we demonstrate that guilt-proneness predicts both benevolence- and integrity-based trustworthy behavior better than the Big Five. We also begin to test the mediating role of interpersonal responsibility. In Study 4, we test this mechanism more precisely and rule out several alternative mechanisms. In Study 5, we manipulate interpersonal responsibility directly and demonstrate that guilt-proneness predicts trustworthiness, but not generosity broadly. In Study 6, we explore whether codes of conduct might increase one's sense of responsibility, and therefore trustworthiness, among both those high and those low in guilt-proneness.

Across our studies, we establish a robust relationship between guilt-proneness and trustworthiness. We employ diverse samples and we use multiple measures of both guilt-proneness and trustworthiness. This approach ensures that we capture a fundamental individual difference and it enables us to make inferences about different types of trustworthiness. In Studies 1, 2, 4, 5 and 6, we use a standard trust game to capture benevolence-based trustworthiness. The trust game primarily captures benevolence because trustworthy behavior in the game entails returning money and honoring the trustor's expectation that the trustee will behave kindly and generously. In Study 3, we use the Rely-or-Verify game to capture integrity-based trustworthiness (Levine & Schweitzer, 2015). The Rely-or-Verify game captures integrity, in addition to benevolence, because trustworthy behavior in the game entails telling the truth, thereby honoring the trustor's expectation that the trustee will behave honestly. We measure trait-level trustworthiness in Study 1 (in addition to examining trust game behavior) by

examining one's intentions to engage in a variety of trustworthy behaviors across time and situations. In contrast to much existing research that has examined how personality influences a broad range of harmful and helpful behaviors, we disentangle trustworthiness from generosity in Study 5. Our approach allows us to assess the robustness and the specificity of the relationship between guilt-proneness and trustworthiness. In each study, we determined our sample size in advance, and we report all variables and conditions collected. All data, syntax, and materials are available on the Open Science Framework at: https://osf.io/k8vt9/.

Study 1: Guilt-proneness and trustworthiness

In Study 1, we document the relationship between guilt-proneness and a newly developed self-report instrument for assessing individuals' trustworthy intentions (i.e., trait trustworthiness). Our measure of trustworthy intentions captures the intention to fulfill others' expectations across a range of trust dilemmas. We provide convergent evidence that guilt-proneness predicts trustworthiness using both the trust game and the newly developed trustworthy intentions scale.

Participants

We set the a priori goal of recruiting 400 U.S. adults to participate in an online study via Amazon Mechanical Turk in exchange for a small payment ($0.45). Participants also earned bonus money based upon their choices. We ultimately ended up with 401 participants (34% female; M_{age} = 33 years, SD = 11.36; $M_{\text{work experience}}$ = 12.29 years, SD = 10.45) who completed the entire study and were eligible for analysis.

Procedure and Materials

Participants responded to a two-part questionnaire. Part one included self-report measures of guilt-proneness and trustworthy intentions and part two included the trust game. We

randomized the order in which participants responded to the different parts of the questionnaire. We also randomized the order in which guilt-proneness and trustworthy intentions were collected within the self-report section of the questionnaire (part two). Adding main and interaction effects of order had no bearing on our key results, and thus we do not discuss order effects further.

Guilt-proneness. We administered the Guilt and Shame Proneness (GASP) scale (Cohen et al., 2011) to assess guilt-proneness. The GASP is a scenario-based measure of guilt-proneness and shame-proneness, which differentiates negative evaluations of one's behavior in response to private transgressions (Guilt-Negative-Behavior-Evaluation), repair behaviors in response to private transgressions (Guilt-Repair), negative self-evaluations in response to public transgressions (Shame-Negative-Self-Evaluation), and withdrawal behaviors in response to public transgressions (Shame-Withdraw). There are four items for each of the four subscales in the GASP. For each item, participants read a short scenario and a possible response to that scenario. Participants then rate how likely they would be to experience that response (1 = *very unlikely* and 7 = *very likely*). For example, one item of the Guilt-Negative-Behavior-Evaluation subscale begins by asking participants to imagine, "You secretly commit a felony" and then asks them, "What is the likelihood that you would feel remorse about breaking the law?" Although the Guilt-Negative-Behavior-Evaluation, Guilt-Repair, and Shame-Negative-Self-Evaluation subscales are moderately to highly correlated with one another, they represent conceptually different responses to transgressions and factor analyses show that they are distinct factors (Cohen et al., 2011). To address potential multicollinearity issues among the subscales, we followed the guidelines outlined by Cohen et al. (2011) and analyzed the four subscales separately in regression analyses.

Trustworthiness. We administered an eight-item trustworthy intentions scale. To create this scale, we adapted items from the Johnson-George and Swap (1982) trust scale so that the items captured the perspective of the trustee rather than the truster. These items capture an individual's intentions to fulfill others' expectations across a variety of activities, such as borrowing money, doing a favor, and providing honest information. For example, we asked participants to rate their agreement with the statement, "If I promised to do a favor for someone, I would follow through" (1 = *never* and 11 = *absolutely all of the time*). Importantly, these items do not refer to a specific truster or circumstance, and thus, reflect trait-level trustworthy intentions. The Appendix describes our scale development analyses and results of exploratory factor analyses. We include the full scale in Table A1 of the Appendix.

Trustworthy behavior. We use the second player's decision in the trust game to measure benevolence-based trustworthy behavior (adapted from Berg, Dickhaut, & McCabe, 1995). In the trust game, the truster (Player 1) is endowed with an amount of money and has the decision to either keep the money for themselves, or pass it to the trustee (Player 2). If the truster passes the money, the amount multiplies and the trustee has the opportunity to keep it all or return some of it to the truster. By passing the money, the truster makes themselves vulnerable to the trustee. Player 1's behavior reflects trust, and Player 2's decision to return money reflects trustworthy behavior (e.g., Buchan et al., 2008; Glaeser et al., 2000). Although this behavior also reflects reciprocity (e.g., Pillutla, Malhotra, & Murnighan, 2003), it captures benevolence-based trustworthiness because it represents the fulfillment of the truster's expectations that the trustee will benevolently share the money with the sender.

We told participants that they were randomly assigned to their roles. In fact, we assigned all participants to the role of Player 2 and paired them with a confederate Player 1. In our version

of the trust game, Player 1 (the confederate) was given $1, which they always chose to pass. That is, we held trusting behavior constant. After Player 1 passed the $1, the amount of money grew to $2.50 and Player 2 could then choose to either "Keep $2.50" or "Return $1.25". Player 2's decision served as our behavioral measure of trustworthiness.

All participants learned about the trust game and had to pass a comprehension check for the game before playing it. A total of 3% of initial participants failed the comprehension check after two attempts. These participants were automatically kicked out of the study; they did not provide complete data and do not appear in any analyses.[1] At the end of the study, we collected demographic information and asked participants their thoughts on the purpose of the study.[2] Participants then received a bonus payment based upon their decisions in the trust game.

Results

For every study, we present the means, standard deviations, and reliability of each of our measures, as well as the bivariate correlations among all of the measures (see Table 1). Across all studies, we standardized all individual difference variables when performing regression analyses to improve the ease of interpretation and we present the raw means of each variable in the correlation matrix.

We use logistic regression to examine the effects of guilt-proneness on trustworthiness; the dependent variable was coded as 1 = returned money in the trust game (i.e. trustworthy behavior) and 0 = kept money in the trust game (i.e. untrustworthy behavior). In this study, and

[1] We followed the same procedure for all studies, such that participants who failed the comprehension check more than once were not eligible to complete the study. A total of 14%, 10.5%, 1.6%, 11.6%, and 4% of initial participants in Studies 2, 3, 4, 5, and 6, respectively, failed the comprehension check, and thus are not included in any analyses.

[2] We used this question to check for suspicion. We had two research assistants code responses to this question in every study in which it was asked (Studies 1, 2, 3, 5, and 6). We find that 0.5% of participants expressed suspicion in Study 1, 0.7% in Study 2, 1.5% in Study 3, 0% in Study 5 and 0.7% in Study 6. Screening out these participants does not influence our results.

all subsequent studies, we present odds ratios *(e^b)* as an indicator of effect sizes in our logistic regressions.

First, we analyzed the relationship between the GASP and our trustworthiness intentions measure (Table 2, Panel A). Individual regression analyses, with each subscale of the GASP entered as independent predictors, demonstrate that both Guilt-Negative-Behavior-Evaluation, $b = .57$, $SE = .06$, $p < .001$, and Guilt-Repair, $b = .69$, $SE = .06$, $p < .001$, are positively associated with trustworthy intentions. These results hold when we enter all GASP subscales as simultaneous predictors (see Table 2, Panel A, Model 5), demonstrating that the cognitive and behavioral elements of guilt-proneness are both uniquely associated with greater trustworthiness.

-- Tables 1 and 2 about here –

Next, we look at the relationship between the GASP and trustworthy behavior in the trust game (Table 2, Panel B). We find that trustworthy behavior is positively predicted by Guilt-Negative-Behavior-Evaluation, $b = .49$, $SE = .11$, $p < .001$, $e^b = 1.64$, and Guilt-Repair, $b = .47$, $SE = .11$, $p < .001$, $e^b = 1.60$. The odds ratios indicate that the likelihood of returning money was 1.64 times as high for those high in Guilt-Negative-Behavior-Evaluation (+1 standard deviation) as compared to those of average Guilt-Negative-Behavior-Evaluation and that that the likelihood of returning money was 1.60 times as high for those high in Guilt-Repair (+1 standard deviation) as compared to those of average Guilt-Repair.

These results hold when we enter all GASP subscales as simultaneous predictors (see Table 2, Panel B, Model 5), again showing that the cognitive and behavioral elements of guilt-proneness are both associated with greater trustworthiness. In other words, individual differences in Guilt-Negative-Behavior-Evaluation, the tendency to anticipate making negative evaluations of one's behavior following private transgressions, and Guilt-Repair, the tendency to anticipate

engaging in repair behaviors following private transgressions, are both associated with trustworthy intentions and behavior.

For illustrative purposes, we depict the frequency of trustworthy behavior at high (top quartile) and low (bottom quartile) levels of Guilt-Negative-Behavior-Evaluation and Guilt-Repair in Figure 2. Individuals with guilt-proneness scores in the upper quartile of Guilt-Negative-Behavior-Evaluation were significantly more likely to return money in the trust game (65.09%, $n = 106$) than were individuals with guilt-proneness scores in the lower quartile of Guilt-Negative-Behavior-Evaluation (36.03%, $n = 111$; $\chi^2 = 18.31$, $p < .001$). Similarly, individuals with guilt-proneness scores in the upper quartile of Guilt-Repair were significantly more likely to return money in the trust game (63.89%, $n = 72$) than were individuals with guilt-proneness scores in the lower quartile of Guilt-Repair (31.07%, $n = 103$; $\chi^2 = 18.48$, $p < .001$).

--- Figure 2 about here ---

Discussion

In Study 1, we provide initial evidence for the link between guilt-proneness and trustworthiness. Compared to those low in guilt-proneness, highly guilt-prone individuals are more likely to hold trustworthy intentions across time and situations, and they are more likely to act in a trustworthy manner in the trust game. In addition to providing compelling evidence in support of our hypothesis, Study 1 also makes an important and novel contribution by introducing a scale to measure trustworthy intentions (i.e., trait-level trustworthiness).

Studies 2 and 3. Guilt-proneness versus The Big-5, and an Initial Test of the Sense of Interpersonal Responsibility Mechanism

In Studies 2 and 3, we replicate the relationship between guilt-proneness and trustworthy behavior and further demonstrate that guilt-proneness predicts trustworthy behavior above and

beyond the Big-5 personality traits. We also explore the mechanisms linking guilt-proneness and trustworthy behavior. We propose that compared to individuals low in guilt-proneness, highly guilt-prone individuals are more attuned to how their behavior affects others and consequently have a heightened sense of interpersonal responsibility for those who trust them. However, it is also possible that the visceral anticipation of guilt and the desire to avoid this negative affective state predicts trustworthiness. We compare the validity of these two mechanisms in Study 2 and find that only sense of interpersonal responsibility mediates the relationship between guilt-proneness and trustworthy behavior.

Study 2. Benevolence-based trustworthy behavior

Participants

We recruited adults to participate in a laboratory study at a U.S. university in exchange for a $10 show-up payment and a bonus payment based on their decisions. We made the a priori decision to recruit as many participants as we could during a three-day laboratory session. We ultimately ended up with 139 participants (70% female; M_{age} = 21 years, SD = 3.97; $M_{work\ experience}$ = 2 years, SD = 3.37) who completed the entire study (the pre-laboratory session survey and the laboratory study) and were eligible for analysis.

Procedure and Materials

Personality measures. We informed participants that they had to complete a quick survey the evening before their session to be eligible for the laboratory session. Each participant received the survey the evening before their laboratory session and was required to complete it before arriving to the laboratory.

The survey contained the Five-Item Guilt-Proneness Scale (GP-5; Cohen, Kim, & Panter, 2014) and the Ten-Item Personality Inventory (Gosling, Rentfrom, & Swann, 2003). The GP-5

includes the four items from the Guilt-Negative-Behavior-Evaluation subscale of the Guilt and Shame Proneness (GASP) scale (Cohen et al., 2011), plus one additional item that was added to increase the reliability of the measure: "Out of frustration, you break the photocopier at work. Nobody is around and you leave without telling anyone. What is the likelihood you would feel bad about the way you acted?" Participants responded to the GP-5 items with a five-point rating scale anchored by 1 = *extremely unlikely* and 5 = *extremely likely* (α = .70).[3] The TIPI captures the Big Five factors: Agreeableness, Emotional Stability, Extraversion, Conscientiousness, and Openness to Experience. Each factor is measured with two items containing a pair of traits (e.g., sympathetic and warm)—respondents indicate the extent to which each pair of traits applies to them. We used the GP-5 and TIPI scales, which are both brief, to limit participant burden and increase compliance with the evening survey. However, because guilt-proneness is focal to the current investigation, we also replicated our main results using the Test of Self-Conscious Affect-3 (TOSCA-3; Tangney, Dearing, Wagner, & Gramzow, 2000), which we report in detail in the online supplemental materials, and using a longer Big Five inventory in Study 3.

Trustworthy behavior. When participants arrived at the behavioral laboratory, we seated them in individual cubicles and had them complete a trust game. As in Study 1, all participants were assigned to the role of the trustee (Player 2), unbeknownst to them.

The trust game we used in Study 2 was similar to the game used in Study 1, except we used lottery tickets rather than monetary payments. In Study 2, Player 1 (the confederate) was given 2 lottery tickets, which would increase to 6 tickets if Player 1 chose to pass them to Player 2. Player 2 could then either choose to either "Keep 6 tickets" or "Return 3 tickets." Each lottery ticket entered the participants into a lottery for a $75 bonus.

[3] For information on how the GP-5 relates to other personality and moral character traits, see Study 3 in Cohen, Panter, Turan, Morse, & Kim (2014).

In Study 2, we asked participants what they would like to do *if* Player 1 passed. That is, we used the strategy method and had participants indicate their preferred course of action before knowing what Player 1's actual choice was. Using this method helps to disentangle trustworthiness from reciprocity. Player 2's decision served as our behavioral measure of trustworthiness.

Sense of interpersonal responsibility. We also asked several questions to assess our proposed mechanism: sense of interpersonal responsibility. Specifically, participants rated their agreement with four statements (1 = *strongly disagree*, 7 = *strongly agree*; $\alpha = .91$): "I feel accountable for my partner's earnings", "I feel a sense of responsibility towards my partner", "I care whether or not my partner earns lottery tickets", and "I feel an obligation to act responsibly towards my partner in the Choice Game." We adapted these items from Salamon and Robinson's (2008) Responsibility Norms scale. These items reflect one's sense of interpersonal responsibility towards a specific person (their partner) during a particular interaction (the trust game).

We counterbalanced the order in which participants responded to these items and made their decision in the trust game. That is, half of the participants rated their sense of interpersonal responsibility towards their partner and then played the trust game, and half of the participants played the trust game and then rated their sense of interpersonal responsibility. Trustworthy behavior and sense of interpersonal responsibility were not influenced by order ($ps > .37$), so we do not discuss order effects further.

Anticipated guilt. We also sought to examine if the affective anticipation of guilt mediated our effects. After participants played the trust game, we asked participants two questions about how much they were thinking about the guilt they might experience when

making their decision in the trust game. We adapted these items from Wiltermuth and Cohen (2014). Specifically, participants indicated the extent to which they were thinking about "How guilty you would feel if you decided to keep all the lottery tickets and your partner decided to pass money to you" and "How bothered you would be if your partner earned 0 lottery tickets"; 1 = *not at all* and 7 = *extremely*. We combined these items to create a measure of anticipated guilt, $r(139) = .76$, $p < .01$.

At the end of the study, we collected demographic information and asked participants what they thought the purpose of the study was. We conducted the lottery one week after the study concluded, and we paid one participant the $75 bonus.

Results

In Table 3, we present the raw means of each variable and the bivariate correlations among the measures, and in Table 4 we present logistic regressions using standardized values of these variables. As indicated in Table 3, none of the Big Five variables measured by the TIPI correlate with trustworthy behavior. However, as indicated in Table 4, we find that guilt-proneness is positively associated with trustworthy behavior, $b = .36$, $SE = .18$, $p = .048$, $e^b = 1.43$ (see Table 4, Model 1). In other words, the likelihood of returning money was 1.43 times as high for those high in guilt-proneness (+1 standard deviation) as compared to those of average guilt-proneness. For illustrative purposes, we depict the frequency of trustworthy behavior at high and low levels of guilt-proneness, which we depict as the lower and upper guilt-proneness quartiles in the sample (see Figure 3). Individuals with guilt-proneness scores in the upper quartile of the sample were marginally more likely to return money in the trust game (54.16%, $n = 48$) than were individuals with guilt-proneness scores in the lower quartile of the sample (35.00%, $n = 40$; $\chi^2 = 3.23$, $p = .072$).

We also find that guilt-proneness significantly correlates with sense of interpersonal responsibility (r = .21, p = .013), but not anticipated guilt (r = .14, p = .11, see Table 3). In light of prior findings and the direction of the correlation, we suspect that anticipated guilt is indeed associated with guilt-proneness (cf. Wiltermuth & Cohen, 2014). Indeed we find a significant correlation between guilt-proneness and anticipated guilt in Study 4.

We used the bootstrap procedure with 10,000 samples to formally test whether sense of interpersonal responsibility mediates the relationship between guilt-proneness and trustworthy behavior (SPSS Process Macro, Model 4, Hayes, 2013). The mediation model included guilt-proneness (GP-5) as the independent variable, sense of interpersonal responsibility as the mediator, and trustworthy behavior as the dependent variable. We find evidence of significant mediation through sense of interpersonal responsibility (Indirect effect = .31, SE = .17, 95% CI [.02, .70]). Once we control for sense of interpersonal responsibility in the model, the relationship between guilt-proneness and trustworthiness becomes non-significant, suggesting full mediation (see Table 4, Model 4).

-- Tables 3 and 4, Figure 3 about here –

Study 3. Integrity-based trustworthy behavior

In Study 3, we extend our investigation by exploring the relationship between guilt-proneness and a measure of integrity-based trustworthiness. We use the Rely-or-Verify game (Levine & Schweitzer, 2015) to operationalize integrity-based trustworthiness.

The Rely-or-Verify game captures two behaviors. First, it assesses the trustee's decision to tell the truth or attempt to exploit the truster by lying. Second, the Rely-or-Verify game assesses the truster's subsequent decision to either rely upon or verify the trustee's claim. The Rely-or-Verify game has the following features: 1) the trustee benefits from successfully

deceiving the truster, 2) the truster can only ascertain the truth by verifying the trustee's claim, 3) relying on the trustee's claim entails risk to the truster, and 4) verifying the trustee's claim is costly. Although the trustee's decision to send an accurate claim in the Rely-or-Verify game reflects benevolent intent towards the trustee, similar to the trust game, it also uniquely captures the tendency to be honest and act with integrity, unlike the trust game. Thus, we interpret behavior within the game as a measure of integrity-based trustworthiness.

We also strengthen our investigation by using a more reliable measure of the Big-5, the full 44-item Big-5 Personality Inventory, and by recruiting a larger sample size, thereby increasing our statistical power relative to Study 2.

Participants

We set the a priori goal of recruiting 400 U.S. adults to participate in an online study via Amazon Mechanical Turk in exchange for a small payment ($0.70). Participants also earned bonus money based upon their decisions. We ultimately ended up with 399 participants (50% female; M_{age} = 35 years, SD = 10.69, $M_{\text{work experience}}$ = 14.1 years, SD = 10.02) who completed the entire study and were eligible for analysis.

Procedure and Materials

We told participants that they would complete two unrelated studies in exchange for payment. We presented one of the studies, "Study W," as a decision-making game (the Rely-or-Verify game), and we presented the other study, "Study L," as a personality questionnaire (guilt-proneness and Big-5 measures). We randomized whether participants completed the decision-making game first or the personality questionnaire first. The order in which participants completed the experiment did not influence our results (ps > .79), and thus we do not discuss order effects further.

The Rely-or-Verify game. "Study W" began with the Rely-or-Verify Game. In our version of Rely-or-Verify, we referred to the trustee as the Red Player and we referred to the truster as the Blue Player. In the game, the Red Player (the trustee) makes the first move; they decide to send a claim that is either *accurate* (trustworthy) or *inaccurate* (untrustworthy). Then, the Blue Player observes the claim, and not knowing the true state of the world, decides to either *Rely* (trust) or *Verify* (not trust) the claim.

In our version of the Rely-or-Verify game, the Red Player had to report whether the amount of money in a jar of coins was odd or even. The true amount in the jar was always even and the Red player knew this. Thus, the Red Player could either send an accurate message ("Even") or an inaccurate message ("Odd"). The Blue Player received this message and could either *Rely* on the message or *Verify* the message. We depict the exact payoffs we used in this version of the Rely-or-Verify game in Figure 4. With this payoff structure for the *Rely-or-Verify* game, there is no pure strategy equilibrium. However, there is a mixed strategy equilibrium in which the Red Player provides accurate information with probability 2/3 and the Blue Player relies on that information with probability 2/5. We use this equilibrium as a benchmark against which to measure trustworthiness; if participants were perfectly rational and risk-neutral, they would behave in a trustworthy way (i.e., send accurate information) 2/3 of the time. The full instructions for the Rely-or-Verify game and the solution for the game's equilibrium are provided in Levine and Schweitzer (2015).

-- Figure 4 here --

Participants read the full instructions of the Rely-or-Verify game and were assigned to the role of the Red Player (the trustee). We paired them with a confederate Blue Player (the truster). Participants then had to pass a comprehension check in order to complete the entire

study. After participants passed the comprehension check, they made a decision as the Red Player in our study.

Sense of interpersonal responsibility. After participants made Rely-or-Verify decisions, they provided their sense of interpersonal responsibility ratings, using similar items to those we used in Study 2. Specifically, participants rated their agreement with four statements (1 = *strongly disagree,* 7 = *strongly agree*; α = .90): "I felt accountable for my partner's earnings", "I felt a sense of responsibility to tell the truth", "I cared whether or not my partner earned money", and "I feel an obligation to act responsibly towards my partner."

Guilt-proneness and the Big-5. "Study L" consisted of our measure of guilt-proneness (the GP-5, as in Study 2) and the 44-item Big-5 Personality Inventory (John, Donahue, & Kentle, 1991). After participants completed the questionnaires, they answered demographic questions and received their base payment for participation. The next day, we followed up with participants to pay them a bonus payment based on their decisions in the study. We calculated bonuses by assuming that participants' partners (a confederate) had played the equilibrium strategy.

Results

In Table 5, we present the raw means of each variable and the bivariate correlations among the measures, and in Table 6 we present logistic regressions using standardized values of these variables. In our logistic regressions, the dependent variable was coded as 1 when the participant sent a truthful message in the Rely-or-Verify game (trustworthy behavior) and 0 when the participant sent an untruthful message in the Rely-or-Verify game (untrustworthy behavior).

We find that guilt-proneness is positively associated with integrity-based trustworthy behavior, b = .35, SE = .11, p = .001, e^b = 1.42 (see Table 6, Model 1); the likelihood of sending

an accurate message in the Rely-or-Verify game was 1.42 times as high for those high in guilt-proneness (+1 standard deviation) compared to those of average guilt-proneness. We depict the frequency of trustworthy behavior at high and low levels of guilt-proneness in Figure 5. Interestingly, we find that individuals who are high in guilt-proneness (i.e., the top quartile of respondents) are significantly more trustworthy (82.35%) than the rational equilibrium would predict and that individuals low in guilt-proneness (i.e., the bottom quartile of respondents) are significantly less trustworthy (56.67%) than the rational equilibrium would predict. For each group, a binomial test of proportions revealed that the proportion of participants who had provided accurate information was significantly different from 66.66% (the rational equilibrium; *p*s < .001). Furthermore, individuals with guilt-proneness scores in the upper quartile of the sample were significantly more likely to send truthful messages than were individuals with guilt-proneness scores in the lower quartile of the sample ($\chi^2 = 15.11$, $p < .001$).

Consistent with our mediation hypothesis, we find that guilt-proneness correlates significantly with sense of interpersonal responsibility ($r = .30$, $p < .001$, see Table 5) and that sense of interpersonal responsibility is associated with trustworthy behavior, $b = 1.32$, $SE = .15$, $p < .001$, $e^b = 3.73$ (see Table 6, Model 2). As in Study 1, we used the bootstrap procedure with 10,000 samples to formally test whether sense of interpersonal responsibility mediates the relationship between guilt-proneness and trustworthy behavior (SPSS Process Macro, Model 4, Hayes, 2013). We find evidence of significant mediation through sense of interpersonal responsibility (Indirect effect = .39, SE = .08, 95% CI [.24, .56]). Once we control for sense of interpersonal responsibility in the model, the relationship between guilt-proneness and trustworthiness becomes non-significant, suggesting full mediation (see Table 6, Model 3).

-- Tables 5 and 6, Figure 5 about here --

As shown in Table 5, agreeableness is significantly correlated with trustworthy behavior, sense of interpersonal responsibility, and guilt-proneness. Nonetheless, the effect of guilt-proneness on trustworthy behavior remains significant and positive, b = .32, SE = .12, p = .009, e^b = 1.38, when controlling for agreeableness, b = .24, SE = .14, p = .089, e^b = 1.27, extraversion, b = .05, SE = .13, p = .735, e^b = 1.05, conscientiousness, b = -.10, SE = .14, p = .45, e^b = .90, openness, b = -.07, SE = .12, p = .544, e^b = .93, neuroticism, b = -.08, SE = .15, p = .567, e^b = .92, and gender, b = .22, SE = .37, p = .048 e^b = 1.25 (see Table 6, Model 4). That is, none of the Big-5 personality traits – agreeableness, extraversion, conscientiousness, openness, or neuroticism – significantly predicts trustworthiness when entered simultaneously with guilt-proneness (agreeableness was marginal).

Discussion

In Studies 2 and 3 (as well as two additional studies described in the online supplement), we identify a robust link between a specific individual difference—guilt-proneness—and trustworthy behavior. We find that guilt-proneness predicts benevolence-based trustworthy behavior in the trust game and integrity-based trustworthy behavior in the Rely-or-Verify game. Importantly, we demonstrate that guilt-proneness predicts trustworthiness beyond the effects of the Big Five personality factors—agreeableness, conscientiousness, extraversion, openness to experience, and neuroticism. Although some past research suggests that agreeableness significantly predicts trustworthy behavior in the trust game (Ben-Ner & Halldorsson, 2010), we find only marginal evidence of this relationship. Our results suggest that guilt-proneness may be a better predictor of trustworthiness than agreeableness, though further research is needed to test the robustness of this result. The Big Five, including agreeableness, are broad dimensions of personality that only tangentially capture sense of interpersonal responsibility. Guilt-proneness,

in contrast, is a facet of personality that directly relates to sense of interpersonal responsibility, and we find that it is closely linked with trustworthy behavior.

We did two things to examine the robustness of these results. First, we examined the effect of gender. Consistent with prior research, we expected gender to be correlated with both guilt-proneness (e.g., Cohen et al., 2011; Lutwak & Ferrari, 1996) and trustworthiness (Buchan et al., 2008). Importantly, controlling for gender does not substantively change any of our findings in any of our studies, indicating that the effects of guilt-proneness and gender on trustworthy behavior are independent. That is, gender and guilt-proneness are unique predictors of trustworthy behavior. We also find no evidence of significant Gender X Guilt-Proneness interactions across our studies.

Second, we ran supplementary studies using different measures of guilt-proneness. Specifically, we examined whether the guilt-proneness scale of the Test of Self-Conscious Affect-3 (TOSCA-3; Tangney, Dearing, Wagner, & Gramzow, 2000) also predicted benevolence-based trustworthy behavior and integrity-based trustworthy behavior. We find robust evidence of the relationship between guilt-proneness and trustworthiness regardless of whether the TOSCA-3, the GASP, or GP-5 are used to measure guilt-proneness. We summarize the results of all additional data in our online supplementary materials.

In Studies 2 and 3, we also document the mediating role of sense of interpersonal responsibility. We find that sense of interpersonal responsibility is more closely associated with trustworthiness than is guilt-proneness, which is to be expected given that sense of interpersonal responsibility should be a more proximal predictor to trustworthiness decisions. The correlation between interpersonal responsibility and trustworthiness is significantly greater than the correlation between guilt-proneness and trustworthiness in both Studies 2 and 3 ($ps < .001$). We

build on this finding by exploring whether priming a sense of responsibility can increase trustworthiness both among people who are high and low in guilt-proneness in Study 6.

Study 4: Ruling out alternative mechanisms

In Study 4, we extend our investigation by ruling out alternative mechanisms that could underlie the relationship between guilt-proneness and trustworthiness. As in Study 2, we examine anticipated guilt as a potential alternative. We also examine the possibility that people who are high in guilt-proneness experience more positive emotions as a result of being trustworthy than those low in guilt-proneness (i.e., experience greater "warm glow," Andreoni, 1990). We also extend our investigation by measuring the potential mechanisms and dependent variable (trustworthiness) at different points in time.

Participants

We recruited adults to participate in a 3-part laboratory study at a U.S. university in exchange for a $10 show-up payment and a bonus payment based on their decisions. We made the a priori decision to recruit as many participants as we could in two laboratory sessions (two, 3-day sessions) and then end data collection. This study contained three parts: an initial survey which assessed guilt-proneness and demographic information, a laboratory study in which we measured our mechanism measures, and a follow-up survey that contained the trust game and our behavioral measure of trustworthiness. A total of 405 adults completed the initial survey, 351 adults participated in the laboratory, and 305 people completed the follow-up survey. A total of 292 participants completed all three parts of the study (76% female; M_{age} = 20 years, SD = 1.39; $M_{\text{work experience}}$ = 2 years, SD = 2.00). We find no differences in guilt-proneness among participants who did and did not complete part 2 (p = .71) or part 3 (p = .60).Given that we

conduct each set of analyses using all available data, there may be a different number of participants included in different analyses).

Procedure and Materials

Guilt-proneness. Participants who were interested in signing up for our laboratory sessions were informed that they had to complete a quick survey the evening before their session to be eligible to participate. Each participant who initially signed up for our laboratory sessions received the survey the evening before their scheduled session and was asked to complete it before arriving at the laboratory. The survey contained the TOSCA-3 (Tangney et al., 2000), as well as demographic information.

Similar to the GASP and the GP-5, the TOSCA-3 is a scenario-based measure. It consists of 11 scenario-based questions that assess reactions to everyday negative events (e.g. accidentally hitting an animal with your car, missing a lunch meeting with a friend) and five scenario-based questions that assess reactions to everyday positive events (e.g. being rewarded for your work team's good performance). After each scenario, the TOSCA-3 presents participants with items designed to measure guilt-proneness, as well as related reactions. Specifically, the TOSCA-3 examines the degree to which individuals anticipate feeling guilt, shame, detachment, externalization, and pride in response to positive and negative scenarios. For every scenario, participants rated the likelihood that they would experience four reactions using a five-point rating scale anchored at 1 = *not likely* and 5 = *very likely*. Items indicative of guilt-proneness focus on feeling badly about one's behavior (e.g., "You'd feel bad you hadn't been more alert driving down the road.") or intentions to engage in behaviors that repair harm caused by the transgression (e.g., "You'd think you should make it up to him as soon as possible."). We

followed the guidelines of Tangney et al. (2000) to analyze the results of the TOSCA-3.[4] Because guilt-proneness and shame-proneness are often correlated, we report the effects of guilt-proneness with and without shame-proneness as a covariate (Schaumberg & Flynn, 2012; Tangney, Miller, Flicker, & Barlow, 1996).

Mechanism measures. When participants arrived at the behavioral laboratory, they were seated in individual cubicles and learned about the trust game. Specifically, they were asked to imagine playing "The Choice Game", which was identical to the game we used in Study 1. In this version of the trust game, Player 1 was given $1, which increased to $2.50 if Player 1 passed the initial $1 to Player 2. Then, Player 2 had the decision to either keep the $2.50 (untrustworthy behavior) or return $1.25 (trustworthy behavior).

After passing a comprehension check on the "Choice Game," participants were asked to imagine that they were Player 2 and that their partner chose to "Pass $1." Then, we asked participants, "To what extent do you think your decision in the Choice Game would be influenced by the following factors," followed by our proposed mechanism items (sense of interpersonal responsibility) and items capturing several alternative mechanisms, in a random order. All items were measured using a 7-point rating scale anchored at *not at all* and *extremely*.

Sense of interpersonal responsibility. We used four items to assess sense of interpersonal responsibility, similar to the items we used in Studies 2 and 3: "How responsible I feel for my partner's earnings", "How obligated I feel to act responsibly towards my partner in the Choice Game", "How much I care about my partner's earnings", and "How morally obligated I feel to share the money" (α = .88).

[4] Although we were primarily interested in guilt-proneness, we administered the full TOSCA-3, consistent with extant research. We did not have any a priori hypotheses about the relationship between other dimensions of the TOSCA-3 and trustworthiness, and the effects of guilt-proneness remain significant when the other TOSCA-3 subscales are included as covariates.

Anticipated guilt. As in Study 1, we also examined the affective anticipation of guilt by measuring, "How guilty I would feel about keeping all the money" and "How distressed I would feel if my partner earned nothing", $r(319) = .68$, $p < .01$.

Warm glow. One additional reason that individuals may behave in generous and trustworthy ways is the desire to feel "warm glow" (Andreoni, 1990), or the desire to feel positive emotions in response to doing good. We collected two items to examine whether individuals who are high in guilt-proneness are motivated by warm glow more than individuals who are low in guilt-proneness (and whether this mediates the relationship between guilt-proneness and trustworthiness): "How happy I would feel about returning half of the money" and "How proud I would feel about returning half of the money" $r(319) = .62$, $p < .01$.

Self-interest. We also asked participants to respond to three items about the extent to which they were influenced by self-interest: "How happy I would feel about keeping all the money", "How important it is for me to earn the largest possible bonus", and "How excited I would feel to earn \$1.50" ($\alpha = .65$).[5]

To mask the purpose of the study, we also asked participants to imagine that they were Player 1 and asked a similar set of questions (whether their decision as Player 1 would be influenced by their sense of responsibility, anticipated guilt, warm glow, and self-interest). We had no hypotheses pertaining to these questions and did not analyze those data.

Trustworthy behavior. Then, we sent participants a new survey the evening after they completed the mechanism measures. This survey simply contained the trust game, which was

[5] The second item of our self-interest scale should have asked participants how excited they would feel to earn \$2.50 (the amount associated with the decision to keep all of the money). However, due to a typographical error, we asked participants how excited they would feel to earn \$1.50. This error makes our estimate of self-interest more conservative, because participants high in self-interest would want to receive more than \$1.50.

nearly identical to the version we used in Study 2. However, in this study each lottery ticket entered the participants into a lottery for a $99 gift card, rather than a $75 bonus. We note that the trust game participants actually played (as part of the evening survey) was different from the game they imagined when responding to our potential mechanism measures. We used different versions of the game to mask the purpose of the study.

We told participants that they were randomly assigned to their roles. In fact, we assigned all participants to the role of Player 2 and paired them with a confederate Player 1, as we did in the prior studies. We conducted the lottery one week after the study concluded, and we paid one participant the $99 bonus.

Results

In Table 7, we present the raw means of each variable and the bivariate correlations among the measures, and in Table 8 we present our key logistic regressions using standardized values of these variables.

Replicating the results of Studies 1 and 2, we find that guilt-proneness is positively associated with benevolence-based trustworthy behavior (i.e., return behavior in the trust game), b = .29, SE = .13, p = .019, e^b = 1.34 (see Table 8, Model 1). The effect of guilt-proneness remains stable, significant and positive, b = .40, SE = .15, p = .007, e^b = 1.50, controlling for shame-proneness, b = -.20, SE = .15, p = .17, e^b = .81 (see Table 8, Model 2).

For illustrative purposes, we depict the frequency of trustworthy behavior at high and low levels of guilt-proneness, which we depict as the lower and upper guilt-proneness quartiles in the sample (see Figure 6). Individuals with guilt-proneness scores in the upper quartile of the sample were marginally more likely to return money in the trust game (52.05%, n = 73) than were

individuals with guilt-proneness scores in the lower quartile of the sample (32.84%, $n = 67$; $\chi^2 = 5.27$, $p = .022$).

We also find that guilt-proneness significantly correlates with sense of interpersonal responsibility ($r = .34$, $p < .001$), as well as anticipated guilt ($r = .29$, $p < .001$) and warm glow ($r = .29$, $p < .001$), but not self-interest ($r = .002$, $p = .97$, see Table 7).

-- Tables 7 and 8, Figure 6 about here --

We used the bootstrap procedure with 10,000 samples to formally test which of the mechanisms above mediates the relationship between guilt-proneness and trustworthy behavior (SPSS Process Macro, Model 4, Hayes, 2013). The mediation model included guilt-proneness as the independent variable, sense of interpersonal responsibility, anticipated guilt, and warm glow as simultaneous mediators, and trustworthy behavior as the dependent variable. We did not include self-interest in the model because it was not significantly correlated with guilt-proneness. We find evidence of significant mediation through sense of interpersonal responsibility (Indirect effect = .26, SE = .10, 95% CI [.09, .50]), but not through anticipated guilt (Indirect effect = 0.02, SE = .07, 95% CI [-.11, .15]), or warm glow (Indirect effect = .04, SE = .05, 95% CI [-.06, .15]). Both with and without including anticipated guilt and warm glow as covariates, once we control for sense of interpersonal responsibility, the relationship between guilt-proneness and trustworthiness becomes non-significant, suggesting full mediation by sense of interpersonal responsibility (see Table 8, Models 3 and 4).

Discussion

Study 4 provides further evidence that the relationship between guilt-proneness and trustworthiness is mediated by a sense of interpersonal responsibility. Although guilt-proneness was associated with a greater anticipation of guilt when behaving in an *un*trustworthy way, as

well as a greater warm glow (anticipation of happiness and pride) when behaving in a trustworthy way, these two mechanisms do not account for the relationship between guilt-proneness and trustworthiness. Guilt-proneness was not associated with lower self-interest. Furthermore, although both self-interest (r = -.24) and sense of interpersonal responsibility (r = .40) were correlated with trustworthiness, the relationship between sense of responsibility and trustworthy behavior was significantly stronger (z = 2.13, p =.016). This provides further insight into why guilt-proneness (which is uniquely associated with sense of responsibility) may be a better predictor of trustworthiness than traits that are broadly associated with self-interest (e.g., Machiavellianism).

Although guilt-proneness, sense of interpersonal responsibility, and trustworthiness were all measured at different time points in Study 4, our evidence for mediation thus far is nonetheless correlational. To test the causal role of interpersonal responsibility, we manipulate it directly in Studies 5 and 6. In Study 5, we directly manipulate our proposed mechanism: one's sense of responsibility for the truster within the trust context. We demonstrate that a trustee's level of responsibility for the truster moderates the relationship between guilt-proneness and trustworthiness, providing further evidence for our proposed theoretical model. In Study 6, we prime a general sense of responsibility and examine whether this can increase trustworthiness both among those who are high and low in guilt-proneness.

Study 5. Manipulating one's responsibility towards the trustee

In Study 5, we manipulate the degree to which the truster makes herself vulnerable to the trustee in the trust game, thereby manipulating how responsible the trustee is for the truster's outcomes. This design achieves two aims. First, it allows us to test our proposed mechanism, interpersonal responsibility, using a causal chain design (Spencer, Zanna, & Fong, 2005).

Second, it allows us to disentangle trustworthiness from generosity. If people who are high (vs. low) in guilt-proneness are generally more generous, then we would expect them to return money to the truster regardless of how much the truster initially passed (i.e., how much the truster trusted the trustee). On the other hand, if people who are high (vs. low) in guilt-proneness are particularly sensitive to the degree to which they are responsible for others, then we would expect the relationship between guilt-proneness and return behavior to be moderated by the truster's initial degree of vulnerability. Specifically, we would expect that highly guilt-prone trustees will return more money to the truster than will less guilt-prone trustees when the truster initially passes a lot of money, but will *not* return more money to the truster than will less guilt-prone trustees when the truster initially passes very little money.

Participants

We set the a priori goal of recruiting 400 U.S. adults to participate in an online study via Amazon Mechanical Turk in exchange for a small payment ($1). Participants also earned bonus money based upon their decisions. We ultimately ended up with 402 participants (51% female; M_{age} = 36 years, SD = 11.0; $M_{\text{work experience}}$ = 16 years, SD = 10.5) who completed the entire study and were eligible for analysis.

Procedure and Materials

As in Study 3, we told participants that they would complete two unrelated studies in exchange for payment. We presented one of the studies, "Study W," as a decision-making game (the trust game), and we presented the other study, "Study L," as a scenario questionnaire (the TOSCA-3 measure of guilt-proneness). Participants always completed the trust game first, before completing our guilt-proneness measure. We randomly assigned participants to one of two conditions in the trust game: high vulnerability or low vulnerability.

The trust game. We assigned all participants to the role of Player 2 in the trust game to capture benevolence-based trustworthy behavior, as in Studies 1, 2, and 4. However, rather than using a dichotomous version of the trust game, in which the truster either chose to "Pass" or "Keep" the money and the trustee chose to either "Return" or "Keep" the passed amount, we used a continuous version of the trust game. We also used larger stakes in this study.

Specifically, in Study 5, Player 1 started with $20 and could pass any amount (between $0 and $20) to Player 2. Whatever amount Player 1 passed to Player 2 was tripled and became part of Player 2's earnings. Then, Player 2 could pass back any amount to Player 1, and would earn the amount that was leftover. Participants read three different examples of Player 1 and Player 2 decisions and then had to pass a comprehension check. Participants who passed the comprehension check were then randomly assigned to learn that Player 1 passed $2 (out of $20) to them or $20 (out of $20) to them. Passing $2 reflects low vulnerability, whereas passing $20 reflects high vulnerability. These values were based on previous manipulations of vulnerability (i.e., levels of initial trust; Pillutla, Malhotra, & Murnighan, 2003). This manipulation influenced the degree to which participants were responsible for the truster's outcomes. In the low vulnerability condition, the truster kept $18, and the trustee was only responsible for allocating at most 25%, $6 ($2 x 3), of the truster's potential total earnings. In the high vulnerability condition, the truster kept $0, and the trustee was responsible for allocating 100% of the truster's potential total earnings.

Participants used a slider scale to decide how much of the tripled amount to pass back to Player 1. The slider scale was anchored at $0 and $6 in the low vulnerability condition, and $0 and $60 in the high vulnerability condition. Participants were informed that 25% of participants would actually earn the bonuses associated with their decisions in the trust game.

Sense of interpersonal responsibility. Participants also provided their sense of interpersonal responsibility ratings. We used similar items to those we used in Studies 2 and 3. Specifically, participants indicated the extent to which their decision was (will be) influenced by the following factors (1 = *not at all*, 7 = *extremely*; α = .94): "How responsible I felt (feel) for my partner's earnings", "How obligated I felt (feel) to act responsibly towards my partner in the Choice Game", "How much I cared (care) about my partner's earnings", and "How morally obligated I felt (feel) to share the money." We randomized the order in which participants made their decision in the trust game and completed these mechanism measures. The order in which participants completed these parts of the experiment did not influence our results (*p*s > .18), and thus we do not discuss order effects further.

Guilt-proneness. After completing "Study W" which included the trust game and questions about participants' sense of responsibility, participants moved on to "Study L," which consisted of our measure of guilt-proneness. We used the TOSCA-3, as in Study 4 and in our supplemental studies.

At the end of the study, we collected demographic information and asked participants their thoughts on the purpose of the study. We then paid 25% of participants the bonus associated with their decisions in the trust game.

Results

Trustworthy behavior. We conceptualize trustworthy behavior as the amount returned by the trustee as a proportion of the amount sent by the truster (money returned divided by money sent). However, using the amount returned by the trustee as a proportion of the trustee's total endowment (money returned divided by three times money sent) yields identical results.

We used OLS regression to analyze the effects of guilt-proneness, manipulated level of vulnerability, and their interaction on trustworthy behavior. There was a main effect of vulnerability, $b = .71$, $SE = .06$, $p < .001$, such that participants passed back more money when trustees made themselves more vulnerable, consistent with prior work (Pillutla, Malhotra, & Murnighan, 2003). There was no main effect of guilt-proneness, $b = .02$, $SE = .04$, $p = .59$. Importantly, however, there was a significant Vulnerability X Guilt-Proneness interaction, $b = .13$, $SE = .06$, $p = .026$. Specifically, when trusters made themselves vulnerable and passed all of their money (and thus, participants were highly responsible for their partners' outcomes), guilt-proneness significantly predicted trustworthy behavior (i.e., returning the money), $b = .15$, $SE = .05$, $p = .003$. However, when trusters did not make themselves vulnerable and passed a very small amount of money (and thus, participants were *not* highly responsible for their partners' outcomes), guilt-proneness had no relationship with trustworthy behavior, $b = .02$, $SE = .03$, $p = .52$. These results suggest that guilt-proneness is uniquely associated with trustworthy behavior. People high in guilt-proneness are more likely to return money than are people low in guilt-proneness, but only when trusters have made themselves vulnerable, thereby signaling high trust and the trustee's responsibility to reciprocate. When others do not make themselves vulnerable, people high in guilt-proneness are no more likely to return money than are people low in guilt-proneness.

We plot the relationship between guilt-proneness and trustworthy behavior at low and high levels of interpersonal vulnerability in Figure 7.

---Figure 7 about here---

Sense of interpersonal responsibility. We used OLS regression to analyze the effects of guilt-proneness, manipulated level of vulnerability, and their interaction on measured sense of interpersonal responsibility.

There was a main effect of vulnerability, $b = .68$, $SE = .09$, $p < .001$, such that participants felt more responsible for their partners when their partners made themselves more vulnerable. There was no main effect of guilt-proneness, $b = .06$, $SE = .06$, $p = .95$. Importantly, however, there was a marginally significant Vulnerability X Guilt-Proneness interaction, $b = .18$, $SE = .09$, $p = .052$, paralleling the trustworthy behavior results: when trusters made themselves vulnerable, guilt-proneness significantly predicted sense of interpersonal responsibility, $b = .24$, $SE = .07$, $p = .001$, but when trusters did not make themselves vulnerable, guilt-proneness had no relationship with sense of interpersonal responsibility, $b = .06$, $SE = .06$, $p = .33$.

Mediation analyses. We ran moderated mediation analyses using the bootstrap procedure with 10,000 samples (SPSS Process Macro, Model 7, Hayes, 2013) to test the processes by which guilt-proneness and vulnerability affect trustworthy behavior. We find significant evidence of moderated mediation (Indirect effect = .07, SE = .04, 95% CI [.003, .15]) such that interpersonal responsibility mediates the relationship between guilt-proneness and trustworthy behavior when vulnerability is high (Indirect effect = .09, SE = .03, 95% CI [.04, .15]), but not when vulnerability is low (Indirect effect = .02, SE = .03, 95% CI [-.03, .07]).

Discussion

In Study 5, we manipulated the degree to which the truster made themselves vulnerable to the trustee. Consistent with Pillutla, Malhotra, and Murnighan (2003), we find that the truster's level of vulnerability *causally* influences trustworthy behavior. Participants return

significantly greater proportions of money to others who have taken substantial risks when trusting them than to those who have not made themselves vulnerable to the risks of trusting.

Importantly, we also find that vulnerability moderates the effect of guilt-proneness on trustworthiness. Thus, we demonstrate that highly guilt-prone individuals are not simply more generous; rather, they are sensitive to social expectations. People who are high in guilt-proneness are more generous when there is an implicit expectation to be generous and they are responsible for others' outcomes, but they are no more generous than those who are low in guilt-proneness when there is not an expectation to be generous and they are not responsible for others' outcomes.

Finally, in addition to disentangling generosity from trustworthiness, our Study 5 findings underscore the importance of interpersonal responsibility in accounting for the effects of guilt-proneness on trustworthiness. As in Studies 2 and 4, we find that trustees who are high in guilt-proneness report a greater sense of interpersonal responsibility than do those trustees who are low in guilt-proneness. This greater sense of interpersonal responsibility among those high (versus low) in guilt-proneness, in turn, is associated with greater trustworthiness.

Study 6. Manipulating responsibility via codes of conduct

In our final study, we extend our investigation by exploring whether priming a broad sense of responsibility through a code of conduct can increase trustworthiness even among those who are not dispositionally inclined to feel a sense of interpersonal responsibility. Participants in this study read a code of conduct (or a set of instructions in the control condition) and then completed the trust game. Our manipulation mirrors corporate codes of conduct that focus on either interpersonal responsibility or individualism. For example, the opening sentence in Bank of America's code of conduct states, "We have the responsibility to do the right thing for our

customers, shareholders, communities and one another." (Bank of America Corporation Code of Ethics, 2009). In contrast, companies such as Bridgewater Associates focus on self-interest; in his list of *Principles,* founder Ray Dalio states, "I believe that pursuing self-interest in harmony with the laws of the universe and contributing to evolution is universally rewarded, and what I call 'good'" (Dalio, 2011). In the present study, we examine how guidelines that prime interpersonal responsibility versus individual self-interest influence trustworthiness among those high in guilt-proneness and among those low in guilt-proneness.

Method

In Study 6, we conducted an initial screening survey to assess guilt-proneness. Then, we recruited participants who were high or low in guilt-proneness to complete the main study, which contained the responsibility induction and our dependent variables.

Initial Screening

We initially recruited 1000 adults to participate in an online study via Amazon Mechanical Turk in exchange for a small payment ($0.30). We recruited 1000 participants for our initial survey because we intended to recruit those whose guilt-proneness scores were in the top third or the bottom third in the initial screening, and we wished to obtain a sample size of roughly 600 participants, after moderate attrition.

In the initial survey, participants completed the GP-5. We informed participants that they could be recruited for future surveys based on their responses. As in our prior studies, the GP-5 was highly reliable despite its brevity ($\alpha = .78$). Participants also provided demographic information in this survey.

Main Study

One day later, we advertised the main study to participants whose guilt-proneness scores were in the top third (M = 4.60, SD = 0.28) or the bottom third (M = 2.78, SD = 0.57) of all participants. We used the top third and bottom third of participants to ensure that the two groups would be distinct, and that our main study would have the appropriate level of power. Because we recruited participants who either scored in the top or bottom third of guilt-proneness, we analyze guilt-proneness as a dichotomous predictor.

Participants. We ultimately ended up with 552 participants from Amazon Mechanical Turk (49% female; M_{age} = 35 years, SD = 11.57; $M_{\text{work experience}}$ = 15 years, SD = 10.58) who completed the entire study and were eligible for analysis.

Sense of interpersonal responsibility manipulation. We randomly assigned participants to read either a code of conduct that emphasized interpersonal responsibility (*high interpersonal responsibility*) or a set of instructions that emphasized individual self-interest (*low interpersonal responsibility*). In our high sense of interpersonal responsibility condition, participants read an "Amazon Mechanical Turk Code of Conduct." In our low interpersonal responsibility condition, participants read "Amazon Mechanical Turk Instructions." We called the low interpersonal responsibility manipulation "Instructions" because we expected that the phrasing "Code of conduct" could itself influence one's sense of interpersonal responsibility. We present the full manipulation in Figure 8.

-- Figure 8 about here --

Dependent variables. After participants read either the code of conduct or the instructions, they learned about the trust game and had to pass a comprehension check. After passing the comprehension check, participants learned that they were assigned to the role of Player 2 in the trust game and then made a decision. Our primary dependent variable was

benevolence-based trustworthiness, as measured by Player 2's return behavior in the trust game. We used the same trust game in Study 6 that we used in Studies 1, 2 and 4, with one modification—trust game decisions entered participants in a $55 lottery.

After participants completed the trust game, they answered the same questions assessing sense of interpersonal responsibility as those we used in Study 2 ($\alpha = .93$). Finally, we collected demographic information and asked participants what they thought the purpose of the study was. One week after we completed the study, we conducted the lottery, selected one participant, and paid them the $55 bonus.

Results

Trustworthy behavior. In Figure 9, we display the percentage of participants who were trustworthy (i.e., returned money in the trust game) in each of our four experimental conditions.

-- Figure 9 about here --

Using logistic regression, we find a main effect of manipulated interpersonal responsibility on trustworthy behavior, $b = .59$, $SE = 0.10$, $p < .001$, $e^b = 1.81$. Specifically, individuals who were asked to behave responsibly were more likely to return money in the trust game than were individuals who were asked to look out for themselves (74.30% vs. 50.37%). We also find a main effect of guilt-proneness on trustworthy behavior, $b = .55$, $SE = 0.10$, $p < .001$, $e^b = 1.73$; individuals high in guilt-proneness were more likely to return money in the trust game than were individuals low in guilt-proneness (73.38% vs. 51.82%).

However, we find no significant Guilt-proneness X Manipulated interpersonal responsibility interaction ($b = .11$, $SE = 0.10$, $p = .28$, $e^b = 1.11$). Interestingly, a simple contrast reveals no difference between the Low Guilt-Proneness/High Interpersonal Responsibility and the High Guilt-Proneness/Low Interpersonal Responsibility conditions (63.01% vs. 60.71%, $\chi^2 =$

.16, $p = .69$), suggesting that some interventions, such as the one used in the current study, may be able to improve the behavior of individuals who are not predisposed to trustworthiness.

Sense of interpersonal responsibility. We used a two-way ANOVA to analyze the effects of guilt-proneness, manipulated sense of interpersonal responsibility, and their interaction on measured sense of interpersonal responsibility. Our manipulation influenced sense of interpersonal responsibility in the trust game, $F(1, 548) = 50.49$, $p < .001$, $d = .53$. Participants who were instructed to behave responsibly reported feeling more responsible for their partner ($M = 5.30$, $SD = 1.52$) than participants who had been instructed to look out for themselves ($M = 4.34$, $SD = 1.90$). Consistent with our prior studies, we also find a main effect of guilt-proneness, $F(1, 548) = 47.19$, $p < .001$, $d = .56$. Individuals high in guilt-proneness felt more responsible for their partners ($M = 5.29$, $SD = 1.67$) than individuals low in guilt-proneness ($M = 4.36$, $SD = 1.80$). We find no Guilt-proneness X Manipulated sense of interpersonal responsibility interaction, $F(1, 548) = .88$, $p = .35$.

Mediation analyses. We ran two separate mediation analyses using the bootstrap procedure, each with 10,000 samples (SPSS Process Macro, Model 4, Hayes, 2013), to test the processes by which guilt-proneness and manipulated sense of interpersonal responsibility affect trustworthy behavior. In the first model, we entered guilt-proneness as the independent variable, measured sense of interpersonal responsibility as the mediator, manipulated sense of interpersonal responsibility as a covariate, and trustworthy behavior as the dependent variable. Controlling for manipulated sense of interpersonal responsibility allows us to isolate the effects of guilt-proneness on trustworthy behavior. We find evidence of significant mediation through measured sense of interpersonal responsibility (Indirect effect = .87, SE = .16, 95% CI [.58, 1.19]).

In the second model, we entered manipulated sense of interpersonal responsibility as the independent variable, measured sense of interpersonal responsibility as the mediator, guilt-proneness as a covariate, and trustworthy behavior as the dependent variable. As before, we find evidence of significant mediation through measured sense of interpersonal responsibility (Indirect effect = .90, SE = .16, 95% CI [.61, 1.23]).

Discussion

In Study 6, we find that codes of conduct that make interpersonal responsibility salient can increase trustworthiness among individuals who are low in guilt-proneness and in individuals who are high in guilt-proneness. Interestingly, unlike in Study 5, we find no interaction between our responsibility manipulation and trait guilt-proneness on trustworthy behavior. We consider two possible explanations for this difference. First, it is possible that our dichotomization of guilt-proneness limited our ability to detect a significant interaction between guilt-proneness and our manipulation. By dichotomizing guilt-proneness – that is, by recruiting participants who were either high (top third) or low (bottom third) in guilt-proneness – we simplified our design, but we may have lost the statistical power to detect the moderating influence of our responsibility manipulation.

Second, it is possible that our manipulations in Study 5 were stronger than our manipulations in Study 6 and more directly interrupted the psychological mechanism linking guilt-proneness to trustworthiness. Specifically, in the low vulnerability condition in Study 5, the truster signaled that they did not expect the participant to return money to them in this particular interaction. Additionally, in the low vulnerability condition, the truster violated norms of fairness (by keeping $18 and allowing the trustee to make at most $6). This could make participants – even those who are high in guilt-proneness – feel that they are no longer responsible for the

truster, nor are they obligated to behave generously. In contrast, in the low responsibility condition of Study 6, although we broadly primed self-interest, it is possible that participants high in guilt-proneness still felt responsible for the specific truster with which they were paired. Indeed, supplemental analyses demonstrate that even in the low responsibility condition, participants who were high in guilt-proneness felt more responsible for their partner (M = 4.73, SD = 1.81) than did those who were low in guilt-proneness (M = 3.90, SD = 1.90; $F(1, 548) = 17.08$, $p < .001$, $d = .44$).

General Discussion

Although trust is fundamental for effective interpersonal relationships, prior work has disproportionately focused on only one half of an inherently dyadic construct. Whereas existing trust research has focused largely on the question of when people are more or less likely to trust others, our work offers insight into who is *worthy* of that trust. Across six studies and sixteen supplemental studies summarized in our supplemental materials, we demonstrate that guilt-proneness (i.e., individual differences in the extent to which people anticipate feeling guilty about wrongdoing) is a key driver of trustworthiness. Specifically, we find that compared to individuals low in guilt-proneness, individuals who are high in guilt-proneness feel a greater sense of interpersonal responsibility when they are entrusted and, as such, are less likely to exploit others' trust. We provide evidence of interpersonal responsibility as an underlying mechanism through both correlational analyses and experimental designs and we rule out several alternative mechanisms. In doing so, we also demonstrate that guilt-proneness is associated with trustworthiness specifically, rather than generosity broadly. Finally, we demonstrate that priming a sense of interpersonal responsibility via codes of conduct can increase trustworthiness among those who are low in guilt-proneness, as well as among those who are high in guilt-proneness.

Our research makes important theoretical and practical contributions. First, our research underscores the importance of investigating trustworthiness. In contrast to extant trust research that has focused on the perspective of the truster (Mayer et al., 1995; Whitener et al., & Werner, 1998), we call for future work to deepen our understanding of the trustee. We know surprisingly little about what makes people more or less trustworthy (as opposed to broadly more generous or ethical), and much of what we know has examined trustworthiness through the lens of trust game behavior and conceptualized trustworthiness as a calculated reaction to a truster's initial level of trust (Ashraf et al., 2006; Ben-Ner & Halldorsson, 2010; Buchan et al., 2008; Gunnthorsdottir, McCabe, & Smith, 2002; Pillutla et al., 2003; Schweitzer, Ho, & Zhang, 2016). In our investigation, we demonstrate that this approach to studying trustworthiness is overly narrow. We show how individual differences in guilt-proneness influence trustworthiness, independent of a truster's perceptions of a trustee, or a truster's initial level of trust. In doing so, we demonstrate that guilt-proneness and its concomitant sense of interpersonal responsibility play a critical role in fostering trustworthy behavior.

Our investigation also makes an important, and related, methodological contribution. We document the relationship between guilt-proneness and trustworthiness with a novel attitudinal measure of trustworthiness, and with two different behavioral measures that represent benevolence-based trustworthiness and integrity-based trustworthiness. By developing a scale that captures trait-level trustworthy intentions, and by examining integrity-based trustworthy behavior (with the Rely-or-Verify game) in addition to benevolence-based trustworthy behavior (with the trust game), we advance the study of trust and trustworthiness.

Furthermore, our findings contribute to a growing body of research that documents the benefits of guilt and guilt-proneness for organizations, relationships, and society (e.g., Bohns &

Flynn, 2012; Cohen, Kim, Jordan, & Panter, 2016; Cohen, Panter & Turan, 2013; Cohen, Panter, Turan, Morse, & Kim, 2013, 2014; De Hooge et al., 2007; Flynn & Schaumberg, 2012; Schaumberg, & Flynn, 2012, 2017; Stuewig et al., 2015; Tangney et al., 2007; Tangney, Stuewig, & Martinez, 2014; Wiltermuth & Cohen, 2014). In doing so, we provide insight into whom individuals *should* (and should not) trust. Whereas extant research has examined when individuals are *perceived* to be trustworthy, we examine whether someone *is* actually trustworthy. Consequently, our findings offer prescriptive advice for avoiding trust violations: Be wary of trusting individuals with low levels of guilt-proneness, and ensure that trustees feel personally responsible for the costs of broken trust.

Future Directions

Future research is needed to examine how both guilt-proneness and trustworthiness relate to a broader set of personality traits and to understand how guilt-proneness influences trustworthiness across time and across different contexts. Although our results suggest that guilt-proneness is a better predictor of trustworthiness than the Big Five personality traits, future work should extend our investigation by further examining personality traits such Machiavellianism (Gunnthorsdottir et al., 2002) and moral identity (Aquino & Reed, 2002), and by examining contextual factors, such as subtle environmental cues of interpersonal responsibility, that could influence trustworthiness.

Future research should also explore how guilt-proneness influences trust *over time* (e.g., Lewicki, Tomlinson, & Gillespie, 2006). When trusters make themselves vulnerable, highly guilt-prone individuals are likely to honor this trust, and thus, are likely to be labeled as trustworthy in future interactions. Consequently, highly guilt-prone individuals may develop positive reputations and develop more effective long-term trusting relationships over time.

Highly guilt-prone individuals may also be able to restore trust more easily following a violation. Trust repair and forgiveness following a transgression represent a significant challenge (Exline, Baumeister, Bushman, Campbell, & Finkel, 2004; Finkel, Rusbult, Kumashiro, & Hannon, 2002; Hannon, Rusbult, Finkel, & Kamashiro, 2010; Haselhuhn, Schweitzer, & Wood, 2010; Rusbult, Kumashiro, Finkel, & Wildschut, 2002; Schweitzer, Hershey & Bradlow, 2006). We know that individuals high in guilt-proneness are more forgiving of others who have wronged them than are individuals low in guilt-proneness (Jordan, Flynn, & Cohen, 2015), and being forgiving of others is likely to be an asset for trust repair. Nonetheless, additional research is needed to better understand the role of guilt-proneness in the trust repair process.

It would also be interesting to examine whether individuals exploit the good intentions of those who are high in guilt-proneness. Specifically, if individuals have insight into the link between guilt-proneness and trustworthiness, it is possible that they could strategically manipulate highly guilt-prone individuals into joining exploitative relationships. On the other hand, in Study 5, we find that guilt-prone individuals were not more trustworthy when their partner was not vulnerable and had violated the norms of fairness, so it is not clear whether high (versus low) levels of guilt-proneness makes a person more susceptible to being defrauded or otherwise taken advantage of. Future work could examine if guilt-prone individuals are sensitive to whether individuals are genuinely trusting them or attempting to manipulate them, and if they adjust their trustworthiness accordingly. It is also possible that untrustworthy people may be able to mimic guilt–proneness and exploit others' trust. Ultimately, future research is needed to understand when not to trust (Yip & Schweitzer, 2015).

Beyond guilt-proneness, future studies in organizations could fruitfully examine the positive and negative consequences of attempting to instill a greater sense of interpersonal

responsibility. For example, organizations may want to emphasize interpersonal responsibility in their company mission statements and have employees review and repeat these statements before activities that require a high level of trustworthiness. If done well, employers may be able to build a sense of interpersonal responsibility and instill trustworthiness over time. If done poorly, however, employers may trigger reactance and harm trustworthiness. These recommendations require further research, however, before they are implemented.

Conclusion

Trust *and* trustworthiness are critical for effective relationships and effective organizations. Individuals and institutions incur high costs when trust is misplaced, but people can mitigate these costs by engaging in relationships with individuals who are trustworthy. Our findings extend the substantial literature on trust by deepening our understanding of trustworthiness: When deciding in whom to place trust, trust the guilt-prone.

References

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, *100*(401), 464-477.

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423.

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193-208.

Bank of America Corporation Code of Ethics. (2009). Retrieved from http://investor.bankofamerica.com/phoenix.zhtml?c=71595&p=irol-govconduct#fbid=_7BS4ZJrRKU

Barkataki, I., Kumari, V., Das, M., Taylor, P., & Sharma, T. (2006). Volumetric structural brain abnormalities in men with schizophrenia or antisocial personality disorder. *Behavioural Brain Research*, *169*(2), 239-247.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin*, *115*(2), 243.

Becker, A., Deckers, T., Dohmen, T., Falk, A., & Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics*, 4, 453-478.

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64-79.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122-142.

Bohns, V. K., & Flynn, F. J. (2012). Guilt by design: Structuring organizations to elicit guilt as an affective reaction to failure. *Organization Science, 24*(4), 1157-1173.

Brooks, A. W., Dai, H., & Schweitzer, M. E. (2014). I'm sorry about the rain! Superfluous apologies demonstrate empathic concern and increase trust. *Social Psychological and Personality Science, 5*(4), 467-474.

Buchan, N. R., Croson, R. T., & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization, 68*(3), 466-476.

Cohen, T. R., Kim, Y., & Panter, A. T. (2014). *The five-item guilt-proneness scale (GP-5).* Carnegie Mellon University, Pittsburgh, PA. http://dx.doi.org/10.13140/RG.2.1.2847.2167

Cohen, T. R., Kim, Y., Jordan, K. P., & Panter, A. T. (2016). Guilt-proneness is a marker of integrity and employment suitability. *Personality and Individual Differences, 92*, 109-112.

Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt-proneness and moral character. *Current Directions in Psychological Science, 21*(5), 355-359.

Cohen, T. R., Panter, A. T., & Turan, N. (2013). Predicting counterproductive work behavior from guilt-proneness. *Journal of Business Ethics, 114*(1), 45-53.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2013). Agreement and similarity in self-other perceptions of moral character. Journal of Research in Personality, 47, 816-830. doi:10.1016/j.jrp.2013.08.009

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology, 107*(5), 943-963.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*(5), 947.

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, *92*(4), 909.

Costa Jr, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*(4), 343-359.

Dalio, Ray. "PRINCIPLES." (2011). Bridgewater Culture and Principles. Bridgewater Associates. Web. 11 May 2016.

De Cremer, D., van Dijk, E., & Pillutla, M.M., 2010. Explaining Unfair Offers in Ultimatum Games and Its effects on Trust: An Experimental Approach. *Business Ethics Quarterly*, 20(1): 107-126

Derfler-Rozin, R., Pillutla, M., & Thau, S. (2010). Social reconnection revisited: The effects of social exclusion risk on reciprocity, trust, and general risk-taking. *Organizational Behavior and Human Decision Processes*, 112(2), 140-150

De Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2007). Moral sentiments and cooperation: Differential influences of shame and guilt. *Cognition and Emotion*, *21*(5), 1025-1042.

De Hooge, I. E., Nelissen, R., Breugelmans, S. M., & Zeelenberg, M. (2011). What is moral about guilt? Acting "prosocially" at the disadvantage of others. *Journal of Personality and Social Psychology*, *100*(3), 462.

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science*, *12*(4), 450-467

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.

Dunn, J., Ruedy, N. E., & Schweitzer, M. E. (2012). It hurts both ways: How social comparisons harm affective and cognitive trust. *Organizational Behavior and Human Decision Processes*, *117*(1), 2-14.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust. *Journal of personality and Social Psychology*, *88*(5), 736.

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*(1), 122.

Dunning, D., Fetchenhauer, D., & Schlösser, T. M. (2012). Trust as a social and emotional act: Noneconomic considerations in trust behavior. *Journal of Economic Psychology*, *33*(3), 686-694.

Eronen, M., Angermeyer, M. C., & Schulze, B. (1998). The psychiatric epidemiology of violent behaviour. *Social Psychiatry and Psychiatric Epidemiology*, *33*(1), S13-S23.

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, 42, 1585-1593.

Exline, J. J., Baumeister, R. F., Bushman, B. J., Campbell, W. K., & Finkel, E. J. (2004). Too proud to let go: narcissistic entitlement as a barrier to forgiveness. *Journal of Personality and Social Psychology, 87*(6), 894.

Finkel, E. J. (2017). The All-Or-Nothing Marriage: How the best marriages work. New York: Dutton

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with betrayal in close relationships: Does commitment promote forgiveness?. *Journal of Personality and Social Psychology*, *82*(6), 956.

Flynn, F. J., & Schaumberg, R. L. (2012). When feeling bad leads to feeling good: Guilt-proneness and affective organizational commitment. *Journal of Applied Psychology*, *97*(1), 124.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, 115(3), 811-846.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528.

Grant, A. M., & Wrzesniewski, A. (2010). I won't let you down… or will I? Core self-evaluations, other-orientation, anticipated guilt and gratitude, and job performance. *Journal of Applied Psychology*, *95*(1), 108.

Gunnthorsdottir, A., McCabe, K., & Smith, V. (2002). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology*, *23*(1), 49-66.

Hannon, P. A., Rusbult, C. E., Finkel, E. J., & Kamashiro, M. (2010). In the wake of betrayal: Amends, forgiveness, and the resolution of betrayal. *Personal Relationships*, *17*(2), 253-278.

Hardin, Russell (2004). Trust and trustworthiness. NY: Russell Sage Foundation.

Haselhuhn, M. P., Schweitzer, M. E., & Wood, A. M. (2010). How implicit beliefs influence trust recovery. *Psychological Science, 21*(5), 645-648.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review, 20*(2), 379-403.

Hodson, G., Book, A., Visser, B. A., Volk, A. A., Ashton, M. C., & Lee, K. (2018). Is the Dark Triad common factor distinct from low Honesty-Humility? *Journal of Research in Personality, 73,* 123–129. https://doi.org/10.1016/j.jrp.2017.11.012

Jakobwitz, S., & Egan, V. (2006). The dark triad and normal personality traits. *Personality and Individual Differences, 40*(2), 331-339.

Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology, 43*(6), 1306-1317.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The "Big Five" inventory: versions 4a and 54.* Berkeley, CA: Institute of Personality Assessment and Research.

Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23*(3), 531-546.

Jones, D. N., & Paulhus, D. L. (2017). Duplicity among the dark triad: Three faces of deceit. *Journal of Personality and Social Psychology, 113*(2), 329-342.

Jordan, J., Flynn, F., & Cohen, T. R. (2015). Forgive them for I have sinned: The relationship between guilt and forgiveness of others' transgressions. *European Journal of Social Psychology, 45(4),* 441-459. doi: http://dx.doi.org/10.1002/ejsp.2101

Kashy, D. A., & DePaulo, B. M. (1996). Who lies?. *Journal of Personality and Social Psychology*, *70*(5), 1037.

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: the effects of apology versus denial for repairing competence-versus integrity-based trust violations. *Journal of Applied Psychology*, *89*(1), 104.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*(1), 569-598.

Larrick, R. P. (2016). The Social Context of Decisions. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 441-467.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88-106.

Lewicki, R. J., & Bunker, B. B. (1995). Trust in relationships: A model of development and decline. Jossey-Bass.

Lewicki, R.J., McAllister, D. and Bies, R.H. (1998) Trust and distrust: New relationships and realities. *Academy of Management Review*. Vol 23, No. 3, 438-458.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991-1022.

Lönnqvist, J.-E., Verkasalo, M., Wichardt, P. C., & Walkowitz, G. (2012). Personality disorder categories as combinations of dimensions: Translating cooperative behavior in borderline personality disorder into the five-factor framework. *Journal of Personality Disorders*, 26, 298-304.

Lount, R. B., Jr. (2010). The impact of positive mood on trust in interpersonal and intergroup

interactions. *Journal of Personality and Social Psychology*, 98(3), 420 – 433.

Lount Jr, R. & Pettit, N. (2012). The social context of trust: The role of status. *Organizational Behavior and Human Decision Processes*, *117*(1), 15-23.

Lount, R. B., Zhong, C. B., Sivanathan, N., & Murnighan, J. K. (2008). Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Personality and Social Psychology Bulletin*, *34*(12), 1601-1612.

Luchies, L. B., Wieselquist, J., Rusbult, C. E., Kumashiro, M., Eastwick, P. W., Coolsen, M. K., & Finkel, E. J. (2013). Trust and biased memory of transgressions in romantic relationships. *Journal of personality and social psychology*, *104*(4), 673.

Lutwak, N., & Ferrari, J. R. (1996). Moral affect and cognitive processes: Differentiating shame from guilt among men and women. *Personality and Individual Differences*, *21*(6), 891-896.

Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes*, *94*(2), 61-73.

Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*, *47*(3), 534-559.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. 1995. An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734.

Miller, P. J., & Rempel, J. K. (2004). Trust and partner-enhancing attributions in close relationships. *Personality and Social Psychology Bulletin*, *30*(6), 695-705.

Molden, D. C., & Finkel, E. J. (2010). Motivations for promotion and prevention and the role of trust and commitment in interpersonal forgiveness. *Journal of Experimental Social Psychology*, *46*(2), 255-268.

Mulder, L. B., Van Dijk, E., De Cremer, D., & Wilke, H. A. (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology*, *42*(2), 147-162.

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The Malevolent Side of Human Nature : A Meta-Analysis and Critical Review of the Literature on the Dark Triad (Narcissism ,. *Perspectives on Psychological Science*, *12*(2), 183–204. https://doi.org/10.1177/1745691616666070

Özer, Ö., Zheng, Y., & Ren, Y. (2014). Trust, trustworthiness, and information sharing in supply chains bridging China and the United States. *Management Science*, *60*(10), 2435-2460.

Özer, Ö., & Zheng, Y. (2017). Establishing Trust and Trustworthiness for Supply Chain Information Sharing. In *Handbook of Information Exchange in Supply Chain Management* (pp. 287-312). Springer International Publishing.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556-563.

Pillutla, M. M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, *39*(5), 448-455.

Pinter, B., Insko, C. A., Wildschut, T., Kirchner, J. L., Montoya, R. M., & Wolf, S. T. (2007). Reduction of interindividual–intergroup discontinuity: The role of leader accountability and proneness to guilt. *Journal of Personality and Social Psychology*, *93*, 250-265. doi:10.1037/0022-3514.93.2.25

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95.

Rempel, J. K., Ross, M., & Holmes, J. G. (2001). Trust and communicated attributions in close relationships. *Journal of Personality and Social Psychology, 81*(1), 57.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Rusbult, C. E., Kumashiro, M., Finkel, E. J., & Wildschut, T. (2002). The war of the roses: An interdependence analysis of betrayal and forgiveness. *Understanding marriage: Developments in the study of couple interaction,* 251-281.

Sah, S., Loewenstein, G., & Cain, D. M. (2011). Insinuation anxiety: fear of signaling distrust after conflict of interest disclosures. *Available at SSRN 1970961*.

Sah, S., Loewenstein, G., & Cain, D. (2013). The burden of disclosure: Increased compliance with distrusted advice. *Journal of Personality and Social Psychology, 104*(2), 289-304.

Salamon, S. D., & Robinson, S. L. (2008). Trust that binds: The impact of collective felt trust on organizational performance. *Journal of Applied Psychology, 93*(3), 593-601.

Schaumberg, R. L., & Flynn, F. J. (2017). Clarifying the link between job satisfaction and absenteeism: The role of guilt proneness. *Journal of Applied Psychology, 102*(6), 982-992.

Schaumberg, R. L., & Flynn, F. J. (2012). Uneasy lies the head that wears the crown: the link between guilt-proneness and leadership. *Journal of Personality and Social Psychology*, *103*(2), 327.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*(1), 1-19.

Schweitzer, M. & Ho, T. (2005). Trust but verify: Monitoring in interdependent relationships, in *Experimental and Behavioral Economics* (Ed. J. Morgan), 13, 87-106.

Schweitzer, M., Ho, T., & Zhang, X. (2016) How monitoring influences trust: A tale of two faces. *Management Science*. Forthcoming.

Schweitzer, M. E., Brooks, A. W., & Galinsky, A. D. (2015). The Organizational Apology. *Harvard Business Review*, September.

Simpson, J. A. (2007). Foundations of interpersonal trust. In A. W. Kruglanski & E. T. Higgins (Eds.), Social psychology: Handbook of basic principles (2nd ed., pp. 587-607). New York: Guilford.

Spencer, S., Zanna, S., & Fong, G. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*, 845-851.

Stuewig, J., Tangney, J. P., Kendall, S., Folk, J. B., Meyer, C. R., & Dearing, R. L. (2015). Children's proneness to shame and guilt predict risky and illegal behaviors in young adulthood. *Child Psychiatry & Human Development, 46*(2), 217-227.

Tangney, J. P. (1996). Conceptual and methodological issues in the assessment of shame and guilt. *Behaviour Research and Therapy, 34*(9), 741–754.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. New York, NY: Guilford Press.

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology, 70(6),* 1256 –1269.

Tangney, J.P., Dearing, R., Wagner, P.E., & Gramzow, R. (2000). *The Test of Self-Conscious Affect – 3 (TOSCA-3).* George Mason University, Fairfax VA.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58*, 345-372.

Tangney, J. P., Stuewig, J., Furukawa, E., Kopelovich, S., Meyer, P. J., & Cosby, B. (2012). Reliability, validity, and predictive utility of the 25-item Criminogenic Cognitions Scale (CCS). *Criminal justice and behavior*, *39*(10), 1340-1360.

Tangney, J. P., Stuewig, J., & Martinez, A. G. (2014). Two faces of shame: The roles of shame and guilt in predicting recidivism. *Psychological Science*, *25*(3), 799-805.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, *58*, 345.

Thielmann, I., & Hilbig, B. E. (2015). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin*, *41*(11), 1523-1536.

Tomlinson, E. C., & Mryer, R. C. (2009). The role of causal attribution dimensions in trust repair. *Academy of Management Review*, *34*(1), 85-104.

Wildschut, T., & Insko, C. A. (2006). A paradox of individual and group morality: Social psychology as empirical philosophy. In P. A. M. Van Lange (Ed.), *Bridging social psychology: Benefits of transdisciplinary approaches* (pp. 377-384). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Wiltermuth, S. S., & Cohen, T. R. (2014). "I'd only let you down": Guilt-proneness and the avoidance of harmful interdependence. *Journal of Personality and Social Psychology*, *107*(5), 925-942.

Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. (1998). Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *Academy of Management Review***,** *23*(3), 513-530.

Yip, J. A., & Schweitzer, M. E. (2015). Trust promotes unethical behavior: excessive trust, opportunistic exploitation, and strategic exploitation. *Current Opinion in Psychology, 6*, 216-220.

Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science, 9*(2), 141-159.

Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review*, 19, 277-302.

Tables

Table 1. Descriptive Statistics and Correlations (Study 1, *N* = 401)

Scale	*M (SD)*	1. Trustworthy Intentions	2. Trustworthy Behavior	3. GASP – Guilt: Neg Behavior Eval	4. GASP – Guilt: Repair	5. GASP – Shame: Neg Self Eval	6. GASP – Shame: Withdraw
1. Trustworthy Intentions	9.29 (1.30)	α = .90					
2. Trustworthy Behavior	.51 (.500)	.231***	--				
3. GASP - Guilt: Neg Behavior Eval	5.25 (1.23)	.436***	.233***	α = .76			
4. GASP - Guilt: Repair	5.50 (.93)	.527***	.224***	.577***	α = .66		
5. GASP - Shame: Neg Self Eval	5.44 (1.05)	.375***	.105*	.613***	.536***	α = .70	
6. GASP - Shame: Withdraw	3.12 (1.09)	-.216***	-.058	-.014	-.161**	-.003	α = .62
7. Male	.66 (.48)	-.212***	-.091^{+}	-.233***	-.190***	-.271***	-.064

Note. ***p < .001, **p < .01, *p < .05, ^{+}p < .10. Means in all descriptive statistic tables reflect raw means.

Table 2. Regressions in Study 1 (*N* = 401)

Panel A: OLS Regression: Trustworthy intentions regressed on the GASP (guilt-proneness)

	Steps in the OLS Regression					
	1	2	3	4	5	6
Constant	9.287***	9.287***	9.287***	10.091***	9.866***	10.075***
	(.059)	(.055)	(.060)	(.193)	(.165)	(.187)
Guilt:NBE	.568***				.241**	.229**
	(.059)				(.073)	(.073)
Guilt:Repair		.687***			.468***	.462***
		(.055)			(.070)	(.070)
Shame:NSE			.488***		.089	.064
			(.060)		(0.071)	(0.071)
Shame:Withdraw				-.259***	-.186***	-.195***
				(.059)	(.050)	(.050)
Male						-.275*
						(.118)
Adjusted R-Squared	.188	.276	.138	.044	.322	.330

Panel B: Logistic Regression: Trustworthy behavior in the trust game regressed on the GASP (guilt-proneness)

	Steps in the Logistic Regression					
	1	2	3	4	5	6
Constant	.051	.053	.055	.055	.052	.193
	(.103)	(.103)	(.100)	(.100)	(.104)	(.184)
Guilt:NBE	.494***				.445**	.436**
	(.109)				(.145)	(.145)
Guilt:Repair		.471***			.338*	.335*
		(.108)			(.138)	(.138)
Shame:NSE			.213*		-.230	-.250+
			(.101)		(.140)	(.142)
Shame:Withdraw				-.116	-.060	-.067
				(.100)	(.106)	(.106)
Male						-.212
						(.229)
Nagelkerke R-Squared	.072	.067	.015	.004	.098	.101

Note. NBE = Negative Behavior Evaluation. NSE = Negative Self-Evaluation. Regressions were performed with standardized means of Guilt and Shame measures.
*** $p \leq .001$, ** $p \leq .01$, *$p < .05$., +$p < .10$.

Table 3. Descriptive Statistics and Correlations (Study 2, *N* = 139)

Scale	*M (SD)*	1. Trustworthy Behavior	2. GP-5	3. Anticipated Guilt	4. Sense of Interpersonal Responsibility	5. Extraversion	6. Agreeableness	7. Conscientiousness	8. Emotional Stability	9. Openness to Experience
1. Trustworthy Behavior	.42 (.50)									
2. GP-5	3.64 (.71)	.170*	α = .70							
3. Anticipated Guilt	3.48 (1.66)	.341***	.136	r = .76						
4. Sense of Interpersonal Responsibility	4.03 (1.46)	.436***	.210*	.605***	α = .91					
5. Extraversion	4.32 (1.25)	.115	.255**	.115	0.132	r = .41				
6. Agreeableness	4.91 (1.17)	.100	.395**	.200*	.256**	.128	r = .24			
7. Conscientiousness	5.52 (1.15)	.019	-.053	.013	.000	.009	.191*	r = .35		
8. Emotional Stability	4.71 (1.33)	-.047	.031	-.061	-.020	.002	.334***	.291**	r = .42	
9. Openness to Experience	5.13 (1.08)	.032	.139	-.028	.055	.244**	.286**	.171*	.183*	r = .14
10. Male	.30 (.46)	-.185*	-.182*	-.176*	-.211*	.011	-.075	-.028	.284**	.124

Note. ***p < .001, **p < .01, *p < .05, ^{+}p < .10. Means in all descriptive statistic tables reflect raw means.

Table 4. Logistic Regression: Trustworthy behavior in the trust game regressed on the GP-5 (guilt-proneness) (Study 2, *N* = 139)

	Steps in the logistic regression					
	1	**2**	**3**	**4**	**5**	**6**
Constant	-.314* (.174)	-.342 (.183)	-.402* (.195)	-.412* (.197)	-.413* (.198)	-.287 (.236)
GP-5	.356* (.180)			.262 (.211)	.262 (.211)	.231 (.213)
Anticipated Guilt		.752*** (.196)			.294 (.240)	.286 (.241)
Sense of Interpersonal Responsibility			1.083*** (.231)	1.063*** (.234)	.894*** (.267)	.861*** (.267)
Male						-.425 (.447)
Nagelkerke R-Squared	.039	.153	.254	.266	.278	.285

Note. *** $p \leq .001$, ** $p \leq .01$, *$p < .05$. Regressions were performed with standardized means of GP-5, Anticipated Guilt, and Sense of Interpersonal Responsibility.

Table 5. Descriptive Statistics and Correlations (Study 3, *N* = 399)

Scale	*M(SD)*	1. Trustworthy Behavior	2. GP-5	3. Sense of Interpersonal Responsibility	4. Extraversion	5. Agreeableness	6. Conscientious-ness	7. Neuroticism	8. Openness to Experience
1. Trustworthy Behavior	.72 (.45)								
2. GP-5	3.84 (.92)	.162***	α = .84						
3. Sense of Interpersonal Responsibility	4.55 (1.66)	.510***	.299***	α = .90					
4. Extraversion	3.81 (1.25)	.031	-.054	.046	α = .89				
5. Agreeableness	5.11 (1.00)	.150**	.343***	.199***	.288***	α = .86			
6. Conscientiousness	5.32 (.99)	.024	.180***	.002	.292***	.374***	α = .88		
7. Neuroticism	3.44 (1.33)	-.076	.051	.015	-.376***	-.445***	-.465***	α = .91	
8. Openness to Experience	4.99 (1.03)	.002	.036	.034	.308***	.260***	.173**	-.151**	α = .88
9. Male	.50 (.50)	.005	-.271***	-.050	.038	-.112*	-.106*	-.168**	-.009

Note. ***$p \leq .001$, **$p < .01$, *$p < .05$, $^{+}p < .10$. Means in all descriptive statistic tables reflect raw means.

Table 6. Logistic Regression: Trustworthy behavior in the Rely-or-Verify game regressed on the GP-5 (guilt-proneness) (Study 3, *N* = 399)

	Steps in the Logistic Regression			
	1	2	3	4
Constant	.965*** (.114)	1.252*** (.144)	1.252*** (.144)	.864*** (.167)
GP-5	.349** (.109)		.040 (.133)	.322** (.123)
Sense of Interpersonal Responsibility		1.317*** (.147)	1.306*** (.151)	
Extraversion				.045 (.132)
Conscientiousness				-.102 (.136)
Agreeableness				.236^{+} (.139)
Openness				-.074 (.123)
Neuroticism				-.083 (.146)
Male				.220 (.247)
Nagelkerke R-Squared	.036	.345	.345	.059

Note. *** $p \leq .001$, ** $p \leq .01$, *$p < .05$, $^{+}p < .10$. Regressions were performed with standardized means of GP-5, Sense of Interpersonal Responsibility, and Big 5 Personality measures.

Table 7. Descriptive Statistics and Correlations (Study 4)

Scale	*M (SD)*	1. Trustworthy Behavior	2. TOSCA: Guilt-Proneness	3. TOSCA: Shame-Proneness	4. TOSCA: Externalization	5. TOSCA: Detachment	6. TOSCA: Alpha Pride	7. TOSCA: Beta Pride	8. Sense of Interpersonal Responsibility	9. Anticipated Guilt	10. Warm Glow	11. Self interest
1. Trustworthy Behavior	.45 (.50)	--										
2. TOSCA: Guilt-Proneness	3.91 (52)	.136*	α = .79									
		n = 302										
3. TOSCA: Shame-Proneness	3.11 (.59)	.011	.551***	α = .79								
		n = 302	n = 405									
4. TOSCA: Externalization	2.31 (.61)	-.133*	-.030	.467***	α = .82							
		n = 302	n = 405	n = 405								
5. TOSCA: Detachment	2.79 (.57)	-.078	-0.099*	.169**	.598***	α = .72						
		n = 302	n = 405	n = 405	n = 405							
6. TOSCA: Alpha Pride	3.85 (.61)	-.138*	.305***	.130**	.145**	.325***	α = .62					
		n = 302	n = 401	n = 401	n = 401	n = 401						
7. TOSCA: Beta Pride	3.90 (.61)	-.071	.350***	.088^{+}	.087^{+}	.291**	.803***	α = .62				
		n = 302	n = 401	n = 401	n = 401	n = 401	n = 401					
8. Sense of Interpersonal Responsibility	4.54 (1.59)	.375***	.300***	.126*	-0.077	-.038	.033	0.086	α = .88			
		n = 295	n = 348	n = 348	n = 348	n = 348	n = 348	n = 348				
9. Anticipated Guilt	4.10 (1.79)	.320***	.252***	.122*	-.003	-.037	.012	.054	.807***	r = .65		
		n = 295	n = 348	n = 348	n = 348	n = 348	n = 348	n = 348	n = 351			
10. Warm Glow	4.16 (1.59)	.296***	.229***	.108*	-.023	-.029	-.001	.043	.672***	.648**	r = .61	
		n = 295	n = 348	n = 348	n = 348	n = 348	n = 348	n = 348	n = 351	n = 351		
11. Self interest	4.66 (1.35)	-.262***	-.013	.108*	.076	.096^{+}	.149**	.140**	-.185***	-.168**$^{+}$	-.069	α = .62
		n = 295	n = 348	n = 348	n = 348	n = 348	n = 348	n = 348	n = 351	n = 351	n = 351	
12. Male	.25 (.43)	-.099	-.328***	-.208***	.144**	.142**	-.100*	-.123*	-.134*	-.131*	-.090	.013
		n = 298	n = 391	n = 391	n = 391	n = 391	n = 391	n = 391	n = 345	n = 351	n = 345	n = 345

Note. ***p < .001, **p < .01, *p < .05, ^{+}p < .10. Means in all descriptive statistic tables reflect raw means.

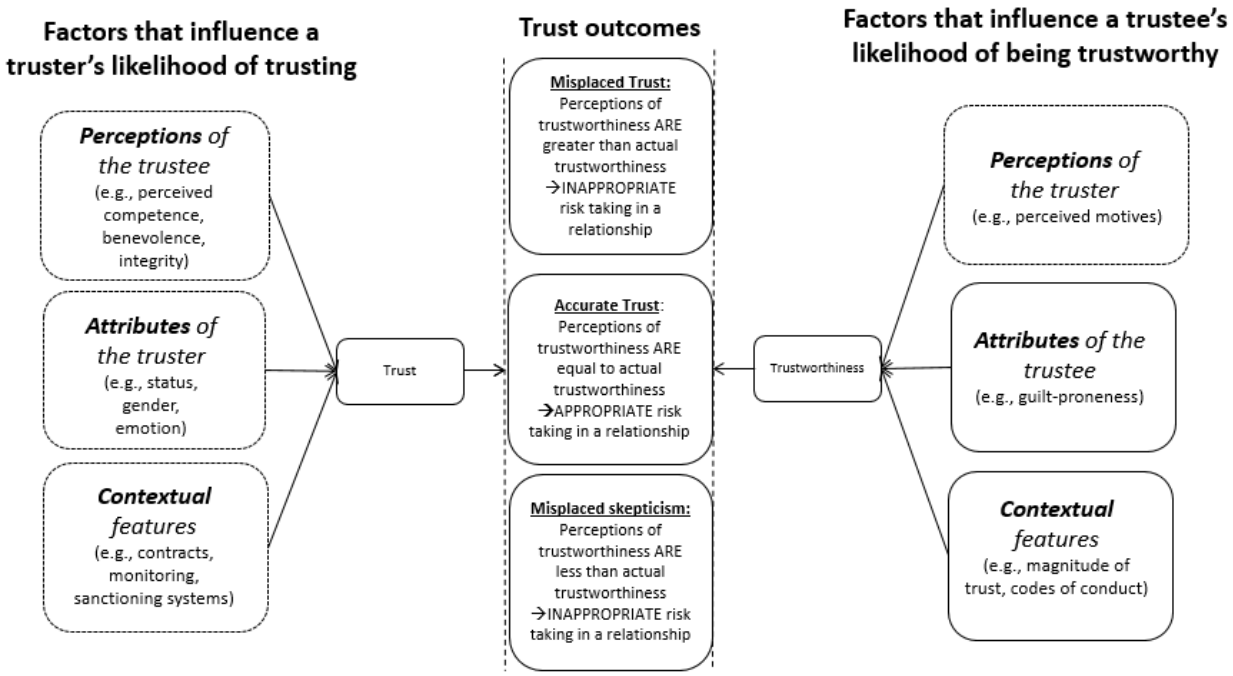
Table 8. Logistic Regression: Trustworthy behavior in the trust game regressed on the TOSCA (guilt-proneness) (Study 4)

	Steps in the Logistic Regression					
	1	2	3	4	5	6
Constant	-.259*	-.262*	-.340**	-.348**	-.369**	-.315*
	(.116)	(.117)	(.128)	(.129)	(.132)	(.158)
Guilt Proneness (TOSCA)	.264*	.435**	-.025	-.039	-.020	.075
	(.120)	(.149)	(.138)	(.140)	(.141)	(.183)
Shame Proneness (TOSCA)		-.285*				-.174
		(.145)				(.165)
Anticipated Guilt				.051	-.007	.015
				(.215)	(.222)	(.226)
Warm Glow				.161	.234	.256
				(.178)	(.184)	(.187)
Self Interest					-.524***	-.526***
					(.139)	(.143)
Sense of Interpersonal Responsibility			.971***	.831***	.816***	.786**
			(.160)	(.241)	(.252)	(.255)
Male						-.321
						(.341)
Nagelkerke R-Squared	.022	.038	.207	.211	.264	.278
N	309	309	309	309	309	302

Note. *** $p \leq .001$, ** $p \leq .01$, *$p \leq .05$, $^{+}p < .10$. Regressions were performed with standardized means of Guilt-Proneness, Shame-Proneness, Sense of Interpersonal Responsibility, Anticipated Guilt, Warm Glow, and Self Interest.

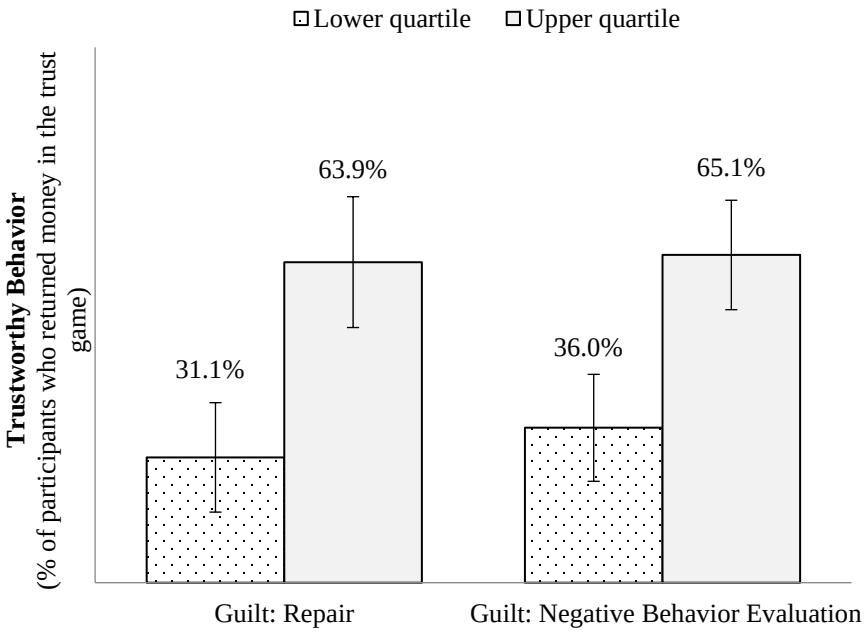
Figures

Figure 1. A complete model of interpersonal trust



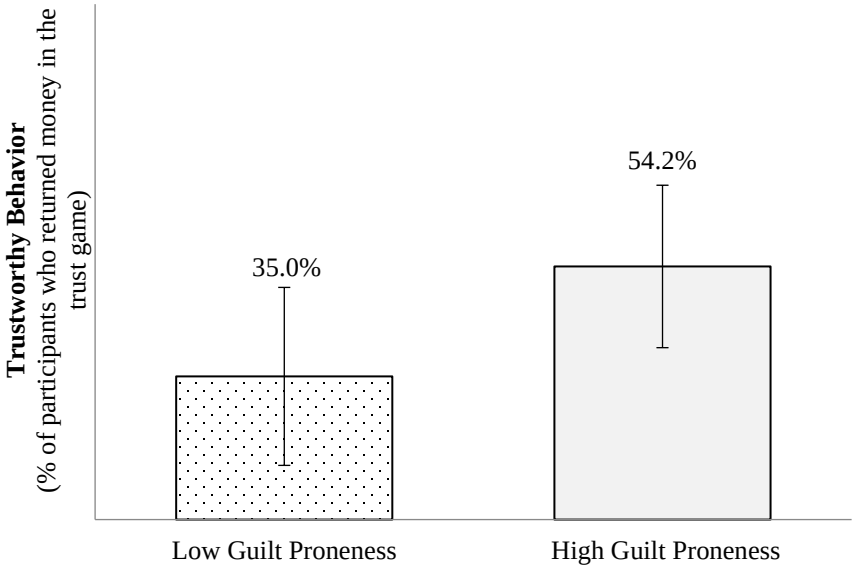
Note. Existing research has primarily examined factors that influence the truster's likelihood of trusting (left-hand side of the model). In the present research, we shift focus to the right-hand side of the model by examining one attribute of the trustee (guilt-proneness) that influences trustworthiness. We encourage future research to examine other factors that influence trustworthiness, such as the trustee's perception of the truster, as well as the outcomes of appropriate and inappropriate levels of trust.

Figure 2. Trustworthy behavior in the trust game as a function of Guilt: Repair and Guilt: Negative Behavior Evaluation (Study 1)



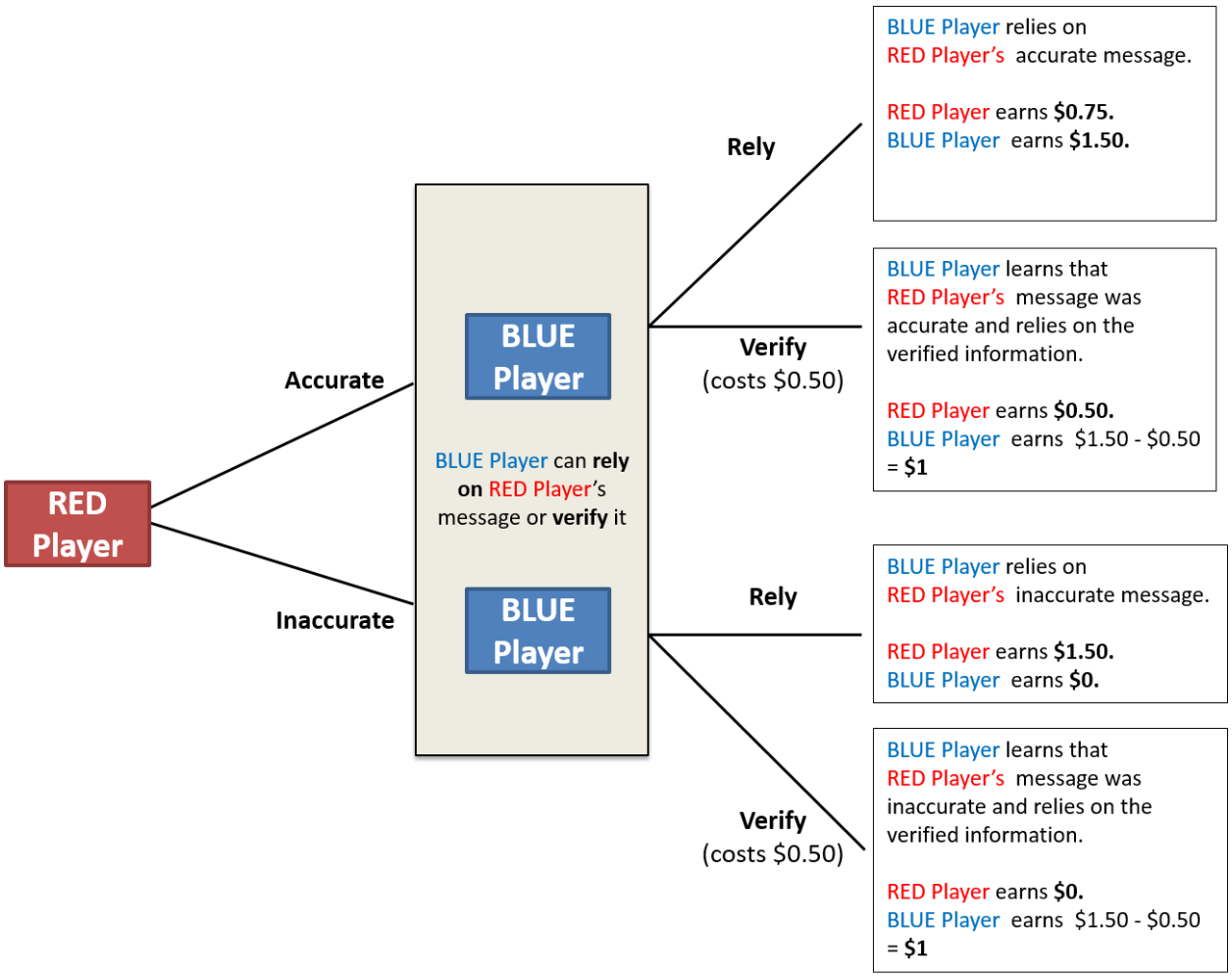
Note. Low and high values of Guilt: Repair reflect participants who scored in the lower quartile (n = 103) and upper quartile (n = 72) of the Guilt: Repair subscale of the GASP. Low and high values of Guilt: Negative Behavior Evaluation reflect participants who scored in the lower quartile (n = 111) and upper quartile (n = 106) of the Guilt: Repair subscale of the GASP. Error bars represent 95% confidence intervals.

Figure 3. Trustworthy behavior in the trust game as a function of guilt-proneness (Study 2)



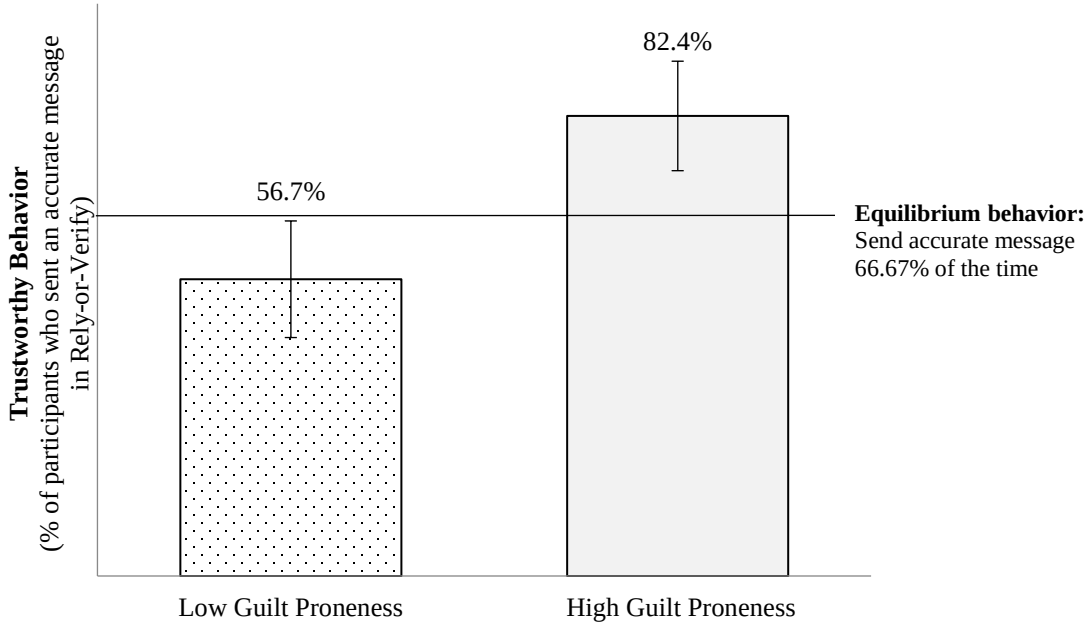
Note. Low and high values of guilt-proneness reflect participants who scored in the lower quartile (n = 40) and upper quartile (n = 48) of guilt-proneness (GP-5). Error bars represent 95% confidence intervals.

Figure 4. The Rely-or-Verify Game



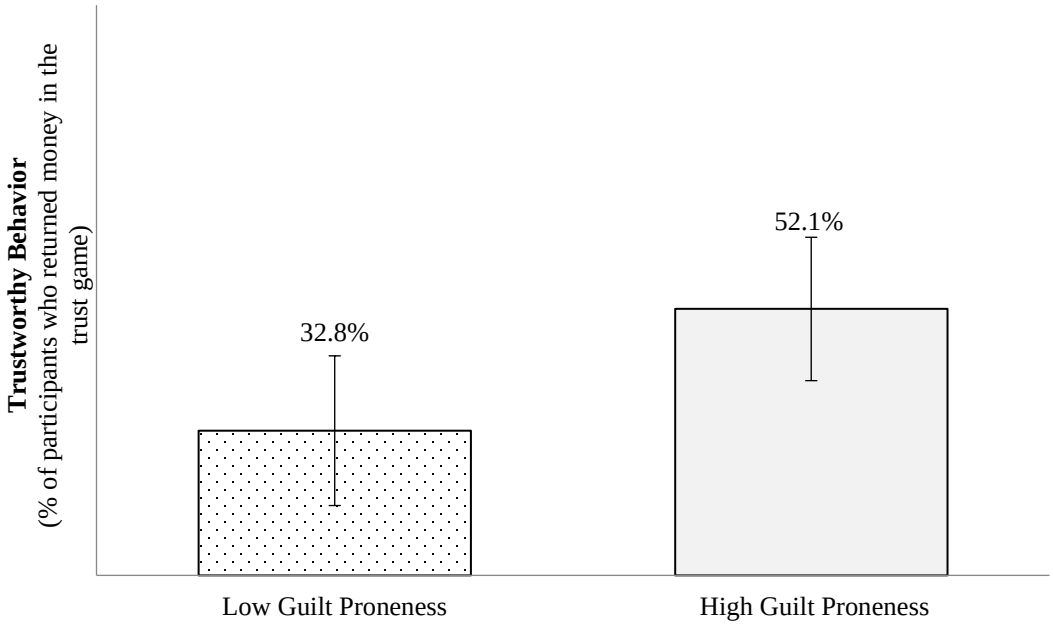
Note. Adapted with permission from: Levine, E. E., & Schweitzer, M. E. (2015) Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes,* 126, 88-106.

Figure 5. Trustworthy behavior in the Rely-or-Verify as a function of guilt-proneness (Study 3)



Note. Low and high values of guilt-proneness reflect participants who scored in the lower quartile (n = 90) and upper quartile (n = 102) of guilt-proneness (GP-5). Error bars represent 95% confidence intervals.

Figure 6. Trustworthy behavior in the trust game as a function of guilt-proneness (Study 4)



Note. Low and high values of guilt-proneness reflect participants who scored in the lower quartile (n = 67) and upper quartile (n = 73) of guilt-proneness (TOSCA-3). Error bars represent 95% confidence intervals.

Figure 7. Trustworthy behavior in the trust game as a function of guilt-proneness and trustee's vulnerability (Study 5, *N* = 402)

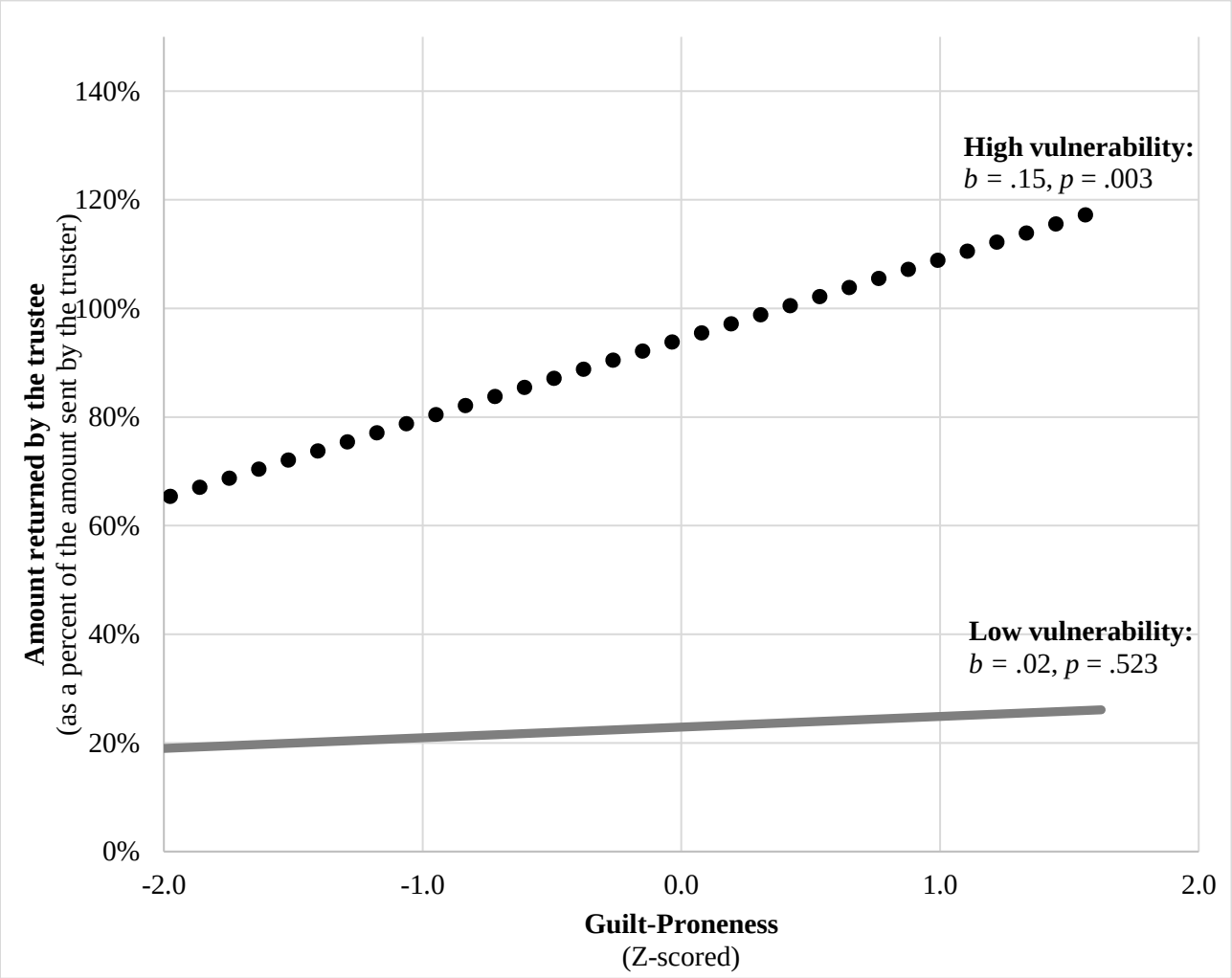


Figure 8. Sense of responsibility manipulation (Study 6)

Panel A. Instructions in the High Responsibility Condition

Amazon Mechanical Turk Code of Conduct

Please read carefully.

MTurk is made up of thousands and thousands of workers, who each rely on the decisions of researchers, employers, and fellow Mturkers, to earn wages.

Please consider the needs of others as you complete your studies today. You have a responsibility to all stakeholders of the Mturk platform. Please treat others responsibly in all of your studies today.

Please type the following statement in the space below:

I will act responsibly on Mturk today.

Panel B. Instructions in the Low Responsibility Condition

Amazon Mechanical Turk Instructions

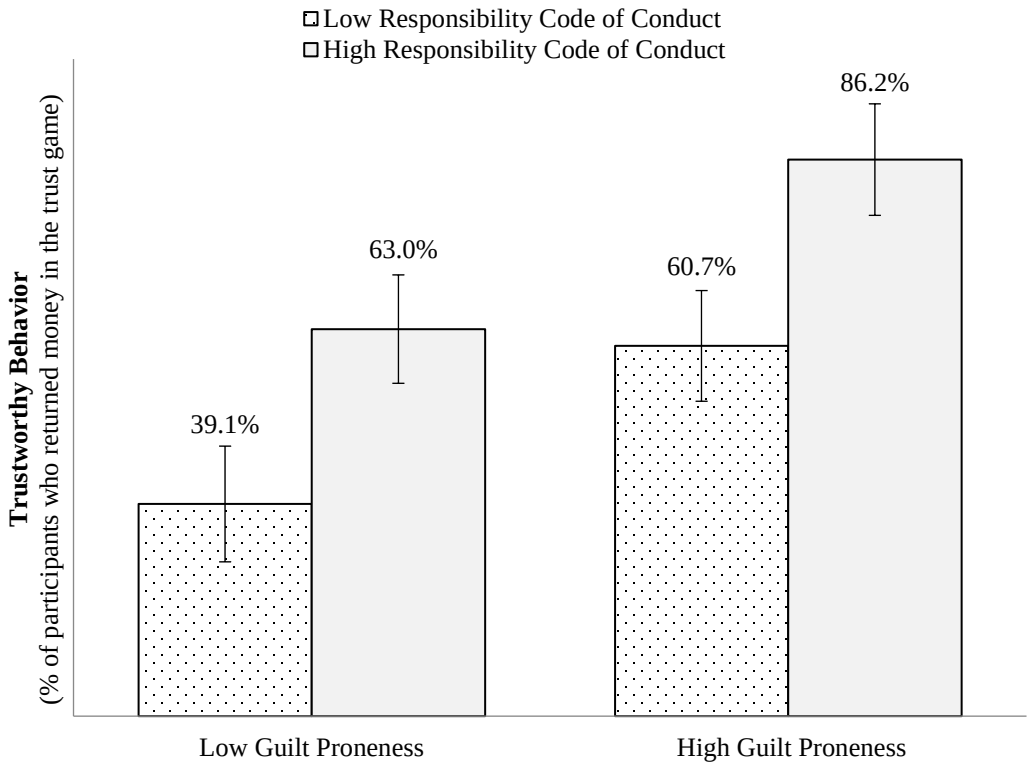
Please read carefully

MTurk is made up of thousands and thousands of workers, who each rely on the decisions of researchers, employers, and fellow Mturkers, to earn wages.

Please consider that every participant is looking out for him or herself when they complete their studies. You should too. Do what is best for yourself in all of your studies today.

Please retype your MTurk ID in the space below:

Figure 9. Trustworthy behavior in the trust game as a function of guilt-proneness and manipulated sense of responsibility via codes of conduct (Study 6, *N* = 552)



Note. Low and high values of guilt-proneness reflect participants who scored in the lower and upper third of guilt-proneness (GP-5) in an initial screening. Error bars represent 95% confidence intervals.

Appendix

Trustworthy Intentions Scale Validation

We conducted a pilot study to validate a new measure of trustworthy intentions, and establish its discriminant validity from trusting attitudes.

Method

We recruited 201 adults (42% female; M_{age} = 31 years, SD = 9.27; $M_{work\ experience}$ = 11.45 years, SD = 8.48) to participate in an online study via Amazon Mechanical Turk in exchange for a small payment ($0.35). Participants completed items designed to measure trusting attitudes and trustworthy intentions, in a random order. We provide the exact items in the Table A1. All items were measured with an 11-point scale anchored at 1 = "Never" and 11 = "Absolutely all of the time."

Trusting attitudes scale. We adapted the Johnson-George & Swap (1982) Specific Interpersonal Trust Scale to measure trusting attitudes. We altered the scale so that each question focused on general trust rather than trust in a specific person. Each of the eight trusting attitudes items reflects the belief that others are trustworthy.

Trustworthy intentions scale. We developed new items to capture trustworthy intentions by adapting each of the eight items in the trusting attitudes scale to focus on the perspective of the trustee (e.g., participants' intentions to behave in a trustworthy manner), rather than perceptions of a target's trustworthiness.

Results

We conducted an exploratory factor analysis in Mplus version 7.2 using maximum likelihood estimation and oblique (geomin) rotation. We examined one-factor, two-factor, three-factor, and four-factor solutions. The first five eigenvalues were 7.63, 2.82, 0.95, 0.65, and 0.56.

A scree plot of these eigenvalues suggested that two or three factors explain the data well. Although we expected the two-factor model to have the best fit, the overall model fit statistics indicated that the three-factor model was a better fit to the data than the two-factor model: $\chi^2_{difference}(14) = 89.51$, $p < .001$. Two-factor model fit: RMSEA = .084 (90% CI = .070 to .099); CFI = .942; TLI = .922; SRMR = .037; $\chi^2(89, N = 201) = 215.88$, $p < .001$. Three-factor model fit: RMSEA = .058 (90% CI = .040 to .076); CFI = .977; TLI = .963; SRMR = .023; $\chi^2(75, N = 201) = 126.37$, $p = .002$.

Examination of the factor loadings indicated that the trustworthy intentions items loaded on their own factor (factor 1), but the trusting attitudes items loaded on two separate factors (see Table A1). The third factor was composed exclusively of double-loadings—no item loaded more strongly on the third factor than on factor 1 (trustworthy intentions) or factor 2 (trusting attitudes). Factor 1 (trustworthy intentions) was moderately correlated with factor 2 (trusting attitudes), $r = .47$, and with factor 3, $r = -.43$, $p < .05$; factor 2 (trusting attitudes) was not significantly correlated with factor 3, $r = -.02$.

The internal consistency reliability of the eight trustworthy intentions items was very good ($\alpha = .94$). Despite the double loadings on factors 2 and 3 in the factor analysis, the internal consistency reliability of the eight trusting attitudes items was also good ($\alpha = .88$).

The mean level of trustworthy intentions in this sample was 9.40 ($SD = 1.44$), and the mean level of trusting attitudes was 7.35 ($SD = 1.63$), suggesting high overall levels of trustworthy intentions and trusting attitudes. Women ($M = 9.88$, $SD = 1.09$) reported greater trustworthy intentions than men ($M = 9.06$, SD = 1.56), $t(199) = 4.12$, $p < .001$. Trusting attitudes were similar for women ($M = 7.31$, $SD = 1.66$) and men ($M = 7.38$, $SD = 1.60$), $t(199) = -0.28$, p = .77.

Table A1. Factor loadings from an exploratory factor analysis of trusting attitudes and trustworthy intentions items.

	Item	Factor 1	Factor 2	Factor 3
Trustworthy Intentions	1. If I promised to do a favor for someone, I would follow through.	**0.906***	-0.010	0.012
	2. If I borrowed something of value and returned it broken, I would offer to pay for the repairs.	**0.780***	0.028	-0.129
	3. If someone loaned me money, I would pay them back as soon as I could.	**0.894***	**-0.125***	-0.054
	4. If I were going to give someone a ride somewhere and didn't arrive on time, I would have a good reason for the delay.	**0.809***	0.015	0.020
	5. If I knew what kinds of things hurt people's feelings, they would never have to worry that I would use them against them.	**0.817***	0.006	**0.352***
	6. If I decided to meet someone for lunch, I would definitely be there.	**0.772***	0.042	-0.092
	7. I would never intentionally misrepresent someone else's point of view to others.	**0.887***	-0.049	0.139
	8. People can expect me to tell them the truth.	**0.840***	0.008	0.079
Trusting Attitudes	9. If someone promised to do me a favor, I believe that the person would follow through.	0.141	**0.702***	-0.179
	10. If someone borrowed something of value and returned it broke, I believe the person would offer to pay for the repairs.	0.162*	**0.641***	-0.012
	11. I would be willing to lend someone almost any amount of money, because I generally believe that others would pay me back as soon as they could.	-0.142	**0.549***	**0.356***
	12. If someone were going to give me a ride somewhere and the person didn't arrive on time, I would generally believe there was a good reason for the delay.	0.066	**0.627***	-0.184
	13. If someone knew what kinds of things hurt my feelings, I generally would not worry that the person would use them against me, even if our relationship changed.	0.074	**0.689***	**0.275***
	14. If I decided to meet someone for lunch, I would be certain the person would be there.	-0.007	**0.755***	**-0.480***

15. Generally, I believe that others would never intentionally misrepresent my point of view to others.	-0.025	**0.802***	**0.308***
16. Generally, I expect that others will tell me the truth.	0.035	**0.763***	0.047

Note. N= 201. All items were measured on 11-point rating scales anchored at 1 = "Strongly disagree" and 11 = "Strongly agree". **p* <. 05