**Valid and Reliable Measures of Generalized Trust:
Evidence from a Nationally Representative Survey and Behavioral Experiment¹**

Blaine G. Robbins
New York University Abu Dhabi

May 24th, 2022

Word Count: 18,323 (abstract, text, references)

Key words: generalized trust, measurement validation, confirmatory factor analysis

Corresponding author: Blaine Robbins, New York University Abu Dhabi, Building A5 1191, PO Box 129188, Abu Dhabi, United Arab Emirates.

Email: bgr3@nyu.edu

¹ I am indebted to Elisa Bienenstock, Jennifer Glanville, and Maria Grigoryeva for helpful conversations and/or suggestions. Participants of the Spanish National Research Council's Institute of Public Goods and Policies seminar series provided invaluable comments, particularly Sara Degli Esposti, Diego Gambetta, and Francisco Herreros. Finally, I would like to thank John Dombrowski, Stefan Subias, and Mark Watts from NORC for their research assistance.

Abstract

Across the social sciences, generalized trust has been one of the most frequently examined constructs since researchers first introduced measures of it in the 1950s. Despite its significance, there is growing consensus that conventional measures of generalized trust are prone to measurement invalidity and non-equivalence, which places sociological knowledge about generalized trust in serious doubt. In this article, I advance trust research in sociology by (1) refining two new self-report measures of generalized trust—the Stranger Face Trust scale (SFT) and the Imaginary Stranger Trust scale (IST)—and (2) assessing their empirical performance on a nationally representative probability sample ($N = 1,264$). I compare the reliability and validity of SFT, IST, and traditional measures of generalized trust across a number of measurement validation tests. Results suggest that SFT provides the most accurate and consistent measure of generalized trust. I conclude by discussing the implications of my findings for sociological knowledge.

Introduction

In the study of beliefs and values, generalized trust has been one of the most continuously examined constructs since researchers first introduced measures of it in the 1950s. Rotter, early on, defined generalized trust as a “generalized expectancy” of a certain level of reliability from strangers that is learned over a lifetime (Rotter 1967: 653). More recent definitions view generalized trust as the optimism and unconditional faith people have in unknown others (Uslaner 2002; Yamagishi and Yamagishi 1994), or the baseline starting point at which people (dis)trust strangers for any number of matters (Bauer and Freitag 2018; Colquitt et al. 2007; Erikson 1950; Robbins 2016b). Simply put, generalized trust is the idea that people possess a “standard estimate of the probability of trustworthiness, p^* , for the average person he [or she] meets (Coleman 1990: 104; see also Hardin 2002, pp. 116-126).”

Throughout this literature, leading scholars consider trust in generalized others as a core feature of civil society, political culture, and social capital (Coleman 1990; Fukuyama 1995; Inglehart 1997; Paxton 2007; Putnam 1995; Uslaner 2002; Verba et al. 1995). In modern society, generalized trust is the “basis of reciprocity, social connectedness, peaceful collective action, inclusiveness, tolerance, gender equality, confidence in institutions, and democracy itself (Delhey et al. 2011: 787).” Because of this—and the fact that generalized trust “lubricate[s] the inevitable frictions of daily life (Putnam 2000: 135)” —countless studies are published annually on the causes and consequences of generalized trust at multiple levels of analysis (see Dinesen et al. 2020; Kramer 1999; Nannestad 2008; Schilke et al. 2021; Smith 2010; Uslaner 2018 for reviews), with some pointing to the deterioration of community and politics in the United States as a result of declines in generalized trust (Paxton 1999; Putnam 2000).

Notwithstanding its importance, problems of measurement invalidity and non-equivalence plague studies of generalized trust. Due to these measurement problems, “it is difficult for researchers to identify whether scientific discoveries using [traditional measures] are—or will be—a function of true patterns and relations or an artifact of measurement (Robbins 2021: 30).” According to some scholars, these issues of validity are so grim that if sociologists continue to use traditional measures, the accumulation of sociological knowledge about generalized trust will undoubtedly suffer (Baur and Freitag 2018; Hardin 2006; Robbins 2021, 2022). Despite attempts by researchers to correct this deficiency over the past two decades (Bauer and Freitag 2018; Delhey et al. 2011; Miller and Mitamura 2003; Reeskens and Hooghe 2008; Sturgis and Smith 2010; Torpe and Lolle 2011; van Hoorn 2014), major advances in the measurement of generalized trust have been rare.

Recently, however, trust scholars have proposed a solution. To properly measure generalized trust, Bauer and Freitag (2018) argue that it is necessary to average trust “across many situations that [entail] a variety of trustees, expected behaviors, and contexts. To avoid confusion, we suggest using the term ‘cross-situational trust’ (p. 29).” In a series of papers, Robbins (2021, 2022) actualized Bauer and Freitag’s (2018) vision by designing two new measures of generalized trust: a Stranger Face Trust Scale (SFT) and an Imaginary Stranger Trust Scale (IST). The first scale, SFT, shows respondents the faces of six strangers drawn from the Chicago Face Database (Ma et al. 2015), strangers who vary according to their perceived gender (men and women) and race (Whites, Blacks, Latinx). Respondents then assess the degree to which they trust each stranger along four matters (e.g., looking after a loved one, repaying a loan, etc.). The second scale, IST, asks respondent to imagine meeting a total stranger for the first time, and then to identify the extent to which they would trust this imaginary stranger along

the four matters previously mentioned. SFT was created as a long-form questionnaire, while IST was fashioned as a short-form questionnaire for easy implementation by the research community.

With these newly proposed scales, Robbins (2021, 2022) assessed their face validity and content validity, and performed various measurement validation tests. Using a mix of cross-sectional and two-wave general population convenience samples as well as a behavioral economic subject pool, Robbins (2021, 2022) observed strong empirical support for the reliability, validity, and measurement invariance of SFT and IST. Robbins (2021, 2022) also detailed how traditional measures of generalized trust are less valid and exhibit greater measurement non-equivalence than SFT and IST.

The current study develops broader versions of SFT and IST, and demonstrates their utility as standard measures of generalized trust. I improve upon previous research by including two new racial categories (East Asian and South Asian) and two new low-cost matters, which produces four short- and long-form versions of SFT and IST: SFT-24 (six faces, four matters), SFT-60 (ten faces, six matters), IST-4 (four matters), and IST-6 (six matters). Analyses of a nationally representative probability sample ($N = 1,264$) and a follow-up study of a one-shot investment game ($N = 600$), support the conclusions of Bauer and Freitag (2018) and Robbins (2021, 2022): operationalizations of “cross-situational trust” provide reliable and valid measures of generalized trust. Specifically, SFT-24 and SFT-60 are more reliable (internal consistency) and valid (convergent, discriminant, and predictive) than IST-4 and IST-6, which are, in turn, more reliable and valid than the traditional measures of generalized trust. The consistency of results across studies indicates the quality in which SFT and IST measure generalized trust. The widespread implementation of SFT and IST will (1) ensure that researchers are validly and reliably measuring the concept of generalized trust, (2) reduce measurement error on the left-

and right-hand side of statistical models, and (3) increase our ability to compare results across groups and time.

A Brief History of Measuring Self-Reports of Generalized Trust

One of the first measures of generalized trust was a single item “most-people trust” question developed by the Office of Public Opinion Research in the United States (Opor 1942). The survey question asked: “Do you think most people can be trusted?” with response options of “yes”, “no”, and “no opinion.” A decade and a half later, Rosenberg (1956) refined this measure and developed a balanced version of the question: “Some people say that most people can be trusted. Others say you can’t be too careful in your dealings with people. How do you feel about it?” This question, along with four other items, was used by Rosenberg (1956) in a “faith in people” Guttman scale. About a decade later, Almond and Verba (1963) employed Rosenberg’s questions in their pioneering comparative study of civic culture. Their study subsequently laid the groundwork for contemporary research on political culture, social capital, and trust.

Modified versions of these questions have been used for decades in some of the largest general population surveys, including the U.S. General Social Survey, European Values Study, Gallup World Poll, and World Values Survey. These datasets inform much of what we know about generalized trust, such as the positive age-trust correlation (Clark and Eisenstein 2013; Putnam 2000; Robinson and Jackson 2001), ethnoracial differences in generalized trust (Douds and Wu 2018; Evangelist 2021; Wilkes 2011), the links between personality and generalized trust (Andersson 2012; Freitag and Bauer 2016; Uslaner 2002), and contextual effects like ethnic heterogeneity, intentional homicide, and income inequality (Alesina and La Ferrara 2002; Marschall and Stolle 2004; ; Messner et al. 2004; Putnam 2000, 2007; Rosenfeld et al. 2007; Rothstein and Uslaner 2005; Uslaner 2002, 2012).

Since Rosenberg's seminal work, interest in the self-report measurement of generalized trust has risen sharply, with important lines of research addressing: (1) the number of items to include; (2) the uni- or multi-dimensionality of generalized trust; (3) the length of answer scales; and (4) the extent to which scales and concepts in questions about generalized trust are interpreted the same way across respondents. I review this work in greater detail below.

First, there is considerable debate over whether to use single-item measures of generalized trust versus multi-item scales. Uslander (2011, 2015) strongly defends the use of standard single-item measures such as the "most-people trust" question, while others—who draw on classic test theory—contend that multi-item scales increase reliability and allow for a correction of measurement error (Freitag and Bauer 2013; Glanville and Paxton 2007).

Assuming that a set of indicators are valid measures of a latent construct, multi-item scales should be used over single-item measures (Kline 2005). In sociology and political science, for instance, generalized trust is often measured using a modified version of Rosenberg's (1956) "faith in people" scale, known as the "misanthropy scale" (Brehm and Rahn 1997; Paxton 1999; Reeskens and Hooghe 2008; Zmerli and Newton 2008), or a multi-item generalized social trust scale asking about trust in people of another nationality, another religion, and people you meet for the first time (Freitag and Bauer 2013; Delhey et al. 2011; Newton and Zmerli 2011; Robbins 2016a). The advantage of these well-known multi-item generalized trust scales is that they adjust for measurement error with a limited number of survey items, making them attractive for large-scale population surveys with limited space. In spite of these advances, a considerable body of work in the social sciences continues to use single-item measures of generalized trust (Alesina and La Ferrara 2002; Kiley and Vaisey 2020; Knack and Keefer 1997; La Porta et al. 1997; Lee

2013; Michalopoulos and Xue 2021; Nunn and Wantchekon 2011; Paxton 2007; Paxton and Glanville 2015; Robbins 2012; Tabellini 2010; You 2012).

The second historical development in the measurement of generalized trust concerns its dimensionality; that is, whether there are different forms of generalized trust that can be captured by measurement models (Delhey et al. 2011; Freitag and Bauer 2013; Freitag and Traunmüller 2009; Newton and Zmerli 2011; Whiteley 2000). For instance, if a subset of indicators (among a larger set) correlates with each other but not with others, this suggests that the subset of indicators captures something different than generalized trust, like particularized trust (Freitag and Traunmüller 2009; Glanville and Paxton 2007), identity-based trust (Freitag and Bauer 2013), or in-group/out-group trust (Delhey et al. 2011). Identifying these alternative forms of trust has allowed scholars to further refine the measurement of generalized trust, and improve statistical estimates.

The third, more recent, development concerns the use of different answer scales (Lundmark et al. 2016; Uslaner 2011, 2015); that is, whether indicators of generalized trust should use binary response options (e.g., “yes” or “no”) or longer answer scales (e.g., “do not trust at all”, “do not trust very much”, “trust somewhat”, etc.). Evidence indicates that longer answer scales are more advantageous than dichotomous or binary response options (Krosnick and Fabrigar 1997; Lundmark et al. 2016). 7- and 11-point answer scales tend to produce stronger correlations with criterion measures than dichotomous answer scales, and increase the detail in which latent factors can be measured.

The fourth and perhaps most contentious debate concerns equivalence of measurement; that is, whether answer scales and terms used in questions of generalized trust are interpreted the same way across groups, respondents, and time (Delhey et al. 2011; Freitag and Bauer 2013;

Reeskens and Hooghe 2008; Uslaner 2002; Sturgis and Smith 2010; Torpe and Lolle 2011).

Issues of measurement equivalence are of fundamental importance. If one group of respondents interprets survey questions about generalized trust differently than another group of respondents, then an observed difference between groups would not be a function of true differences in generalized trust. These differences, instead, would be an artifact of measurement. As Reeskens and Hooghe (2008) ask: “Does the concept of ‘most people’ have the same meaning for a respondent in the middle of the metropolitan areas in London or Paris, as it has for a respondent in some remote village in the north of Sweden? (p. 516)”

Most research in this area suggests that traditional measures of generalized trust exhibit measurement non-equivalence (or measurement non-invariance). Early studies using think-aloud approaches or open-ended questions found that the “most-people trust” question “brings up general evaluations of society (Uslaner 2002: 19)”, while later research showed that the concept of “trusting most people” has different meanings for different respondents (Miller and Mitamura 2003) and that the interpretation of “most people” may lead to biased responses (Sturgis and Smith 2010; see also Bauer and Freitag 2018). Likewise, research using multigroup confirmatory factor analysis has found that the three-item misanthropy scale is not sufficiently equivalent across European societies (Reeskens and Hooghe 2008), while items from the generalized social trust scale exhibit measurement equivalence across some ethnic groups and countries (Freitag and Bauer 2013). Relatedly, research by Delhey et al. (2011) and Torpe and Lolle (2011) reveals that the radius of generalized trust—or how wide a circle of generalized others respondents imagine as “most people”—varies considerably across countries (see also van Hoorn 2014).

In summary, ever since the original “most-people trust” question, measures of generalized trust have gone through phases of refinement. In spite of these developments, issues

of reliability, validity, and measurement non-equivalence remain, raising doubts about the associations and trends observed in the literature as well as the scientific community's ability to make comparisons across studies, countries, and time.

Recent Advances in the Measurement of Generalized Trust

In a series of recent papers, Blaine Robbins (2021, 2022) offers solutions to many of the debates surrounding the measurement of generalized trust. Inspired by the idea that generalized trust is best measured as “an average across many situations that entails a variety of trustees, expected behaviors, and contexts (Bauer and Freitag 2018: 29)”, Robbins (2022) proposed two new self-report measures of generalized trust: a Stranger Face Trust scale (SFT) and an Imaginary Stranger Trust scale (IST). The first scale, SFT, shows respondents the faces of six strangers drawn from the Chicago Face Database (Ma et al. 2015), and asks respondents to assess each stranger along four matters for which trust is placed. The faces vary according to gender (men and women) and race (Whites, Blacks, Latinx), while the matters concern pecuniary and nonpecuniary behaviors of general interest to individuals (e.g., looking after a loved one, repaying a loan, etc.). The second scale, IST, asks respondent to imagine meeting a total stranger for the first time and to identify the extent to which they would trust said imaginary stranger along the four matters previously mentioned (e.g., repaying a loan). SFT was created as a long-form questionnaire, while IST was fashioned as a short-form questionnaire for easy implementation by the research community.

Robbins (2022) bases the development of SFT and IST on two key assumptions:

- *Generalized trust is best assessed in a person-specific manner:* If person A is a high generalized truster, he or she should trust persons C,..., Z for any number of matters to a

greater extent than person B who is a low generalized trustor (Bauer and Freitag 2018; Coleman 1990; Cook et al. 2005; Hardin 2002; Robbins 2016b).

- *Generalized trust is best assessed in a domain-specific manner*: If person A is a high generalized trustor, he or she should trust person C for any number of matters to a greater extent than person B who is a low generalized trustor (Bauer and Freitag 2018; Coleman 1990; Cook et al. 2005; Hardin 2002; Robbins 2016b).

With these two assumptions as well as ideas from classical test theory, Robbins (2022) treats “generalized trust as a latent construct in which observed indicators of generalized trust manifest as assessments of trustworthiness for *particular matters* (in the case of SFT and IST) regarding *specific strangers* (in the case of SFT) (p. 8).” If generalized trust is a baseline starting point at which people (dis)trust strangers for any number of matters (Bauer and Freitag 2018; Colquitt et al. 2007; Erikson 1950; Robbins 2016b), then “the average of the assessments of trustworthiness across particular matters and specific strangers is the ideal measurement of generalized trust (Robbins 2022: 312).” By providing respondents with *specific strangers* and *particular matters*, Robbins (2022) argues that issues of measurement equivalence, such as interpretations of “most people” (Bauer and Freitag 2018; Miller and Mitamura 2003; Reeskens and Hooghe 2008; Sturgis and Smith 2010), the “radius of trust” problem (Delhey et al. 2011; Torpe and Lolle 2011; van Hoorn 2014), and *what* respondents trust “most people” to do (Robbins 2016b), are alleviated.

Robbins (2022) contends, as a result, that SFT and IST have greater face validity and content validity than traditional measures of generalized trust. Over a series of measurement validation tests, Robbins (2021, 2022) establishes the reliability (internal consistency, test-retest reliability), validity (factorial, convergent, discriminant, concurrent, predictive, known-groups),

and measurement equivalence of SFT and IST across sociodemographic groups, samples, and time. He also reveals that common generalized trust scales—the Misanthropy scale (MST) and Generalized Social Trust scale (GST)—appear to be less valid, less reliable, and less equivalent across sociodemographic groups than SFT and IST.

In spite of these advances, a number of questions and tests remain. MST and GST are established measures of generalized trust that have been used in representative random samples for decades, both nationally and cross-nationally. The representation of SFT and IST has yet to be established to such a degree. One key question is whether Robbins' (2021, 2022) findings hold when applied to a representative probability sample of U.S. respondents. For instance, is the negative correlation observed between age and SFT representative of U.S. adults or a product of sampling bias (see Robbins 2022)? Another key question is whether SFT and IST correlate with state-level measures of intentional homicide, ethnic fractionalization, and income inequality, as previously observed in the literature using traditional measures of generalized trust (Dinesen et al. 2020; Messner et al. 2004; Putnam 2000; Rosenfeld et al. 2007; Uslander 2002; Rothstein and Uslander 2005). Investigating whether these state-level correlations replicate for SFT and IST is an important research endeavor that will help adjudicate between existing accounts of generalized trust.

Beyond sample representation, previous validation tests in support of SFT and IST are limited to three racial categories (White, Black, Latinx) and/or high-cost pecuniary and non-pecuniary matters (e.g., keeping a secret, watching after a loved one, etc.). The current paper broadens the empirical scope of SFT and IST by featuring a greater diversity of faces (in the case of SFT) and trust domains (in the case of SFT and IST).

Finally, while the validity of SFT and IST have been established, Robbins (2022) assessed the discriminant validity of his measures on a small number of variables and tested the predictive validity of his scales on a small student sample (Robbins 2021). Given that generalized trust is a belief (Bauer and Freitag 2018; Coleman 1990; Hardin 2002; Robbins 2016), do SFT and IST discriminate from a broader set of psychological variables that measure attitudes and preferences, such as altruism and betrayal aversion (Falk et al. 2018)? And does the predictive validity of SFT and IST replicate in a larger, general population probability sample of U.S. respondents?

Data and Methods

Sample

An online survey and a behavioral economic experiment were coded and administered by the University of Chicago's National Opinion Research Center (NORC). A general population probability sample of U.S. adults age 18 years and older were selected from NORC's AmeriSpeak Panel for this study.¹ The sample was selected using sampling strata based on age, race/Hispanic ethnicity, education, and gender (48 sampling strata in total). The size of the selected sample per sampling stratum was determined by the population distribution of each stratum. In addition, sample selection took into account expected differential survey completion rates by demographic groups so that the set of panel members with a completed interview for the study was a representative sample of the target population. NORC invited 3,727 panelists to

¹ For AmeriSpeak, randomly selected U.S. households are sampled using area probability and address-based sampling, with a known, non-zero probability of selection from the NORC National Sampling Frame. These sampled households are then contacted by U.S. mail, telephone, and field interviewers (face-to-face). The panel provides sample coverage of approximately 97 percent of the U.S. household population. Those excluded from the sample include people with P.O. Box only addresses, some addresses not listed in the USPS Delivery Sequence File, and some newly constructed dwellings.

complete the online survey on May 21st, 2020. 1,264 panelists completed the survey by June 17th, 2020, yielding a survey completion rate of 33.9 percent and a weighted cumulative response rate of 6.8 percent.² Panelists who completed the online survey were compensated with 3,000 AmeriPoints (or the cash equivalent of \$3 USD). On June 12th, the incentive was increased to 5,000 AmeriPoints (or \$5USD) to help collect the remaining completed cases. The median time to complete the online survey was 16 minutes.³

Six-and-a-half months after the survey field period, a sub-sample of 875 panelists who completed the online survey were invited to participate in a one-shot, online investment game (Berg et al. 1995). In total, 604 panelists participated in the investment game between the dates of January 4th and January 27th, 2021, yielding a survey completion rate of 91.9 percent and a weighted cumulative response rate of 15.5 percent. Panelists who completed the online investment game were compensated as a function of the choices made in the game. Incentives ranged between 1,000 and 24,000 AmeriPoints. All panelists, even those whose game resulted in them receiving 0 points, were awarded at least 1,000 AmeriPoints. The median time to complete the online investment game was 5 minutes.

Participants

While slight differences exist, benchmark comparisons with the February-March 2020 Current Population Surveys revealed that the sample characteristics were representative of the

² The target sample size was 1,250. At the end of recruitment, NORC finished with an extra 14 completed surveys.

³ NORC applied cleaning rules to the survey data for quality control. In total, 25 cases were removed from the final set of completed interviews based on three cleaning rules: (1) Speeders, or those who completed the survey in less than one-third the median duration; (2) High refusal rates, or those that skipped or refused more than 50 percent of the eligible questions; and (3) Straight-liners, or those that straight-lined grid item questions. Of the 25 cases removed, 22 were marked with one of the three flags, 2 were marked with two of the three flags, and 1 was marked with three of the three flags.

U.S. population from which the AmeriSpeak panelists were randomly drawn (see Tables S1 and S2 in the Supplemental Materials online). In terms of relative majorities, 51.9 percent of the participants who completed the online survey were female, 44.1 percent had an Associate's degree or some college, 63.5 percent were non-Hispanic White, and 32.3 percent were between the ages of 18 and 34 ($M = 47.46$, $SD = 17.05$, $\min = 19$, $\max = 95$). Relative majorities were roughly the same for respondents who participated in the online investment game. For more information about sociodemographic characteristics and summary statistics for all variables used in the present study, see Table A1 in the Appendix.

Online Survey

The AmeriSpeak survey consisted of ten blocks: two—an introductory screen and a debriefing screen—were anchored at the beginning and end of the survey, respectively, while the remaining eight blocks were shown in random order from respondent-to-respondent. Each block contained a thematic set of questions, ranging from the traditional generalized trust questions to dispositional optimism scales to social value orientation scales to Big-5 personality scales. Within each block, I randomized the order of questions, the order of grid items, and the order of response scales (e.g., agree-disagree versus disagree-agree). Sociodemographic characteristics were asked by NORC prior to participation in the online survey, and were provided as deliverable preloads to the data set.

The first goal of this survey is to establish the internal consistency, factor structure, and convergent validity of SFT and IST. I do this for (a) the original scales developed by Robbins (2021, 2022), where SFT and IST consist of four trust domains (or matters) and SFT consists of six faces (male and female White, Black, and Latinx faces), and (b) new long-form versions of SFT and IST that include two new trust domains and four new faces (male and female East and

South Asian faces). Following prior research, I use MST and GST as benchmarks for convergent validity and discriminant validity. The second goal of this survey is to investigate the discriminant validity of SFT and IST with a broad set of psychological variables, including measures of economic preferences, particularized trust, and political trust. The third goal of the survey is to explore known-groups validity with personality traits and sociodemographic characteristics measured at the individual-level, and with intentional homicide, ethnic heterogeneity, and income inequality measured at the state-level. Known-groups validity is the degree to which scores on a newly proposed scale confirm hypothesized differences between known groups (e.g., confirming gender differences in self-reports of depression). The fourth and final goal is to replicate Robbins (2021) and investigate the measurement invariance of SFT-60 and IST-6 across sociodemographic groups. Below, I review and summarize the survey items and scales measured in this survey.

Stranger trust measures. In a recent set of papers, Robbins (2021, 2022) developed and validated two new scales of generalized trust: the Stranger Face Trust scale (SFT) and the Imaginary Stranger Trust scale (IST). SFT, in its original structure, consists of six faces drawn from the Chicago Face Database (Ma et al. 2015), which consists of 597 high-resolution photographs of male and female human faces of varying age and ethnicity (<https://chicagofaces.org/>). Robbins (2022) selected the six faces based on four criteria: age (median U.S. ages), race (Caucasian, African American, and Latino/a), gender (male and female), and perceived trustworthiness (neutral evaluations) normed by an independent rater sample.⁴

⁴ These criteria are important. For SFT to tap into generalized trust, it requires a diversity of faces that vary according to race and gender. Median ages capture the “median” or “average” person one might encounter in the United States. SFT uses neutral facial expressions, neutral evaluations of perceived trustworthiness, and median ages to hold such factors constant, so that trust ratings are not upwardly

For each human face, respondents are asked to “imagine meeting the following stranger for the first time. Please identify how much you would trust this stranger for each of the following.” For IST, respondents are not shown human faces. Instead, respondents are asked to “imagine meeting a total stranger for the first time. Please identify how much you would trust this stranger for each of the following.” SFT and IST contain four domains (or matters) of trust for which respondents would rate each of the six human faces (SFT) and the imaginary stranger (IST): (a) “To keep a secret that is damaging to your reputation” (SECRET); (b) “To repay a loan of one thousand dollars” (LOAN); (c) “To look after a child, family member, or loved one while you are away” (CHILD); and (d) “To provide advice about how best to manage your money” (ADVICE).⁵ Each trust domain is measured using a 4-point scale, ranging from *Do not trust at all* to *Trust completely*, with *Do not trust very much* and *Trust somewhat* in-between, with a *Don’t know* option at the end of the scale.

In the present paper, I modify the original SFT and IST scales in two ways. First, I add four new faces to SFT: male and female East and South Asian faces. Note that SFT was originally developed for respondents in the United States. The three predominant racial-ethnic groups in the United States are non-Hispanic Whites, Hispanics, and Blacks. While a strength within the U.S. context, these three racial categories may limit generalizability of SFT to the Americas and Europe. With the inclusion of East Asian and South Asian ethnoracial groups, the five racial categories (White, Black, Latinx, East Asian, and South Asian) encompass the majority of ethnoracial groups observed in the world, which strengthens the cross-cultural

biased (in the case of happy faces or high perceived trustworthiness) or downwardly biased (in the case of angry faces or low perceived trustworthiness). By relying on these four criteria, Robbins (2021, 2022) has repeatedly shown that SFT is a valid and reliable measure of generalized trust, and that traditional measures, such as the misanthropy scale or the “most people” trust question, are less valid and reliable.

⁵ Robbins (2022) used three criteria to select the four trust domains: generality, commonality, and cost. Kindly see Robbins (2022) for a longer discussion of these selection criteria.

equivalence of SFT. The two East Asian faces were drawn from the Chicago Face Database. The neutral expression photographs used were AF-200 (age = 32, number of raters = 28, perceived Asian = 100%, perceived trustworthiness = 3.92 on 7-point scale) and AM-205 (age = 34, number of raters = 24, perceived Asian = 96%, perceived trustworthiness = 3.87 on 7-point scale). Since the Chicago Face Database does not contain South Asian faces, I selected two faces from the FaceResearch.org data base (DeBruine 2021). The Face Research lab does not norm photographs by an independent rater sample; it does, however, provide the gender and self-reported race of each photographed face. See the Supplemental Materials online for headshots of all ten human faces.

Second, I add two new trust domains to SFT and IST. Three criteria were used to select the four original trust domains (Robbins 2022): generality, commonality, and cost. Given these three criteria, the four trust domains were intended to (1) span specific instances of a particular matter (*generality*), (2) concern matters for which people have trusted, frequently do trust, or can easily imagine trusting others to maintain (*commonality*), and (3) capture matters that are salient to respondents if trust was broken or betrayed (*costly*). By using four trust domains that are general, common, and costly, Robbins (2022) argues that “the research community can better discriminate between high and low social trusters (p. 42)” while maintaining measurement equivalence (Robbins 2021). In spite of these criteria, one drawback of the four trust domains is that they concern matters for which people typically trust friends, family, and acquaintances, not necessarily strangers. I thus add two new trust domains that concern matters for which people commonly (dis)trust strangers (Cohn et al. 2019): one explicitly pecuniary domain (e) “To return a lost wallet or purse” (WALLET), and one non-pecuniary (f) “To help change a flat tire on your passenger car” (TIRE).

Measures of convergent validity. Convergent validity is determined by assessing (a) the degree of confidence we have that a latent construct is well measured by its observed indicators (Campbell and Fiske 1959), and (b) the degree to which an operationalization is similar to (or converges on) other operationalizations that it theoretically should be similar to (Pedhazur and Schmelkin 1991). Factorial validity (i.e., internal consistency, reliability, and factor structure) is commonly used to establish the former (a), while correlations between latent constructs are frequently employed to determine the latter (b).

Two instruments common to the General Social Survey and World Values Survey were used as benchmark comparisons for SFT/IST and to establish convergent validity between closely related latent constructs: a 3-item *Misanthropy Scale* (MST; Brehm and Rahn 1997; Glanville et al. 2013; Paxton 1999; Mewes et al. 2021) and a 3-item *Generalized Social Trust* scale (GST; Delhey et al. 2011; Newton and Zmerli 2011; Robbins 2016a). MST consists of 3 items scored on a dichotomous answer scale. Item 1 (TRUST) asks “Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?” with *most people can be trusted* and *need to be very careful in dealing with people* as anchors with a *don’t know* option. Item 2 (FAIR) asks “Do you think that most people would try to take advantage of you if they got the chance or would they try to be fair?” with *most people would try to take advantage of me* and *most people try to be fair* as anchors with a *don’t know* option. Item 3 (HELP) asks “Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves?” with *people mostly try to be helpful* and *people mostly look out for themselves* as anchors with a *don’t know* option. *Don’t know* responses were treated as system missing and all 3 items were recoded to parallel polarity (i.e., 1 = most people can be trusted, most people try to be fair, and people mostly try to be helpful).

GST consists of 3 items scored on a 4-point scale anchoring at *Do not trust at all* and *Trust completely* with *Do not trust very much* and *Trust somewhat* in-between the anchors and a *Don't know* option at the end of the scale. The question asked “Could you tell me for each whether you trust people from this group completely, somewhat, not very much, or not at all?” with grid items being “people you meet for the first time” (FIRST), “people of another religion” (RELI), and “people of another nationality” (NATION). *Don't know* responses were treated as system missing.

Measures of discriminant validity. Discriminant validity is determined by assessing the degree to which an operationalization is not similar to (or diverges from) other operationalizations that it theoretically should not be similar to (Pedhazur and Schmelkin 1991). Strictly speaking, discriminant validity requires that “a test not correlate too highly with measures from which it is supposed to differ (Campbell 1960: 548).” If evidence for discriminant validity is weak, then “constructs [have] an influence on the variation of more than just the observed variables to which they are theoretically related” and “researchers cannot be certain results confirming hypothesized structural paths are real or whether they are a result of statistical discrepancies (Farrell 2010: 324).”

Four instruments capturing particularized trust, political trust, economic preferences, and social value orientations were used to assess discriminant validity: a 3-item *Particularized Social Trust* scale (PST; Freitag and Bauer 2013; Freitag and Traunmüller 2009); a 4-item *Political Trust* scale (POT; Newton and Zmerli 2011; Zmerli and Newton 2008); 9 items capturing economic preferences (Falk et al. 2018); and a 9-item *Prosocial Orientation* scale (Van Lange et al. 1997).

Prior research shows that measures of generalized trust and particularized trust are moderately correlated (Freitag and Traunmüller 2009; Glanville and Shi 2020; Newton and Zmerli 2011). Recent models, however, suggest that generalized trust and particularized trust are distinct constructs with different roots, and should, in theory, be weakly correlated (Cao and Galinsky 2020; Robbins 2022). To measure PST, I used 3 items scored on a 4-point scale anchoring at *Do not trust at all* and *Trust completely* with *Do not trust very much* and *Trust somewhat* in-between the anchors and a *Don't know* option at the end of the scale. The question asked “Could you tell me for each whether you trust people from this group completely, somewhat, not very much or not at all?” with grid items being “your family” (FAMILY), “your neighborhood” (NEIGH), and “people you know personally” (KNOW). *Don't know* responses were treated as system missing.

Decades of political theory suggest that generalized trust and political trust mutually reinforce each other (Almond and Verba 1963; Putnam 1993; Uslaner 2002), with recent empirical research supporting an association between the two constructs (e.g., Dinesen et al. 2021).⁶ Using GST, Newton and Zmerli (2011) observed weak-to-moderate correlations between GST and POT across countries (Zmerli and Newton 2008 found similar strengths of association for different generalized trust measures).⁷ Likewise, Robbins (2022) showed that POT was moderately correlated with SFT-24 ($r = .345$), IST-4 ($r = .416$), MST ($r = .454$), and GST ($r = .435$). Given prior empirical research—and that generalized trust and political trust are distinct, theoretical constructs—I expect a valid measure of generalized trust to yield a weak-to-moderate

⁶ See Glanville and Paxton (2007) for a skeptical take on the association. For them, generalized trust and political trust are both influenced by a common cause, such as an underlying predisposition or personality trait.

⁷ The smallest correlation observed was in Spain ($r = .042$), while the largest was in India ($r = .347$; see Newton and Zmerli 2011, p. 182).

correlation with POT. To measure POT, I used 4 items scored on a 4-point scale anchoring at *None at all* and *A great deal* with *Not very much* and *Quite a lot* in-between the anchors and a *Don't know* option at the end of the scale. The question asked “We are going to list a number of organizations. For each one, could you tell us how much confidence you have in them: Is it a great deal of confidence, quite a lot of confidence, not very much confidence, or none at all?” with grid items being “the police” (POLICE), “the courts” (COURTS), “the government” (GOVERN), and “political parties” (PARTIES). *Don't know* responses were treated as system missing.

Many theories in the social sciences assume that a set of preferences motivate individual behavior and action. This includes preferences for risk, the timing of costs and benefits, as well as reciprocity and altruism. Falk et al. (2018) recently showed that heterogeneity in economic preferences exists between *and* within countries. The authors also found that various economic preferences are correlated, giving rise to distinct “preference profiles” of groups of countries (e.g., prosocial traits of positive reciprocity, altruism, and trust correlate). Nevertheless, Falk et al. (2018) show that their single-item measure of trust (“I assume that people have only the best intentions” on an 11-point scale) is weakly correlated with measures of positive reciprocity and altruism at the individual-level (see Falk et al.’s Online Appendix C). This finding is intuitive since trust is a belief and not a preference. For instance, scholars frequently conflate trust with perceptions of risk (Gambetta 1988). Yet, *preferences for risk* are theoretically different than *perceptions of risk* (Hardin 2002; Robbins 2016b). As a result, generalized trust should be weakly correlated with preferences for risk as well as other economic preferences like negative reciprocity (Robbins 2016b, 2022).

To measure economic preferences, I used 9 items from Falk et al. (2018), which capture time preferences, risk preferences, positive and negative reciprocity, and altruism. The first item, Patience, is used to measure time preferences. The measure of patience is a self-report of willingness to wait on an 11-point scale anchoring at *Completely unwilling to do so* and *Very willing to do so*, with a *Don't know* option at the end of the scale. The question asked “How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?” Risk preferences were also elicited through self-reports. The measure of risk preferences, Risk Taking, asks for the respondents’ self-assessment of their willingness to take risks on an 11-point scale (“In general, how willing or unwilling are you to take risks?”). The scale anchored at *Completely unwilling to take risks* and *Very willing to take risks*, with a *Don't know* option at the end of the scale.

Positive reciprocity consists of 2 items: (1) Gift Exchange, and (2) Return a Favor. For gift exchange, respondents were asked to imagine that they got lost in an unfamiliar area and that a stranger—when asked for directions—offered to take them to their destination. Respondents were then asked which out of six hypothetical presents (worth between USD \$5 and \$30) they would give to the stranger as a “thank you” (with a *No present* and *Don't know* option anchored at either end of the scale). For return a favor, respondents were asked to provide a self-report of how willing they are to return a favor on an 11-point scale (“When someone does me a favor, I am willing to return it”) anchoring at *Does not describe me at all* and *Describes me perfectly*, with a *Don't know* option at the end of the scale.

Negative reciprocity was elicited through three self-assessments: (1) Take Revenge, (2) Punish Unfair Behavior (Self), and (3) Punish Unfair Behavior (Other). First, respondents were asked how willing they are to take revenge on an 11-point scale: “If I am treated very unjustly, I

will take revenge at the first occasion, even if there is a cost to do so.” The response scale was anchored at *Does not describe me at all* and *Describes me perfectly*, with a *Don’t know* option at the end of the scale. The second and third items probed respondents about their willingness to punish someone for unfair behavior, either toward themselves or a third person. The question asked: “Even if there are costs for you, how willing are you to punish someone who treats the following people unfairly?”, with “You” and “Others” as grid items. Both grid items anchored at *Completely unwilling to do so* and *Very willing to do so*, with a *Don’t know* option at the end of the scale.

Altruism was measured with two items: (1) Donation Decision, and (2) Give to Good Cause. For donation decision, respondents were asked to imagine a hypothetical situation where they unexpectedly received USD \$1,000 and to state how much of this amount they would donate (“Imagine the following situation: Today you unexpectedly received \$1,000. How much of this amount would you donate to a good cause?”). Respondents were allowed to enter any amount between \$0 and \$1,000 into a numerical box. For give to good cause, respondents were asked how willing they would be to give to good causes without expecting anything in return on an 11-point scale (“How willing are you to give to good causes without expecting anything in return?”). Response options were anchored at *Completely unwilling to do so* and *Very willing to do so*, with a *Don’t know* option at the end of the scale.

In addition to economic preferences, I investigated social preferences, particularly prosocial values (Van Lange et al. 1997). Prosocial orientations are key dimensions of morality (Simpson and Willer 2015), and, according to Uslaner (2002), morality is the basis of generalized trust. This expectation, however, is misguided. Generalized trust is a situation-independent expectation of generalized others, while social preferences concern an interest in the

outcomes of self *and* others. A measure of prosocial orientations, then, should be uncorrelated with a measure of generalized trust given that both are theoretically distinct constructs (belief versus preference). Recent work by Robbins (2022) supports this proposition.

To measure Prosocial Orientation, I used a scale that consisted of nine hypothetical decision scenarios (Van Lange et al. 1997), where participants decide for each scenario how to divide resources between themselves and a hypothetical stranger, or “other”. Each scenario includes three options corresponding to one of three social values or preferences (Deutsch 1960): a cooperative choice, which maximizes joint gain; an individualist choice, which maximizes personal gain without regard to the other's outcome; and a competitive choice, which maximizes the difference between gains to self and other. The Prosocial Orientation scale was constructed by summing the total number of cooperatives choices made across the nine hypothetical decision scenarios.

Measures of known-groups validity. Known-groups validity (or known-groups differences) is determined by assessing the degree to which an operationalization can discriminate between two groups known to differ on the variable of interest (Davidson 2014). It is, in other words, the degree to which scores on latent constructs differ between known groups. In this approach to construct validation, measurement items are administered to groups “that are known to have...different levels of the construct to confirm whether the hypothesized difference is reflected in the scores (Davidson 2014: 3481)” of the different groups. Given that knowledge about differences between known-groups is a function of traditional measures, which may be biased and invalid, known-groups validity is not a test of validity *per se*. It is an assessment tool used to confirm or reject prior findings. It should be used *if and only if* new instruments exhibit superior validity and reliability over traditional measures. Previous and current findings suggest

this is the case (see Robbins 2021, 2022). I examine known-groups validity at the individual- and state-levels.

At the individual-level, I included a number of measures to explore their associations with SFT/IST and MST/GST, since previous studies indicate that personality traits (Freitag and Bauer 2016) and sociodemographic characteristics (Alesina and La Ferrara 2002) predict levels of generalized trust. In terms of personality traits, I used the 6-item Life Orientation Test-Revised scale to measure *Dispositional Optimism* (Scheier et al. 1994), and a 15-item short-form version of the *Big-5 Personality* scale (Hahn et al. 2012).

I investigate the effects of dispositional optimism on generalized trust since theories suggest that trust is a function of optimism (Uslaner 1998, 2002), with recent empirical research supporting this proposition (Andersson 2012). Dispositional optimism consists of 5 items scored on a 5-point agree-disagree scale anchoring at *Completely agree* and *Completely disagree* with *Somewhat agree*, *Neither agree nor disagree* and *Somewhat disagree* in-between the anchors and a *Don't know* option at the end of the scale. The question asked respondents to be “as honest and accurate” as they can and to not let their responses “to one statement influence [their] responses to other statements”. Grid items included “In uncertain times, I usually expect the best”, “If something can go wrong for me, it will”, and “Overall, I expect more good things to happen to me than bad.” All items were recoded so that the scale has the same direction (higher values indicate greater optimism). *Don't know* responses were treated as system missing.

Given prior research, I expect the personality traits of conscientiousness (-), openness (+), and agreeableness (+) to correlate with generalized trust (Freitag and Bauer 2016); that is, individuals who vary according to their levels of conscientiousness, openness, and agreeableness should differ in their levels of generalized trust. The Big-5 personality scale was generated

according to the BFI-S items (Hahn et al. 2012). This is a short-form version of the standard BFI comprising 15 items, with three items related to each personality trait of extraversion, conscientiousness, openness, agreeableness, and neuroticism. Respondents were presented with 15 statements, and asked whether each statement applies to them (“I see myself as someone who...”) using a 5-point agree-disagree scale (see the dispositional optimism rating scale). Statements included “...is talkative”, “...is original, comes up with new ideas”, “worries a lot”, and “does things efficiently.” All items were recoded so that the scale has the same direction (higher values indicate greater agreement with the statement). *Don’t know* responses were treated as system missing. From these 15 items, I constructed a scale for each of the five personality traits based on their respective three items.

With respect to sociodemographic characteristics, I included measures of age (in years), gender (female, male), race-ethnicity (non-Hispanic White, Black, Hispanic, Other Race), education (no HSD, HSD, some college or AA, BA +), marital status (married, widowed, divorced, separated, never married, living with partner), employment status (paid employee, self-employed, not working), household income (logged per capita household income), region (Northeast, Midwest, South, West), and metro area (metro area, non-metro area).⁸ Previous research shows that all of these sociodemographic characteristics have yielded statistically significant associations with generalized trust (Alesina and La Ferrara 2002; Evangelist 2021; Freitag and Traunmüller 2009; Hooghe et al. 2009; Mewes et al. 2021; Paxton 2007; Robbins 2011; Simpson 2006). To conserve space, I review one consistent finding. A robust observation in the literature is that age and generalized trust—particularly in the United States—are

⁸ Northeast (CT, MA, ME, NH, NJ, NY, PA, RI, VT), Midwest (IA, IL, IN, KS, MI, MN, MO, ND, NE, OH, SD, WI), South (AL, AR, DC, DE, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV), and West (AK, AZ, CA, CO, HI, ID, MT, NM, NV, OR, UT, WA, WY).

positively correlated (Clark and Eisenstein 2013; Robinson and Jackson 2001; Schwadel and Stout 2012; Wilkes 2011), indicating that generalized trust is higher among older cohorts than younger cohorts. I attempt to replicate this finding—as well as other findings—with MST/GST, and assess the extent to which it holds with the analysis of SFT/IST.

At the state-level, I included three common correlates of generalized trust to explore their associations with aggregate measures of SFT, IST, MST, and GST: intentional homicide, ethnic heterogeneity, and income inequality. I focus on these state-level correlations for two reasons. First, safety and security (Messner et al. 2004; Putnam 2000; Rosenfeld et al. 2007), diversity (Cao and Galinsky 2020; Dinesen et al. 2020), and fairness (Uslaner 2002; Rothstein and Uslaner 2005) have featured heavily in theoretical models of generalized trust for decades. Second, there is strong empirical support for the negative effects of intentional homicide, ethnic heterogeneity, and income inequality on generalized trust (Dinesen and Sønderskov 2015; Hastings 2018; Smith 1997; Uslaner 2002; Wu 2020).

Intentional homicide for each state is measured by averaging the number of deaths (all homicides) per 100,000 over a seven-year period (2013-2019). Data are drawn from the Center for Disease Control and Prevention's National Center for Health Statistics. For ethnic heterogeneity, data were obtained through the Integrated Public Use Microdata Series USA, which integrates U.S. census data and the American Community Surveys. With this data, I computed the ethnic fractionalization index (Alesina et al. 2003). This measure is interpreted as the probability that two randomly selected individuals within a state will *not* belong to the same ethnic group. For each state, I took the average of the ethnic fractionalization index over a five-year period (2014-2018). Finally, I use the Gini coefficient to measure income inequality. The Gini coefficient ranges from 0, indicating perfect equality, to 100, perfect inequality. Data for the

Gini coefficient was drawn from the United States Census Bureau, and averaged over a five-year period for each state (2014-2018).

Online Investment Game

I use the investment game to assess the predictive validity of the four generalized trust scales. Prior to the online experiment, respondents provided their informed consent, read instructions, and were asked to complete a few exercises to confirm their understanding of the game. Respondents then played one round of the classic investment (or trust) game (Berg et al. 1995). At the beginning of the experiment, respondents were randomly assigned to one of two roles: player A or player B. Respondents were informed of these two roles, and were told that they would be randomly matched to another AmeriSpeak participant. Both players started the experiment with 6000 AmeriPoints (hereafter “points”). Respondents were told that player A moved first, and that player A had the opportunity to transfer all, some, or none of their 6000 points to player B in increments of 500 points. Each point that player A sent to player B was tripled. If player A sent 500 points to player B, for example, player B would receive 1500 points. The amount player A transferred to player B would be displayed on both players’ screens. Player B then had the possibility to return points to player A. Player B could choose any amount between zero and the tripled amount player A transferred to player B in increments of 500 points. The amount player B returned to player A would not be tripled. The 6000 points player B started the game with could not be sent to player A. All respondents were informed of and aware of these rules. The experiment ended after player B submitted the amount of points he or she would like to return to player A.

Measures of predictive validity. Trust and trustworthiness are two different constructs. Trust is often conceptualized as person A’s belief about the competence and motivation of

person B with respect to matter Y (i.e., what is of interest to person A), while trustworthiness is person B's actual level of competence and degree of motivation to achieve matter Y on behalf of person A (Bauer 2021; Cook et al. 2005; Hardin 2002; Robbins 2016b). A valid measure of generalized trust, then, should positively predict trust measured in an investment game, and predict trustworthiness to a negligible or small degree. Trust in the investment game was measured as the proportion of points sent to player B. Trustworthiness in the investment game was measured as the proportion of points returned to player A. Trust measured in the investment game was used as a criterion variable for the validation of SFT, IST, MST, and GST. Trustworthiness measured in the investment game was used to assess discriminant validity.

Results

Analytic Strategy

For all measurement validation tests presented below, analyses were estimated in Mplus 8.5 (Muthen and Muthen 2017) or Stata 15.2. Robust full-information maximum likelihood estimation (MLR) was used throughout. Since models estimated with MLR are relatively robust against deviations from the assumption of normality (DiStefano 2002; Muthén and Kaplan 1985), I model all indicators as continuous variables. As a robustness check, I used mean-and-variance adjusted weighted least squares (WLSMV) to estimate models where the indicators were treated as categorical. Although not shown, models using MLR for continuous indicators and WLSMV for categorical indicators are substantively similar. With respect to missing data, under the assumption of data missing completely at random or missing at random, MLR for missing data provides asymptotically unbiased estimates (Allison 2001). I have no reason to believe that the data are missing not at random given the non-sensitive nature of the questions. Finally, all models are estimated without post-stratification survey weights. However, I estimated

models with post-stratification survey weights to ensure the findings are robust to non-response error and adjustment error (Arel-Bundock and Pelc 2018; Gelman 2007; Groves et al. 2009; Pepinsky 2018). Results are largely consistent with the unweighted estimates presented in the main text (see the Supplemental Materials online).

Convergent Validity

Robbins (2021, 2022) found that SFT was best represented by a single latent factor. For SFT, a higher-order confirmatory factor analysis (CFA) in which trust domain indicators (SECRET, LOAN, CHILD, ADVICE) loaded on first-order factors for each stranger's face—White female (WF), White male (WM), Black female (BF), Black male (BM), Latin female (LF), Latin male (LM)—which in turn loaded on a second-order factor (SFT), yielded model fit similar to a single-factor model. In a departure from Robbins (2021, 2022), I use latent factor scores for WF, WM, BF, BM, LF, and LM instead of row mean scales. I use this alternative estimation strategy to adjust for measurement error and the smaller factor loadings of some indicators. I refer to this original scale as SFT-24, and treat it as a 6-item single factor in which SFT-24 loads on the latent factor scores of WF, WM, BF, BM, LF, and LM in the present manuscript.

I also follow this measurement strategy for the two new matters (WALLET, TIRE) and four new faces: East Asian female (EF), East Asian male (EM), South Asian female (SF), and South Asian male (SM). For this new single-factor model—which I refer to as SFT-60—the six trust domains (SECRET, LOAN, CHILD, ADVICE, WALLET, TIRE) of each face (WF, WM, BF, BM, LF, LM, EF, EM, SF, SM) were used to estimate latent factor scores. I thus treat SFT-60 as a 10-item single factor in which SFT-60 loads on the latent factor scores of WF, WM, BF, BM, LF, LM, EF, EM, SF, and SM (where each face constitutes a 6-item scale). Results using

Robbins' (2021, 2022) original estimation strategy yield similar findings (see the Supplemental Materials online).

Robbins (2021, 2022) also found that IST was best represented by a single-factor latent construct in which the four trust domain indicators (SECRET, LOAN, CHILD, ADVICE) loaded on IST. To distinguish the original IST measurement model from the new measurement model, I refer to the former as IST-4 (SECRET, LOAN, CHILD, ADVICE) and the latter as IST-6 (SECRET, LOAN, CHILD, ADVICE, WALLET, TIRE).

To establish convergent validity of the two old (SFT-24, IST-4) and two new (SFT-60, IST-6) latent factors, I first assess the degree to which all four are well measured by their observed indicators. This is done by examining the factorial validity of SFT-24, SFT-60, IST-4, and IST-6 (i.e., internal consistency, factor loadings, average variance extracted, etc.). I then assess the degree to which all four are similar to (or converge on) other operationalizations of generalized trust (i.e., MST and GST). I examine correlations between latent constructs to gauge convergent validity. Below, I review and then execute these various tests.

I first examine Cronbach's α , which assesses internal consistency of the observed indicators. Cronbach's α values range between 0 and 1, with values greater than .70 indicating acceptable reliability (Nunnally 1978). I then test the four-factor structure of SFT-24, IST-4, MST, and GST with a CFA, followed by a separate test of the four-factor structure of SFT-60, IST-6, MST, and GST. To do this, I examine the standardized factor loadings. Factor loadings greater than .70 suggest acceptable convergent validity since more than 50 percent (i.e., $.70^2$) of the variance in an indicator is explained by the latent construct to which the indicator is theoretically related (Campbell and Fiske 1959; Kline 2005). I then calculate the average variance extracted (AVE). AVE measures the level of variance captured by a latent construct

versus the level due to measurement error (Fornell and Larcker 1981). The AVE for a latent construct is calculated by summing the variances extracted (i.e., r^2) from each indicator over the total number of indicators. AVE values greater than .50 are considered acceptable (Cheung and Wang 2017; Fornell and Larcker 1981).

To assess overall model fit and the factorial validity of each four-factor CFA, several indices of fit were used, including the root mean squared error of approximation (RMSEA), the comparative fit index (CFI), and the standardized root mean square residual (SRMR). The RMSEA is a sample-free index that measures the difference between the estimated model and the population covariance matrix. RMSEA ranges between 0 and 1, with values less than .08 indicating adequate model fit (Hu and Bentler 1999). The CFI, in contrast, adjusts for issues of sample size inherent to the χ^2 test of model fit, and investigates the difference between the data and the estimated model. CFI ranges from 0 to 1, with values greater than .90 suggesting adequate model fit (Hu and Bentler 1999). The SRMR is a standardized measure of the square root of the difference between the sample covariance matrix and the model covariance matrix. Like the RMSEA, SRMR ranges between 0 and 1, with values less than .08 indicating adequate model fit (Muthén and Muthén 2017).

Next, I evaluate correlations between latent constructs. Correlations greater than .85 stipulate convergence, while correlations less than .85 indicate that two latent constructs diverge and likely measure different underlying concepts. New measures should be correlated with existing measures but not above a correlation of .85, which would suggest that the new measure may not be an improvement over the old one (Kline 2005). Following Fornell and Larcker (1981), I also establish discriminant validity by comparing AVEs to squared correlations between latent constructs. According to the Fornell-Larcker criterion, each latent construct's

AVE must be greater than or equal to squared correlations with other constructs in the model. If not, then the two measures may be measuring the same underlying construct.

Descriptive statistics and reliability. Item means, standard deviations, number of observations, and Cronbach α are shown in Table 1. For SFT-24 and SFT-60, Cronbach's α was .970 and .983, respectively, indicating good internal consistency; the four-to-six trust domains of each face (e.g., WF) also exhibited good internal consistency, with Cronbach's α between .860 and .884. IST-4 ($\alpha = .829$) and IST-6 ($\alpha = .822$) yielded similar levels of internal consistency. The Cronbach α 's for MST ($\alpha = .744$) and GST ($\alpha = .760$) also yielded acceptable internal consistency.

Factorial validity and intercorrelations between latent constructs. In order to identify a CFA, I constrained the first unstandardized factor loading of each factor to 1. Factor variances as well as intercepts and residual variances of the observed indicators were freely estimated and statistically significant (not shown but available upon request). Results are provided in Table 2. In model 1, the four-factor CFA of SFT-24, IST-4, MST, and GST provided a good fit to the data, $\chi^2(98) = 351.586$, RMSEA = .045, CFI = .966, and SRMR = .054. For SFT-24 and IST-4, all standardized factor loadings were equal to or above the standard cutoff value of .70 (save for IST-4 on ADVICE). The AVEs for SFT-24 and IST-4 were also above standard cutoffs ($> .50$). In contrast, some standardized factor loadings were below .70 for MST and GST, while others were above .70. Likewise, the AVE for GST was above the acceptable cutoff of .50, while MST was not. This suggests that the level of variance captured by MST due to measurement error is greater than standard cutoffs. Overall, the convergent validity of observed indicators of SFT-24 and IST-4 were superior to that of MST and GST.

With respect to discriminant validity, the intercorrelations between latent constructs were all less than .85, while the AVEs for each latent construct were greater than the squared correlations between latent constructs.⁹ This set of findings reveals latent construct discrimination between SFT-24, IST-4, MST, and GST. Replicating Robbins (2022), the correlations between SFT-24 and IST-4 ($r = .737$) and MST and GST ($r = .537$) were greater than all other correlations between latent constructs, showing that both sets of constructs exhibit more conceptual overlap with each other than the other set. Overall, the findings indicate discriminant validity between latent constructs.

In model 2, the four-factor CFA of SFT-60, IST-6, MST, and GST provided a good fit to the data, $\chi^2(203) = 800.705$, RMSEA = .048, CFI = .950, and SRMR = .065. Overall, the analysis of convergent and discriminant validity yielded similar results to those found in model 1, save for three key differences. First, two factor loadings for IST-6, WALLET and TIRE, were below the standard cutoff of .70, although both were near lowest acceptable limits (Cheung and Wang 2017). Second, the AVE for IST-6 was slightly below .50, showing some measurement error. Third, the squared correlation between SFT-60 and IST-6 ($r^2 = .538$) was slightly greater than the AVE for IST-6 (AVE = .458). Taken together, the results indicate that SFT-60 and IST-6 exhibit considerable construct overlap but do not exhibit convergent validity.

Given comparable factorial validity and convergent validity between SFT-24 and SFT-60 as well as IST-4 and IST-6, all remaining measurement validation tests—discriminant validity, known-groups validity, and predictive validity—will only include SFT-60 and IST-6, with MST and GST serving as benchmarks. Validation tests for SFT-24 and IST-4 as well as indicators of MST (TRUST, FAIR, and HELP) can be found in the Supplemental Materials online. Finally,

⁹ This was not the case for IST-4, which yielded an AVE of .543: a value slightly smaller than the r^2 of .546 between SFT-24 and IST-4.

Robbins (2022) found that SFT-24 trust evaluations were not solely driven by cultural stereotypes, identity effects, or social desirability bias. I replicate these findings in the analysis of SFT-60.

Discriminant Validity

To assess discriminant validity, I used structural equation models (SEM) to analyze the degree to which SFT-60, IST-6, MST, and GST correlate with PST, POT, indicators of economic preferences, and the prosocial orientation scale. In these SEMs, a single endogenous latent factor (SFT-60, IST-6, MST, GST) is regressed on a single exogenous latent factor (PST, POT) or observed indicator (e.g., risk taking). Standardized path coefficients serve as an estimate of the strength of association.

Figure 1 presents the standardized path coefficients—or correlations—and their respective 95% CIs for SFT-60 and IST-6. Overall, the results show that SFT-60 *and* IST-6 are significantly and positively correlated with PST, POT, patience, and risk taking albeit weakly. For SFT-60 and IST-6, most measures of positive reciprocity and negative reciprocity were statistically non-significant (with some exceptions), whereas SFT-60 and IST-6 are weakly correlated with some measures of altruism and the prosocial orientation scale. Taken together, the various tests of association indicate that SFT-60 and IST-6 exhibit discriminant validity.

Figure 2 presents the standardized path coefficients—or correlations—and their respective 95% CIs for MST and GST. Unlike SFT-60 and IST-6, MST and GST are moderately-to-strongly correlated with PST and POT. For all other variables, MST and GST reveal negligible-to-weak correlations, but do exhibit statistically significant associations with more criterion variables than SFT-60 and IST-6 (e.g., measures of positive and negative reciprocity). In short, MST and GST discriminate from some criterion variables but converge

upon others (PST, POT), suggesting that SFT-60 and IST-6 have greater discriminant validity than MST and GST.

Note that SFT-24 and IST-4 yield comparable estimates. These additional tests can be found in the Supplemental Materials online.

Known-Groups Validity

Individual-level. At the individual-level, I estimated four SEMs where the endogenous latent factors of SFT-60, IST-6, MST, and GST were regressed on dispositional optimism, Big-5 personality scales, sociodemographic characteristics, and controls. In Table 3, I present the results of the estimated models.

Models 1 and 2 reveal strikingly consistent findings. Conscientiousness (-) and agreeableness (+) are significantly related to SFT-60 and IST-6 in the anticipated direction (Freitag and Bauer 2016), while both latent constructs exhibit variation across regions of the United States (the West has the highest levels of generalized trust). All other key individual-level variables are statistically unrelated to SFT-60 and IST-6. This is a significant finding, which shows that generalized trust, measured as SFT-60 and IST-6, does not exhibit differences by age (Clark and Eisenstein 2013; Putnam 2000; Robinson and Jackson 2001), gender (Alesina and La Ferrara 2002; Mewes 2014), race and ethnicity (Douds and Wu 2018; Evangelist 2021; Wilkes 2011), and socioeconomic status (e.g., education; Oskarsson et al. 2017).

Although model 3 shows that MST is significantly correlated with conscientiousness (-) and agreeableness (+), it also departs from estimates found in Models 1 and 2. Notably, MST is correlated with dispositional optimism (+), age (+), race and ethnicity (non-Hispanic Whites > Blacks and Hispanics), region (the Northeast has the highest levels of generalized trust), and metro area (non-metro area > metro area). I fail to observe these associations for SFT-60 and

IST-6. While model 3 replicates many of the known-groups differences found in the literature, model 4 does not. In fact, the only variable that is significantly correlated with GST is dispositional optimism (+). All other variables render statistically non-significant effects.

Taken together, not one variable is consistently associated with the four generalized trust scales. In spite of this heterogeneity, there are some noticeable patterns. Two of the Big-5 personality scales—conscientiousness and agreeableness—are associated with three of the four generalized trust scales; SFT-60 and IST-6 show similar regional patterns; and MST and GST are both positively correlated with dispositional optimism. In other words, there are key similarities across the two new scales of generalized trust, as well as clear differences between the old and new scales. All of this suggests that some previous findings, like the positive age-trust correlation (Clark and Eisenstein 2013; Putnam 2000; Robinson and Jackson 2001) or ethnoracial differences in generalized trust (Douds and Wu 2018; Evangelist 2021; Wilkes 2011), should be reevaluated in light of the superior validity and reliability of SFT/IST (see current measurement validation tests as well as previous findings by Robbins 2021, 2022).

These estimates largely parallel those found using SFT-24 and IST-4 (see the Supplemental Materials online).

State-level. At the state-level, I estimated a series of correlations between state-level indicators (intentional homicide, ethnic fractionalization, and income inequality) and aggregated state-level latent factor scores of SFT-60, IST-6, MST, and GST. In Figure 3, I present scatter plots, 95% CIs, and tests of association.

Row one in Figure 3 (Panels A, B, C, and D) shows that all four generalized trust scales are negatively correlated with income inequality (Gini coefficient). While there is variation in the strength of associations (weak-to-moderate correlations) and statistical significance across scales,

the results nonetheless support the classic proposition that social cleavages and class divisions make it difficult for generalized trust to flourish in society (Uslaner 2002; Rothstein and Uslaner 2005).

In row two (Panels E, F, G, and H), Figure 3 shows that SFT-60 is statistically unrelated to ethnic heterogeneity ($r = -.016$), while the other three scales appear to be negatively correlated with ethnic heterogeneity ($r > -.240$ but $< -.250$) albeit weakly at the $p < .10$ level. This is an important finding which likely stems from measurement. Note that IST-6, MST, and GST require respondents to assess imaginary trustees (e.g., “imagine a complete stranger” for IST-6, “most people” for MST, and “people of” for GST). These results suggest that respondents in ethnically homogeneous contexts may think of “most people” or “complete strangers” as someone who is like themselves (Simpson et al. 2007). By contrast, respondents in ethnically heterogeneous settings may think of out-group members as “most people”, thereby reducing generalized trust. This evidence, however tentative it may be, suggests that the diversity-trust link might be an artifact of measurement (Dinesen et al. 2020).

In row three (Panels I, J, K, and L), Figure 3 shows that SFT-60 and IST-6 are negatively correlated with intentional homicide but to a negligible degree ($r = -.018$ for SFT-60; $r = -.091$ for IST-6). By comparison, MST ($r = -.298$, $p = .035$) and GST ($r = -.285$, $p = .044$) are both negatively correlated with intentional homicide but to a larger degree. These results cast doubt on Putnam’s (2000) classic observation that places with low generalized trust may become trapped in a “vicious circle in which low levels of trust and cohesion lead to higher levels of crime, which lead to even lower levels of trust and cohesion (p. 317).” What my results indicate is that once measures of generalized trust consider specific strangers (e.g., White female) and

particular matters (e.g., keeping a secret), respondents tend to trust at similar levels regardless of heterogeneity in degrees of exposure to state-level violence.

State-level correlations for SFT-24 and IST-4 are comparable to that of SFT-60 and IST-6, respectively. These additional tests can be found in the Supplemental Materials online.

Predictive Validity

To assess discriminant validity, I used structural equation models (SEM) to analyze the degree to which SFT-60, IST-6, MST, and GST correlate with the proportion of tokens sent and returned in an investment game. In these SEMs, a single endogenous observed indicator (proportion of tokens sent and returned) is regressed on a single exogenous latent factor score (SFT-60, IST-6, MST, GST). Standardized path coefficients serve as an estimate of the strength of association.

Figure 4 shows that all four generalized trust scales are positively correlated with the proportion of points sent in the investment game, but that SFT-60 ($\beta = .20$) and IST-6 ($\beta = .12$) are the only statistically significant scales at the $p < .05$ level. By contrast, the positive correlation between MST/GST and the proportion of tokens sent was weak and statistically non-significant ($p > .05$). Figure 5 shows that SFT-60, IST-6, and GST are negatively correlated with the proportion of points returned in the investment game, while MST is positively correlated with the proportion of tokens returned. Out of the four scales, only the correlation between IST-6 and the proportion of tokens returned was statistically significant at the $p < .05$ level. Taken together, SFT-60 fully meets expectations (positively predicts trust but not trustworthiness), IST-6 partially meets expectations (positively predicts trust and negatively predicts trustworthiness), and neither MST nor GST meet the criteria for predictive validity.

The findings presented here are robust to statistical controls save for IST-6.¹⁰ Furthermore, SFT-24 and IST-4 yield comparable, albeit weaker, estimates. These additional tests can be found in the Supplemental Materials online.

Measurement Invariance

Although not a central focus of the paper given that Robbins (2021) has established the measurement invariance of SFT-24 and IST-4 across sociodemographic groups, I replicate his analysis for SFT-60 and IST-6. Results confirm the measurement invariance of SFT-60 and IST-6 across gender, race, education, and age-groups. See the Supplemental Materials online for this additional analysis of measurement invariance.

Robustness of SFT to Cultural Stereotypes and Identity Effects

To explore the role of cultural stereotypes and social identities in ratings of SFT, I first examine each matter of trust separately (results available upon request). I find a gender of face effect for two of the six matters: female faces yield higher mean ratings for CHILD, while male faces yield higher mean ratings for TIRE. For the other four matters (SECRET, LOAN, ADVICE, WALLET), there is a negligible, but statistically significant difference, in favor of female faces. Yet, I do not observe a co-gender effect; that is, for all six matters of trust, male and female faces are rated similarly by male and female respondents.

For race of faces, Black faces yield the highest mean ratings across all matters of trust, while White faces yield the lowest mean ratings, which is contrary to stereotypic expectations in

¹⁰ For the proportion of tokens sent by Player A: SFT-60 ($\beta = .14$, $SE = .05$, $p = .01$), IST-6 ($\beta = .06$, $SE = .05$, $p = .28$), MST ($\beta = .07$, $SE = .05$, $p = .16$), and GST ($\beta = .03$, $SE = .05$, $p = .50$). For the proportion of tokens returned by Player B: SFT-60 ($\beta = .02$, $SE = .06$, $p = .72$), IST-6 ($\beta = -.09$, $SE = .06$, $p = .09$), MST ($\beta = .11$, $SE = .05$, $p = .05$), and GST ($\beta = -.07$, $SE = .06$, $p = .23$). The control variables included all measures of economic preferences and the Prosocial Orientation scale.

the United States. In fact, the Black male face tends to yield the highest mean ratings across all matters, while the White male face tends to yield the lowest mean ratings. This finding is consistent with Robbins (2022) as well as the Chicago Face Database independent rater sample (Ma et al. 2015). Robbins (2022) also shows that these differences in ratings are not a function of social desirability bias.¹¹

In terms of ethnoracial social identity, I do observe co-race effects, with slight in-group favoritism for three of the six matters: SECRET, CHILD, and TIRE. These co-race effects, however, are inconsistent across matters (e.g., Hispanic respondents exhibit in-group favoritism for CHILD but not for SECRET); are unique to specific ethnoracial pairings (e.g., for SECRET, only Black respondents exhibit in-group favoritism); apply to other out-groups (e.g., for TIRE, Black respondents trust Black faces, Hispanic faces, East Asian faces, and South Asian faces more than White faces); go unobserved for some races (e.g., Asian respondents); and are sensitive to post-stratification survey weights (e.g., SECRET). Second, when regressing WF,

¹¹ Social desirability bias is the tendency to underreport (overreport) socially undesirable (desirable) attitudes and behaviors on surveys. In relation to SFT, social desirability bias—if present—would lead to upwardly biased trust ratings of certain races and genders if respondents harbored negative attitudes toward those races and genders, and were aware that such prejudice was socially undesirable. Six pieces of evidence suggest that social desirability exerts minimal influence on SFT. First, online surveys create social distance between the researcher and respondent, which decreases interviewer effects and dampens social desirability bias (Dillman 2009). Second, social desirability distortions are less severe in samples of the general population than university student populations (Peterson 2001), with respondents from the general population willing to answer surveys in socially unpopular and politically incorrect ways (Sears 2004; Stark et al. 2022). Third, confidentiality and anonymity dampen social desirability bias (Bradburn et al. 1978), both of which AmeriSpeak safeguard by providing researchers with deidentified data. Fourth, the six domains of trust—SECRET, LOAN, CHILD, ADVICE, WALLET, and TIRE—are non-sensitive in nature: neither SFT nor IST directly ask respondents to report on socially sensitive topics like racism, prejudice, or sexism. Fifth, even though IST is much less likely to suffer from social desirability bias than SFT, both scales are highly correlated ($r > .700$, see Table 2 and Robbins 2021, 2022). Sixth, Robbins (2022) finds that the short-form version of the Marlow-Crowne Social Desirability Scale is weakly correlated with SFT ($r = 0.139$), and to a smaller degree than IST ($r = 0.212$), MST ($r = 0.181$), and GST ($r = 0.171$).

WM, BF, BM, LF, LM, EF, EM, SF, and SM on the race and gender of respondents in a multivariate regression, I fail to observe consistent co-race or co-gender effects.

Overall, the differences I observe are minor as well as expected. Measuring “generalized trust as an average across many situations that entails a variety of trustees, expected behaviors, and contexts (Bauer and Freitag 2018: 29)” undoubtedly summons specific cultural stereotypes and social identities. But by showing respondents a standardized set of faces and matters that vary to a great extent, SFT treats cultural stereotypes and social identities about gender and race as measurement error in its estimate of generalized trust.

Discussion and Conclusion

Summary of Results

To investigate generalized trust, social scientists working across a variety of disciplines rely on a small set of measurement tools—operationalizations largely built from the classic “most-people trust” question. These tools inform most of what we know about the causes and consequences of generalized trust (Abascal and Baldassarri 2015; Fairbrother and Martin 2013; Glanville et al. 2013; Herreros 2004; Kuwabara 2015; Messner et al. 2004; Mewes 2014; Rosenfeld et al. 2001; Paxton 2002, 2007; Robbins 2012, 2016a), as well as trends in generalized trust through time (Clark and Eisenstein 2013; Mewes et al. 2021; Paxton 1999; Putnam 2000; Robinson and Jackson 2001; Schwadel and Stout 2012; Uslaner 2002; Wilkes 2011). Traditional measures of generalized trust, however, have come under increasing empirical scrutiny, with some scholars suggesting that they severely bias statistical estimates and compromise tests of sociological theory. As a solution, Robbins (2021, 2022) proposed two new self-report measures of generalized trust—SFT and IST—which were shown to be more valid, and exhibit greater measurement equivalence, than traditional measures.

I argue that SFT and IST are required to rigorously test theories of—and advance sociological knowledge about—generalized trust, both of which require further refinement and validation. To do so, I first expand the ethnicity of faces shown in SFT and the matters of trust under evaluation in IST and SFT. All of which yield four short- and long-form versions of the scales: SFT-24 (six faces, four matters), SFT-60 (ten faces, six matters), IST-4 (four matters), and IST-6 (six matters). Second, I test the validity of these new scales alongside traditional measures of generalized trust in a nationally representative probability sample ($N = 1,264$) and follow-up study of a one-shot investment game ($N = 600$).

My results support the proposition that the most accurate and consistent measures of generalized trust constitute “an average across many situations that entails a variety of trustees, expected behaviors, and contexts (Bauer and Freitag 2018: 29).” I find that long- and short-form versions of SFT are more reliable and valid than IST (-4 and -6), which, in turn, are more reliable and valid than MST and GST. The evidence in favor of SFT and IST is particularly strong given that both sets of scales discriminate from economic preferences and other forms of trust, exhibit convergent validity across a number of tests, display measurement equivalence across sociodemographic groups, and predict measures of behavioral trust. The fact that the performance of SFT and IST is consistent with prior research indicates just how well they measure generalized trust.

Traditional measures of generalized trust fare much worse. I found that MST and GST weakly discriminate from scales of particularized trust and political trust, and are significantly correlated with a number of economic preferences. MST and GST are also imperfectly measured by their observed indicators, and neither predict measures of behavioral trust. In spite of some differences, the present findings largely replicate and reinforce prior research (Robbins 2021,

2022), and—when taken together—indicate that sociologists who use MST and GST cannot be certain of whether their statistical estimates are real or an artifact of measurement.

This raises the question of how to interpret previous findings based on traditional measures of generalized trust. My analysis of known-groups differences provides preliminary and tentative answers. At the individual-level, it may be worthwhile to reconsider prior findings linking generalized trust to age (Clark and Eisenstein 2013; Putnam 2000; Robinson and Jackson 2001), race-ethnicity (Douds and Wu 2018; Evangelist 2021; Wilkes 2011), and dispositional optimism (Andersson 2012; Uslaner 2002). The state-level analysis suggests that previous estimates of ethnic heterogeneity (Alesina and La Ferrara 2002; Marschall and Stolle 2004; Putnam 2007; Uslaner 2012) and intentional homicide (Putnam 2000; Messner et al. 2004; Rosenfeld et al. 2007) might be upwardly biased. Other individual- and state-level variables, however, appear to yield relatively consistent results across the four generalized trust scales, namely Big-5 personality (Freitag and Bauer 2016) and income inequality (Uslaner 2002; Rothstein and Uslaner 2005). Below, I review implications of these results for research on trust.

Implications for Sociological Research

The implications of this study for sociological research are abundant. First, the results provide strong support for previous findings while simultaneously refuting others. In his agenda-setting book, *The Moral Foundations of Trust* (2002), Eric Uslaner claims that generalized trust is a moral value—or stable disposition—that solidifies in early adulthood (see also Vaisey and Lizardo 2016; Kiley and Vaisey 2020). According to Uslaner, generalized trust is a core feature of one's personality that depends more on early childhood experiences and the intergenerational transmission of values (Erikson 1964; Rotenberg 1995), and less on the major events or mundane occurrences that one experiences throughout life. Uslaner also contends that declining optimism

and growing economic inequality account for the well-documented decay in generalized trust over the past 50 years in the United States (see also Andersson 2012; Paxton 2002; Putnam 2000; Robinson and Jackson 2001).

By using newly validated measures of generalized trust, the present paper refutes some of Uslaner's (2002) central claims. Robbins (2022) and I find that SFT and IST are weakly correlated with measures of moral motivations and dispositional optimism (the opposite holds true for MST and GST). We also show that SFT and IST are statistically unrelated to age (i.e., older people are not more trusting than younger people), that the new scales are equivalent across age-categories (Robbins 2021), and that generalized trust is malleable and open to change (Robbins 2021). These findings, by contrast, do not hold for MST or GST. Together, this evidence raises serious doubts about Uslaner's (2002) core propositions, save for one: the link between generalized trust and income inequality. The robustness of the inequality-trust link suggests that rising income inequality accounts for the decline of generalized trust in the United States (Hastings 2018). Yet, the severity of this decline is relatively unknown due to the scientific community's reliance on MST and the "most-people trust" question over the past 50 years. In all likelihood, traditional measures of generalized trust have been measuring declines in values, optimism, trustworthiness, *and* generalized trust. Because of such confounding, and the weak correlations I observe between SFT/IST and income inequality, declines in generalized trust may not be as severe as previously observed (Mewes et al. 2021; Paxton 1999; Putnam 2000; Robinson and Jackson 2001).

This study also recommends new theoretical and empirical possibilities for understanding ethnoracial differences in generalized trust. Alongside age and education, scholars have focused on race as "*the* life experience that has the biggest impact on trust (Uslaner 2002: 91)." As such,

many studies find significant ethnoracial gaps in generalized trust (Alesina and La Ferrara 2002; Douds and Wu 2018; Evangelist 2021; Nunnally 2002; Patterson 1999; Smith 1997; Uslander 2002; Wilkes 2011), with Black-White differences being the most striking and prominent (Patterson 1999). Smith (2010) attributes these gaps to historical and contemporary discrimination, neighborhood context, and ethnoracial socialization. Using MST and GST, Robbins (2022) and I replicate the finding that Blacks have lower levels of trust than Whites. But when employing more valid and reliable scales, SFT and IST fail to generate similar ethnoracial differences in generalized trust.

Why might this be the case? The main reason has to do with measurement. Traditional measures of generalized trust capture more than just the “standard estimate of the probability of trustworthiness...for the average person he [or she] meets (Coleman 1990: 104)”, and fail to discriminate from other closely related—but fundamentally different—concepts, which may drive the previously observed ethnoracial gaps in generalized trust. Moreover, Robbins (2021) shows that Whites and non-Whites differ in how they interpret and understand the survey items used in traditional measures. MST and GST, in other words, are non-equivalent across ethnic groups. Like Simpson et al. (2007), I contend that Blacks, Latinos, and other racial groups are not less trusting than Whites because traditional measures “are not well suited to understanding how trust varies with social categories such as race (Simpson et al. 2007: 525).” While Simpson et al. (2007) expect greater trust within ethnoracial groups than across, Robbins (2022) and I find little support for this assertion: estimates of trust in SFT are not consistently stronger for same-race pairings (i.e., respondents and faces of a similar race). So, when Uslander (2002, p. 107) claims that Blacks are much more trusting of other Blacks than “most people” (70 percent versus 23 percent), the source of these ethnoracial differences in trust, I contend, is some unobserved

stable attitude or disposition that is (a) a product of systemic racism and histories of discrimination, and (b) correlated with traditional measures of generalized trust but not SFT or IST.

Beyond ethnoracial differences in trust, a large literature has also tackled the question of whether ethnic diversity erodes generalized trust (Abascal and Baldassari 2015; Dinesen and Sønderskov 2018; Putnam 2007; Schaeffer 2014; van der Meer and Tolsma 2013). A recent review and meta-analysis by Dinesen et al. (2020) clarifies some of the core findings of this body of research. Dinesen et al. (2020) find a consistent negative relationship between ethnic diversity and trust, and show that this “relationship is stronger for trust in neighbors and when ethnic diversity is measured more locally (p. 441).” When using traditional measures of generalized trust, I am able to replicate this negative association at the state-level. But when using SFT, I find that the association between ethnic diversity and generalized trust is negligible, suggesting that social scientists should view the diversity-trust link with caution.

Like the ethnoracial gap in trust, I contend that this null effect is an artifact of measurement: people residing in ethnically homogenous communities may interpret traditional measures of generalized trust in a systematically different way than people living in diverse communities. The ambiguity in the “most people” or “people of” referent categories lead to different interpretations—and ultimately different levels of generalized trust—across ethnically homogeneous and heterogenous social contexts. SFT overcomes this measurement problem by standardizing the strangers and matters of trust under evaluation. As Dinesen et al. (2020) point out, there is a great deal of variation in local levels of ethnic diversity within states, and the diversity-trust link becomes stronger as geographic units become smaller. Thus, my results are tentative. They refute the diversity-trust link previously observed at the state-level, but a more

fine-grained evaluation must be conducted with local-level measures of diversity to contest whether ethnic diversity erodes generalized trust.

Another important finding in the literature concerns the geographic distribution of generalized trust in the United States. Some contend that generalized trust is lower in the South because of collectivist social structures deeply entrenched in these states (Simpson 2006). While my findings support this claim across all four generalized trust scales, I do find that SFT and IST yield different levels of generalized trust across other U.S. regions, namely the Midwest and West. Again, the likely culprit behind this heterogeneity across measures is invalidity and non-equivalence.

The differences in measurement across U.S. regions as well as numerous other findings in the present manuscript, illustrate the fundamentally generative nature of this study. For the current project, the relationship between theory and empirics is synergistic, with theoretical insights stimulating novel empirical tools, which, in turn, spur new hypotheses and propositions. My hope is that these new measurement tools offer a variety of ways for sociologists to engage with—and advance how we think about—generalized trust.

Second, this study recommends against tried-and-true scales of generalized trust, namely MST, GST, and the “most-people trust” question. These traditional measures probably *do* tap generalized trust to some degree, but risk failing to (a) discriminate from other psychological mechanisms, and (b) measure trust the same way across groups. MST and GST, for instance, require researchers to control for measurement confounding by including additional—sometimes burdensome—measures and scales in surveys and statistical models. They also force researchers to contend with, and justify whether, a finding is real or a function of measurement non-

equivalence. An important implication of this study is that both of these hurdles are difficult to not only justify, but overcome.

Consequently, the present study provides a compelling case to incorporate SFT and IST into research on generalized trust. In light of the magnitude of gains to reliability, validity, and measurement equivalence, SFT-60 should be used whenever generalized trust is treated as a key independent *or* dependent variable. The gains to statistical precision are too great to sacrifice in favor of MST, GST, or one of their respective indicators. When survey length is an issue, SFT-24 is an equally reliable and valid substitute. One possible downside is that aspects of these measures are difficult to implement in some survey modes. SFT requires survey takers to see pictures of faces, the order of which should be randomly assigned along with the order of trust domains for each face. While these requirements can be accommodated by most survey modes, telephone surveys present a key hurdle. In the case of telephone surveys, my suggestion is to use IST-6 over MST or GST, since IST-6 is a more valid and reliable measure of generalized trust than common alternatives. This also applies to situations where generalized trust is a mere control variable in a study. Overall, the benefits of SFT and IST far outweigh the possible costs.

Directions for Future Research

Important work has been done investigating the reliability, validity, and measurement equivalence of SFT and IST, but more work is required. Two avenues of research would further solidify SFT and IST as *the* standardized tools for measuring generalized trust. First, MST and the “most-people trust” question have been used in general population surveys for decades, both nationally and cross-nationally. The representation of SFT and IST has yet to be established to such a degree. Doing so requires a major research program, where the primary goal is to establish the reliability and validity of SFT and IST in representative samples from countries

across the globe. This is an important endeavor given the non-equivalence of traditional measures of generalized trust both within and between regions of the world (Delhey et al. 2011; Lolle and Torpe 2011; Reeskens and Hooghe 2008). As Davidov et al. (2014) write: “Determining whether people in certain countries score differently in measurements of interest or whether concepts relate differently to each other across nations can indisputably assist in testing theories and advancing our sociological knowledge (p. 55).” Currently, it is an outstanding question whether—and to what degree—SFT and IST are valid and reliable outside of Western contexts.

Second, it remains to be seen whether the present findings on generalized trust hold when applied to general population surveys in non-Western societies. For instance, is the negative correlation observed between age and SFT-60 unique to the United States or does it exist in other cultural contexts? Another key question is whether SFT and IST correlate with social, political, economic, and cultural indicators of a country. For instance, safety and security (Putnam 2000; Messner et al. 2004), diversity (Cao and Galinsky 2020; Dinesen et al. 2020), and fairness (Uslaner 2002; Rothstein 2011) feature heavily in theoretical models of generalized trust. Investigating whether country-level correlations between generalized trust and intentional homicide, ethnic fractionalization, and income inequality replicate for SFT and IST is an important next step. The results of which will help adjudicate between classic theoretical accounts of generalized trust.

Conclusion

Social scientists have made significant contributions to the measurement of generalized trust. Most of these methodological innovations, however, have yet to match concept and theory with operationalization and measurement. In an effort to bridge this gap, Robbins (2021, 2022)

developed two new measures of generalized trust—SFT and IST—that take the average values of trust across particular matters *and* specific strangers. In the present study, I provide support for the reliability and validity of these new measures with a nationally representative survey and behavioral experiment. Given the current and prior findings, my recommendation is to minimize the continued use of traditional measures in favor of SFT and IST, so that future theories of generalized trust can stand on solid empirical foundations.

Table A1. Summary Statistics and Descriptions of Variables

Name	Definition	Mean	SD	Min.	Max.	N
<i>Trust Scales</i>						
SFT-60	Stranger Face Trust scale (10 faces, 6 matters per face)	0	0.536	-0.742	1.811	1195
SFT-24	Stranger Face Trust scale (6 faces, 4 matters per face)	0	0.550	-0.608	1.883	1157
IST-6	Imaginary Stranger Trust scale (6 matters)	0	0.486	-0.602	1.912	1127
IST-4	Imaginary Stranger Trust scale (4 matters)	0	0.504	-0.458	1.999	1207
MST	Misanthropy scale (3 items)	0	0.260	-0.343	0.324	1233
GST	Generalized Social Trust scale (3 items)	0	0.368	-1.112	0.744	1203
<i>Discriminant Validity</i>						
PST	Particularized Trust scale	0	0.392	-1.654	0.560	1253
POT	Political Trust scale	0	0.368	-1.112	0.744	1203
Patience	Willingness to wait (0 = completely unwilling, 10 = very willing)	7.503	2.120	0	10	1213
Risk Taking	Willingness to take risks (0 = completely unwilling, 10 = very willing)	5.802	2.167	0	10	1233
Gift Exchange	Gift giving choice scenario (gift one of six presents worth between USD \$5 and \$30)	20.479	9.435	0	30	1158
Return a Favor	Willingness to return a favor (0 = does not describe me, 10 = describes me perfectly)	8.963	1.518	0	10	1246
Take Revenge	Willingness to take revenge (0 = does not describe me, 10 = describes me perfectly)	2.534	2.607	0	10	1240
Punish Unfair Behavior (Self)	Willingness to punish unfair behavior (0 = completely unwilling, 10 = very willing)	5.627	2.984	0	10	1162
Punish Unfair Behavior (Others)	Willingness to punish unfair behavior (0 = completely unwilling, 10 = very willing)	5.759	2.744	0	10	1135
Donation Decision	Donation choice scenario (donate between USD \$0 and \$1,000 to a good cause)	148.190	211.953	0	1000	1242
Give to Good Cause	Willingness to give to good cause (0 = completely unwilling, 10 = very willing)	7.991	2.364	0	10	1237
Prosocial Orientation	Resource division choice scenario (0 = none, 9 = nine cooperative choices)	6.570	2.489	0	9	1264
<i>Known-Groups Validity, Individual-Level</i>						
Dispositional Optimism	Dispositional Optimism scale (5 items)	0	0.472	-1.174	0.815	1254
Big-5 Personality Scales	BFI-S (15 items with 3 items per scale)					
Extraversion		0	0.863	-1.912	1.453	1256
Conscientiousness		0	0.431	-1.734	0.541	1255
Openness		0	0.490	-1.797	0.732	1252
Agreeableness		0	0.458	-1.618	0.541	1256
Neuroticism		0	0.824	-1.447	1.456	1254
Age	Age in years	47.461	17.054	19	95	1264
Gender						
Female (omitted)	Female respondent	0.519	---	0	1	656
Male	Male respondent	0.481	---	0	1	608
Race						
Non-Hispanic white (omitted)	Non-Hispanic White respondent	0.635	---	0	1	803
Black	Non-Hispanic Black respondent	0.127	---	0	1	161
Hispanic	Hispanic respondent	0.172	---	0	1	218
Other	Other race or 2+ races respondent	0.064	---	0	1	82
Education						
No HSD (omitted)	No high school diploma or equivalent	0.045	---	0	1	57
HSD	High school diploma or equivalent	0.173	---	0	1	219
Some college or AA	Some college (no degree) or an Associate's degree	0.441	---	0	1	558
BA +	Bachelors, Masters, Professional, or Doctorate degree	0.340	---	0	1	430
Marital Status						
Married (omitted)		0.485	---	0	1	614
Widowed		0.041	---	0	1	52
Divorced		0.134	---	0	1	170
Separated		0.015	---	0	1	20
Never married		0.229	---	0	1	290
Living with Partner		0.093	---	0	1	118
Employment Status						
Paid Employee (omitted)		0.548	---	0	1	693
Self-employed		0.094	---	0	1	120
Not Working		0.356	---	0	1	451
ln(per capita HH income)	natural logged per capita household income	9.837	0.999	6.032	12.561	1264
Region						
Northeast (omitted)	Proportion of panel respondents living in the Northeast region	0.613	---	0	1	204
Midwest	Proportion of panel respondents living in the Midwest region	0.227	---	0	1	287
South	Proportion of panel respondents living in the South region	0.340	---	0	1	430
West	Proportion of panel respondents living in the West region	0.271	---	0	1	343
MSA Status						
Metro (omitted)	Proportion of panel respondents living in metro areas	0.841	---	0	1	1063
Non-metro area	Proportion of panel respondents living in non-metro areas	0.159	---	0	1	201
<i>Known-Groups Validity, State-Level</i>						
Income Inequality	Average Gini coefficient over five years (2014-2018)	46.573	2.043	42.37	52.83	51
Ethnic Heterogeneity	Average ethnic fractionalization over five years (2014-2018)	35.302	15.313	9.2	79.02	51
Intentional Homicide	Average number of deaths (all homicides) per 100,000 over seven years (2013-2019)	5.636	3.212	1.85	18.26	50
<i>Predictive Validity</i>						
Points Sent to Player B	Proportion of points sent to Player B	.419	.339	0	1	302
Points Returned to Player A	Proportion of points returned to Player A	.463	.299	0	1	289

Note: All scales are latent factor scores

References

- Abascal, Maria and Baldassarri. 2015. "Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined." *American Journal of Sociology* 121: 722-782.
- Alesina, Alberto and Eliana La Ferrara. 2002. "Who Trusts Others?" *Journal of Public Economics* 85: 207-234.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8: 155-194.
- Allison, Paul D. 2001. *Missing Data*. London, UK: Sage Publications.
- Almond, Gabriel A. and Sidney Verba. 1963. *The Civic Culture*. Princeton, NJ: Princeton University Press.
- Andersson, Matthew A. 2012. "Dispositional Optimism and the Emergence of Social Network Diversity." *Sociological Quarterly* 53: 92-115.
- Arel-Bundock, Vincent and Krzysztof J. Pelc. 2018. "When Can Multiple Imputation Improve Regression Estimates?" *Political Analysis* 26: 240-245.
- Bauer, Paul. 2021. "Clearing the Jungle: Conceptualizing Trust and Trustworthiness." In Raquel Barradas de Freitas and Sergio Lo Iacono, eds., *Trust Matters: Cross-Disciplinary Essays*. Bloomsbury.
- Bauer, Paul and Markus Freitag. 2018. "Measuring Trust." In Eric Uslaner, ed., *Oxford Handbook of Social and Political Trust*, pp. 15-36. Oxford University Press.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10: 122-142.
- Bradburn, Norman M., Seymour Sudman, Ed Blair, and Carol Stocking. 1978. "Question Threat and Response Bias." *Public Opinion Quarterly* 42: 221-234.
- Brehm, John and Wendy Rahn. 1997. "Individual-Level Evidence for the Causes and Consequences of Social Capital." *American Journal of Political Science* 41: 999-1023.
- Campbell, Donald T. 1960. "Recommendations for APA Test Standards Regarding Construct, Trait, or Discriminant Validity." *American Psychologist* 15: 546-553.
- Campbell, Donald T. and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56: 81-105.
- Cao, Jiyin and Adam D. Galinsky. 2020. "The Diversity-Uncertainty-Valence (DUV) Model of Generalized Trust Development" *Organizational Behavior and Human Decision Processes* 161: 49-64.
- Cheung, Gordon W. and Chang Wang. 2017. "Current Approaches for Assessing Convergent and Discriminant Validity with SEM: Issues and Solutions." *Academy of Management Proceedings* 1.
- Clark, April K. and Marie A. Eisenstein. 2013. "Interpersonal Trust: An Age-Period-Cohort Analysis Revisited." *Social Science Research* 42: 361-375.
- Cohn, Alain, Michel André Maréchal, David Tannenbaum, and Christian Lukas Zünd. 2019. "Civic Honesty Around the Globe." *Science* 365: 70-73.
- Coleman, James. 1990. *Foundations of Social Theory*. Harvard University Press.
- Colquitt, Jason A., Brent A. Scott, and Jeffery A. LePine. 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: 909-27.
- Cook, Karen S., Russell Hardin, and Margaret Levi. 2005. *Cooperation without Trust?* Russell Sage.

- Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt, and Jaak Billiet. 2014. "Measurement Equivalence in Cross-National Research." *Annual Review of Sociology* 40: 55-75.
- Davidson, Megan. 2014. "Known-Groups Validity." In *Encyclopedia of Quality of Life and Well-Being Research*, edited by A.C. Michalos (p. 3481). Springer.
- DeBruine, Lisa. 2021. "Face Research Lab London Set." *Figshare*.
https://figshare.com/authors/Lisa_DeBruine/3235773.
- Delhey, Jan and Kenneth Newton. 2005. "Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism?" *European Sociological Review* 21: 311-327.
- Delhey, Jan, Kenneth Newton, and Christian Welzel. 2011. "How General Is Trust in 'Most People'? Solving the Radius of Trust Problem." *American Sociological Review* 76:786-807.
- Deutsch, Morton. 1960. "Trust, Trustworthiness, and the F Scale." *Journal of Abnormal and Social Psychology* 61: 138-140.
- Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. Wiley.
- Dinesen, Peter Thisted and Kim Mannemar Sønderskov. 2015. "Ethnic Diversity and Social Trust: Evidence from the Micro-Context." *American Sociological Review* 80: 550-573.
- Dinesen, Peter T., Merlin Schaeffer, and Kim Mannemar Sønderskov. 2020. "Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review." *Annual Review of Political Science* 23: 441-465.
- Dinesen, Peter T., Kim Mannemar Sønderskov, Jacob Sohlberg, and Peter Esaiasson. 2021. "Close (Causally Connected) Cousins? Evidence on the Causal Relationship between Political Trust and Social Trust." *Public Opinion Quarterly*, *Forthcoming*.
- DiStefano, Christine. 2002. "The Impact of Categorization with Confirmatory Factor Analysis." *Structural Equation Modeling* 9: 327-346.
- Douds, Kiara and Jie Wu. 2018. "Trust in the Bayou City: Do Racial Segregation and Discrimination Matter for Generalized Trust?" *Sociology of Race and Ethnicity* 4: 567-584.
- Erikson, Erik. H. 1950. "Growth and Crisis of the 'Healthy Personality'." In M. J. E. Senn, ed., *Symposium on the Health Personality*, pp. 91-146. Josiah Macy, Jr. Foundation.
- Evangelist, Michael. 2021. "Narrowing Racial Differences in Trust: How Discrimination Shapes Trust in a Racialized Society." *Social Problems*, *forthcoming*.
- Fairbrother, Malcolm and Isaac W. Martin, 2013. "Does Inequality Erode Social Trust? Results from Multilevel Models of US States and Counties." *Social Science Research* 42: 347-360.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics* 133: 1645-1692.
- Farrell, Andrew M. 2010. "Insufficient Discriminant Validity: A Comment on Bove, Pervan, Beatty, and Shiu (2009)." *Journal of Business Research* 63: 324-327.
- Fornell, Claes and David F. Larcker. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error." *Journal of Marketing Research* 18: 39-50.
- Freitag, Markus and Paul Bauer. 2013. "Testing for Measurement Equivalence in Surveys: Dimensions of Social Trust across Cultural Contexts." *Public Opinion Quarterly* 77: 24-44.
- Freitag, Markus and Paul C. Bauer. 2016. "Personality Traits and the Propensity to Trust Friends and Strangers." *The Social Science Journal* 53: 467-476.
- Freitag, Markus and Richard Traummüller. 2009. "Spheres of Trust: An Empirical Analysis of the Foundations of Particularized and Generalized Trust." *European Journal of Political Research* 48: 782-803.

- Fukuyama, Francis. 1995. *Trust: The Social Virtues and the Creation of Prosperity*. Free Press.
- Gelman, Andrew. 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22: 153-164.
- Glanville, Jennifer L. and Pamela Paxton. 2007. "How do we Learn to Trust? A Confirmatory Tetrad Analysis of the Sources of Generalized Trust." *Social Psychology Quarterly* 70: 230-242.
- Glanville, Jennifer L. and Qianyi Shi. 2020. "The Extension of Particularized Trust to Generalized and Out-Group Trust: The Constraining Role of Collectivism." *Social Forces* 98: 1801-1828.
- Glanville, Jennifer L., Matthew A. Andersson, and Pamela Paxton. 2013. "Do Social Connections Create Trust? An Examination using New Longitudinal Data." *Social Forces* 92: 545-562.
- Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology, 2nd Edition*. Wiley.
- Hahn, Elisabeth, Juliana Gottschling, and Frank M. Spinath. 2012. "Short Measurements of Personality – Validity and Reliability of the GSOEP Big Five Inventory (BFI-S)." *Journal of Research in Personality* 46: 355-359.
- Hardin, Russell. 2002. *Trust & Trustworthiness*. Russell Sage.
- Hardin, Russell. 2006. *Trust*. Polity.
- Hastings, Orestes P. 2018. "Less Equal, Less Trusting? Longitudinal and Cross-Sectional Effects of Income Inequality on Trust in U.S. States, 1973-2012." *Social Science Research* 74: 77-95.
- Herreros, Francisco. 2004. *The Problems of Forming Social Capital. Why Trust?* Palgrave.
- Hu, L. and P. M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling* 6: 1-55.
- Inglehart, Ronald. 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press.
- Kiley, Kevin and Stephen Vaisey. 2020. "Measuring Stability and Change in Personal Culture Using Panel Data." *American Sociological Review* 85: 477-506.
- Kline, Rex B. 2005. *Principles and Practice of Structural Equation Modeling (2nd ed.)*. New York, NY: Guilford Press.
- Knack, Stephen and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *The Quarterly Journal of Economics* 112: 1251-1288.
- Kramer, Roderick M. 1999. "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions." *Annual Review of Psychology* 50: 569-598.
- Krosnick, Jon A. and Leandre R. Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys." In Lars Lyberg et al., eds., *Survey Measurement and Process Quality*, pp. 97-106. Wiley.
- Kuwabara, Ko. 2015. "Do Reputation Systems Undermine Trust? Divergent Effects of Enforcement Type on Generalized Trust and Trustworthiness." *American Journal of Sociology* 120: 1390-1428.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1997. "Trust in Large Organizations." *American Economic Review* 87: 333-338.
- Lee, Cheol-Sung. 2013. "Welfare States and Social Trust." *Comparative Political Studies* 46: 603-630.

- Lundmark, Sebastian, Mikael Gilljam, and Stefan Dahlberg. 2016. "Measuring Generalized Trust: An Examination of Question Wording and the Number of Scale Points." *Public Opinion Quarterly* 80: 26-43.
- Ma, Debbie S., Joshua Correll, and Bernd Wittenbrink. 2015. "The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data." *Behavior Research Methods* 47: 1122-1135.
- Marschall, Melissa J. and Dietlind Stolle. 2004. "Race and the City: Neighborhood Context and the Development of Generalized Trust." *Political Behavior* 26: 125-153.
- Messner, Steven F, Richard Rosenfeld, and Eric P. Baumer. 2004. "Dimensions of Social Capital and Rates of Criminal Homicide." *American Sociological Review* 69: 882-903.
- Mewes, Jan. 2014. "Gen(d)realized Trust: Women, Work, and Trust in Strangers." *European Sociological Review* 30: 373-386.
- Mewes, Jan, Malcolm Fairbrother, Giuseppe Nicola Giordano, Cary Wu, and Rima Wilkes. 2021. "Experiences Matter: A Longitudinal Study of Individual-Level Sources of Declining Social Trust in the United States." *Social Science Research* 95: 102537.
- Michalopoulos, Stelios and Melanie Meng Xue. 2021. "Folklore." *Quarterly Journal of Economics*, *Forthcoming*.
- Miller, Alan S. and Tomoko Mitamura. 2003. "Are Surveys on Trust Trustworthy?" *Social Psychology Quarterly* 66: 60-70.
- Muthén, Bengt O. and David Kaplan. 1985. "A Comparison of Some Methodologies for the Factor Analysis of Non-Normal Likert Variables." *British Journal of Mathematical and Statistical Psychology* 38: 171-189.
- Muthén, Bengt O. and Linda K. Muthén. 2017. *Mplus 8* [Computer software]. Los Angeles: Authors.
- Newton, Kenneth and Sonja Zmerli. 2011. "Three Forms of Trust and Their Association." *European Political Science Review* 3: 169-200.
- Nunn, Nathan and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101: 3221-3252.
- Nunnally, Jum. 1978. *Psychometric Theory*. McGraw-Hill.
- OPOR. 1942. *Office of Public Opinion Research, Survey 213 'War'*.
- Oskarsson, Sven, Peter Thisted Dinesen, Christopher T. Dawes, Magnus Johannesson, and Patrik K. E. Johannesson. 2017. "Education and Social Trust: Testing a Causal Hypothesis Using the Discordant Twin Design." *Political Psychology* 38: 515-531.
- Patterson, Michael O. 1999. "Liberty against the Democratic State: On the Historical and Contemporary Sources of American Distrust." Pp. 151-207 in *Democracy and Trust*. Mark E. Warren, editor. Cambridge University Press.
- Paxton, Pamela. 1999. "Is Social Capital Declining in the United States? A multiple indicator assessment." *American Journal of Sociology* 105: 88-127.
- Paxton, Pamela. 2007. "Association Memberships and Generalized Trust: A Multilevel Model across 31 countries." *Social Forces* 86: 47-76.
- Paxton, Pamela, and Jennifer Glanville, 2015. "Is Trust Rigid or Malleable? A Laboratory Experiment." *Social Psychology Quarterly* 78: 194-204.
- Pedhazur, Elazar J. and Liora Pedhazur Schmelkin. 1991. *Measurement, Design, and Analysis: An Integrated Approach*. Lawrence Erlbaum Associates.
- Pepinsky, Thomas J. 2018. "A Note on Listwise Deletion Versus Multiple Imputation." *Political Analysis* 26: 480-488.

- Peterson, Robert. 2001. "On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-Analysis." *Journal of Consumer Research* 28: 450-461.
- Putnam, Robert. 1993. *Making Democracy Work*. Princeton University Press.
- Putnam, Robert. 2000. *Bowling Alone: The Collapse and Revival of American Community*. Free Press.
- Putnam, Robert. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-First Century." *Scandinavian Political Studies* 30: 137-174.
- Reeskens, Tim and Marc Hooghe. 2008. "Cross-Cultural Measurement Equivalence of Generalized Trust. Evidence from the European Social Survey (2002 and 2004)." *Social Indicators Research* 85: 515-532.
- Robbins, Blaine. 2012. "Institutional Quality and Generalized Trust: A Nonrecursive Causal Model." *Social Indicators Research* 107: 234-258.
- Robbins, Blaine. 2016a. "From the General to the Specific: How Social Trust Motivates Relational Trust." *Social Science Research* 55: 16-30.
- Robbins, Blaine. 2016b. "What is Trust? A Multidisciplinary Review, Critique, and Synthesis." *Sociology Compass* 10: 972 - 986.
- Robbins, Blaine G. 2022. "Measuring Generalized Trust: Two New Approaches." *Sociological Methods & Research* 51: 305-356.
- Robbins, Blaine. 2021. "An Empirical Comparison of Four Generalized Trust Scales: Test-Retest Reliability, Measurement Invariance, Predictive Validity, and Replicability." *Sociological Methods & Research* (First View).
- Robinson, Robert V. and Elton F. Jackson. 2001. "Is Trust Declining in America? An Age-Period-Cohort Analysis." *Social Science Research* 30: 117-145.
- Rosenberg, M. 1956. "Misanthropy and Political Ideology." *American Sociological Review* 21: 690-695.
- Rosenfeld, Richard, Eric P. Baumer, and Steven F. Messner. 2007. "Social Trust, Firearm Prevalence, and Homicide." *Annals of Epidemiology* 17: 119-125.
- Rotenberg, Ken J. 1995. "The Socialisation of Trust: Parents' and Children's Interpersonal Trust." *International Journal of Behavioral Development* 18: 713-726.
- Rothstein, Bo and Eric Uslaner. 2005. "All for All: Equality, Corruption, and Social Trust." *World Politics* 58: 41-72.
- Rotter, J. B., 1967. "A New Scale for the Measurement of Interpersonal Trust." *Journal of Personality* 35: 651-665.
- Schaeffer, Merlin. 2014. *Ethnic Diversity and Social Cohesion: Immigration, Ethnic Fractionalization, and Potentials for Civic Action*. Ashgate.
- Scheier, Michael F., Charles S. Carver, and Michael W. Bridges. 1994. "Distinguishing Optimism from Neuroticism (and Trait Anxiety, Self-Mastery, and Self-Esteem): A Reevaluation of the Life Orientation Test." *Journal of Personality and Social Psychology* 67: 1063-1078.
- Schilke, Oliver, Martin Reimann, and Karen S. Cook. 2021. "Trust in Social Relations." *Annual Review of Sociology* 47: 239-259.
- Schwadel, Philip and Michael Stout. 2012. "Age, Period and Cohort Effects on Social Capital." *Social Forces* 91: 233-252.
- Sears, D. O. and P. J. Henry. 2005. "Over Thirty Years Later: A Contemporary Look at Symbolic Racism and Its Critics." *Advances in Experimental Social Psychology* 37: 95-150.

- Simpson, Brent. 2006. "The Poverty of Trust in the Southern United States." *Social Forces* 84: 1625-1623.
- Simpson, Brent and Robb Willer. 2015. "Beyond Altruism: Sociological Foundations of Cooperation and Prosocial Behavior." *Annual Review of Sociology* 41: 43-63.
- Simpson, Brent, Tucker McGrimmon, and Kyle Irwin. 2007. "Are Blacks Really Less Trusting than Whites? Revisiting the Race and Trust Question." *Social Forces* 86: 525-552.
- Smith, Tom W. 1997. "Factors Relating to Misanthropy in Contemporary American Society." *Social Science Research* 26: 170-196.
- Smith, Sandra S. 2010. "Race and Trust." *Annual Review of Sociology* 36: 453-475.
- Stark, Tobias H., Floor M. van Maaren, Jon A. Krosnick, and Gaurav Sood. 2022. "The Impact of Social Desirability Pressures on Whites' Endorsement of Racial Stereotypes: A Comparison Between Oral and ACASI Reports in a National Survey." *Sociological Methods & Research* 51: 605-631.
- Sturgis, Patrick and Patten Smith. 2010. "Assessing the Validity of Generalized Trust Questions: What Kind of Trust are we Measuring?" *International Journal of Public Opinion Research* 22: 74-92.
- Tabellini, Guido, 2010. "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association* 8: 677-716.
- Torpe, Lars and Henrik Lolle. 2011. "Identifying Social Trust in Cross-Country Analysis: Do we Really Measure the Same?" *Social Indicators Research* 103: 481-500.
- Uslaner, Eric M. 1998. "Social Capital, Television, and the "Mean World": Trust, Optimism, and Civic Participation." *Political Psychology* 19: 441-467.
- Uslaner, Eric M. 2002. *The Moral Foundations of Trust*. Cambridge: Cambridge University Press.
- Uslaner, Eric M. 2011. "Generalized Trust Questions." In John H. Aldrich and Kathleen M. McGraw, eds., *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*, pp. 101-114. Princeton University Press.
- Uslaner, Eric M. 2015. "Measuring Generalized Trust: In Defense of the 'Standard' Question." In Fergus Lyon, Guido Möllering, and Mark N.K. Saunders, eds., *Handbook of Research Methods on Trust*, 2nd Edition, pp. 97-106. Elgar.
- Uslaner, Eric M, ed. 2018. *The Oxford Handbook of Social and Political Trust*. Oxford University Press.
- Vaisey, Stephen and Omar Lizardo. 2016. "Cultural Fragmentation or Acquired Dispositions? A New Approach to Accounting for Patterns of Cultural Change." *Socius* 2: 1-15.
- Van Der Meer, Tom and Jochem Tolsma. 2014. "Ethnic Diversity and its Effects on Social Cohesion." *Annual Review of Sociology* 40: 459-478.
- Van Hoorn, Andre. 2014. "Trust Radius versus Trust Level: Radius of Trust as a Distinct Trust Construct." *American Sociological Review* 79: 1256-1259.
- Van Lange, Paul A. M., Wilma Otten, Ellen M. N. De Bruin, and Jeffrey A. Joireman. 1997. "Development of Prosocial, Individualistic, and Competitive Orientations: Theory and Preliminary Evidence." *Journal of Personality and Social Psychology* 73: 733-46.
- Verba, Sidney, Kay Lehman Scholzman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Harvard University Press.
- Whiteley, Paul F. 2000. "Economic Growth and Social Capital." *Political Studies* 48: 443-466.
- Wilkes, Rima. 2011. "Re-thinking the Decline in Trust: A Comparison of Black and White Americans." *Social Science Research* 40 1596-1610.

- Wu, Cary. 2020. "Does Migration Affect Trust? Internal Migration and the Stability of Trust Among Americans." *Sociological Quarterly* 61: 523-543.
- Yamagishi, Toshio and Midori Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18: 129-166.
- You, Jong-sung. 2012. "Social Trust: Fairness Matters More than Homogeneity." *Political Psychology* 33: 701-721.
- Zmerli, Sonja and Kenneth Newton. 2008. "Social Trust and Attitudes Toward Democracy." *Public Opinion Quarterly* 72: 706-724.

Table 1. Descriptive Statistics of SFT-24, SFT-60, IST-4, IST-6, MST, and GST

Variables	α	M	SD	Min	Max	N
Convergent Validity						
<i>Stranger face trust scale</i>						
SFT-24	.970					
White female (WF)	.878	0.75	0.73	0	3	1127
White male (WM)	.871	0.60	0.66	0	3	1128
Black female (BF)	.884	0.77	0.75	0	3	1123
Black male (BM)	.874	0.75	0.73	0	3	1119
Latin female (LF)	.881	0.75	0.74	0	3	1121
Latin male (LM)	.872	0.70	0.70	0	3	1117
SFT-60	.983					
White female (WF)	.859	1.01	0.69	0	3	1164
White male (WM)	.867	0.91	0.67	0	3	1165
Black female (BF)	.872	1.02	0.71	0	3	1161
Black male (BM)	.875	1.08	0.72	0	3	1171
Latin female (LF)	.868	1.01	0.70	0	3	1161
Latin male (LM)	.872	1.02	0.69	0	3	1165
East Asian female (EF)	.860	1.03	0.71	0	3	1166
East Asian male (EM)	.877	1.02	0.70	0	3	1168
South Asian female (SF)	.867	1.02	0.71	0	3	1161
South Asian male (SM)	.880	0.97	0.70	0	3	1167
<i>Imaginary stranger trust scale</i>						
IST-4	.829					
IST-6	.822					
Secret (SECRET)		0.50	0.73	0	3	1141
Loan (LOAN)		0.58	0.77	0	3	1128
Child (CHILD)		0.40	0.70	0	3	1170
Advice (ADVICE)		0.81	0.81	0	3	1121
Wallet (WALLET)		1.23	0.86	0	3	1140
Tire (TIRE)		1.66	0.79	0	3	1173
<i>Misanthropy scale</i>						
MST	.744					
Trust (TRUST)		0.35	0.47	0	1	1196
Fair (FAIR)		0.61	0.48	0	1	1101
Helpful (HELP)		0.53	0.49	0	1	1156
<i>Generalized social trust scale</i>						
GST	.760					
Meet first time (FIRST)		1.27	0.75	0	3	1150
Another religion (RELI)		1.88	0.67	0	3	1072
Another nationality (NATION)		1.87	0.65	0	3	1047

Note: For SFT-24 and SFT-60, face scales (e.g., WF) were generated using row mean methods. In the analysis, I use latent factor scores of each face.

Table 2. Confirmatory Factor Analysis Assessing Convergent and Discriminant Validity

Latent Factor	Indicator	Model 1: SFT-24, IST-4			Model 2: SFT-60, IST-6		
		Loading	VE	AVE	Loading	VE	AVE
SFT				.852			.861
	WF	0.921***	.848		0.912***	.831	
	WM	0.885***	.783		0.877***	.769	
	BF	0.915***	.837		0.916***	.839	
	BM	0.938***	.879		0.936***	.876	
	LF	0.933***	.870		0.934***	.872	
	LM	0.947***	.896		0.945***	.893	
	EF				0.935***	.874	
	EM				0.949***	.900	
	SF				0.944***	.891	
	SM				0.934***	.872	
IST				.565			.458
	SECRET	0.762***	.580		0.724***	.524	
	LOAN	0.804***	.646		0.795***	.632	
	CHILD	0.745***	.555		0.716***	.512	
	ADVICE	0.694***	.481		0.717***	.514	
	WALLET				0.606***	.367	
	TIRE				0.449***	.201	
MST				.483			.483
	TRUST	0.662***	.438		0.668***	.446	
	FAIR	0.725***	.526		0.724***	.524	
	HELP	0.698***	.487		0.693***	.480	
GST				.600			.600
	FIRST	0.597***	.356		0.601***	.361	
	RELI	0.829***	.687		0.828***	.685	
	NATION	0.870***	.757		0.869***	.755	
		<i>r</i>	<i>r</i> ²		<i>r</i>	<i>r</i> ²	
	<i>Corr</i> (SFT, IST)	.737	.543		.734	.538	
	<i>Corr</i> (SFT, MST)	.299	.089		.354	.125	
	<i>Corr</i> (SFT, GST)	.313	.097		.380	.144	
	<i>Corr</i> (IST, MST)	.284	.080		.357	.127	
	<i>Corr</i> (IST, GST)	.247	.061		.327	.106	
	<i>Corr</i> (MST, GST)	.537	.288		.538	.289	
		<i>N</i> = 1259			<i>N</i> = 1259		
		χ^2 (98) = 351.586, <i>p</i> < .001			χ^2 (203) = 800.705, <i>p</i> < .001		
		RMSEA = .045			RMSEA = .048		
		90% C.I. RMSEA = [.040, .050]			90% C.I. RMSEA = [.045, .052]		
		CFI = .966			CFI = .950		
		SRMR = .054			SRMR = .065		

*** *p* < .001

Notes : standardized factor loadings. Indicators for SFT-24, SFT-60, IST-4, IST-6, MST, and GST treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). VE = variance extracted (i.e., the R^2 or the Squared Multiple Correlation); AVE = average variance extracted (the average of the VEs for each latent construct). All correlations between latent constructs were statistically significant at the *p* < .05 level.

Table 3. Structural Equation Models Assessing Known-Groups Validity of SFT-60, IST-6, MST, and GST

	Model 1	Model 2	Model 3	Model 4
	SFT-60	IST-6	MST	GST
<i>Latent factors</i>				
Dispositional optimism	0.035 (0.046)	0.043 (0.050)	0.346*** (0.051)	0.242*** (0.050)
<i>Big-5 personality scales</i>				
Extraversion	0.066 (0.039)	0.043 (0.041)	0.044 (0.042)	0.039 (0.041)
Conscientiousness	-0.256*** (0.066)	-0.237*** (0.068)	-0.348*** (0.073)	-0.065 (0.073)
Openness	0.073 (0.058)	0.011 (0.063)	0.073 (0.059)	0.031 (0.062)
Agreeableness	0.187** (0.065)	0.187** (0.068)	0.197** (0.068)	0.090 (0.074)
Neuroticism	-0.007 (0.050)	-0.059 (0.055)	-0.064 (0.056)	0.067 (0.054)
<i>Observed covariates</i>				
Age	-0.033 (0.041)	-0.013 (0.045)	0.175*** (0.047)	0.075 (0.045)
Female (ref.)				
Male	0.016 (0.034)	0.052 (0.037)	0.016 (0.038)	-0.026 (0.036)
Non-Hispanic white (ref.)				
Black	0.048 (0.034)	0.025 (0.040)	-0.117** (0.036)	-0.012 (0.036)
Hispanic	-0.066 (0.033)	-0.034 (0.036)	-0.099** (0.037)	-0.066 (0.037)
Other	-0.004 (0.029)	-0.002 (0.031)	-0.024 (0.072)	-0.038 (0.034)
No HSD (ref.)				
HSD	-0.018 (0.058)	-0.014 (0.069)	-0.024 (0.072)	-0.024 (0.070)
Some college or AA	0.074 (0.073)	0.023 (0.086)	0.023 (0.091)	-0.028 (0.087)
BA +	0.091 (0.073)	0.043 (0.085)	0.116 (0.090)	0.013 (0.085)
Married (ref.)				
Widowed	-0.004 (0.029)	-0.027 (0.037)	-0.040 (0.032)	-0.008 (0.035)
Divorced	-0.023 (0.030)	-0.046 (0.031)	-0.052 (0.033)	-0.028 (0.033)
Separated	0.013 (0.026)	0.003 (0.036)	-0.037 (0.031)	-0.004 (0.036)
Never married	-0.037 (0.036)	-0.009 (0.038)	-0.024 (0.039)	-0.026 (0.039)
Living with partner	0.024 (0.031)	-0.005 (0.035)	-0.026 (0.033)	0.009 (0.033)
Paid employee (ref.)				
Self-employed	-0.010 (0.032)	0.043 (0.035)	-0.035 (0.032)	-0.017 (0.030)
Not working	-0.013 (0.034)	-0.018 (0.039)	0.014 (0.037)	0.049 (0.035)
ln(per capita HH income)	0.062 (0.036)	-0.008 (0.042)	0.046 (0.040)	0.047 (0.038)
Northeast (ref.)				
Midwest	0.099* (0.042)	0.061 (0.043)	-0.010 (0.041)	0.029 (0.044)
South	0.006 (0.043)	-0.026 (0.044)	-0.091* (0.043)	-0.068 (0.049)
West	0.170*** (0.043)	0.095* (0.044)	-0.074 (0.042)	0.042 (0.045)
Metro area (ref.)				
Non-metro area	-0.022 (0.029)	-0.037 (0.032)	0.083** (0.030)	0.017 (0.031)
	<i>N</i> = 1264	<i>N</i> = 1264	<i>N</i> = 1264	<i>N</i> = 1264

* $p < .05$, ** $p < .01$, *** $p < .001$

Notes : Standardized coefficients (robust standard errors). Indicators for all latent factors treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). Exogenous latent factors and observed covariates were allowed to freely covary (not shown). Control variables included household internet access, home ownership, building type of residence, telephone service for the household, survey duration, and survey mode device.

Figure 1. Standardized Coefficients Between Generalized Trust Scales (SFT-60 and IST-6) and Measures of Economic Preferences

Notes: Bars around standardized coefficients reflect 95% CIs.

Figure 2. Standardized Coefficients Between Generalized Trust Scales (MST and GST) and Measures of Economic Preferences

Notes: Bars around standardized coefficients reflect 95% CIs.

Figure 3. Scatter Plots of State-Level Indicators and Generalized Trust Scales

Notes: Solid red lines are best linear unbiased predictions. Dashed red lines reflect 95% CIs. Solid blue markers represent each state.

Figure 4. Standardized Coefficients Between Generalized Trust Scales and Proportion of Points Sent by Player A

Notes: Green bars around standardized coefficients reflect 95% CIs.

Figure 5. Standardized Coefficients Between Generalized Trust Scales and Proportion of Points Returned by Player B

Notes: Brown bars around standardized coefficients reflect 95% CIs.

Supplemental Material

Valid and Reliable Measures of Generalized Trust: Evidence from a Nationally Representative Survey and Behavioral Experiment

Contents

1. Original SFT Faces	
1.1. White Female	2
1.2. White Male	3
1.3. Black Female	4
1.4. Black Male	5
1.5. Latin Female	6
1.6. Latin Male	7
2. New SFT Faces	
2.1. East Asian Female	8
2.2. East Asian Male	9
2.3. South Asian Female	10
2.4. South Asian Male	11
3. Representativeness	
3.1. Table S1. Representativeness of Survey Sample Characteristics	12
3.2. Table S2. Representativeness of Trust Game Sample Characteristics	13
4. Post-Stratification Survey Weights	
4.1. Table S3. CFA assessing convergent and discriminant validity with survey weights	14
4.2. Table S4. Known-groups validity with survey weights	15
4.3. Figure S1. Discriminant validity of SFT-60 and IST-6 with survey weights	16
4.4. Figure S2. Discriminant validity of MST and GST with survey weights	17
4.5. Figure S3. Predictive validity of points sent by Player A with survey weights	18
4.6. Figure S4. Predictive validity of points returned by Player B with survey weights	19
5. Original Estimation Strategy	
5.1. Table S5. CFAs using Robbins (2019, 2021) original estimation strategy	20
6. Validation Tests for SFT-24 and IST-4	
6.1. Table S6. Known-groups validity	21
6.2. Figure S5. Discriminant validity	22
6.3. Figure S6. Predictive validity of points sent by Player A	23
6.4. Figure S7. Predictive validity of points returned by Player B	24
6.5. Figure S8. Scatter plots of state-level indicators and generalized trust scales	25
7. Measurement Invariance of SFT-60 and IST-6	
7.1. Table S7. MI of SFT-60 by gender, race, education and age-groups	28
7.2. Table S8. MI of IST-6 by gender, race, education and age-groups	29
8. Validation Tests of MST Indicators (TRUST, FAIR, and HELP)	
8.1. Table S9. Known-groups validity	30
8.2. Figure S9. Discriminant validity	31
8.3. Figure S10. Predictive validity of points sent by Player A	32
8.4. Figure S11. Predictive validity of points returned by Player B	33
8.5. Figure S12. Scatter plots of state-level indicators and MST indicators (TRUST, HELP, and FAIR)	34
9. CFA of Latent Constructs	
9.1. Table S10. CFA of particularized social trust (PST)	35
9.2. Table S11. CFA of political trust (POT)	36
9.3. Table S12. CFA of dispositional optimism (OPT)	37

9.4. Table S13. CFA of Big-5 personality scales.....	38
10. References	
10.1.	39

1. Original SFT Faces

IMAGE 1 – White Female

IMAGE 2 – White Male

IMAGE 3 – Black Female

IMAGE 4 – Black Male

IMAGE 5 – Latin Female

IMAGE 6 – Latin Male

2. New SFT Faces

IMAGE 7 – East Asian Female

IMAGE 8 – East Asian Male

IMAGE 9 – South Asian Female

IMAGE 10 – South Asian Male

3. Representativeness

Table S1. Comparison of Unweighted and Weighted Sample Characteristics to Population Benchmark, Survey Sample

	Unweighted	Weighted	Benchmark	Difference
Household Income				
Less than \$29,999	23.6	24.1	17.5	6.6
\$30,000 to \$74,999	41.6	38.7	33.1	5.6
\$75,000 to \$124,999	22.9	23.1	24.6	-1.5
\$125,000 Plus	11.9	14.1	24.9	-10.8
Age				
18 - 34	32.3	29.3	29.3	0
35 - 49	22.4	24.3	24.3	0
50 - 64	24.8	24.9	24.9	0
65 Plus	20.6	21.5	21.5	0
Race-Ethnicity				
Non-Hispanic White	63.5	62.8	62.8	0
Non-Hispanic Black	12.7	11.9	11.9	0
Hispanic	17.2	16.7	16.7	0
Non-Hispanic Asian and Pacific Islander	2.7	6.1	6.4	-0.3
Non-Hispanic Others	3.8	2.6	2.2	0.4
Education Status				
Less than High School	4.5	9.8	9.8	0
High School Equivalent	17.3	28.2	28.2	0
Some College or Associate Degree	44.1	27.7	27.7	0
Bachelor's Degree	20.6	19.8	21.8	-2
Graduate Degree	13.4	14.5	12.4	2.1
Household Ownership				
Owner Occupied	63.6	71.4	67.5	3.9
Renter Occupied and Other	36.4	28.6	32.5	-3.9
Children in Household				
With 1+ Under 18 Years	29	27.4	33.1	-5.7
Without Children Under 18	71	72.6	66.9	5.7
Marital Status				
Currently Married	48.6	49.4	52.6	-3.2
Separated, Divorced, Widowed, or Single	51.4	50.6	47.5	3.2
Sex				
Male	48.1	48.3	48.3	0
Female	51.9	51.7	51.7	0

Note: Survey sample ($N = 1,264$). The "difference" column refers to the numerical difference between weighted and benchmark values. Benchmark values were drawn from the Current Population Survey. Weighted Cumulative Response Rate is 6.8%.

Table S2. Comparison of Unweighted and Weighted Sample Characteristics to Population Benchmark, Trust Game Subsample

	Unweighted	Weighted	Benchmark	Difference
Household Income				
Less than \$29,999	22.7	26.3	15.1	11.2
\$30,000 to \$74,999	42.5	40.1	30.3	9.8
\$75,000 to \$124,999	24.3	23.5	24.2	-0.7
\$125,000 Plus	10.4	10.2	30.4	-20.2
Age				
18 - 34	30.1	29.5	29.5	0
35 - 49	26.8	24.2	24.2	0
50 - 64	23.7	24.6	24.6	0
65 Plus	19.4	21.7	21.7	0
Race-Ethnicity				
Non-Hispanic White	63.2	62.8	62.8	0
Non-Hispanic Black	13.6	12	12	0
Hispanic	16.9	16.7	16.7	0
Non-Hispanic Asian and Pacific Islander	3.5	6.5	6.5	0
Non-Hispanic Others	2.8	2.1	2.2	-0.1
Education Status				
Less than High School	3	9.8	9.8	0
High School Equivalent	18	27.8	27.8	0
Some College or Associate Degree	42.9	27.6	27.6	0
Bachelor's Degree	20.7	19.9	22.1	-2.2
Graduate Degree	15.4	14.9	12.7	2.2
Household Ownership				
Owner Occupied	65.4	70.8	70.1	0.7
Renter Occupied and Other	24.6	29.2	29.9	-0.7
Children in Household				
With 1+ Under 18 Years	31.1	28.7	33.2	-4.5
Without Children Under 18	68.9	71.3	66.8	4.5
Marital Status				
Currently Married	51.5	50.1	52.8	-2.7
Separated, Divorced, Widowed, or Single	48.5	49.9	47.2	2.7
Sex				
Male	50.7	48.5	48.5	0
Female	49.3	51.5	51.5	0

Note: Trust game subsample ($N = 604$). The "difference" column refers to the numerical difference between weighted and benchmark values. Benchmark values were drawn from the Current Population Survey. Weighted Cumulative Response Rate is 15.5%.

4. Post-Stratification Survey Weights

Table S3. Confirmatory Factor Analysis Assessing Convergent and Discriminant Validity with Post-Stratification Survey Weights

Latent Factor	Indicator	Model 1: SFT-24, IST-4			Model 2: SFT-60, IST-6		
		Loading	VE	AVE	Loading	VE	AVE
SFT				.845			.859
	WF	0.915***	.837		0.904***	.817	
	WM	0.880***	.775		0.878***	.772	
	BF	0.920***	.828		0.909***	.827	
	BM	0.944***	.891		0.944***	.891	
	LF	0.929***	.863		0.934***	.873	
	LM	0.936***	.876		0.934***	.872	
	EF				0.939***	.882	
	EM				0.953***	.908	
	SF				0.939***	.882	
	SM				0.932***	.868	
IST				.568			.464
	SECRET	0.755***	.570		0.715***	.512	
	LOAN	0.811***	.657		0.798***	.637	
	CHILD	0.747***	.558		0.716***	.513	
	ADVICE	0.700***	.490		0.727***	.529	
	WALLET				0.602***	.363	
	TIRE				0.485***	.235	
MST				.487			.487
	TRUST	0.649***	.422		0.655***	.430	
	FAIR	0.706***	.499		0.708***	.501	
	HELP	0.735***	.541		0.728***	.530	
GST				.598			.599
	FIRST	0.559***	.312		0.565***	.319	
	RELI	0.835***	.697		0.834***	.695	
	NATION	0.886***	.786		0.885***	.783	
		<i>r</i>	<i>r</i> ²		<i>r</i>	<i>r</i> ²	
	<i>Corr</i> (SFT, IST)	.750	.562		.743	.552	
	<i>Corr</i> (SFT, MST)	.277	.076		.324	.104	
	<i>Corr</i> (SFT, GST)	.296	.087		.368	.135	
	<i>Corr</i> (IST, MST)	.264	.069		.330	.108	
	<i>Corr</i> (IST, GST)	.245	.060		.326	.106	
	<i>Corr</i> (MST, GST)	.513	.263		.515	.265	
		<i>N</i> = 1259			<i>N</i> = 1259		
		χ^2 (98) = 254.499, <i>p</i> < .001			χ^2 (203) = 570.192, <i>p</i> < .001		
		RMSEA = .036			RMSEA = .038		
		90% C.I. RMSEA = [.030, .041]			90% C.I. RMSEA = [.034, .042]		
		CFI = .963			CFI = .952		
		SRMR = .056			SRMR = .066		

*** *p* < .001

Notes : standardized factor loadings. Indicators for SFT-24, SFT-60, IST-4, IST-6, MST, and GST treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). VE = variance extracted (i.e., the R^2 or the Squared Multiple Correlation); AVE = average variance extracted (the average of the VEs for each latent construct). All correlations between latent constructs were statistically significant at the *p* < .05 level.

Table S4. Structural Equation Models Assessing Known-Groups Validity of SFT-60, IST-6, MST, and GST with Post-Stratification Survey Weights

	Model 1	Model 2	Model 3	Model 4
	SFT-60	IST-6	MST	GST
<i>Latent factors</i>				
Dispositional optimism	0.016 (0.054)	0.003 (0.063)	0.292*** (0.063)	0.280*** (0.056)
<i>Big-5 personality scales</i>				
Extraversion	0.070 (0.046)	0.066 (0.050)	0.056 (0.060)	0.066 (0.050)
Conscientiousness	-0.308*** (0.087)	-0.252** (0.084)	-0.174 (0.103)	-0.120 (0.086)
Openness	0.055 (0.073)	-0.025 (0.081)	0.049 (0.078)	0.051 (0.075)
Agreeableness	0.199*** (0.087)	0.230* (0.096)	0.115 (0.095)	0.075 (0.090)
Neuroticism	0.017 (0.057)	-0.024 (0.068)	-0.061 (0.077)	0.150* (0.064)
<i>Observed covariates</i>				
Age	-0.052 (0.053)	0.004 (0.066)	0.231*** (0.060)	0.134* (0.055)
Female (ref.)				
Male	-0.006 (0.043)	0.044 (0.052)	0.006 (0.056)	-0.011 (0.042)
Non-Hispanic white (ref.)				
Black	0.033 (0.039)	0.072 (0.057)	-0.134** (0.047)	-0.031 (0.045)
Hispanic	-0.091* (0.040)	-0.062 (0.044)	-0.109* (0.047)	-0.054 (0.047)
Other	-0.016 (0.038)	0.060 (0.057)	0.035 (0.057)	-0.028 (0.041)
No HSD (ref.)				
HSD	-0.032 (0.074)	-0.044 (0.094)	-0.056 (0.097)	-0.090 (0.087)
Some college or AA	0.068 (0.069)	0.012 (0.091)	0.008 (0.096)	-0.117 (0.082)
BA +	0.061 (0.078)	0.007 (0.100)	0.113 (0.105)	-0.048 (0.085)
Married (ref.)				
Widowed	0.022 (0.036)	0.034 (0.072)	-0.042 (0.045)	0.020 (0.041)
Divorced	-0.032 (0.036)	-0.068 (0.036)	-0.064 (0.040)	-0.067 (0.048)
Separated	0.009 (0.030)	-0.005 (0.042)	-0.019 (0.034)	0.037 (0.033)
Never married	-0.012 (0.046)	0.036 (0.055)	0.036 (0.054)	0.031 (0.045)
Living with partner	0.044 (0.040)	-0.012 (0.048)	-0.015 (0.042)	0.037 (0.040)
Paid employee (ref.)				
Self-employed	-0.006 (0.038)	0.009 (0.041)	-0.032 (0.044)	-0.017 (0.035)
Not working	-0.030 (0.042)	-0.074 (0.052)	0.002 (0.051)	0.000 (0.043)
ln(per capita HH income)	0.146** (0.047)	0.082 (0.058)	0.013 (0.062)	0.079 (0.051)
Northeast (ref.)				
Midwest	0.135* (0.052)	0.045 (0.048)	0.016 (0.051)	0.061 (0.052)
South	0.016 (0.052)	-0.006 (0.054)	-0.070 (0.055)	-0.023 (0.060)
West	0.168* (0.050)	0.106* (0.053)	-0.010 (0.051)	0.038 (0.053)
Metro area (ref.)				
Non-metro area	-0.047 (0.038)	-0.031 (0.040)	0.088* (0.043)	0.031 (0.041)
	<i>N</i> = 1264	<i>N</i> = 1264	<i>N</i> = 1264	<i>N</i> = 1264

* $p < .05$, ** $p < .01$, *** $p < .001$

Notes : Standardized coefficients (robust standard errors). Indicators for all latent factors treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). Exogenous latent factors and observed covariates were allowed to freely covary (not shown). Control variables included household internet access, home ownership, building type of residence, telephone service for the household, survey duration, and survey mode device.

Figure S1. Standardized Coefficients Between Generalized Trust Scales (SFT-60 and IST-6) and Measures of Economic Preferences

Notes: Bars around standardized coefficients reflect 95% CIs with Survey Weights

Figure S2. Standardized Coefficients Between Generalized Trust Scales (MST and GST) and Measures of Economic Preferences

Notes: Bars around standardized coefficients reflect 95% CIs with Survey Weights

Figure S3. Standardized Coefficients Between Generalized Trust Scales and Proportion of Points Sent by Player A

Notes: Green bars around standardized coefficients reflect 95% CIs with survey weights

Figure S4. Standardized Coefficients Between Generalized Trust Scales and Proportion of Points Returned by Player B

Notes: Brown bars around standardized coefficients reflect 95% CIs with survey weights

5. Original Estimation Strategy

Table S5. Confirmatory Factor Analysis Assessing Convergent and Discriminant Validity

Latent Factor	Indicator	Model 1: SFT-24, IST-4			Model 2: SFT-24, IST-4		
		Loading	VE	AVE	Loading	VE	AVE
SFT				.851			.852
	WF	0.920***	.846		0.921***	.848	
	WM	0.883***	.779		0.885***	.784	
	BF	0.913***	.835		0.915***	.836	
	BM	0.938***	.880		0.938***	.881	
	LF	0.932***	.868		0.933***	.870	
	LM	0.939***	.900		0.947***	.898	
IST				.568			.565
	SECRET	0.763***	.581		0.762***	.580	
	LOAN	0.802***	.644		0.804***	.646	
	CHILD	0.746***	.557		0.745***	.555	
	ADVICE	0.700***	.491		0.694***	.482	
MST				.483			.483
	TRUST	0.662***	.438		0.622***	.449	
	FAIR	0.725***	.526		0.725***	.525	
	HELP	0.698***	.487		0.698***	.475	
GST				.600			.601
	FIRST	0.597***	.356		0.597***	.365	
	RELI	0.829***	.687		0.829***	.686	
	NATION	0.870***	.757		0.870***	.751	
		<i>r</i>	<i>r</i> ²		<i>r</i>	<i>r</i> ²	
	<i>Corr</i> (SFT, IST)	.739	.546		.737	.543	
	<i>Corr</i> (SFT, MST)	.299	.089		.299	.089	
	<i>Corr</i> (SFT, GST)	.316	.099		.313	.097	
	<i>Corr</i> (IST, MST)	.285	.081		.284	.080	
	<i>Corr</i> (IST, GST)	.248	.061		.247	.061	
	<i>Corr</i> (MST, GST)	.537	.288		.537	.288	
<i>N</i> = 1259				<i>N</i> = 1259			
χ^2 (98) = 347.695, <i>p</i> < .001				χ^2 (98) = 351.586, <i>p</i> < .001			
RMSEA = .045				RMSEA = .045			
90% C.I. RMSEA = [.040, .050]				90% C.I. RMSEA = [.040, .050]			
CFI = .966				CFI = .966			
SRMR = .055				SRMR = .054			

*** *p* < .001

Notes : standardized factor loadings. Indicators for SFT-24, IST-4, MST, and GST treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). VE = variance extracted (i.e., the R^2 or the Squared Multiple Correlation); AVE = average variance extracted (the average of the VEs for each latent construct). In model 1, SFT-24 estimated with row-mean scores; in model 2, SFT-24 estimated with latent factor scores. All correlations between latent constructs were statistically significant at the *p* < .05 level.

6. Validation Tests for SFT-24 and IST-4

Table S6. Structural Equation Models Assessing Known-Groups Validity of SFT-24 and IST-4

	Model 1	Model 3
	SFT-24	IST-4
<i>Latent factors</i>		
Dispositional optimism	0.023 (0.46)	0.014 (0.050)
Big-5 personality scales		
Extraversion	0.065 (0.040)	0.038 (0.041)
Conscientiousness	-0.251*** (0.066)	-0.228** (0.068)
Openness	0.078 (0.060)	-0.015 (0.063)
Agreeableness	0.175** (0.066)	0.176* (0.069)
Neuroticism	0.000 (0.050)	-0.052 (0.055)
<i>Observed covariates</i>		
Age	-0.014 (0.042)	-0.001 (0.045)
Female (ref.)		
Male	0.027 (0.035)	0.057 (0.038)
Non-Hispanic white (ref.)		
Black	0.062 (0.035)	0.052 (0.040)
Hispanic	-0.056 (0.033)	-0.021 (0.036)
Other	-0.006 (0.031)	0.001 (0.032)
No HSD (ref.)		
HSD	-0.007 (0.055)	-0.013 (0.070)
Some college or AA	0.064 (0.069)	0.012 (0.088)
BA +	0.076 (0.070)	0.034 (0.087)
Married (ref.)		
Widowed	-0.006 (0.031)	-0.026 (0.038)
Divorced	-0.028 (0.030)	-0.040 (0.032)
Separated	0.015 (0.025)	0.012 (0.034)
Never married	-0.011 (0.025)	0.000 (0.038)
Living with partner	0.022 (0.032)	0.005 (0.036)
Paid employee (ref.)		
Self-employed	-0.016 (0.033)	0.051 (0.036)
Not working	-0.022 (0.032)	-0.012 (0.039)
ln(per capita HH income)	0.045 (0.037)	-0.016 (0.043)
Northeast (ref.)		
Midwest	0.090* (0.043)	0.055 (0.043)
South	-0.008 (0.043)	-0.014 (0.044)
West	0.145* (0.044)	0.086 (0.045)
Metro area (ref.)		
Non-metro area	-0.024 (0.030)	-0.049 (0.032)
	<i>N</i> = 1264	<i>N</i> = 1264

* $p < .05$, ** $p < .01$, *** $p < .001$

Notes : Standardized coefficients (robust standard errors). Indicators for all latent factors treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). Indicators for SFT-24 estimated as latent factor scores. Exogenous latent factors and observed covariates were allowed to freely covary (not shown). Control variables included household internet access, home ownership, building type of residence, telephone service for the household, survey duration, and survey mode device.

Figure S5. Standardized Coefficients Between Generalized Trust Scales (SFT-24 and IST-4) and Measures of Economic Preferences

Notes: Bars around standardized coefficients reflect 95% CIs

Figure S6. Standardized Coefficients Between Generalized Trust Scales and Proportion of Points Sent by Player A

Notes: Green bars around standardized coefficients reflect 95% CIs

Figure S7. Standardized Coefficients Between Generalized Trust Scales and Proportion of Points Returned by Player B

Notes: Brown bars around standardized coefficients reflect 95% CIs

Figure S8. Scatter Plots of State-Level Indicators and Generalized Trust Scales (SFT-24 and IST-4)

Notes: Solid red lines are best linear unbiased predictions. Dashed red lines reflect 95% CIs. Solid blue markers represent each state.

7. Measurement Invariance of SFT-60 and IST-6

Measurement invariance (MI) across gender (male, female), race (non-Hispanic White, Black, Hispanic, Other Race), education (no HSD, HSD, some college or AA, BA +), and age-groups (18-34, 35-49, 50-64, 65+) was examined by testing and comparing three nested models with multiple group analysis (Liu et al. 2017; Putnick and Bornstein 2016; Vandenberg and Lance 2000). In multiple group analysis, each successive model includes the previous model restrictions plus additional constraints. Model 1, the configural invariance model, tests the equivalence of the factor structure. Model 2, the metric invariance model, tests the equivalence of the factor loadings. Model 3, the scalar invariance model, tests the equivalence of measurement intercepts. The nested models are tested with χ^2 difference tests, but since the test is sensitive to sample size I rely on changes to the CFI, RMSEA, and SRMR across models. Following Cheung and Rensvold (2002) and Chen (2007), I use criteria of a .01 change in CFI paired with changes in RMSEA and SRMR of .015. If nested model comparisons reject the χ^2 difference test, yield $\Delta\text{CFI} > .01$, $\Delta\text{RMSEA} > .015$, and $\Delta\text{SRMR} > .015$, then the null hypothesis of invariance should be rejected.

The analysis of MI across gender, race, education, and age-groups for SFT-60 and IST-6 are reported in Table S7 and Table S8, respectively. In each table, I report nested model comparisons between models that estimated configural invariance, metric invariance, and scalar invariance by group (i.e., gender, race, education, and age-group). To identify the configural invariance model, I follow Vandenberg (2002) and Vandenberg and Lance (2000) and constrained the factor means and variances to 0 and 1, respectively, across groups. Factor loadings as well as item intercepts and residuals variances were freely estimated across groups. To identify the metric invariance model, I constrained the factor loadings to equality across groups, constrained the factor variance to 1 for the first group but freely estimated the factor variances for all other groups, constrained the factor means to 0 across groups, and freely estimated item intercepts and residual variances across groups. To identify the scalar invariance model, I constrained the factor loadings and item intercepts to equality across groups, constrained the factor variance and mean to 1 and 0, respectively, for the first group but freely estimated the factor variances and means for all other groups, and freely estimated item residual variances across groups.

Starting with SFT-60 in Table S7, some of the χ^2 difference tests were statistically significant but none of the changes to SRMR, RMSEA, and CFI indicated that the null hypothesis of MI should be rejected. In other words, SFT-60's factor loadings and item intercepts were invariant by gender, race, education, and age-group. Results for IST-6 are reported in Table S8. Like SFT-60, some of the χ^2 difference tests were statistically significant but, again, none of the changes to SRMR, RMSEA, and CFI indicated that the null hypothesis of MI should be rejected.

In short, SFT-60 and IST-6 exhibit MI by gender, race, education, and age-groups.

Table S7. Tests of Measurement Invariance of SFT-60 by Gender, Race, Education, and Age-Groups

	<i>df</i>	χ^2	SRMR	RMSEA	CFI	Δdf	$\Delta\chi^2$	Δ SRMR	Δ RMSEA	Δ CFI
<i>Gender (N = 1195)</i>										
1. Configural Invariance	70	144.932*	.014	.042	.981					
1 versus 2						9	6.162	.002	.000	-.002
2. Metric Invariance	79	160.566*	.016	.042	.979					
2 versus 3						9	23.414*	.001	.000	-.003
3. Scalar Invariance	88	80.671*	.017	.042	.976					
<i>Race (N = 1195)</i>										
1. Configural Invariance	140	276.908*	.018	.057	.976					
1 versus 2						27	34.999	.007	-.001	-.003
2. Metric Invariance	167	324.108*	.025	.056	.973					
2 versus 3						27	58.392	.004	.000	-.005
3. Scalar Invariance	194	378.771*	.029	.056	.968					
<i>Education (N = 1195)</i>										
1. Configural Invariance	140	302.290*	.019	.062	.970					
1 versus 2						27	33.437	.006	-.001	-.004
2. Metric Invariance	167	351.640*	.025	.061	.966					
2 versus 3						27	46.940*	.003	-.001	-.004
3. Scalar Invariance	194	404.821	.028	.060	.962					
<i>Age (N = 1195)</i>										
1. Configural Invariance	140	308.838*	.017	.064	.965					
1 versus 2						27	18.647	.004	-.003	-.004
2. Metric Invariance	167	355.133*	.021	.061	.961					
2 versus 3						27	35.773	.001	-.001	-.004
3. Scalar Invariance	194	404.631*	.022	.060	.957					

Note : SRMR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; CFI = comparative fit index; $\Delta\chi^2$ = Satorra-Bentler scaled χ^2 difference test; * $p < .05$

Table S8. Tests of Measurement Invariance of IST-6 by Gender, Race, Education, and Age-Groups

	<i>df</i>	χ^2	SRMR	RMSEA	CFI	Δdf	$\Delta\chi^2$	Δ SRMR	Δ RMSEA	Δ CFI
<i>Gender (N = 1227)</i>										
1. Configural Invariance	18	219.106*	.069	.135	.878					
1 versus 2						5	7.323	.003	-.013	-.005
2. Metric Invariance	23	232.179*	.072	.122	.873					
2 versus 3						5	6.294	.000	-.009	-.005
3. Scalar Invariance	28	248.175*	.072	.113	.867					
<i>Race (N = 1227)</i>										
1. Configural Invariance	36	252.190*	.071	.140	.867					
1 versus 2						15	6.987	.003	-.024	.003
2. Metric Invariance	51	262.300*	.074	.116	.870					
2 versus 3						15	64.058*	.008	-.002	-.032
3. Scalar Invariance	66	330.204*	.082	.114	.838					
<i>Education (N = 1227)</i>										
1. Configural Invariance	36	251.008*	.071	.140	.875					
1 versus 2						15	8.975	.003	-.022	-.001
2. Metric Invariance	51	268.335*	.074	.118	.874					
2 versus 3						15	45.835*	.007	-.005	-.023
3. Scalar Invariance	66	322.554*	.081	.113	.851					
<i>Age (N = 1227)</i>										
1. Configural Invariance	36	257.697*	.072	.142	.872					
1 versus 2						15	16.304	.006	-.021	-.004
2. Metric Invariance	51	279.859*	.078	.121	.868					
2 versus 3						15	59.243*	.008	-.004	-.028
3. Scalar Invariance	66	343.268*	.086	.117	.840					

Note : SRMR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; CFI = comparative fit index; $\Delta\chi^2$ = Satorra-Bentler scaled χ^2 difference test; * $p < .05$

8. Validation Tests of TRUST, FAIR, and HELP

Table S9. Structural Equation Linear Probability Models Assessing Known-Groups Validity of TRUST, FAIR, and HELP

	Model 1	Model 2	Model 3
	TRUST	FAIR	HELP
<i>Latent factors</i>			
Dispositional optimism	0.109* (0.044)	0.320*** (0.044)	0.260*** (0.045)
<i>Big-5 personality scales</i>			
Extraversion	0.029 (0.037)	0.009 (0.038)	0.069 (0.039)
Conscientiousness	-0.254*** (0.064)	-0.230*** (0.063)	-0.258*** (0.065)
Openness	0.066 (0.053)	0.026 (0.052)	0.054 (0.055)
Agreeableness	0.163** (0.058)	0.150* (0.060)	0.115 (0.061)
Neuroticism	-0.177*** (0.049)	0.027 (0.049)	-0.025 (0.049)
<i>Observed covariates</i>			
Age	0.011 (0.041)	0.153*** (0.042)	0.180*** (0.041)
Female (ref.)			
Male	0.040 (0.033)	0.064 (0.033)	-0.063 (0.033)
Non-Hispanic white (ref.)			
Black	-0.123*** (0.031)	-0.080* (0.034)	-0.055 (0.033)
Hispanic	-0.079** (0.030)	-0.073* (0.033)	-0.061 (0.032)
Other	-0.028 (0.030)	-0.012 (0.032)	-0.004 (0.030)
No HSD (ref.)			
HSD	-0.004 (0.060)	0.002 (0.061)	-0.053 (0.064)
Some college or AA	0.009 (0.075)	0.057 (0.077)	-0.016 (0.081)
BA +	0.107 (0.075)	0.130 (0.076)	0.023 (0.081)
Married (ref.)			
Widowed	-0.006 (0.030)	-0.026 (0.028)	-0.040 (0.028)
Divorced	-0.057 (0.030)	-0.042 (0.030)	-0.012 (0.029)
Separated	-0.007 (0.028)	-0.017 (0.028)	-0.051* (0.025)
Never married	-0.004 (0.034)	-0.004 (0.036)	-0.027 (0.036)
Living with partner	-0.021 (0.028)	-0.009 (0.031)	-0.015 (0.031)
Paid employee (ref.)			
Self-employed	0.012 (0.030)	-0.033 (0.028)	-0.043 (0.029)
Not working	-0.003 (0.033)	0.030 (0.034)	-0.007 (0.032)
ln(per capita HH income)	-0.007 (0.035)	0.051 (0.035)	0.034 (0.035)
Northeast (ref.)			
Midwest	0.022 (0.039)	-0.008 (0.037)	-0.044 (0.038)
South	-0.068 (0.041)	-0.062 (0.039)	-0.085* (0.040)
West	-0.019 (0.040)	-0.030 (0.038)	-0.125** (0.038)
Metro area (ref.)			
Non-metro area	0.038 (0.029)	0.074 (0.027)	0.052 (0.028)
	<i>N</i> = 1264	<i>N</i> = 1264	<i>N</i> = 1264

* $p < .05$, ** $p < .01$, *** $p < .001$

Notes : Standardized coefficients (robust standard errors). Indicators for all latent factors treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). Exogenous latent factors and observed covariates were allowed to freely covary (not shown). Control variables included household internet access, home ownership, building type of residence, telephone service for the household, survey duration, and survey mode device.

Figure S9. Standardized Coefficients Between MST Indicators (TRUST, FAIR, and HELP) and Measures of Economic Preferences

Notes: Bars around standardized coefficients reflect 95% CIs

Figure S10. Standardized Coefficients Between MST Indicators (TRUST, FAIR, and HELP) and Proportion of Points Sent by Player A

Notes: Green bars around standardized coefficients reflect 95% CIs

Figure S11. Standardized Coefficients Between MST Indicators (TRUST, FAIR, and HELP) and Proportion of Points Returned by Player B

Notes: Brown bars around standardized coefficients reflect 95% CIs

Figure S12. Scatter Plots of State-Level Indicators and MST Indicators (TRUST, FAIR, and HELP)

Notes: Solid red lines are best linear unbiased predictions. Dashed red lines reflect 95% CIs. Solid blue markers represent each state.

9. CFA of Latent Constructs

Table S10. Confirmatory Factor Analysis of Particularized Social Trust (PST)

Latent Factor	Indicator	Loading	VE	AVE
PST				.425
	FAMILY	0.668***	.446	
	NEIGH	0.540***	.291	
	KNOW	0.735***	.540	
<hr/>				
$N = 1253$				
$\chi^2 (0) = 0.000$				
RMSEA = .000				
90% C.I. RMSEA = [.000, .000]				
CFI = 1.000				

*** $p < .001$

Notes : standardized factor loadings. Indicators for PST treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). VE = variance extracted (i.e., the R^2 or the Squared Multiple Correlation); AVE = average variance extracted (the average of the VEs for each latent construct).

Table S11. Confirmatory Factor Analysis of Political Trust (POT)

Latent Factor	Indicator	Loading	VE	AVE
POT				.384
	POLICE	0.574***	.329	
	COURTS	0.760***	.577	
	GOVERN	0.656***	.430	
	PARTIES	0.450***	.202	
<hr/>				
$N = 1243$				
$\chi^2 (2) = 69.910, p < .001$				
RMSEA = .165				
90% C.I. RMSEA = [.133, .200]				
<hr/>				
CFI = .924				

*** $p < .001$

Notes : standardized factor loadings. Indicators for POT treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). VE = variance extracted (i.e., the R^2 or the Squared Multiple Correlation); AVE = average variance extracted (the average of the VEs for each latent construct).

Table S12. Confirmatory Factor Analysis of Dispositional Optimism (OPT)

Latent Factor	Indicator	Loading	VE	AVE
OPT				.469
	UNCERTAIN	0.477***	.227	
	WRONG	0.736***	.541	
	OPTIMISTIC	0.523***	.273	
	MYWAY	0.811***	.657	
	GOOD	0.774***	.599	
<hr/>				
$N = 1243$				
$\chi^2 (5) = 303.986, p < .001$				
RMSEA = .218				
90% C.I. RMSEA = [.198, .240]				
CFI = .863				

*** $p < .001$

Notes : standardized factor loadings. Indicators for OPT treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). VE = variance extracted (i.e., the R^2 or the Squared Multiple Correlation); AVE = average variance extracted (the average of the VEs for each latent construct).

Table S13. Confirmatory Factor Analysis of Big-5 Personality Scales

Latent Factor	Indicator	Big-5 Personality		
		Loading	VE	AVE
EXTRA				.477
	TALK	0.670***	.448	
	OUTGOING	0.919***	.844	
	RESERVED	0.375***	.140	
CONSC				.340
	THOROUGH	0.722***	.521	
	LAZY	0.359***	.128	
	EFFICIENT	0.611***	.373	
OPEN				.409
	ORIGINAL	0.741***	.549	
	ARTISTIC	0.568***	.322	
	IMAGINE	0.597***	.356	
AGREE				.350
	FORGIVE	0.567***	.321	
	RUDE	0.366***	.133	
	CONSIDERATE	0.773***	.597	
NEURO				.469
	WORRY	0.669***	.447	
	NERVOUS	0.914***	.835	
	CALM	0.355***	.126	
		<i>r</i>	<i>r</i> ²	
	<i>Corr</i> (EXT, CON)	.314	.098	
	<i>Corr</i> (EXT, OPE)	.373	.139	
	<i>Corr</i> (EXT, AGR)	.376	.141	
	<i>Corr</i> (EXT, NEU)	-.345	.119	
	<i>Corr</i> (CON, OPE)	.549	.301	
	<i>Corr</i> (CON, AGR)	.617	.380	
	<i>Corr</i> (CON, NEU)	-.172	.029	
	<i>Corr</i> (OPE, AGR)	.527	.277	
	<i>Corr</i> (OPE, NEU)	-.184	.033	
	<i>Corr</i> (AGE, NEU)	-.090	.008	
<i>N</i> = 1256				
χ^2 (80) = 772.770, <i>p</i> < .001				
RMSEA = .083				
90% C.I. RMSEA = [.078, .088]				
CFI = .836				

*** *p* < .001

Notes : standardized factor loadings. Indicators for all latent factors treated as continuous (factor variances, item intercepts, and residual variances freely estimated but not shown). VE = variance extracted (i.e., the R^2 or the Squared Multiple Correlation); AVE = average variance extracted (the average of the VEs for each latent construct). All correlations between latent constructs were statistically significant

10. References

- Chen, Fang Fang. 2007. "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling* 14: 464-504.
- Cheung, Gordon W. and Roger B. Rensvold. 2002. "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance." *Structural Equation Modeling* 9: 233-255.
- Liu, Yu, Roger E. Millsap, Stephen G. West, Jenn-Yun Tein, Rika Tanaka, and Kevin J. Grimm. 2017. "Testing Measurement Invariance in Longitudinal Data with Ordered-Categorical Measures." *Psychological Methods* 22: 486-506.
- Putnick, Diane L. and Marc H. Bornstein. 2016. "Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research." *Developmental Review* 41: 71-90.
- Vandenberg, Robert J. 2002. "Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures." *Organizational Research Methods* 5: 139-158.
- Vandenberg, Robert J. and Charles Lance. 2000. "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research." *Organizational Research Methods* 3: 4-70.