From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2012

Building, Betraying, and Buffering Trust in Interracial and Same-Race Friendships

Bergsieker, Hilary Burbank

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01mg74qm137

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

BUILDING, BETRAYING, AND BUFFERING TRUST IN INTERRACIAL AND SAME-RACE FRIENDSHIPS

HILARY BURBANK BERGSIEKER

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF

PSYCHOLOGY

ADVISOR: J. NICOLE SHELTON

JUNE 2012

© Copyright by Hilary Burbank Bergsieker, 2012. All rights reserved.

TABLE OF CONTENTS

Acknowledgements v

Abstract vi

Chapter 1: Introduction 1

Real-World Relevance 2

Intergroup and Interracial Trust 4

Challenges of Investigating Trust 12

Research Overview 16

Chapter 2: Defining Trust 19

Conceptual Approaches 19

Discriminant Validity 20

Pilot Studies: Scale Development 23

Chapter 3: Dynamics of Trust in Close Relationships 28

Trust Development 28

Study 1: Trust and Betrayal Between Friends 30

Chapter 4: Betrayal and Trust Repair 53

Individual Differences 53

Group-based Asymmetries 55

Study 2a: Virtual Interracial Trust Betrayals 57

Study 2b: Virtual Interracial Trust Betrayals 64

Chapter 5: Buffering Trust with Induced Closeness 69

Closeness and Forgiveness 69

Closeness in Intergroup Contexts 70

Study 3: Trust Betrayals in a Prisoner's Dilemma Game 71

Chapter 6: General Discussion 87

Strengths and Limitations 89

Implications and Future Directions 91

References 95

Appendices 106

ACKNOWLEDGEMENTS

I am profoundly grateful for the encouragement, support, and inspiration that my advisor Nicole Shelton has graciously provided throughout my time at Princeton. Nicole's extraordinary skill, insight, patience, and kindness have redefined scholarship and mentorship for me. I am deeply indebted to Susan Fiske and Andy Conway for invaluable theoretical and statistical training, as well as their contributions to my dissertation work. I also thank my other committee members—Stacey Sinclair and Matt Botvinick—and the members of the SSP and Fiske labs for their feedback. I am grateful for the vibrant intellectual community of the Psychology Department and especially the friendship of my colleagues Jan Alegre, Crystal Hall, Deborah Holoien, Drew Jacoby-Senghor, Lu Lu Kuang, Tan Lee, Ann Marie Russell, Avani Sood, Mary Steffel, Valerie Taylor, and Matt Trujillo.

Financial support for this research came from fellowships awarded by NSF and Princeton's Joint Degree Program in Social Policy, as well as SPSSI and APAGS dissertation grants. I greatly appreciated the opportunity to share these findings at the 2011 IZA Workshop on Discrimination and Ethnicity and 2012 SPSP Meeting. This dissertation also benefited greatly from Hila Calev's research assistance, RoseMarie Stevenson's coordination of the departmental participant pool, and Tamara Thatcher's patience with my many questions about dissertation procedures.

Words cannot express the extent of my gratitude to my husband Steven, who has walked beside me (and at times carried me) for every step of my graduate school journey, offering boundless encouragement, support, and love. I also thank my parents—Doug and Rachel Burbank, Rick and Chris Bergsieker—for enabling me to commence and complete this endeavor, as well as our son Phil for the joy, wonder, and affection that he has added to our family in the latter part of this adventure.

Above all, I thank and praise God for giving me the grace I needed to complete my Ph.D. and blessing me in this season with the fellowship of a loving Christian community that has sustained and enlightened me from my first weeks in graduate school through today.

ABSTRACT

Trust is theorized to be essential for close relationship functioning, and close friendships hold special promise for improving intergroup relations. This dissertation examines how trust (not mere liking) is attained, maintained, and regained between Black and White individuals as a function of closeness (inclusion of other in the self) in friendship contexts. Trust placed in outgroup (vs. same-race) friends is hypothesized to be shallower, asymmetric, and more vulnerable to disruption. In Study 1, a dyadic investigation over time, White and Black students who described interracial (vs. same-race) friendships initially reported less trust, closeness, stimulating companionship, reliable alliance, self-disclosure, perceived other disclosure, perceived help/support, and perceived understanding, followed by less perceived other disclosure and less contact a year later. Participants' initial trust and closeness (not liking) mediated these effects. Initial trust predicted fewer subsequent friend betrayals and greater closeness to an outgroup friend's racial group one year later. In Studies 2a and 2b, White and Black participants imagined a White or Black friend engaging in relational behaviors over time, sometimes including a betrayal. Black participants reported less cultural trust (not liking) after imagining a betrayal by a White (vs. Black) friend. No cultural trust gap emerged for White participants, or for Black participants who were high in subjective closeness or not betrayed. In Study 3, White and Black strangers completed either control or friendship-inducing closeness tasks in dyads before a prisoner's dilemma game that sometimes simulated a partner defection. After partner defections, Black participants reported lower trust (not liking) for White than Black partners, an effect partially mediated by negative other-directed affect, in the control condition. Black participants in the closeness condition and no-betrayal condition trusted White and Black partners equally, as did White participants. These studies (a) suggest that Black individuals trust ingroup more than outgroup individuals, particularly after betrayals, except when subjective closeness is high, and (b) underscore the centrality of trust (not liking) and closeness for successful interracial friendships.

CHAPTER 1: INTRODUCTION

This examination of trust-related processes in same-race and interracial friendships integrates theory and methodology from intergroup relations and close relationships. This work builds on compelling theoretical and empirical claims advocating friendship as a particularly potent form of intergroup contact (Pettigrew, 1997, 1998), inclusion of minority perspectives (Shelton, 2000), and investigation of interaction dynamics using a relational, dyadic approach (Shelton & Richeson, 2006b). Indeed, the past decade has seen a dramatic increase in research assessing intergroup friendships or interpersonal dynamics of dyadic intergroup interactions. Within this body of work, however, measures of intergroup friendship vary broadly, often focusing on the absolute number or proportion of intergroup friends rather than specific interpersonal properties of these relationships (see Davies, Tropp, Aron, Pettigrew & Wright, 2011), and few studies incorporate reliable measures of trust per se (see Dovidio et al., 2008; Hewstone et al., 2008; Kramer, 2004; Tropp, 2008).

By contrast, although research on close relationships has historically neglected trust (see J. Simpson, 2007b), scholars of close relationships have come to increasingly assert its importance. Trust has been deemed a "cardinal construct" that transforms relationship perceptions (J. Simpson, 2007a, p. 604), "one of the most desired qualities in any close relationship" (Rempel, Holmes, & Zanna, 1985, p. 95), and "the single most important ingredient for the development and maintenance of happy, well-functioning relationships" (J. Simpson, 2007b, p. 264). As such, trust in close relationships has attracted a surge of attention in recent years, though most work in this area focuses on romantic relationships. Although pioneering analysis of friendship (e.g., Fehr, 1996) has spurred some work on trust and friendship (e.g., Crawford, 1977; Sapadin, 1988; Tesch & Martin, 1983), research on the interpersonal dynamics of intergroup friendships remains scarce (see Turner & Feddes, 2011), and almost no intergroup friendship studies assess trust (but see Davies, 2009).

Real-World Relevance

Interracial and intergroup trust merit investigation because they pivotally influence a broad array of noteworthy societal and individual outcomes, ranging from actions toward aggregate groups to dynamics of interpersonal encounters. Moreover, existing evidence suggests that trust can have spillover effects that extend beyond the original trusted target. Distrust of racial outgroups can undermine dyadic interracial interactions (e.g., Dovidio et al., 2008; Shelton & Richeson, 2006a), just as trusting interpersonal relationships with outgroup individuals can improve behavioral intentions toward entire groups (e.g., Nadler & Leviatan, 2006; Tam, Hewstone, Kenworthy, & Cairns, 2009).

Collective Outcomes

Many aspects of macro-level intergroup relations, ranging from support for group-targeted policies to waging war, hinge upon intergroup trust. Intergroup distrust constitutes a substantial barrier to reconciliation between ethnic, national, and religious groups currently or recently engaged in conflict (e.g., Muslims and Serbs in Croatia: Cehajic, Brown, & Castano, 2008; Arabs and Jews in Israel: Saguy, Tausch, Dovidio, & Pratto, 2009; Protestants and Catholics in Northern Ireland: Tam et al., 2009). Similarly, higher levels of intergroup distrust predict increased support for social policies that maintain group hierarchy, such as the death penalty (Soss, Langbein, & Metelko, 2003) and criminal sentencing (Sidanius, Levin, & Pratto, 1998). Even when individuals are not in direct contact with individual members of other groups, intergroup trust shapes relations between them.

Conversely, aggregation of individual encounters can influence group-level outcomes. Trust, especially interracial trust, is theorized to underly "collective efficacy" (Sampson, Raudenbush, & Earls, 1997) and "social capital" (Bourdieu, 1983), the networks of interpersonal associations and trust or reciprocity norms that enable collective action. Among 10 indices of social capital—social trust, interracial trust, associational involvement, informal socializing, electoral politics, protest politics, civic leadership, faith-based engagement, diversity of friendships, and charitable

behaviors—assessed in the Social Capital Community Benchmark Survey, trust and associational networks emerged as the two underlying factors that accounted for the most variance (Brown & Ferris, 2007). Social capital and collective efficacy are associated with numerous collective benefits (e.g., improved economic performance, Putnam, 1993; reduced violent crime, Sampson et al., 1997), highlighting the positive potential downstream effects of increased interracial trust for society.

Individual Outcomes

In addition to producing collective gains, increased trust is theorized to benefit individuals by facilitating more efficient interpersonal exchanges (see Gulati & Sytch, 2008) and improving close interpersonal relationships (see J. Simpson, 2007a). Organizational research links trust to lower negotiation costs, reduced conflict, improved information sharing, as well as more cooperation and better organizational performance (e.g., Dyer & Chu, 2003; Uzzi, 1997; Zaheer, McEvily, & Perrone, 1998). Close relationship theories posit that trust can increase emotional security (J. Simpson, 2007b) and facilitate pursuit of connection rather than self-protection goals (see Murray & Holmes, 2009).

In interracial contexts, distrust has indeed been linked to lower utility in economic exchanges and more aversive interpersonal encounters. For instance, poor Black individuals often opt for a transaction with a known partner over a more lucrative transaction with a stranger (Hall, 2008). In economic games that offer the possibility of increasing earnings by choosing to trust another person with an "investment," individuals may pass up this opportunity with outgroup (vs. ingroup) partners (Eckel & Wilson, 2003; B. Simpson, McGrimmon, & Irwin, 2007). Especially for minorities, interracial distrust is associated with lower-quality interactions between doctors and patients (e.g., Gordon, Street, Sharf, Kelly, & Souchek, 2006), police and citizens (Tyler & Huo, 2002), and professors and students (Mendoza-Denton, Downey, Purdie, Davis, & Pietrzak, 2002). Increasing interracial trust thus holds potential for enabling more efficient and enjoyable intergroup exchanges.

Intergroup and Interracial Trust

Psychological research and public opinion polling data both document substantial variation in the extent to which members of different groups—especially racial and ethnic groups—trust people generally, other groups collectively, and individual outgroup members (Smith, 2010). Such trust differentials are particularly pronounced for members of lower-status groups (Tropp, 2008). Several existing theoretical perspectives and empirical findings provide insight into the antecedents, implications, and potential resolution of group-based differences in trust. The following discussion focuses primarily on trust between different racial and ethnic groups, especially Whites and Blacks.

Theoretical Accounts

The general tendency to trust ingroup members and distrust outgroup members (Insko & Schopler, 1998) may arise from many sources. According to Social Identity Theory (Tajfel, 1970), merely perceiving and identifying with groups can give rise to processes that undermine trust. People generally demonstrate and expect more resource sharing from ingroup (vs. outgroup) members (de Dreu, 2010; Tajfel, 1970). People tend to attribute more hostile intent to groups than individuals (Abelson, Dasgupta, Park, & Banaji, 1998; Frey & Tropp, 2006), particularly in competitive contexts (Insko, Schopler, Hoyle, Dardis, & Graetz, 1990). Another identity-based model (Brewer, 2001) posits trade-offs (a) between benefits of cooperative interdependence versus costs of extending trust indiscriminately and (b) between opposing needs for inclusion versus differentiation. Extending "depersonalized" trust only to members of a bounded ingroup allows individuals to reap the rewards of mutual cooperation and sustain ingroup norms against exploitation, while feeling included in their own group and distinct from outgroups.

In addition to group biases arising from mere identification with an ingroup, interracial dynamics are shaped by societal stratification, with some groups occupying positions of historical (and enduring) privilege and others occupying subordinate positions. Social dominance theory

(Pratto, Sidanius, Stallworth & Malle, 1994) asserts that this stratification is not accidental. Instead, hierarchy arises and persists due to a general orientation to perceive relentless, zero-sum competition between groups and to desire one's ingroup to be superior to and dominate over outgroups. Moreover, ingroup identification and social dominance orientation (SDO) interact to produce especially high levels of discrimination against minimal outgroups (Sidanius, Pratto, & Mitchell, 1994). Relative to Black Americans, White Americans report higher levels of SDO (Jost & Thompson, 2000) and tend to view racial bias in zero-sum terms that paint Whites increasingly as losing out to Blacks (Norton & Sommers, 2011), implying that Whites on average may be more likely to display negative, domineering behaviors toward Blacks than vice-versa.

Another prominent theory in intergroup relations, aversive racism theory (Dovidio & Gaertner, 2004), outlines another source of friction in interracial interactions that may lead Blacks to distrust Whites. According to this perspective, modern forms of intergroup bias tend to be subtle and rooted in deep-seated negative affective responses to outgroup members, rather than driven by conscious intentions or ideology. Even Whites who consider themselves egalitarian may display negative non-conscious behaviors toward Blacks and treat them unfairly in ambiguous situations. Thus, perceiving individual Whites to be well intentioned may not inspire trust that they will indeed act in Blacks' best interests (see Dovidio, Gaertner, Kawakami, & Hodson, 2002).

Taken together, these theoretical perspectives imply that interracial interactions in general may constitute strain-test situations, particularly for members of lower-status, historically subjugated groups. Trying to differentiate among Whites who are (a) genuine egalitarians, (b) subtly prejudiced aversive racists, or (c) proponents of ingroup superiority or dominance presents an attributional dilemma for minorities. In contexts of attributional ambiguity, Blacks may distrust Whites' motives and discount their ostensibly positive behaviors (Crocker, Voelkl, Testa, & Major, 1991), much as

people low in chronic trust discount romantic partners' motives and underestimate their responsiveness (Holmes & Rempel, 1989; Campbell, Simpson, Boldry, & Rubin, 2010).

Generalized and Group-based Trust

Convergent studies by sociologists, political scientists, and economists reveal large race-based differences in generalized trust levels (e.g., Alesina & La Ferrara, 2002; Marschall & Stolle, 2004; Putnam, 2007; Uslaner, 2002), prompting the observation that "Race is the life experience that has the biggest impact on trust" (Uslaner, 2002, p. 91). The most commonly used measure of generalized trust, from the General Social Survey (GSS), a recurring representative poll of American adults, asks: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" In the 2006 GSS (Davis, Smith, & Mardsen, 2006), Blacks scored lowest among the 40 ethnic groups surveyed on generalized trust. Although other minority groups—Asians, Native Americans, and Latinos—also reported lower generalized trust levels relative to Whites, the Black-White gap was the most pronounced: Far more Black (80%) than White (53%) respondents indicated that they did not trust people in general. Controlling for social class differences accounts for the Latino-White trust gap (Kiecolt, Morris, & Toussaint, 2006), but the Black-White trust gap persists: The wealthiest Blacks report trust levels comparable to the poorest Whites (Uslaner, 2002). Analogous Black-White trust gaps have been observed among college students (e.g., Glaeser, Laibson, Scheinkman, & Soutter, 2000; B. Simpson et al., 2007). Prevalent explanations for the Black-White trust gap involve (a) actual and perceived discrimination against Blacks by individuals and institutions, (b) more neighborhood disorder and fewer structural resources in Black neighborhoods, and (c) parental socialization preparing children to cope with bias (see Smith, 2010).

Stark differences also emerge in group-based interracial trust, defined as trust placed collectively in members of racial ingroups versus outgroups. Among Black respondents, 7%

indicated that they trusted most White people, as opposed to some (81%) or no (11%) White people (Davis et al., 2006). Combining all racial groups surveyed, 75% of respondents indicated that the people they trusted were "almost all" or "mostly" member of their own racial group. Notably, Blacks do not report high levels of distrust for ingroup members: 70% of Blacks report that other Blacks can be trusted (Uslaner, 2002). In research that did not explicitly ask about issues of trust but instead solicited Whites' and Blacks' open-ended reports of their racial attitudes (Monteith & Spicer, 2000), 8-9% of Blacks in two samples spontaneously mentioned distrusting Whites, whereas fewer than 1% of Whites mentioned distrusting Blacks, suggesting that group-based interracial distrust may be particularly salient for Blacks.

Correlates and Consequences of Group-based Trust

Several studies have examined antecedents and outcomes of group-based trust, typically guided by Allport's (1954) contact hypothesis that positive intergroup contact improves outgroup attitudes and testing trust and other trust-related processes as possible mediators of this link. Notably, although a recent meta-analysis of contact studies revealed stronger effects of contact on attitudes among majority than minority individuals (Tropp & Pettigrew, 2005), none of the following studies tested majority/minority asymmetries in group-based trust. Measures of contact with outgroup members (e.g., contact quality, number of outgroup friends) are positively associated with outgroup trust between Protestants and Catholics in Northern Ireland (Paolini, Hewstone, Cairns, & Voci, 2004; Tam et al., 2007; Tam et al., 2009; Tausch, Tam, Hewstone, Kenworthy, & Cairns, 2007), Bosnian Muslims in contact with Serbians (Cehajic et al., 2008), Whites and South Asians in Britain (Turner, Hewstone, & Voci, 2007), and Arabs in contact with Jews in Israel (Saguy et al., 2009). Significant mediators of the contact-trust link include reduced intergroup anxiety (Tam et al., 2007; Tausch et al., 2007) and increased reciprocal self-disclosure (Turner et al., 2007). Trust significantly mediated the associations between reciprocal self-disclosure and more positive

intergroup attitudes (Turner et al., 2007), as well as the association between intergroup contact and increased intergroup forgiveness (Cehajic et al., 2008; Tam et al., 2007), and positive (vs. negative) behavioral tendencies toward the outgroup (Tam et al., 2009).

Collectively, these contact studies highlight not only the link between intergroup contact and trust, but also several interpersonal processes potentially undergirding this association. The positive effects of reduced intergroup anxiety and increased self-disclosure on outgroup trust parallel theorizing about the effects of uncertainty reduction and reciprocal self-disclosure on building trust in close relationships (Holmes & Rempel, 1989). The finding that greater group-based trust predicted increased positive and decreased negative behavioral tendencies more so than general evaluation (or liking) of the outgroup (Tam et al., 2009) supports the claim of the present research that trust is distinct from liking and consequential in intergroup contexts.

Complementing these correlational findings are experimental studies of the associations between group-based trust, individual outgroup members' behaviors (e.g., self-disclosure), and intergroup interaction dynamics. One such study tested the impact of self-disclosure on outgroup trust by having an Islamic fundamentalist (or conservative) confederate make a personalized self-disclosure to secular Turkish (or liberal) participants (Ensari & Miller, 2002). Perceptions of outgroup members as trusting increased if the confederate appeared to be a typical (vs. atypical) outgroup member or if her group membership was salient (vs. not salient). Another study tested whether initial group-based trust moderated the influence of an outgroup member's positive behavior on subsequent behavioral intentions toward the outgroup (Nadler & Liviatan, 2006). Jews higher (vs. lower) in trust for Palestinians responded more favorably to a Palestinian target's empathetic (vs. non-empathic) apology, expressing more willingness to reconcile and reciprocate empathy as well as more positive assessments of future Israeli-Palestinian relations.

In addition to influencing broader outgroup attitudes, group-based distrust negatively influences the interpersonal and intrapersonal dynamics of intergroup interactions. Lower levels of interracial trust predicted a pattern of cardiovascular reactivity consistent with physiological threat, heightened stress appraisals, and higher ratings of partner bias among Latina participants who received positive feedback from a White (vs. Latina) confederate (Sawyer et al., 2008). Minorities' distrust of Whites predicted decreases in contact, positive perceptions of friendship, and comfort self-disclosing with a White (vs. Black) friend; as well as decreased contact with, closeness to, and positive affect over time toward a White (vs. minority) roommate (Shelton & Richeson, 2006a). Among Black (but not White) participants, greater interracial distrust predicted decreased liking for a White interaction partner, increased perceptions of the partner as prejudiced, and decreased interaction enjoyment (Shelton & Richeson, 2006a). These findings have implications for trust's trajectory in intergroup and interracial relationships: Minorities who trusted Whites showed less physiological threat and greater feelings of closeness and comfort self-disclosing, both of which should further reinforce trust over time. These studies show that interracial trust shapes interpersonal interactions and can have asymmetric outcomes for Whites and minorities.

Trust in Specific Outgroup Others

The pervasiveness of group-based intergroup and interracial distrust need not imply that individual members of other racial groups are inevitably viewed with suspicion. Relative to groups, individuals are typically perceived to be less intrinsically greedy or competitive and more trusting, eliciting less fear and more cooperation from others (Insko & Schopler, 1998; Insko et al., 1990). Many studies have used experimental mixed-motive games (e.g., the "trust game"; Berg, Dickhaut, & McCabe, 1995) to assess whether individuals display less trusting behavior toward outgroup than ingroup strangers. Behavioral trust findings for strangers from different racial groups remain inconclusive (see Burns, 2006). In some studies, people gave more money to ingroup (Eckel &

Wilson, 2003; B. Simpson et al., 2007) or White (Burns, 2006) partners, whereas no discrimination occurred in other studies (e.g., Glaeser et al., 2000). Similarly, at times ingroup (Glaeser et al., 2000) or Black (B. Simpson et al., 2007) partners returned more money than other partners, but in other instances partners did not discriminate (Eckel & Wilson, 2003). When differences have emerged, they reflect ingroup favoritism, parallel to work in minimal groups (e.g., Tajfel, 1970).

Returning to more cognitive definitions of trust—the focus of the present research—several studies illustrate processes by which the mixed or ambiguously neutral behavior of an outgroup individual in an intergroup encounter or relationship can undermine trust if that behavior is construed negatively. After completing a task with an aversive racist (vs. nonprejudiced or prejudiced) White participant, Black participants reported less trust in their White partner (Dovidio, 2001). Although aversive racist Whites thought they had displayed friendly behavior, coders assessed it as less nonverbally friendly than that of other Whites, which may have signaled lack of trustworthiness. Similarly, Black (but not White) patients reported decreased trust in non-Black doctors after an initial appointment, a drop fully mediated by perceptions of receiving less support, partnership, and information from doctors during the clinical interaction (Gordon et al., 2006). Such decreases in trust for an outgroup individual after perceiving that person to be unfriendly or unresponsive parallel the processes shaping trust in close relationships (Holmes & Rempel, 1989; Reis, 2007; Wieselquist, Rusbult, Forster, & Agnew, 1999).

Although a growing body of research examines the consequences of interpersonal trust in specific close relationship partners, virtually no work investigates the effects of trusting a specific outgroup individual on other interpersonal aspects of that intergroup relationship or for intergroup relations more broadly. (Many of the aforementioned intergroup contact studies have assessed aspects of individuals' friendships with outgroup members, but they have not focused on specific relationships or trust.) A noteworthy—perhaps solitary—exception is dissertation work by Davies

(2009) with retrospective and longitudinal studies contrasting same-race and interracial friendship formation. In the retrospective study, participants reported recollections of the early stages of their friendship with a same- or different-race person, current perceptions of the relationship, and outgroup attitudes. In the longitudinal study, participants reported outgroup attitudes and perceptions of a recently commenced same-race or interracial friendship early in the semester then again 5 and 10 weeks later. Though limited by some methodological concerns (partially random assignment to condition, single-item measures, an *n* of 16 same-race friendships over time), this work observes several interesting effects, some of which parallel the current research. For participants retrospectively describing interracial (vs. same-race) friendships, mean levels of initial trust, current closeness, and current inclusion of other in the self (IOS; Aron, Aron, & Smollan, 1992) were lower, and initial trust was more strongly associated with initial similarity and current time spent together. Participants describing interracial (vs. same-race) friendships over time reported less intimate self-disclosure at Times 1 and 3; moreover, frequency of self-disclosure at Time 1 predicted perceptions of the friendship's durability at Time 3. Participants in interracial friendships who perceived their friend to be more trustworthy at Time 3 admired their friend's racial group more. The current research extends Davies' work by resolving the aforementioned methodological limitations, collecting relationship perceptions from both individuals in each friendship dyad, tracking relationship processes for a longer time, assessing betrayal behaviors, comparing the effects of predictors simultaneously to identify the independent impact of trust versus liking on subsequent relationship dynamics, and experimentally manipulating betrayals as well as closeness or IOS.

Closeness and Inclusion of Other in the Self

Subjective interpersonal closeness and IOS (Aron et al., 1992) are convergent constructs that both assess a sense of shared "oneness" between people. IOS represents the extent of overlap in the mental representation of self and other, potentially incorporating the other person's resources,

perspectives, and characteristics in an expanded self-concept (Aron & Aron, 1986; Aron et al., 1992). The IOS scale, a single-item pictorial measure using progressively more overlapping circles labeled "Self" and "Other," is interpreted by participants as illustrating interpersonal union, closeness, or connection (Aron et al., 1992).

Such self-other convergence parallels a broadly recurring theme in the literature on trust in both interpersonal and intergroup contexts (Cook, Hardin, & Levi, 2005; McLaughlin-Volpe, 2004), whether in interactions between strangers or close relationship partners (Aron, Melinat, Aron, Vallone, & Bator, 1997; Clark, Graham, Williams, & Lemay, 2008). Several models of trust emphasize the merging (or "encapsulation") of self- and other-interest or the integration of own and partner goals, needs, and desires as integral to trust building (Hardin, 2002; Holmes & Rempel, 1989; Rempel et al., 1985; J. Simpson, 2007a, 2007b; Wieselquist et al., 1999). In the present research, closeness or IOS is expected to positively predict trust across all studies.

Challenges of Investigating Trust

Numerous studies in behavioral economics and organizational science have examined trust between strangers in experimental game settings (see Cook & Cooper, 2003) and sociologists have a longstanding interest in generalized trust as a central component in social capital (see Putnam, 2000). In contrast, psychologists have not yet given trust extensive theoretical and empirical attention in interpersonal and especially intergroup contexts (see Dovidio et al., 2008; Hewstone et al., 2008; Kramer, 2004; Tropp, 2008). One researcher recently observed that "relatively little is known about how and why interpersonal trust develops, is maintained, and unravels when betrayed" (J. Simpson, 2007b, p. 264), because trust is (a) complex and multidimensional, hence difficult to operationalize, (b) variable across stages of relationship development, and (c) active in "strain tests" situations that are tricky to observe and study. Each of these challenges informed the present work.

Complexity

As a multidimensional construct (Kramer & Carnevale, 2001), trust is defined and measured differently both across and within social science literatures. Organizational research alone contains over 125 distinct trust measures, of which only 10-20 percent have been replicated, with roughly 50 multi-dimensional measures that propose 38 different dimensions of trust (Lewicki, Tomlinson, & Gillespie, 2006; McEvily & Tortoriello, 2011). Moreover, researchers have construed diverse positive outcomes as "trust" or conceptually conflated trust with related but multiply determined phenomena (e.g., cooperation; Deutsch, 1958; see Ferrin, Bligh, & Kohles, 2007). For example, a recent review of trust and intergroup contact (Tropp, 2008) bases its discussion of trust on empirical findings that link contact with anxiety, closeness, prejudice, liking, and positive outgroup emotions, but not trust per se. It is not uncommon for researchers to discuss prejudice findings in terms of intergroup trust (e.g., Czopp, 2007), but actual measures of trust remain relatively rare. For instance, work commonly cited as demonstrating a connection between trust and liking (i.e., Petty & Mirels, 1981) actually shows that reciprocal, selective self-disclosure and liking are associated. In contrast, the present research attempts to measure trust directly, rather than infer it from a related construct.

Timecourse

In contrast to "fast" trust between strangers in one-shot games (e.g., Glaeser et al., 2000) or split-second trustworthiness judgments (e.g., Oosterhof & Todorov, 2008), emergent or relational trust is theorized to develop between specific people gradually over time. Practically speaking, most research on trust has focused exclusively on either short-term interactions between strangers (often using the experimental game paradigm) or long-term close relationships (often examining dating or married couples), without synthesizing findings across domains. Some theorists argue that initial trust between strangers and established trust in relationships reflect qualitatively distinct types of trust (e.g., transactional vs. relational trust, Lewicki et al., 2006), whereas others conceptualize trust

broadly enough (e.g., as encapsulated interest, Hardin, 2002) to encompass both domains. Researchers have developed an increasing number of experimental paradigms capable of filling in this gap. For instance, iterated economic games (e.g., Lount, Zhong, Sivanathan, & Murnighan, 2008) can model "extended situations" that unfold in close relationships (see Rusbult & Van Lange, 2003), and zero-acquaintance interactions with a short-term follow-up (e.g., one week later, Reis et al., 2010) have been used to validate theories of close relationship processes. The present research draws theoretical and empirical insights from prior work on trust as it shapes both short-term interactions between strangers and long-term close relationships. Each study reported here incorporates a temporal dimension, whether by assessing trust over time (Study 1), simulating a series of interpersonal interactions (Studies 2a & 2b), or leading participants to engage in joint tasks followed by an iterative prisoner's dilemma (Study 3).

Interdependence

A broad consensus exists among theorists that trust requires interdependence (see Kelley & Thibaut, 1978) and vulnerability (see Kramer & Carnevale, 2001). By definition, depending on another person implies vulnerability (Rusbult & Van Lange, 2003). Potential risks to the self are particularly pronounced in trust-diagnostic or "strain-test" situations (Holmes, 1981; Kelley, 1979) in which the interests of two or more interdependent people are noncorrespondent (i.e., not aligned), such that the best outcome for one person involves sacrifice or cost for the other. If both parties' interests completely coincide, self-interested actions are mutually beneficial but trust is irrelevant (Hardin, 2003; J. Simpson, 2007b), because a partner is "merely someone from whom [one] can expect beneficial actions" (Hardin, 2002, p. 25). Trust-diagnostic situations, in contrast, typically involve "moderately corresponding interests" (J. Simpson, 2007b, p. 265) that afford partners the option to act egoistically and lead actors to experience feelings of vulnerability arising from uncertainty about partners' actual motives, intentions, and actions (Kramer & Carnevale, 2001).

Numerous scenarios that occur routinely in close relationships (e.g., sharing chores) qualify as strain-test situations (for a typology, see Kelley, Holmes, Kerr, Reis, Rusbult, & Van Lange, 2003), but they are not so ubiquitous as to occur spontaneously in the lab or be observed naturalistically, rendering their empirical assessment difficult (J. Simpson, 2007b). Typically, researchers have created strain-test situations by manipulating functional relations between people (e.g., instantiating me-versus-we trade-offs such as the prisoner's dilemma game) or prompting discussions of conflicts or potential partner sacrifices (e.g., Rempel, Ross, & Holmes, 2001; Shallcross & Simpson, 2012). The present work incorporates the concept of strain-test situations in several respects. For one, the primary measure used to assess trust includes items addressing the other's perceived willingness to sacrifice for one's benefit. Participants also reported instances in which their friend disappointed them or failed to follow through (Study 1), imagined a scenario in which a friend broke a promise to them for self-serving reasons (Studies 2a & 2b), or completed an iterated prisoner's dilemma pitting their interests against those of a partner after a friendship induction (Study 3).

Relational Properties

In addition to the aforementioned challenges, a fourth factor has limited trust research in intergroup domains: psychologists' predominant theoretical focus on prejudice and stereotyping, primarily among White or dominant group members, from the field's inception (e.g., Allport, 1954; Katz & Braly, 1933) until very recently. This failure to use a relational approach by incorporating perspectives of both minority and majority members who are involved in interpersonal interactions or relationships with one another (see Shelton & Richeson, 2006b) has hindered investigation of emergent relational variables such as trust (see Tropp, 2008). Now, however, researchers increasingly appreciate the value of examining trust in intergroup contexts. For example, scholars have asserted that as trust is "different from overall attitude and can be seen as a more demanding criterion of interpersonal or intergroup relations than liking because it potentially puts the self or the in-group

directly at risk" (Hewstone et al., 2008, p. 211), and is essential for coordinated action to address intergroup inequality (Dovidio, 2008). The current research assesses trust among participants from diverse backgrounds (Studies 1-3), most often within a dyadic framework (Studies 1 & 3).

Research Overview

The rationale for the current work rests on two premises. If establishing, maintaining, and repairing interpersonal trust is indeed essential for successful close relationship functioning (Holmes & Rempel, 1989; J. Simpson, 2007a, 2007b) and if interracial close relationships—especially friendships—provide a particularly powerful means of improving intergroup relations (Davies et al., 2011; Pettigrew, 1997, 1998), understanding the antecedents and consequences of trust in interracial friendship contexts is crucial. The aims of the present research include: (a) clearly distinguishing trust-related phenomena from those tied to general liking or evaluation in interracial (and same-race) friendship contexts, (b) examining the role of subjective interpersonal closeness or IOS as an antecedent and moderating variable for interracial trust, (c) investigating whether established asymmetries in group-level interracial trust translate to the level of one-on-one interpersonal interactions, and (d) probing the process of trust repair following betrayals in interracial and same-race friendship contexts.

Chapters 2 and 3 of this dissertation draw most heavily on prevalent theory and methods in research on close relationships. The pilot work develops reliable measures of trust and liking that display both construct and divergent validity for specific interpersonal relationships. Study 1 then applies these measures in a dyadic study of established interracial and same-race friendships over time. This study examines whether trust, controlling for liking, can uniquely predict (a) differences between same-race and interracial dyads in relationship dynamics (e.g., levels of self-disclosure), (b) changes over time in friendship functions and interaction behaviors, (c) subsequent contact, as well trust-building and trust-betraying behaviors, and (d) inclusion of the friend's racial group in the self.

Chapters 4 and 5 transition from correlational to experimental approaches, manipulating rather than measuring interpersonal betrayals in Studies 2a, 2b, and 3, as well as subjective closeness or IOS in Study 3. Building on the association tested in Study 1 between initial trust and subsequent friend betrayal behaviors, Studies 2a, 2b, and 3 invert this relationship to test whether experiencing interpersonal betrayals decreases subsequent trust, consistent with virtually all theories of trust, and whether a race-based asymmetry emerges for trust following a betrayal. Paralleling a divergence evident in sociological work on race and generalized trust (see Smith, 2010), Black participants who experienced a betrayal by an imagined or experimentally induced "friend" were expected to trust this person less when he or she was White as opposed to Black. In Study 3 this trust gap was predicted to emerge as a function of Black participants' increased negative other-directed affect (e.g., anger, resentment, hostility) following betrayal by a White (vs. Black) person. Across all Part II studies, subjective interpersonal closeness or IOS—whether measured or experimentally induced—was tested as a moderator of this race-based trust asymmetry. Black participants who reported including their fictive friend in their concept of self to a greater (vs. lesser) extent (in Studies 2a & 2b) or who completed IOS-inducing (vs. control) tasks prior to a betrayal were expected to extend comparable levels of trust to a White or Black individual after a betrayal.

This program of research seeks to contribute to the intergroup relations and close relationships literatures with a multi-method investigation of the specific interpersonal processes surrounding trust, as opposed to other types of interpersonal cognition or affect, in interracial and same-race friendship contexts. Some noteworthy features of this research that collectively distinguish it from prior work include: examining trust through a dyadic lens (Studies 1 & 3), consideration of—even emphasis on—minority perspectives, a focus on friendship as opposed to other interpersonal relationships, differentiation of trust from other related or potentially confounded positive interpersonal cognitions, assessment over a longer-than-typical time period

(Study 1), and experimental manipulation of both closeness (Study 3) and interpersonal betrayals (Studies 2a, 2b, & 3). The broader goal of this work is to explain, predict, and ultimately shape the processes by which trust can persist in interracial friendships and other close relationships against the backdrop of day-to-day negative interpersonal events or potential trust disruptions. Interracial—as opposed to same-race—relationships and interactions are especially vulnerable to disruption from a variety of sources, including concerns about prejudice (e.g., Shelton, Richeson, & Salvatore, 2005), pluralistic ignorance (Shelton & Richeson, 2005), negative nonverbal behaviors (e.g., Dovidio et al., 2002), and even brief conversational disfluencies (Pearson, West, Dovidio, Powers, Buck, & Henning, 2008). Given the wealth of relationship resources theorized to flow from interpersonal trust—increased forgiveness (Molden & Finkel, 2010), accommodation and collaboration (Shallcross & Simpson, 2012), commitment (Holmes & Rempel, 1989), partner responsiveness (Murray & Holmes, 2009), and felt security (J. Simpson, 2007b), for example—trust may be the "missing piece," relationally speaking, that enables individuals from different racial backgrounds to overcome a multitude of potentially disruptive factors and maintain thriving relationships that have spillover benefits for intergroup relations.

CHAPTER 2: DEFINING TRUST

Scholars have proposed highly varied definitions of trust, describing it variously as a type of cognition versus affect, and emphasizing its dispositional versus relational dimensions, so specifying the working conceptualization of trust is an important component of any investigation of trust. Behavioral measures of cooperation or giving in interdependent situations that involve conflicting interests (e.g., the prisoner's dilemma game) can be considered a reliable behavioral indicator of trust (see Glaeser et al., 2000); however, such behavior does not constitute trust per se (see Ferrin et al., 2007) or render phenomenological definitions of trust based in mental states superfluous.

Conceptual Approaches

Theorists disagree to some extent about whether trust is primarily a cognitive or affective construct. Most describe trust in cognitive terms, for instance as a belief, expectation, or conscious intention (Cook et al., 2005; Deutsch, 1973; Holmes & Rempel, 1989; Rousseau, Sitkin, Burt, & Camerer, 1998; J. Simpson, 2007a, 2007b; Yamagishi & Yamagishi, 1994), but a few categorize trust as an emotion (e.g., Tam et al., 2009), or distinguish between affect- and cognition-based trust (e.g., Li, 2008; McAllister, 1995). The present work conceptualizes trust as primarily cognitive, consistent with the bulk of prior theorizing on trust and the assertion that trusting hinges on beliefs about another person's behavior toward oneself.

Early trust work employed a predominantly dispositional conception of trust, whereas more recent work has adopted a relational definition. Dispositional approaches to trust emphasize individuals' tendencies to consider non-specific others generally reliable (e.g., Rotter, 1967, 1971, 1980). Dispositional or "generalized" trust is a "belief in the benevolence of human nature in general and thus is not limited to particular objects" (Yamagishi & Yamagishi, 1994, p. 139). In contrast to individual difference accounts of trust, relational or dyadic approaches examine trust as an emergent dependent variable in specific relationships between individuals. One such approach defines trust as

the specific belief held by individual A that individual B will perform action X in situation S (Cook et al., 2005). These approaches address why (a) people trust some individuals but not others, (b) some people but not others trust a given individual, and (c) people trust an individual in some situations but not others. The current work conceptualizes trust as a joint function of individuals' dispositional tendencies (which may be shaped by membership in a dominant vs. stigmatized group), their perceptions of their partners' character and responsiveness, and the context of the interaction.

In the present work trust is defined as "a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behaviors of another" (Rousseau et al., 1998; p. 395). Key elements in this definition of trust are (a) willingness to accept vulnerability or risk and (b) expectations of benign treatment by the other; notably, vulnerability depends on expectations. This definition, which increasingly predominates in organizational research (see McEvily & Tortoriello, 2011), emphasizes interpersonal trust in specific individuals.

Discriminant Validity

A central claim in the current work is that trust differs from other forms of interpersonal evaluation (e.g., liking) or general intergroup attitudes (e.g., prejudice). Relative to trust, liking (or prejudice) may constitute a more affective, hence primary (see Zajonc, 1980), response upon encountering others, functioning in many ways like interpersonal attraction. Liking a person (or a painting, for that matter) does not require inferences about the mental or volitional state of a liked object, whereas trust requires consideration of another person's likely intentions and behavior toward the self (see Rousseau et al., 1998). If trust, more so than liking, requires consideration of another person's perspective (see Hewstone et al., 2008), it should be a more inherently relational construct than liking. Many theorists posit that trust also depends on knowledge of the person's past behavior, reputation, and interests (see Hardin, 2003), suggesting that the informational and experiential threshold would thus be higher for establishing trust as opposed to liking.

Social Cognition

Empirical work on interactions among strangers suggests a moderate positive association between trusting and liking or admiring a person. In the person-perception tradition, trusting and liking are both thought to stem from perceptions of targets as "warm" (e.g., Fiske, Cuddy, & Glick, 2007). That said, other research indicates that morality (including trustworthiness) and sociability (including likeability) are distinct dimensions of social perception (Leach, Ellemers, & Barreto, 2007). Perceivers appear to distinguish needs to trust versus like others, reporting that they would prefer trustworthy traits over warmth traits in a romantic partner (Koleva, 2012) and value trustworthiness more than agreeableness in prospective relationship partners such as close friends, teammates, family members, romantic partners, and employees (Cottrell, Neuberg, & Li, 2007).

Moreover, impressions of sparse social stimuli (e.g., faces of individuals, brief vignettes, names of social groups) as trustworthy and likable may or may not converge initially, and levels of felt trust and liking can diverge as dyadic interactions or close relationships unfold. Even minimal encounters with targets can lead to divergent trust and liking: For instance, brief exposure to a video of a depleted (vs. non-depleted) target person led to decreases in trust, but not liking (Righetti & Finkenauer, 2011). In an elegant study that distinguished effects of interpersonal interactions on trust versus liking (Reis et al., 2010), participants who either self-disclosed to a responsive and engaged confederate or completed in a fun drawing task with an engaged confederate subsequently displayed increased liking and amusement relative to those in the control condition, but only participants in the responsive self-disclosure condition reported elevated trust and willingness to disclose on highly intimate topics.

Close Relationships

Trust and liking (or love) typically covary in close relationships, but are not redundant constructs and can diverge (e.g., loving an unfaithful spouse or trusting a disagreeable colleague). In

developing measures of trust, researchers have taken care to demonstrate divergent validity between trust and liking or love. For example, Johnson-George and Swap (1982) reported that items on their trust scale loaded on distinct factors from the Rubin (1973) liking and love scale items when participants rated a trusted specific person of their choice. Among dating and married couples, the Rubin (1973) love (but not liking) scale correlated positively with the Rempel et al. (1985) faith (r = .46, p < .001) and dependability (r = .25, p < .05) trust subscales. Studies of processes unfolding in close relationships have also demonstrated that trust can predict unique variance in key outcomes such as forgiveness following transgressions, controlling for commitment levels (Molden & Finkel, 2010), as well as satisfaction with friendship, parent-child, romantic, sibling, and roommate relationships, controlling for liking, loving, and closeness (Conley, Moors, Ziegler, & Feltner, 2011). Collectively, these findings imply that in close relationships trust and positive evaluations such as liking are associated yet distinct.

Intergroup Contexts

Studies of intergroup relations have used sparse and highly variable measures of trust. In the past, researchers have more often collapsed trust with liking (or other positive intergroup attitudes) than systematically differentiated these constructs. Some variously combine, differentiate, and omit trust (relative to other attitudinal measures) even within a single research article (e.g., Paolini et al., 2004; Turner et al., 2007). Nevertheless, emerging intergroup findings support the claim that trust is distinct from liking or intergroup attitudes more broadly. For example, in interracial interactions with an experimentally induced disruption (a 1-second communication delay between partners) trust and liking correlated positively for same-race dyads (r = .44, p < .002), but not interracial dyads (r = .03, *ns*; Pearson et al., 2008). Also, among pro-socially oriented individuals, the tendency to allocate more resources to the ingroup and fewer resources to the outgroup was associated with trust more so than with liking (de Dreu, 2010). Correlational data on intergroup contact also distinguish

between trust and general evaluation of outgroups, although they are positively associated: outgroup attitudes predicted outgroup trust for Arabs in Israel (Saguy et al., 2009) and outgroup trust predicted outgroup attitudes for a British sample (Turner et al., 2007). Compared with positive intergroup attitudes, intergroup trust more strongly predicted self-reported behavioral tendencies toward outgroups among Protestants and Catholics in Northern Ireland (Tam et al., 2009) and British students imagining contact with asylum seekers (Turner, West, & Christie, in press).

Methodologically, most intergroup trust measures assess trust toward outgroup members collectively rather than people in general or specific individuals. For instance, the cultural mistrust inventory assesses Blacks' distrust in Whites with items such as "Probably the biggest reason whites want to be friendly with Blacks is so they can take advantage of them" (Terrell & Terrell, 1981). Other measures rely on semantic differentials of attitudes toward outgroups (e.g., *suspicious—trusting*; Paolini et al., 2004; Shelton & Richeson, 2006a), or reference specific behaviors and issues, such as "You must be wary and not rely on Jews" (Saguy et al., 2009) or "I do not believe in the peaceful intentions of the Palestinians" (Nadler & Liviatan, 2006). The current research seeks to extend prior work by examining trust placed in specific outgroup members, rather than groups as a whole. Moreover, the present work contrasts trust and liking using reliable measures that share a common format, rather than assessing outgroup attitudes with semantic differential trait attributes and trust with statements about expected outgroup behavior (e.g., Tam et al., 2009), or using a two-item (r = .29, p = .09) outgroup trust measure (e.g., Turner et al., in press).

Pilot Studies: Scale Development

Two pilot studies tested sets of items designed to index liking and trust for a specific other person, intended to reliably differentiate between a liked and a trusted target. The current trust scale encompassed several theoretically significant components of interpersonal trust. Items came from the Johnson-George and Swap (1982) Specific Interpersonal Trust Scale for friends, which includes

general, emotional, and reliableness trust subscales, and the Rempel et al. (1985) trust scale for close relationships, which includes predictability, dependability, and faith subscales. Prior research on trust between recently acquainted individuals (e.g., Reis et al., 2010) has also integrated items from these two trust scales. Novel scale items attempted to tap "strain-test" trust (Holmes, 1981), theorized to be a particularly deep form of trust that extends beyond merely liking another person to encompass the expectation that that other person will make sacrifices on your behalf or act benevolently toward you when your interests conflict.

The "cultural" trust subscale aimed to capture concerns or uncertainties regarding self-disclosure and perceived other responsiveness in the specific context of sensitive intergroup topics. This scale included items such as "If we disagreed over a cultural difference, this person would respect my opinions and position" and intended to tap trust deficits that may be topically specific, namely, trusting a person in general, but not to understand one's feelings on racial issues.

The items assessing liking were designed to be face valid (e.g., "Generally speaking, I really like this person") and index the sense of being drawn to someone (e.g., "I sometimes try to avoid spending time with this person," reverse coded) and finding that person pleasant (e.g., "I enjoy being around this person"). These items overlap conceptually with several items (e.g., "I liked the other participant" and "I would like to interact with the other participant again") that have been used to assess liking, as a construct distinct from trust, in prior work (Reis et al., 2010). Table 1 reports the final trust, cultural trust, and liking items used in Studies 1-3, with sources noted for adapted items.

Method

Participants. Demographic information is not available for Pilot 1 participants (N = 39) due to experimenter error. Pilot 2 participants (N = 65) were students with a mean age of 20.3, including 16 men and 38 women, who self-identified as White (n = 37), Black (n = 9), Asian (n = 17), Latino (n = 1), or another racial group (n = 1).

Table 1: *Trust, Cultural Trust, and Liking Scale Items*

Scale	Items
Trust	*Adapted from Rempel et al. (1985):*
	I can count on this person to be concerned about my well-being.
	I know this person will always be ready and willing to offer me strength and support.
	I cannot rely on this person to keep the promises he/she makes to me. (R)
	Sometimes I worry that this person may take advantage of me. (R)
	This person is usually dependable, especially for things that really matter to me.
	I sometimes avoid relying on this person because he/she is unpredictable. (R)
	Adapted from Johnson-George and Swap (1982):
	If we arranged to meet, I am confident that this person would be there.
	If I asked this person to call me at a certain time, I could count on receiving the call.
	This person would never intentionally misrepresent my point of view to others.
	I can confide in this person and know that he/she would not discuss my concerns with others.
	I could count on this person to return an overdue book or mail an important letter for me if I could not go myself.
	If this person came late to meet me, I would guess there was a good reason for the delay.
	If this person knew what kinds of things hurt my feelings, I would worry that he/she might use them against me if our relationship changed. (R)
	If this person unexpectedly laughed at something I did or said, I would wonder if he/she was being critical or unkind. (R)
	Novel items:
	Even if it required a personal sacrifice, this person would support me when I needed help.
	This person might treat me badly in order to get ahead socially, academically, or professionally. (R)
	I think that this person is occasionally insincere or less than fully honest with me. (R)
	In most matters, I trust this person completely.
Cultural trust	This person would not second-guess my reaction if I found someone's comments culturally offensive.
	If I said something that might seem culturally insensitive, this person would give me the benefit of the doubt.
	This person would be supportive if I avoided an activity because of my upbringing or background.
	If we disagreed over a cultural difference, this person would respect my opinions and position.
Liking	I sometimes dislike this person. (R)
	I sometimes find this person irritating or unpleasant. (R)
	My interactions with this person are typically pleasant and enjoyable.
	I enjoy being around this person.
	I sometimes try to avoid spending time with this person. (R)
	Generally speaking, I really like this person.

Note. "(R)" denotes reverse-coded items.

Procedure. In both pilot studies, participants were randomly assigned to think of "someone you know personally whom you [like more than you trust/trust more than you like]" and record that person's initials. Next, they rated their agreement with a series of statements about their trust and liking for that person.

Materials. The trust and liking scales proved reliable: respective αs = .92 and .86 in Pilot 1 and .91 and .91 in Pilot 2. The cultural trust scale was unreliable in Pilot 1 (α = .35), so two items "This person would not second-guess my reaction if someone's comments hurt or offended me" and "If I said something that might be offensive, this person would give me the benefit of the doubt" were revised to explicitly reference cultural issues: "This person would not second-guess my reaction if I found someone's comments culturally offensive" and "If I said something that might seem culturally insensitive, this person would give me the benefit of the doubt" in Pilot 2 (α = .56).

Results

Participants reported significantly higher scores on the liking items for the liked (M = 5.59, SD = 0.86) than the trusted (M = 4.62, SD = 1.06) target in Pilot 1, $t(37) = 3.15$, $p = .003$, and in Pilot 2, $t(62.8) = 3.14$, $p = .003$, respective Ms = 5.19 and 4.26, SDs = 0.94 and 1.42. Similarly, participants reported higher scores on the trust items for the trusted (M = 5.14, SD = 0.82) than the liked (M = 3.92, SD = 0.72) target in Pilot 1, $t(37) = 4.95$, $p < .001$, and in Pilot 2, $t(63) = 3.53$, $p = .001$, respective Ms = 4.96 and 4.15, SDs = 1.09 and 0.75. Mean scores on cultural trust items were slightly, but not significantly higher for trusted (M = 4.95, SD = 0.67) than liked (M = 4.83, SD = 0.67) targets in Pilot 1, $t(37) < 1$, and in Pilot 2 $t(63) = 1.33$, $p = .188$, respective Ms = 4.99 and 4.68, SDs = 0.91 and 1.00. A discriminant function analysis confirmed that the trust and liking items reliably distinguished liked versus trusted targets ($p < .001$). Analyses for gender and race subgroups in Pilot 2 were not conducted due to low numbers of men (16) and Black participants (9).

Next, a confirmatory factor analysis was conducted to analyze parcels of items from the liking and trust scales for the pooled results of both pilot studies. This analysis revealed better fit for a two-factor solution, $\chi^2(1) < 1$, $p = .558$, $\chi^2/df = 0.34$, $RMSEA < .05$, that distinguished between trust and liking items, than a single-factor solution, $\chi^2(2) = 98.03$, $p < .001$, $\chi^2/df = 49.02$, $RMSEA = .68$, that equated trust and liking, $\Delta\chi^2(1) = 97.69$, $p < .001$. This result suggests that the trust and liking scales tap into distinct constructs, despite their positive correlation, $r(102) = .21$, $p = .031$. Cultural trust correlated with trust, $r(102) = .51$, $p < .001$, more than liking, $r(102) = .35$, $p < .001$.

Discussion

This study confirmed that the trust and liking scales used in the present research assess empirically distinct aspects of interpersonal relationships and can distinguish between trusted versus liked targets. In studies that do not involve explicitly nominating targets who are differentially liked versus trusted, trust and liking are expected to be more strongly correlated, but not multicollinear. The cultural trust scale appeared to be more associated with trust than liking, and revising its items improved its reliability from Pilot 1 to Pilot 2. Adding more items to this 4-item scale might have raised its reliability, but would have risked drawing increased attention to issues of culture (and by extension, race) in the subsequent studies, all of which were designed to incorporate race as a subtle feature of their design. Cultural trust may also assume more meaning or cohesion when rating outgroup targets, which neither pilot study directly assessed, or considering ambiguous behaviors.

CHAPTER 3: DYNAMICS OF TRUST IN CLOSE RELATIONSHIPS

Trust theorists typically assert that interpersonal trust develops gradually over time in the course of repeated interactions (Hardin, 2002; Holmes & Rempel, 1989; J. Simpson, 2007a, 2007b; Wieselquist et al., 1999) and requires accumulated evidence to establish (Worchel, Cooper, & Goethals, 1991). Whereas initial interactions between strangers involve calculative trust, in which individuals take risks by trusting each other in specific domains while continually verifying information about each other and monitoring risks and losses (Rousseau et al., 1998), close relationships are theorized to involve relational trust, in which information from *within* the relationship achieves primary importance and affection—based on frequent, longer-term interactions that foster attachment, care, and concern—infuse and sustain the relationship. The present investigation focuses on relational trust in same-race and interracial friendship over time.

Trust Development

Four prominent models describe processes shaping dyadic trust over time in close relationships: a stage-based model (Holmes & Rempel, 1989), a mutual cyclical growth model (Wieselquist et al., 1999), a dyadic model (J. Simpson, 2007b), and a dyadic risk-management theory of mutual responsiveness (Murray & Holmes, 2009). Each model includes feedback loops or cross-lagged effects that emphasize partners' mutual influence on one another and the role of trust as both a start- and end-state. High initial trust is theorized to induce a cascading sequence of positive motivations, actions, and attributions that in turn increase later trust, whereas low initial trust triggers negative processes that reduce trust. Holmes and Rempel (1989) emphasize first "reciprocal reassurance" (e.g., reciprocal self-disclosure) that conveys mutual attachment and reduces uncertainty about risks of interdependence, followed by prospective analysis of relationships based on impressions of partners' responsiveness and motives, ultimately enabling partners high in trust to meet one another's needs and find integrative solutions to conflicts. Wieselquist et al. (1999) assert

that actors' dependence levels drive their commitment and their "pro-relationship behavior" (i.e., sacrificing self-interest for relationship or partner goals), which boosts partners' trust levels; then partners' increased dependence, commitment, and pro-relationship behavior increase actors' trust in turn. J. Simpson (2007b) emphasizes mutuality in the sequence of steps and dispositional influences: Partners enter strain-test situations, make mutually beneficial [or selfish] decisions, leading to positive [negative] attributions, emotions, and future expectancies, which in turn increase [decrease] trust and felt security, which shapes how future strain-test situations unfold. Murray and Holmes (2009) posit distinct trust-related processes after activation of connection versus self-protection goals. Connection goals lead actors to (a) increase dependence on partners, such that partners' responsiveness boosts actors' trust and commitment, or (b) justify their own commitment, leading them to become more response to partners, which boosts partners' trust and commitment. Self-protection goals lead actors to (a) increase partners' dependence in an effort to boost partners' responsiveness or (b) withhold commitment to reduce their vulnerability to non-responsive partners.

Model Convergence

These models all incorporate actors' vulnerability and partners' responsiveness. Actors take risks via increased vulnerability by entering into strain-test situations or accommodative dilemmas that entail escalating dependence on partners, whether via self-disclosure (Holmes & Rempel, 1989; see Reis & Shaver, 1988) or asking partners to make sacrifices (e.g., Shallcross & Simpson, 2012). In strain-test situations, partner responsiveness entails behaviors that benefit the partner or relationship over the self. Insofar as *perceived* partner responsiveness—the "belief that relationship partners are cognizant of, sensitive to, and behaviorally supportive of the self" (Reis, 2007, p. 9)—diverges from actual partner responsiveness, the former is theorized to be more critical for trust maintenance (e.g., J. Simpson, 2007b; Wieselquist et al., 1999). Likewise, in the present work, participants' trust levels are expected to depend on their own perceptions of processes more than on partners' perceptions.

Empirical Evidence

Recent empirical work has confirmed that actors' vulnerability and partners' responsiveness indeed boost trust. Asking partners to consider making large sacrifices increases state trust felt toward partners, particularly for individuals who are initially high in trust (Shallcross & Simpson, 2012). Self-disclosure tends to be reciprocal, insofar as individuals more often disclose to people who disclose to them (Derlega, Metts, Petronio, & Margulis, 1993), and disclosing to a responsive partner increases trust more than completing a fun task unrelated to disclosure or disclosing to a non-responsive partner (Reis et al., 2010; see Reis, 2006, for a review). Both actual and perceived levels of partner responsiveness influence relationship functioning: In support-seeking situations, relationship satisfaction is associated both with objectively observed partner responsiveness (Collins & Feeney, 2000) and perceived partner responsiveness, which is shaped by projections of one's own perceived responsiveness (Lemay, Clark, & Feeney, 2007). Attributions for partner responsiveness and relationship investment matter: Trust correlates positively with perceptions of partners' intrinsic but not extrinsic or instrumental motives (Rempel et al., 1985) and of partners' pro-relationship transformation of motivation (Yovetich & Rusbult, 1994). The present research incorporates several common elements in these trust models, measuring vulnerability-increasing and relationship-benefiting behaviors (e.g., self-disclosure, help/support provided), as well as perceived partner responsiveness (e.g., felt understanding, help/support received, other disclosure) and reliable alliance (a friendship function tied to perceptions of friends' relationship commitment), over time in Study 1. In addition, participants were led to imagine an interaction sequence involving mutual self-disclosure in Study 2 and were given opportunities for dyadic self-disclosure in Study 3.

Study 1: Trust and Betrayal Between Friends

The goals of Study 1 were threefold. The first aim was to establish the divergent validity and unique predictive ability of trust, liking, and closeness in the context of Blacks' and Whites' real-

world same-race and interracial friendships. The second aim was to test for race-based asymmetries or divergences in the characteristics of these friendships. The third aim was to investigate the role of trust, liking, and closeness in accounting for (a) differences between same-race and interracial dyads in relationship dynamics (e.g., levels of self-disclosure), (b) changes over time in friendship functions and interaction behaviors, (c) subsequent contact, as well trust-building and trust-betraying behaviors, and (d) inclusion of the friend's racial group in the self.

Trust (and cultural trust) was predicted to be lower in interracial (vs. same-race) friendships, consistent with theorizing about pro-ingroup generalized trust bias (see Brewer, 1997) and empirical findings of low intergroup (vs. intragroup) trust (e.g., Davis et al., 2006). Because trust and closeness (or IOS) are theorized to be closely related (e.g., see Holmes & Rempel, 1989), reported closeness levels were predicted to be lower in interracial than same-race friendships, parallel to findings of lower intimacy levels in interracial relative to same-race friendships among elementary children (Aboud, Mendelson, & Purdy, 2003) and college students (Shelton, Trail, West, & Bergsieker, 2010). The race-related predictions for the other dependent measures (liking, friendship functions, self-disclosure, perceived other disclosure, own and perceived help/support, self-expression, and perceived understanding) were less clear, due to lack of consistency in prior research on friendship dynamics: Some work has found comparable levels of emotional security and helping behaviors in interracial and same-race friendships (Aboud et al., 2003), whereas other work has found more negative outcomes for self-disclosure, perceived other disclosure, and perceived responsiveness in interracial (vs. same-race) friendships among Black participants (Shelton et al., 2010), and lower liking in interracial (vs. same-race) friendships among Black and White participants (Shelton & Bergsieker, 2008). In interracial friendships, intergroup trust was hypothesized to be asymmetric, with Black participants reporting less trust in White friends than vice-versa (see Smith, 2010). Consistent with self-expansion theory (Wright, Aron, & Tropp, 2002) and prior work in which

priming high-quality interracial (vs. same-race) friendships led participants to include their friend's racial outgroup in the self more (Page-Gould, Mendoza-Denton, Alegre, & Siy, 2010), primary relationship variables associated with high-quality friendship—closeness, liking, and (especially) trust—were expected to positively predict inclusion of outgroup in the self one year later.

Trust was predicted to be correlated with but still distinct from liking and closeness, and each of these variables was expected to show divergent validity over time. Controlling for liking and closeness, trust was hypothesized to significantly predict higher perceived friendship durability (i.e., reliable alliance) and emotional security, as well as less frequent betrayals by one's friend in the following 12 months. In contrast, because liking is theorized to be related to enjoying another person's company and attraction (as opposed to vulnerability, sacrifice, and interdependence), liking was expected to be closely related to self-expression and stimulating companionship.

Method

This study used a 2 (participant race: White, Black) × 2 (dyad race: same-race, interracial) dyadic design over time. Random assignment determined which participants nominated a same- versus different-race friend.

Participants. In the first wave of data collection, 202 self-identified Black and White Princeton undergraduates completed an online friendship questionnaire about a same- or different-race friend of the same gender for partial course credit. (Participants had to select a White or Black friend to be included in this study.) Analyses excluded 24 participants who nominated ineligible friends (20 duplicate friends, 1 graduate student, 1 non-Princeton student, 1 opposite-gender friend, and 1 friend whom the participant declined to invite). The friends nominated by the remaining 178 participants were invited to complete the same online friendship study for $8, and 104 did so (5 of the 104 dyads had to be excluded because one person misperceived the other's race). Only complete dyads (including data from both the original participant and nominated friend) were retained for

analysis; notably, the friend participation rate did not differ for same-race (57%) and interracial (57%) dyads, $\chi^2(1, N = 173) < 1$. The final sample of 99 complete dyads (12 Black/Black, 24 Black/White, 63 White/White) comprised 48 Black and 150 White participants, including 140 women and 58 men, with a mean age of 19.5.

One year later in the second wave of data collection, participants in the 99 complete dyads were invited to complete a shorter follow-up online questionnaire for $5. For 82 of 99 dyads, at least one member participated in the follow-up study, and 138 (70%) of the original 198 individual respondents completed at least part of the follow-up study, including 35 Black and 103 White participants. The attrition rate observed was comparable to that in prior longitudinal friendship research (e.g., 28% in Turner & Feddes, 2011). Moreover, no evidence of selective attrition emerged: Individuals' participation rates did not differ significantly for same-race (71%) versus interracial (65%) friendships, $\chi^2(1, N = 198) < 1$; for original participants (69%) versus nominated friends (71%), $\chi^2(1, N = 198) < 1$; for Black (73%) versus White (69%) students, $\chi^2(1, N = 198) < 1$; or for those who were sophomores (75%), juniors (64%), seniors (71%), versus recent graduates (69%) during the Time 2 data collection, $\chi^2(3, N = 198) = 2.20$, $p = .532$. Neither friendship length nor participants' trust, liking, or closeness at Time 1 significantly predicted subsequent participation at Time 2, respective βs = 0.02, -0.04, -0.02, and 0.00, all ps > .36.

Procedure. Participants completed an online friendship questionnaire for a "Connections and Conversations Between Friends" study. Participants were randomly assigned to select a friend of the same gender and (approximate) age attending the same college and whose race or ethnicity was described either as the "same as you" or "different from you." This procedure closely resembles approaches used in prior research, such as the nomination instructions: "in terms of ethnicity, sexuality, religion, or nationality […] pick your closest friend who you consider to belong to a [similar/different] social group to you" (Turner & Feddes, 2011, p. 917), but embeds the primary

independent variable (race) in a list of distracters (i.e., college, gender, age) to lower its salience. The sample included 3 participants who mistakenly selected a friend whose race did not match their assigned condition but who were retained for analyses because this error rate did not significantly vary by condition, $\chi^2(1, N = 99) < 1$. When the friends nominated by the original participants were invited to participate, they were instructed to complete the questionnaire with respect to the person who selected them. In the second wave of data collection, the original participants and invited friends were invited via email to complete an online "Friendship Follow-up" study about their relationship with one another "even if you and [friend name] are no longer friends."[1]

Materials and measures. The initial questionnaire was administered via Remark Web Survey and the follow-up via Qualtrics.com. Unless otherwise indicated, participants responded using a 7-point scale from 1 (*strongly disagree* or *not at all*) to 7 (*strongly agree*, *extremely*, or *very much*). Each measure assessed at both Time 1 and Time 2 will henceforth be labeled "initial" and "final" (e.g., "initial trust") respectively, for ease of reference, although these two time points do not necessarily represent the start or end of participants' friendships.

Background measures. First, participants reported basic demographic information—gender, class year, race/ethnicity—about themselves and their friend. Participants indicated how long they had known each other: "a few weeks" (0%), "about a month" (2%), "a few months" (15%), "6 months to less than a year" (17%), "a year to less than 2 years" (32%), or "2 years or longer" (33%).

Closeness. Next, participants rated their interpersonal closeness, or self-other overlap, using the IOS scale (Aron et al., 1992). This scale displays seven pairs of circles, one representing the self and one representing the other person, with increasing degrees of overlap. Instructions asked participants to "select the pair of circles below that you feel best represents how close you are to this

[1] One participant reported that her friend from Time 1 was "no longer a friend," but because her (former) friend did not share that perception, their data was retained for analysis. The low rate of non-friend designations (1 out of 138 responses) precluded formal analysis of friendship dissolution.

person." Greater circle overlap indicates greater interpersonal closeness. Participants answered two additional questions assessing closeness: "Relative to all your other friendships, how close are you to this person?" (from the Subjective Closeness Index of Berscheid, Snyder, & Omoto, 1989) and "How well do you know this person?". Averaging these three items yielded a closeness index (Time 1 α = .88, Time 2 α = .91).

Trust, cultural trust, and liking. Participants responded to 28 items (see Table 1) assessing their trust, cultural trust, and liking for their friend. The trust, cultural trust, and liking scales proved reliable, respective αs = .93, .70, and .85 (Time 1) then .95, .76, and .93 (Time 2).

Friendship functions. The next measure included 8 items from three dimensions of the McGill Friendship Questionnaire-Friend's Functions (MFQ-FF; Mendelson & Aboud, 1999). The reliable alliance dimension indicates the perceived durability of a friendship, based on the perception of a friend "remaining available and loyal" (Mendelson & Aboud, 1999, p. 130). The reliable alliance measure includes items assessing whether this person would "want to continue our relationship even if we didn't see each other for a few months," "continue our relationship even if other people did not like me," and "still want to continue our relationship even if we argued" (Time 1 α = .84, Time 2 α = .88). The stimulating companionship dimension assesses whether a friendship is pleasurable, insofar as it entails "doing enjoyable, amusing, or exciting things together" (Mendelson & Aboud, 1999, p. 130). Stimulating companionship items assess whether this person "makes me laugh," "is fun to sit and talk with," and "has good ideas about entertaining things to do" (Time 1 α = .78, Time 2 α = .88). The emotional security dimension reflects whether a friendship provides "comfort and confidence in novel or threatening situations" (Mendelson & Aboud, 1999, p. 130), as indicated by agreeing that this person "would make me feel better if I were worried" and "makes me feel better when I'm upset" (Time 1 α = .88, Time 2 α = .94).

Interaction behaviors. Participants also reported on "typical interactions" with their friend, indicating the extent to which they suppress their emotions or self-disclose to their friend, as well as perceptions that their friend self-discloses to them, cares about them, and understands them. Self-expression was assessed with 4 items: "It is easy to express myself during interactions with this person," "I feel I have to change myself to fit in with this person" (reverse-coded), "I often have to hold back my thoughts and feelings to avoid upsetting this person" (reverse-coded), and "I am completely myself around this person," (Time 1 α = .76, Time 2 α = .84). Self-disclosure was assessed with 3 items asking how much "you express your feelings to this person," "you share your thoughts with this person," and "personal information (e.g., information about you and your views) do you disclose to this person" (Time 1 α = .91, Time 2 α = .95). Perceived other disclosure was measured with 3 parallel items asking how much "of his/her feelings does this person express to you," "does this person share his/her thoughts with you," and "personal information (e.g., information about him/her and his/her views) does this person disclose" (Time 1 α = .88, Time 2 α = .96). The question "How much do you help or support this person?" assessed help and support given to one's friend. Similarly, the question "How much does this person help or support you?" assessed perceived help and support received from one's friend. Perceived understanding was assessed with 3 items stating that this person "really understands me: he/she sees the same virtues in me as I see in myself," "is an excellent judge of my character," and "sees the real me" (Time 1 α = .87, Time 2 α = .93).

Contact quantity and quality. The Time 2 follow-up study asked participants to report the quantity (absolute and relative to other friendships) and quality of their contact with this particular friend in the 12 months since they had completed the initial Time 1 questionnaire. Absolute contact quantity, both remote and in-person, was assessed with two items, $r(134) = .67$: "In the past 12 months, how often have you been in contact remotely (e.g., over the phone, IM, Facebook, texts,

emails)" and "[…] in person (e.g., face-to-face, spending time together, joint activities)," with these response options: "Never," "Less than monthly," "Once a month," "2-3 times a month," "Once a week," "2-3 times a week," and "Daily" (scored 1-7). Relative contact quantity was assessed with two parallel items, $r(134) = .72$, with the stem changed to "Relative to your level of contact with other friends, in the past 12 months, how much contact have you had with this person […]" and response options "Much less," "Less," The same," "More," and "Much more" (scored 1-5). The two contact quality items, $r(134) = .68$, used the stem "How satisfied have you been with the quality of your contact with this person in the past 12 months […]" and the response options "Very dissatisfied," "Dissatisfied," "Neutral," "Satisfied," and "Very satisfied" (scored 1-5).

Trust-building or -betraying behaviors. The follow-up study also asked participants to report the frequency with which their friend had enacted trust-related behaviors toward them in the previous 12 months. Specifically, 6 items asked "in the past 12 months, how often" this person performed trust-building behaviors (supported you in a difficult situation, sacrificed time or effort for your sake, done a favor for you; $\alpha = .88$) and trust-betraying behaviors (done something that upset you, argued or fought with you, disappointed you or failed to follow through; $\alpha = .81$). The response options were "Never," "Rarely," "Sometimes," "Quite often," and "Very often" (scored 1-5).

Group-based generalization. At the end of the follow-up study, three questions probed participants' explicit inclusion of their friend's racial group in the self. Thus, depending on whether participants had been assigned to report on a same-race versus interracial friendship, these questions assessed inclusion of one's racial ingroup versus outgroup in the self, respectively. This measure of explicit outgroup inclusion in the self (from Page-Gould et al., 2010) asked participants: "How close do you feel to the African American/Black [White] people that you know?", "To what extent do you identify with African American/Black [White] people?", and "To what extent do you feel strong ties with African American/Black [White] people?" ($\alpha = .85$).

Results

Because participants were nested in 99 dyads, dyadic mixed-model analyses controlled for non-independence and other dyad-level effects. Analyzing both interracial and same-race dyads made participant race a mixed variable (varying within and across dyads), so all analyses treated the two members of each dyad as indistinguishable (Kenny, Kashy, & Cook, 2006). When analyzing indistinguishable dyads, the variances of the two dyad members are treated as homogeneous (Kenny et al., 2006). Adjusting for interdependence in dyad members' responses may yield fractional degrees of freedom. A 2 (participant race: -1 = Black, 1 = White) × 2 (dyad race: -1 = interracial, 1 = same-race) factorial actor-partner interdependence model (APIM; West, Popp, & Kenny, 2008) initially tested three predictors: participant race, dyad race, and the participant race × dyad race interaction. Interclass correlations (ICCs), means, standard deviations, and correlations are given in Appendix A.

Three primary models were estimated for each dependent variable (see Table 2), with standardized friendship length as a covariate. The first model contrasted same-race and interracial dyads, respectively coded +1 and -1. The second model regressed each dependent measure on initial trust and liking simultaneously to directly compare these constructs' predictive ability. The third "full" model tested dyad race and the primary relationship variables—participants' and friends' initial trust, liking, and closeness—as predictors of each outcome variable. Initial levels of trust and liking, trust and closeness, and liking and closeness were correlated, respective *r*s(196) = .67, .56, and .36, all *p*s < .001, but not multicollinear. (Cultural trust, a trust subscale, was excluded as a predictor because it was more strongly associated with trust, $\beta = .32$, $p = .001$, than liking, $\beta = .11$, $p > .10$, or closeness, $\beta = .06$, $p > .10$, when regressed on these three predictors simultaneously, so including it would tend to underestimate trust's effect on outcomes.) In the full model, "actor" effects describe the influence of a participant's predictor variable on the participant's own outcome variable, whereas "partner" effects describe the influence of the friend's predictor on the participant's outcome.

Asymmetries based on role or race. Supplemental analyses probed within-dyad asymmetries, because this study included two potentially distinguishable types of dyad members: original participants and their invited friends. A difference score was computed for each dyad on each measure by subtracting the score of the invited friend from that of the original participant, to control for covariation in the dyad members' scores and test whether participants and friends differed from each other systematically. Significant differences between original participants and invited friends would render the dyads distinguishable, contrary to the assumptions of APIM analyses with mixed variables. None of the within-dyad difference scores significantly differed from zero at Time 1, all ts(98) < 1.2, ps > .05, or Time 2, all ts(50-55)[2] < 1.9, ps > .05.

Factorial APIM estimates and tests of within-dyad difference scores probed for interracial asymmetries between White and Black friends nested within the same (interracial) dyads. Participant race and the interaction of participant race and dyad race explained little variance; instead, extensive interdependence between dyad members emerged for almost all variables (see *ICC*s in Appendix A). Significant interactions of participant race and dyad race emerged for final cultural trust, B = .22, SE = .10, t(132.9) = 2.23, p = .028, and final perceived help/support, B = .20, SE = .10, t(127.0) = 2.12, p = .036.[3] However, submitting the within-dyad difference scores for these variables to a 2 (original participant race: Black or White) × 2 (dyad race: same-race vs. interracial) analysis of covariance (ANCOVA) controlling for friendship length revealed no significant effects of participant race, dyad race, or their interaction for final cultural trust, all Fs(1, 51) < 1, or final perceived help/support, all Fs(1, 46) < 1.60, ps > .21. Because no race-based asymmetries in interracial friendships were robust across both analytic techniques, subsequent analyses dropped participant race as a predictor.

[2] The *df* for Time 2 variables range from 50 to 55 because those placed earlier in the questionnaire were filled out by more participants (56 complete dyads) than those later on (51 complete dyads).
[3] Blacks reported less final cultural trust than Whites in same-race, t(69.5) = 2.53, p = .014, but not interracial, t(72.7) < 1, dyads. Whites' and Blacks' final perceived help/support did not differ significantly in same-race, t(66.7) = 1.25, p = .215, or interracial, t(57.1) = 1.76, p = .083, dyads.

Background measures. Participants' reports of how long they had known each other were highly correlated (*ICC* = .94, $p < .001$) and were averaged to create a friendship length value for each dyad. A Mann-Whitney *U* test, the non-parametric equivalent of the *t* test when examining ordinal data, confirmed that friendship length did not differ significantly for same-race versus interracial dyads, $U = 708.5$, $p = .103$, both of which had a median friendship length of a year to less than 2 years. All reported analyses include friendship length (standardized) as a covariate.

Gender differences. Six measures revealed gender differences, controlling for friendship length. Relative to men, women reported higher initial levels of emotional security (marginal *M*s = 5.01 & 5.75, *SD*s = 1.21 & 1.21), $t(96) = 3.65$, $p < .001$, self-disclosure (*M*s = 4.78 & 5.46, *SD*s = 1.21 & 1.22), $t(96) = 2.93$, $p = .004$, perceived other disclosure (*M*s = 4.79 & 5.61, *SD*s = 1.06 & 1.05), $t(96) = 4.28$, $p < .001$, help/support (*M*s = 4.83 & 5.68, *SD*s = 1.33 & 1.34), $t(96) = 3.37$, $p = .001$, and (marginally) perceived help/support (*M*s = 5.27 & 5.71, *SD*s = 1.37 & 1.37), $t(96) = 1.74$, $p = 085$. Reported contact quality in the subsequent 12 months was marginally higher for men ($M = 4.13$, $SD = 0.87$) than women ($M = 3.79$, $SD = 0.88$), $t(76.7) = 1.83$, $p = .071$. Participant gender did not significantly qualify any effects of initial trust, all *p*s > .10, so gender was dropped from analysis.

Trust, liking, and closeness. Longitudinal analyses using participants' trust, liking, and closeness measured at Time 1 to predict their trust, liking, and closeness measured at Time 2 enable assessment of these three constructs' divergent validity. As expected, when comparing participants' initial trust, liking, and closeness simultaneously, only initial trust significantly predicted final trust (and final cultural trust), only initial liking significantly predicted final liking, and only initial closeness significantly predicted final closeness (see Table 2, Model 3). Thus, the distinctiveness of trust, liking, and closeness appears to persist over time. No partner (i.e., friend) effects were significant, although a marginal tendency emerged for participants whose friends reported higher initial liking to report greater final closeness to these friends.

Table 2: *Standardized Regression Estimates for Dyad Race, Trust, Liking, and Closeness at Time 1 Predicting the Dependent Measures in Study 1*

Model	1	2		3						
Dependent measure	DR	T	L	DR	T_A	L_A	C_A	T_P	L_P	C_P
Primary relationship variables										
T2 Trust	-.03	.52*	.04	-.10	.54*	.05	-.03	.00	.06	-.06
T2 Cultural trust	-.14	.28*	.13	-.21†	.26*	.15	.07	-.05	.14	.06
T2 Liking	-.18†	.02	.44*	-.19†	.07	.42*	-.01	.07	.01	-.11
T2 Closeness	.11	.18*	.08	.02	.02	.12	.47*	.01	.15†	.04
Friendship functions										
T1 Stimulating companionship	.24*	.33*	.36*	.05	.12†	.37*	.37*	.02	.05	.06
T2 Stimulating companionship	-.03	.02	.29*	-.06	-.05	.29*	.24*	.06	.05	-.02
T1 Emotional security	.15	.56*	.09	-.06	.42*	.11	.19*	-.04	-.09	.27*
T2 Emotional security	.02	.37*	.08	-.04	.28*	.07	.25*	.00	.04	-.05
T1 Reliable alliance	.33*	.72*	-.05	.13*	.60*	-.03	.16*	-.04	.03	.13†
T2 Reliable alliance	.02	.21*	.21*	-.04	.20*	.22*	.12	.10	.06	-.06
Interaction behaviors										
T1 Self-expression	.10	.31*	.46*	-.04	.21*	.48*	.15†	.02	-.13†	.08
T2 Self-expression	-.07	.16	.21*	-.09	.13	.19†	.18	.09	.02	-.17
T1 Self-disclosure	.36*	.34*	.10	.14*	.10	.05	.57*	.00	-.08	.20*
T2 Self-disclosure	.12	.19*	.09	.06	.07	.09	.37*	.06	.02	.03
T1 Perceived other disclosure	.27*	.49*	-.23*	.06	.23*	-.22*	.50*	-.06	.05	.18*
T2 Perceived other disclosure	.31*	.31*	-.15	.23*	.14	-.10	.39*	.05	.10	-.03
T1 Help/support	.31*	.43*	-.02	.07	.15*	-.05	.66*	-.01	.00	.09
T2 Help/support	.12	.29*	.02	.06	.07	.04	.50*	-.04	.10	-.03
T1 Perceived help/support	.28*	.50*	.05	.06	.31*	.02	.46*	.18*	-.08	-.03
T2 Perceived help/support	.13	.35*	.00	.04	.29*	.01	.20†	.11	.06	-.03
T1 Perceived understanding	.21*	.48*	.15*	-.03	.21*	.13*	.57*	.15*	-.06	-.01
T2 Perceived understanding	.03	.30*	.17†	-.03	.23*	.17†	.21*	.04	.03	-.05
Interactions in last 12 months										
T2 Contact quantity (absolute)	.22†	.18*	-.02	.14	.06	.09	.25*	-.09	.15	.22*
T2 Contact quantity (relative)	.24*	.20*	.00	.20†	.03	.08	.25*	-.17	.23*	.12
T2 Contact quality	.11	.39*	-.05	.05	.33*	-.06	.20†	.02	.17	-.18
T2 Trust-building behaviors	.10	.33*	-.02	.02	.14	-.01	.48*	.02	.13	-.02
T2 Trust-betraying behaviors	.18	-.20†	-.01	.15	-.41*	.02	.47*	.01	.10	.06

Note. T1 = Time 1, T2 = Time 2, DR = dyad race (effects coded: -1 interracial, 1 same-race), T = inital trust, L = initial liking, and C = initial closeness. All models covaried for friendship length. $_A$ denotes actor (own) data and $_P$ denotes partner (friend) data in Model 4. $^*p < .05$, $^†p < .10$

As predicted, trust and closeness levels differed in same-race versus interracial friendships (see Figure 1). Participants describing same-race (vs. interracial) friendships reported greater initial levels of trust and closeness, respective βs = .22 and .27, *p*s < .05, but not cultural trust or liking, respective βs = .12 and .09, *p*s > .10. One year later same-race and interracial friendships did not differ in final trust or closeness (see Table 2, Model 1).

Interpersonal closeness is theorized to be a key predictor of trust, so an additional set of path analyses examined whether the data provided more support for a model in which (a) initial closeness mediated the effects of dyad race on initial trust, as hypothesized, or (b) initial trust mediated the effects of dyad race on initial closeness. Both indirect paths proved significant, respective *Z*s = 2.74 and 2.41, *p*s = .006 and .016, but only the former model demonstrated full mediation. The direct path from dyad race to initial trust (β = .22, *p* = .014) dropped to non-significance (β = .06, *p* = .441) when controlling for initial closeness, whereas the direct path from dyad race to initial closeness (β = .27, *p* = .005) remained significant (β = .18, *p* = .027) when controlling for initial trust. Analyses of trust and closeness over time showed that initial closeness predicted final trust more strongly (β = .37, *p* < .001) than initial trust predicted final closeness (β = .24, *p* < .001). Taken together, these results suggest that same-race (vs. interracial) friendships tended to be closer and thus more trusting, rather than more trusting and thus closer.

Friendship functions. Participants describing same-race (vs. interracial) friendships reported significantly higher initial levels of stimulating companionship and reliable alliance, but not emotional security; no dyad-race differences emerged for final friendship functions (see Table 2, Model 1). Model 2 directly compared participants' initial trust and liking as predictors of friendship functions: Initial trust positively predicted reliable alliance and emotional security at both time points and final stimulating companionship, whereas initial liking positively predicted stimulating companionship at both time points and final reliable alliance.

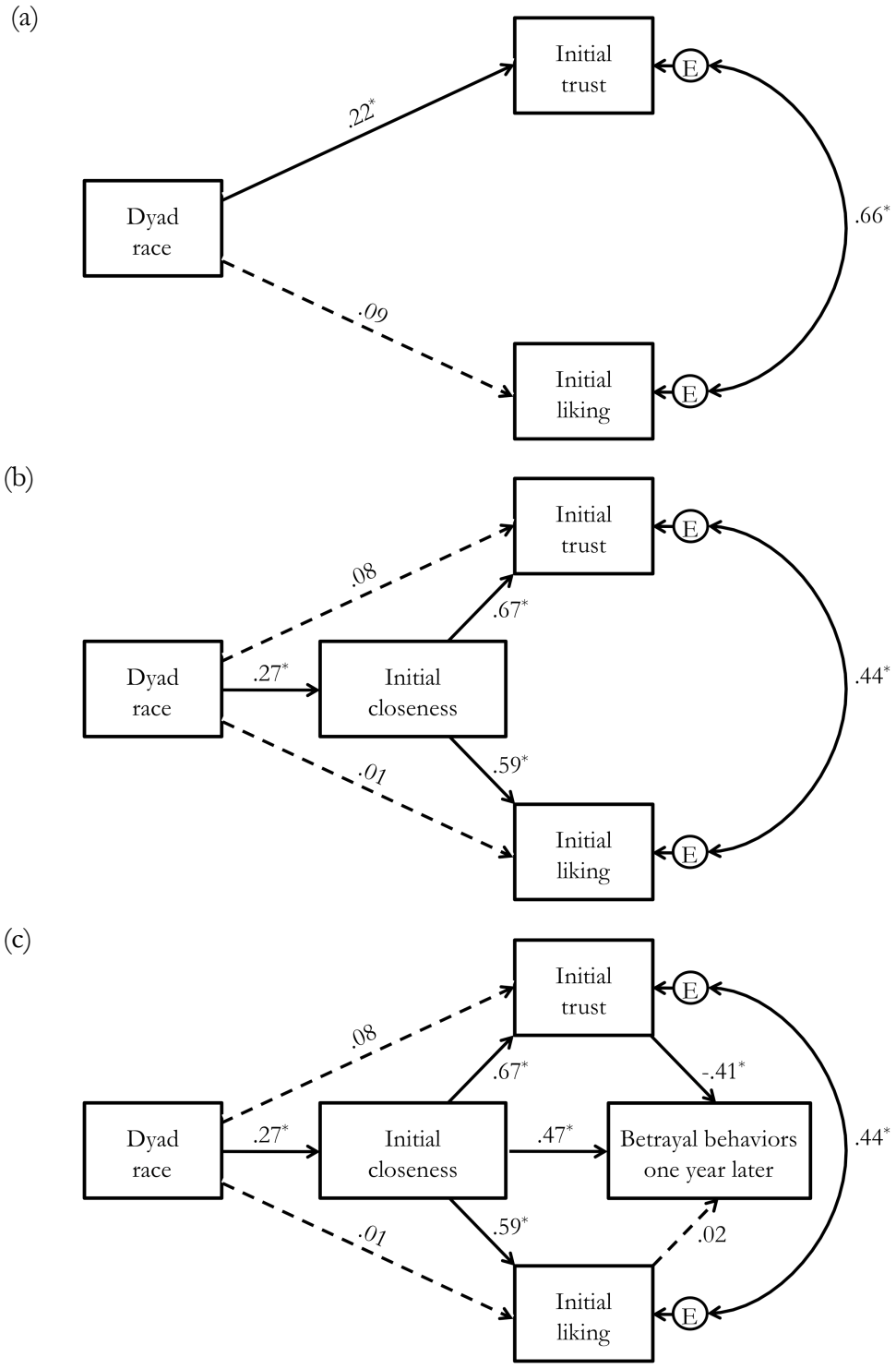


Figure 1. Simplified models showing (a) dyad-race effects, $\chi^2(5) = 8.45$, $p = .133$, $\chi^2/df = 1.69$; (b) mediation by initial closeness, $\chi^2(6) = 12.48$, $p = .052$, $\chi^2/df = 2.08$; and (c) initial trust, closeness, and liking predicting trust-betraying behaviors a year later (due to incomplete data, paths to betrayal were fitted using APIM mixed models, not SEM) in Study 1. Dashed lines indicate non-significant paths dropped from the model without reducing model fit. Paths involving friendship length (all panels) and partner effects for closeness (panel b) were significant but are not displayed. $^*p < .05$

The full model included participants' and friends' initial closeness, as well as friends' initial trust and liking, as predictors (see Table 2, Model 3). Participants' initial closeness was positively associated with all three functions: stimulating companionship, emotional security, and (at Time 1) reliable alliance. Moreover, friends' initial closeness also made significant independent contributions to increasing participants' initial levels of emotional security and (marginally) reliable alliance. No other partner (i.e., friend) effects significantly predicted friendship functions. Participants' initial trust and liking showed largely consistent associations with friendship functions over time. At both time points, initial trust was positively associated with emotional security and reliable alliance; initial liking, with stimulating companionship. These results are theoretically congruent because emotional security—the expectation that another will be supportive when one is emotionally vulnerable—and reliable alliance—perceived friendship durability—are theorized to be key precursors and outcomes of trust, while stimulating companionship involves fun and entertainment, which provides a basis for enjoying someone's company and liking them.

In the full model, the dyad-race differences for initial stimulating companionship and reliable alliance decreased in size. Same-race and interracial dyads no longer differed significantly on initial stimulating companionship, and although participants describing same-race (vs. interracial) friendships reported higher levels of initial reliable alliance, the effect size diminished. Follow-up path analyses tested whether the indirect paths from dyad race to friendship functions through participants' initial trust or closeness were significant. (Initial liking did not differ significantly between same-race and interracial dyads and was thus not tested as a potential mediator.) Notably, the paths from the proposed mediators to the outcomes were based on Model 3—which controls for friendship length, participants' initial liking, and friends' initial trust, closeness, and liking—providing a relatively conservative estimate of the indirect paths. Initial closeness, $Z = 2.43$, $p = .015$, but not initial trust, $Z = 1.40$, $p > .35$, mediated the dyad-race difference in initial stimulating

companionship. The dyad-race difference in initial reliable alliance was mediated by initial trust, $Z = 2.43$, $p = .015$, and (marginally) initial closeness, $Z = 1.85$, $p = .064$.

A final set of analyses probed the predictive potency of participants' initial trust over time for emotional security and reliable alliance, the two friendship functions significantly associated with initial trust at both time points. These analyses added a lagged predictor variable in each model while retaining friendship length as a covariate. For example, these analyses included participants' initial emotional security as a predictor of final emotional security. (All Time 2 variables correlated positively with their Time 1 counterparts, $p < .001$, see Appendix A.) Because the lagged initial predictor variables explain extensive variance in the final measures, any other variable that accounts for added unique variance in such models demonstrates substantial predictive potency. When including initial trust as a lagged predictor of final trust, neither emotional security ($\beta = -.08$, $p = .240$) nor reliable alliance ($\beta = -.08$, $p = .423$) significantly predicted final trust. In contrast, when including a lagged friendship function as a predictor, initial trust significantly predicted final levels of emotional security ($\beta = .18$, $p = .019$) and reliable alliance ($\beta = .29$, $p = .005$). These results suggest that higher trust levels lead to greater subsequent emotional security and reliable alliance more so than vice-versa.

Interaction behaviors. As predicted, participants describing same-race (vs. interracial) friendships reported higher initial levels of self-disclosure, perceived other disclosure, perceived help/support, and perceived understanding, all ps $< .05$. One year later, most of these dyad-race differences dropped to non-significance, one persisted, and none reversed (see Table 2, Model 1).

In the Model 2 comparison between initial trust and liking as predictors, initial trust was significantly associated with all interaction behaviors except final self-expression, whereas initial liking was associated primarily with self-expression. Controlling for initial liking, participants who reported initially trusting their friends more had significantly higher levels of self-expression (at Time

1 only), self-disclosure, perceived disclosure, help/support, perceived help/support, and perceived understanding. Controlling for initial trust, greater initial liking was associated with higher levels of self-expression and perceived understanding (marginally at Time 2), as well as lower levels of initial perceived other-disclosure.

In the full model, participants' initial trust and/or initial closeness showed stronger associations than did initial liking with all interaction behaviors except self-expression (see Table 2, Model 3). Initial trust correlated significantly with initial levels of perceived other disclosure, help/support, and self-expression, as well as perceived help/support and perceived understanding at both time points. In contrast, initial liking showed significant correlations only with initial levels of self-expression, perceived understanding, and (inversely) perceived other disclosure. (Initial liking marginally predicted final levels of self-expression and perceived understanding.) Participants' initial closeness correlated at least marginally with all measures except final self-expression. Four significant partner (i.e., friend) effects emerged for interaction behaviors: Friends' initial closeness positively predicted participants' initial levels of self-disclosure and perceived other disclosure, and friends' initial trust positively predicted participants' initial levels of perceived help/support and perceived understanding. No significant partner effects emerged for friends' initial liking.

When accounting for variation in participants' and friends' trust, liking, and closeness, most differences between same-race and interracial dyads from Model 1 were substantially attenuated (see Table 2, Model 3). Same-race and interracial dyads no longer differed significantly in initial levels of perceived other disclosure, help/support, perceived help/support, or perceived understanding. Controlling for trust, liking, and closeness, participants describing same-race (vs. interracial) friendships reported higher levels of initial self-disclosure and final perceived other disclosure, but the size of these effects decreased. Path analyses confirmed that initial closeness significantly mediated all the dyad race differences observed for interaction behaviors, namely initial levels of

self-disclosure, $Z = 2.72$, $p = .007$, perceived other disclosure, $Z = 2.58$, $p = .010$, help/support, $Z = 2.74$, $p = .006$, perceived help/support, $Z = 2.59$, ps= .010, and perceived understanding, $Z = 2.72$, $p = .006$, as well as final perceived other disclosure, $Z = 2.26$, p =.024. Notably, initial trust also significantly mediated initial levels of perceived help/support and perceived understanding, respective Zs = 2.14 and 2.02, ps = .032 and .043.

Finally, lagged analyses probed two interaction behaviors—perceived help/support and perceived understanding—that were significantly associated with initial trust at both time points. In the models including initial trust as a lagged predictor of final trust, neither perceived help/support ($\beta = .00$, $p = .996$) nor perceived understanding ($\beta = -.04$, $p = .531$) significantly predicted final trust. In contrast, in each model that included a lagged interaction behavior as a predictor, initial trust significantly predicted final levels of perceived help/support ($\beta = .20$, $p = .009$) and perceived understanding ($\beta = .23$, $p = .004$). Thus, higher trust appeared to foster greater subsequent perceived help/support and perceived understanding more so than vice-versa.

Contact quantity and quality. Among the contact measures collected at Time 2 to assess friends' interactions over the intervening 12 months, contact quantity—both in absolute terms and relative to contact with other friends—but not quality showed significant dyad-race effects. Participants in same-race (vs. interracial) friendships reported having significantly higher relative contact quantity and marginally more absolute contact quantity, $p = .093$ (see Table 2, Model 1). In Model 2, controlling for initial liking, participants who reported placing more initial trust in their friends had significantly higher levels of contact quantity and quality in the intervening 12 months.

When accounting for participants' and friends' trust, liking, and closeness in the full model, participants describing same-race (vs. interracial) friendships reported marginally higher levels of relative contact quantity and no difference in absolute contact quality. Participants' initial trust and closeness—not liking—predicted contact quality and quantity. Specifically, participants who

reported greater initial closeness to their friend subsequently reported having experienced higher quantity (absolute and relative) and quality (marginally) of contact with their friend in the intervening 12 months. Participants' initial trust emerged as the only significant (not marginal) predictor of contact quality. Two partner effects emerged: Friends' initial closeness and initial liking positively predicted absolute versus relative contact quantity, respectively.

Trust-building or -betraying behaviors. No dyad-race differences emerged for reporting trust-building or trust-betraying behaviors enacted by friends over the past 12 months (see Table 2, Model 1). Controlling for initial liking in Model 2, participants who initially trusted their friends more reported experiencing more trust-building and fewer trust-betraying behaviors. The full model showed significant influences of participants' initial trust and closeness—but not liking—on trust-related behaviors. Participants who reported greater initial closeness to their friend subsequently reported having experienced both more trust-building and more trust-betraying behaviors enacted by their friend in the intervening 12 months. As hypothesized, participants' initial trust emerged as a significant negative predictor of their friend's trust-betraying behaviors.

Group-based generalization. Although this study primarily examined the internal dynamics of same-race versus interracial friendships, one final measure assessed whether the initial quality of these friendships (interracial friendships, in particular) predicted later explicit inclusion of the friends' group in the self, as self-expansion theory would suggest. Due to space constraints, this measure was only assessed with respect to the friend's racial group (rather than both Blacks and Whites), so it tapped distinct constructs for participants describing same-race versus interracial friendships and was thus omitted from the aforementioned models. Separate sets of analyses were conducted for participants reporting on same-race versus interracial friendships.[4] All analyses

[4] These two phenomena—inclusion of ingroup versus outgroup in the self—are not necessarily comparable, so analyzing same-race and interracial friendships separately represents a more conservative approach. That said, alternate analyses collapsing across conditions revealed a main

controlled for friendship length while including participant race and the primary relationship variables—initial trust, liking, and closeness—as simultaneous predictors.

For participants describing same-race friendships, no significant effects emerged. Black and White participants reported comparable levels of inclusion of ingroup in the self, respective *M*s = 5.57 and 5.57, *SD*s = 0.93 and 1.31, $t(52.9) < 1$. Inclusion of ingroup in the self at Time 2 was marginally correlated with initial closeness, $\beta = .22$, $t(68.3) = 1.93$, $p = .057$, but not with trust, $t(95.3) < 1$, or liking $t(92.1) < 1$. For participants describing interracial friendships, however, two significant effects emerged. Black participants reported greater inclusion of outgroup in the self ($M = 5.09$, $SD = 1.20$) than did White participants ($M = 4.05$, $SD = 1.31$), $t(17.5) = 3.06$, $p = .007$. Moreover, participants who initially trusted their outgroup friend more reported including that outgroup in the self to a greater extent one year later, $\beta = .44$, $t(19.2) = 2.31$, $p = .032$, whereas initial liking and closeness were not related to later inclusion of outgroup in the self, respective *t*s < 1. Follow-up analyses testing effects of participant race, initial trust, and the interaction of participant race and initial trust (controlling for friendship length) showed a slightly stronger relationship between trust and inclusion of outgroup in the self for Whites, $\beta = .66$, $t(22.3) = 2.19$, $p = .039$, than for Blacks, $\beta = .53$, $t(22.8) = 2.52$, $p = .019$, which parallels prior meta-analytic findings that intergroup contact effects tend to be larger for Whites than minorities (Tropp & Pettigrew, 2005), but in this instance the difference in slopes for Whites and Blacks was not significant, $t(20.5) < 1$.

Discussion

This study documents divergence in trust, closeness, friendship functions, interaction behaviors, and interaction experiences in same-race versus interracial real-world friendships. As

---

effect of dyad race, due to greater inclusion of ingroup (vs. outgroup) in the self, $t(70.4) = 2.94$, $p = .004$. A participant race by dyad race interaction qualified this effect, $t(118.9) = 2.24$, $p = .027$: Black (vs. White) participants included the outgroup in the self more (see main text). Initial trust predicted inclusion of friend's racial group in the self, $\beta = .26$, $t(123.4) = 2.22$, $p = .028$, but this relationship was weaker for inclusion of ingroup (vs. outgroup) in the self, $\beta = -.21$, $t(123.4) = 1.86$, $p = .066$.

predicted, participants generally evaluated interracial (vs. same-race) friendships less favorably, reporting significantly lower initial levels of trust, closeness, stimulating companionship, reliable alliance, self-disclosure, perceived other disclosure, perceived help/support, and perceived understanding, followed a year later by less perceived other disclosure and lower contact quantity. Contrary to predictions, no asymmetries based on participant race emerged in interracial (vs. same-race) friendships. The tendency for Black (vs. White) participants to rate friendships slightly more negatively was not significantly affected by dyad race and thus not relevant for comparing interracial and same-race friendships. When accounting for variation in participants' and friends' trust, liking, and closeness, most dyad-race effects dropped to non-significance. All dyad-race effects were mediated (at least marginally) by initial closeness, with initial trust significantly mediating initial levels of reliable alliance, perceived help/support, perceived other disclosure (marginally), and perceived understanding, and no mediation by initial liking.

In addition to probing race-related effects, this study also verified trust's divergent validity relative to liking (and closeness) and its ability to predict unique variance in relationship processes. Participants' initial trust and liking exclusively predicted final trust and liking, respectively. In a direct comparison, liking significantly predicted self-expression and stimulating companionship, whereas trust significantly predicted these variables at Time 1 and all other outcomes at both time points. In the full model including dyad race and friendship length—plus participants' and friends' initial trust, liking, and closeness—as predictors, participants' closeness correlated with almost all outcomes and trust still independently predicted emotional security, reliable alliance, perceived other disclosure, help/support, perceived help/support, perceived understanding, contact quality, and frequency of experiencing trust-betraying behaviors.

Most friendship processes showed positive associations with trust, liking, and closeness, with two notable exceptions: perceived other disclosure and betrayals. Though perceived other disclosure

had positive bivariate correlations with initial trust and liking (see Appendix A), it was predicted positively by initial trust and *negatively* by initial liking in Models 2 and 3. If trust requires interdependence, whereas liking arises from enjoying the other, a partner's disclosure of personal information—signaling an intimate, mutually vulnerable relationship—should be more positively linked to trust than liking. Insofar as initial trust and liking covary, jointly reflecting overall relationship positivity, they should correlate positively overall with perceived other disclosure, but it follows that trust apart from liking corresponded to more disclosure and liking without trust was tied to less disclosure. Similarly, participants' higher initial trust levels predicted experiencing trust-betraying behaviors less frequently in the intervening 12 months, whereas greater initial closeness predicted experiencing trust-betraying (and trust-building) behaviors *more* frequently. These opposing unique effects of trust and closeness on trust betrayals suggests that although more (vs. less) intimate friendships may entail a higher incidence of both constructive and destructive behaviors—much as closer relationships do not differ in positive versus negative emotional tone (Berscheid et al., 1989)—only higher trust, not liking or closeness, predicts experiencing fewer trust violations. This finding identifies a meaningful real-world correlate of trust, highlighting the need to study trust per se, not just closeness or liking.

Finally, although the correlational design of this study limits conclusions about causation, a series of path analyses and lagged analyses investigated possibly causal relationships. Path analyses implied that same-race (vs. interracial) friendships tended to be closer and thus more trusting, not more trusting and thus closer. Notably, the positive correlation observed in interracial friendships between trust and inclusion of friend's racial group in the self, controlling for interpersonal closeness to one's friend, suggests an interesting potential sequence of processes in which interpersonal closeness leads to interpersonal trust, which in turn leads to intergroup closeness in the context of interracial friendships. Holding friendship length constant, participants felt closer to same-race (vs.

cross-race) friends, and this greater inclusion of other in the self accounted for their greater trust reported for same-race friends; however, insofar as participants trusted their cross-race friends, over time they increasingly included that friend's racial group in their self-concept.

In addition, lagged analyses suggested that trust plays a central role in maintaining several key friendship functions and interaction behaviors—emotional security, reliable alliance, perceived help/support, and perceived understanding—over time, more so than vice-versa. The persistent association of trust with reliable alliance—the perception that one's friendship is durable because one's friend will continue it despite infrequent contact, disagreements, or social disapproval—acquires added meaning because interracial friendships tend to be more fragile than same-race friendships (Aboud et al., 2003; Hallinan & Williams, 1987). Interventions designed to strengthen interracial trust, as opposed to merely increasing liking by reducing prejudice, may thus hold greater promise for creating enduring interracial friendships that do not involve frequent betrayals of trust.

CHAPTER 4: BETRAYAL AND TRUST REPAIR

To extend the finding that initial trust significantly predicted whether individuals would report experiencing trust-betraying behaviors (e.g, failure to follow through on commitments) at the hands of a friend in the subsequent year, the present chapter focuses specifically on responses to violations, or betrayals, of trust. Betrayals—defined as being "unfaithful or disloyal" (Steinmetz, 1993, p. 63) or as "the perceived violation by a partner of an implicit or explicit relationship-relevant norm" (Finkel, Rusbult, Kumashiro, & Hannon, 2002, p. 957)—are theorized to undermine future trust. Leading models of trust development in close relationships (Holmes & Rempel, 1989; Murray & Holmes, 2009; J. Simpson, 2007b; Wieselquist et al., 1999) also make predictions regarding trust disruption. Just as favorable responses in strain-test situations build trust, unfavorable responses (e.g., self-serving actions) or perceived non-responsiveness can undermine trust. Notably, the impact of positive versus negative actions on partners' trust appears asymmetric: Trust is more easily broken than built (Rothbart & Park, 1986) and is difficult to restore following a betrayal (Lount et al., 2008; Webb & Worchel, 1986). In mixed-motive games, for instance, the negative impact of two consecutive selfish moves on partners' cooperation is larger and more persistent than the positive impact of two unselfish moves (e.g., Lount et al., 2008).

Individual Differences

Responses to trust violations may vary across individuals (or members of different groups) due to differences in the dispositional or chronic tendency to trust or distrust partners (Murray & Holmes, 2009; J. Simpson, 2007b). Although individual differences in generalized trust (Rotter, 1967; Wrightsman, 1964) have not proven effective at predicting trust in specific close relationships (e.g., Johnson-George & Swap, 1982; Larzelere & Huston, 1980), individuals' attachment style does influence dyadic trust. For instance, *both* partners in a relationship report less dyadic trust if either one has an anxious attachment style (Wieselquist et al., 1999), and people high in attachment-related

anxiety typically construe partners' caregiving as less responsive and helpful (than do independent observers), decreasing felt support (Campbell, Simpson, Boldry, & Kashy, 2005; Collins & Feeney, 2004; J. Simpson, Rholes, & Nelligan, 1992, J. Simpson, Rholes, & Phillips, 1996).

Individuals' dispositional or chronic (dis)trusting tendencies produce divergent attributions for partner behavior and divergent downstream trust. In one study, couples high, medium, or low in dyadic trust discussed (a) resolved or (b) unresolved relationship conflicts or (c) contemplated conflicts (Holmes & Rempel, 1989). Low-trust couples perceived partners' behaviors and motives negatively in all conditions. Medium-trust couples showed a discounting pattern, perceiving partners' behaviors positively but motives negatively after discussing resolved conflicts (relative to the control condition). High-trust couples displayed a buffering pattern, perceiving partners' behavior and motives especially positively after discussing unresolved conflicts. In a study of conflict discussions, high-trust couples made more attribution statements about positive relationship aspects, whereas medium-trust couples focused on negative explanations and low-trust couples voiced positive attributions in public but made negative attributions in private (Rempel, Ross, & Holmes, 2001).

Recurring concerns with rejection can heighten individuals' sensitivity to trust violations, real or perceived. Chronically anticipating interpersonal rejection is associated with inferring negative intent from a new partner's hypothetical behaviors (Downey & Feldman, 1996). Similarly, higher attachment-related anxiety about acceptance predicts attributing hypothetical or actual partner misdeeds and daily conflicts to waning commitment (Campbell et al., 2005; Collins, 1996; Mikulincer, 1998). Less trusting individuals may ironically undermine trust via attempts to sustain it:

> They often become trapped in approach–avoidance conflict situations in which positive partner behaviors are viewed as hopeful signs of possible relationship improvement, but any hint of negative behavior is taken as clear evidence that relationship difficulties might be imminent. As a result, less trusting individuals who are involved in established relationships

> closely monitor their relationships for evidence of their partner's care, concern, and responsiveness. This hypervigilance, however, may lead them to perceive or create the negative relationship outcomes they wish to avoid, given their overreliance on the diagnostic value of negative relationship information. (Campbell et al., 2010, p. 15)

Individuals chronically low in trust more often resort to a defensive attribution process to determine whether genuine caring drives partners' actions (see Holmes & Rempel, 1989). This tendency may reflect a prevention focus (see Gable & Reis, 2001) involving self-protection rather than connection goals (Murray & Holmes, 2009). Managing relationship risks via self-protection implies a negative trust trajectory, such that "when partners are exposed to excessive situational risk, have chronic difficulty trusting, or both […] procedural rules for self-protection are likely to be chronically accessible and habitually enacted—frustrating the growth of trust and stalling commitment" (Murray & Holmes, 2009, p. 925). The present work does not assess attachment style or goal pursuit, but the claim that chronic concerns about others' trustworthiness increase vigilance to trust betrayals informs the current prediction that members of lower-status, historically victimized (vs. higher-status, historically dominant) groups will repair trust less readily following an outgroup betrayal.

Group-based Asymmetries

Several lines of work support the prediction that White and Black individuals may differ in their responses to betrayal by an outgroup (vs. ingroup) member. A growing body of work documents divergences in the perspectives and experiences of majority versus minority group members, spanning cognitive, affective, and behavioral components of interactions (for a review, see Shelton & Richeson, 2006b). Key divergences emerge in the levels and bases of Whites' and Blacks' levels of cognitive depletion (Salvatore & Shelton, 2007), positive and negative affect (Shelton & Richeson, 2006a), prejudice concerns (Shelton, 2003), nonverbal behaviors (Dovidio et al., 2002), discussion of similarities versus disparities (Saguy et al., 2009), and preferred group representations

(Dovidio et al., 2008). These findings point to both dispositional and situational factors that may make minority individuals especially vigilant to and hesitate to repair trust after trust violations.

Although Blacks on average report higher levels of perceived discrimination and distrust toward basic social institutions than do Whites (see Dovidio et al., 2008), considerable individual variation exists among minorities in their concern with encountering discrimination. Personal and community disadvantage each independently contribute to lower trust in others (Ross, Mirowsky, & Pribesh, 2001, 2002), as does parental preparation for bias (Smith, 2010), and these experiences are more common among Blacks than Whites, but are not universal. Prior work suggests that minorities who chronically worry about race-based rejection or distrust Whites' motives are more likely to see the actions of individual White interaction partners as prejudiced (e.g., Shelton & Richeson, 2006a), even if those actions are objectively positive (e.g., Sawyer et al., 2008). For Whites, who more often worry about perpetrating prejudice (Shelton, 2003), a partner's betrayal does not directly confirm their fears, whereas for minorities concerned about being the target of prejudice, trust violations may serve as potent cues of potential discriminatory treatment by an outgroup member.

Situational factors, including structural differences in the status and resources that Whites versus Blacks typically hold in interracial contexts, may lead to divergent vigilance and trust recovery. Attention and vigilance tend to flow up, not down, power hierarchies (see Fiske, 1993). Because Whites, as members of powerful, high-status groups, may not feel that their outcomes are affected by the actions of minorities, they may attend less closely to negative interaction cues (see Shelton & Richeson, 2006b), including trust violations. Indeed, relative to Whites, Blacks show a stronger association of distrust with lack of control in adult samples (Mabry & Kiecolt, 2005) and with threat among college students (Dovidio et al., 2008). Thus, Black participants are hypothesized to repair trust less readily following an outgroup (vs. ingroup) trust violation, relative to Whites.

Study 2a: Virtual Interracial Trust Betrayals

Among the more notable findings of Study 1 is the significant relationship—controlling for liking and closeness—between initially trusting one's friend and subsequently not experiencing betrayal behaviors enacted by that friend. Conceivably, however, this effect might be partially attributable to individual differences in how relationships and behaviors are perceived (e.g., positive illusions). Studies 2a and 2b essentially invert this relationship by testing the extent to which an experimentally manipulated trust betrayal by a hypothetical friend—allowing for standardization of both behavior and friend characteristics—lowers subsequent reported trust levels. This study also brings intergroup factors into sharper focus by investigating whether betrayals by racial outgroup versus ingroup members have different implications for subsequent trust.

Specifically, Study 2 examines trust, cultural trust, and liking in the context of a simulated same-race or interracial friendship in which the friend displayed either a betrayal or consistently positive behavior. Participants, especially Blacks, were hypothesized to report lower levels of trust following trust violations by outgroup as opposed to ingroup members. Intergroup interactions involve substantially more initial uncertainty than intragroup interactions (see Stephan & Stephan, 1985; Vorauer, 2006), creating attributional ambiguity for behavior that may lead to negative intergroup dynamics (see Crocker et al., 1991). Because the target person's behavior is occasionally negative but predominantly positive, overall (general) trust may nevertheless be repaired for all targets. Cultural trust, however, which is theorized to reflect trust that the other person will behave benevolently toward oneself in dealing with culturally sensitive, often ambiguous issues, may be less readily repaired when (White) outgroup members display behavior that is not consistently positive. Because trust and liking are understood to be theoretically and empirically distinct constructs, with liking less sensitive to isolated instances of negativity, no systematic group-based liking differences were predicted. Finally, consistent with the claim that subjective closeness is a key determinant of

trust and can attenuate intergroup trust deficits, closeness was expected to moderate the group-based cultural trust gap for Blacks.

Method

The study used a 3 (behavior: no betrayal, early betrayal, late betrayal) × 2 (target: ingroup, outgroup) × 2 (participant race: Black, White) design. Participants were assigned at random to imagine a same- versus different-race target and interactions that did versus did not include a trust violation. White and Black participants were asked to imagine a series of six detailed interpersonal interactions with a fictive friend, then report the extent to which they trusted and liked this person.

Participants. Originally, 134 Black and White Princeton undergraduate users of the Facebook.com social networking website participated for partial course credit or $5. Analyses excluded 10 participants: 4 failed a manipulation check, 3 reported suspicion that the study examined race, 1 failed to read the instructions, 1 reported being distracted and disengaged, and 1 gave invariant responses ($SD = 0$) across multiple scales. The final sample ($N = 124$) comprised 38 Black and 86 White participants, including 85 women and 39 men, with a mean age of 19.6.

Procedure. Participants completed an online study, entitled "Friendship and Facebook: Virtual Interpersonal Interactions," ostensibly investigating "interpersonal interactions in the age of new media" and specifically "reactions to interpersonal behavior observed in online contexts such as Facebook.com." A Facebook.com-style interface was adopted to bolster the cover story and increase engagement, but instructions clearly stated that the actual Facebook.com website was not used. The study began with an "Internet Use Profile" in which participants reported their age and gender, then answered filler questions regarding their duration of Internet use, types of content sought, preferred services for social purposes, and typical Facebook.com activity.

Next, participants were asked to familiarize themselves with a Facebook.com profile, ostensibly of Princeton sophomore Devon Walker, whose gender matched that of the participant.

Target race was manipulated via the profile photo, which showed either a White or Black teenager. This paradigm—using a fictive Facebook.com user profile to manipulate perceived target race—has been used successfully by other researchers (e.g., Cwir, 2011; Klik & Butz, 2012; Nadolny, Cwir, & Spencer, 2012). In contrast to work by Cwir and colleagues that manipulated participants' similarity to targets via profiles tailored to either contain ideographic information that matched participants' previously reported tastes or not, the present study was designed to maximize felt similarity for all participants by listing a number of interests and tastes ascertained to be widely shared by Princeton undergraduates in the profiles, which varied only with regard to the target's race and gender evident in the profile photo. After viewing the profile, all participants indicated that they had not met Devon previously, so they were instructed to "please imagine that you and Devon are already acquainted—based on being in the same residential college and taking a class or two together—and are currently Facebook friends" for purposes of the subsequent simulated interactions.

All participants were presented with a series of six hypothetical social interactions involving Facebook.com: Five described trust-building behaviors (held constant across conditions), and one described either a trust violation (early- and late-betrayal conditions) or a neutral behavior (no-betrayal condition). Participants in the betrayal conditions read about Devon skipping a planned work-related night meeting with the participant, then appearing in Facebook.com photos from a party that same night. In the sequence of six interactions, this trust violation came respectively first or fourth in the early- and late-betrayal conditions. In the no-betrayal control condition, Devon asked about a paper's due date in the fourth (neutral) interaction.

After the final imagined interaction, participants indicated their closeness (IOS), liking, trust, and cultural trust toward the target (Devon). Last, participants completed manipulation checks about Devon's race and behavior, suspicion probes, demographics, questions about their engagement level,

and background measures of social desirability and general affect toward social groups (e.g., Facebook users, Princeton students, Whites/European Americans, Blacks/African Americans).

Materials and measures. All study materials were presented via the Qualtrics.com website.

Profile photos. The profile photos used to manipulate target race came from color headshots of seniors in an out-of-state 1990s high school yearbook. Pilot ratings (N = 34) of 40 photos yielded a set of 8 photos (2 Black and 2 White of each gender) matched on attractiveness, friendliness, familiarity, and suitability for use as a Facebook.com profile photo (see Appendix B). The 8 stimulus photos were rated as moderately attractive (M = 5.0) and friendly (M = 5.8) on a scale from 0 (*not at all*) to 10 (*extremely*), as well as unfamiliar to all participants and at least moderately suitable to be a Facebook profile photo. All photos were edited to use a standard background and non-descript clothing. The photos were embedded in a fictive Facebook.com user profile, presented as a cached image, that included various interests, TV shows, movies, books, and groups currently popular among both Black and White Princeton undergraduates of both genders (see Appendix C).

Behaviors. The imagined interactions described the target, Devon, enacting various behaviors toward the participant (see Appendix D). Pilot testing (N = 26) assessed the positivity, trustworthiness, believability/realism, and masculinity of 15 behaviors on a scale from 0 (*not at all*) to 100 (*extremely*) with 50 (*moderately*) as the midpoint. The 7 behaviors (5 trust-building, 1 trust-neutral, 1 trust-betraying) retained in the final study were all rated as at least moderately realistic (i.e., not significantly below the midpoint; M = 53.4), all ps > .30, and comparably masculine (M = 36.4), all ps > .23. The 5 trust-building behaviors did not differ from each other on positivity (M = 69.1), all ps > .38, and were rated as more positive than the trust-neutral (M = 42.8) and trust-betraying (M = 11.2) behaviors, both ps < .001. Likewise, the 5 trust-building behaviors did not differ from each other on trustworthiness (M = 61.1), all ps > .99, and received higher trustworthiness ratings than the trust-neutral behavior (M = 48.5), p = .04, and the trust-betraying behavior (M = 13.5), p < .001.

Closeness. Interpersonal closeness, or self-other overlap, was assessed using the IOS scale (Aron et al., 1992). Instructions asked participants to "select the pair of circles below that you feel best represents how close you are to Devon."

Trust, cultural trust, and liking. Participants responded to 28 items assessing the extent to which they thought they would trust or like Devon on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). These items, listed in Table 1, were presented interspersed on one page and modified from Study 1 by substituting "Devon" for "this person" and using the subjunctive tense. The trust, cultural trust, and liking scales proved reliable, respective αs = .90, .74, and .84.

Manipulation checks, engagement items, and suspicion probes. After all dependent measures, manipulation check questions asked participants to indicate Devon's race and most negative behavior (if any). To assess engagement, two questions asked participants to report how distracted they felt during the study and how seriously they took the study, with instructions encouraging candor and emphasizing that their responses would not affect compensation. The next questions assessed participants' understanding of the study purpose, whether they were suspicious about any aspects of the study, and if so, which ones.

Background measures. Participants completed a 6-item short form (see Fischer & Fick, 1993) of the Marlowe-Crowne social desirability measure (e.g., "I have never intensely disliked anyone"; Crowne & Marlowe, 1960) with *true* (1) and *false* (2) as response options. A feeling thermometer elicited ratings of general affect toward social groups (e.g., Facebook users, Princeton students, Whites/European Americans, Blacks/African Americans). Participants used slider bars to "provide a number between 0 and 100 degrees to indicate your overall evaluation of typical members of a target group (from cold to warm)."

Results

Participants' liking, trust, and cultural trust for Devon were each entered into a 2 (behavior: no betrayal, betrayal)[5] × 2 (target: ingroup, outgroup) × 2 (participant race: Black, White) ANCOVA, with participant gender,[6] social desirability (M = 1.32), pro-White affect (M = 64.92), and pro-Black affect (M = 63.06) as covariates (see Table 3 for all means and SDs by condition). For liking, the interaction of behavior, target, and participant race was not significant, $F(1, 112) < 1$, $\eta^2_p < .01$, as predicted, nor did any lower-order significant effects emerge, all Fs(1, 112) < 2.4, ps > .10. For trust, only the main effect of behavior was significant: Participants trusted Devon more in the no-betrayal (M = 5.00) than betrayal (M = 4.60) conditions, $F(1, 112) = 4.90$, $p = .029$, $\eta^2_p = .04$. The hypothesized three-way interaction of behavior, target, and participant race failed to attain significance for trust, $F(1, 112) < 1$, although the means were in the predicted direction for Black participants in the betrayal condition (outgroup = 4.54 vs. ingroup = 4.88).

Table 3: *Mean (SD) Trust, Cultural Trust, Liking, and IOS by Target Behavior, Participant (P) Race, and Target Group in Study 2a*

Target behavior	P race	Target group	*n*	Trust	Cultural trust	Liking	IOS
No betrayal	White	Ingroup	13	5.14 (0.84)	4.83 (0.88)	4.79 (1.02)	3.77 (1.32)
		Outgroup	18	5.11 (0.83)	4.46 (0.87)	5.16 (1.00)	4.10 (1.30)
	Black	Ingroup	4	4.84 (0.84)	4.64 (0.88)	5.34 (1.02)	4.66 (1.32)
		Outgroup	8	4.90 (0.87)	4.84 (0.91)	5.43 (1.05)	4.10 (1.36)
Betrayal	White	Ingroup	31	4.36 (0.86)	4.16 (0.90)	4.83 (1.03)	4.28 (1.34)
		Outgroup	24	4.62 (0.84)	4.36 (0.88)	4.95 (1.01)	4.35 (1.32)
	Black	Ingroup	12	4.88 (0.84)	4.88 (0.88)	5.29 (1.01)	4.43 (1.31)
		Outgroup	14	4.54 (0.90)	4.16 (0.94)	5.03 (1.08)	3.76 (1.40)

Note. Adjusted means covary for participant gender, social desirability, pro-White affect, and pro-Black affect.

[5] Orthogonal contrasts testing the distinct effects of betrayal presence/absence and early/late timing revealed no systematic differences between the early- and late-betrayal conditions. Adding betrayal timing to the model did not significantly increase the R^2 for liking, trust, or cultural trust, all Fs(4, 108) < 1.2. All subsequent analyses thus collapse across the two betrayal conditions.

[6] Gender differences emerged for liking (women outscored men), $F(1, 112) = 4.17$, $p = .044$, $\eta^2_p = .04$, but not general or cultural trust, Fs(1, 112) < 1. Gender did not qualify the behavior × target × participant race interaction for liking, trust, or cultural trust, Fs(1, 105) < 1.

The three-way interaction of behavior, target, and participant race was significant for cultural trust, $F(1, 112) = 3.92$, $p = .050$, $\eta^2_p = .03$. Cultural trust also showed a trend for participants to trust Devon less in general following a betrayal ($M = 4.39$) than no betrayal ($M = 4.69$), $F(1, 112) = 2.59$, $p = .110$, and no other effects approached significance. To interpret this significant three-way interaction, follow-up ANCOVAs confirmed that the target by participant race simple interaction was significant in the betrayal conditions, $F(1, 112) = 4.65$, $p = .033$, $\eta^2_p = .04$, but not in the no-betrayal condition, $F(1, 112) < 1$, $\eta^2_p < .01$. White participants in the betrayal conditions did not report significantly different cultural trust levels for ingroup (i.e., White) and outgroup (Black) targets, $F(1, 112) < 1$, $\eta^2_p < .01$; whereas Black participants in the betrayal conditions trusted ingroup targets more than outgroup targets, $F(1, 112) = 4.09$, $p = .045$, $\eta^2_p = .04$ (see Figure 2).

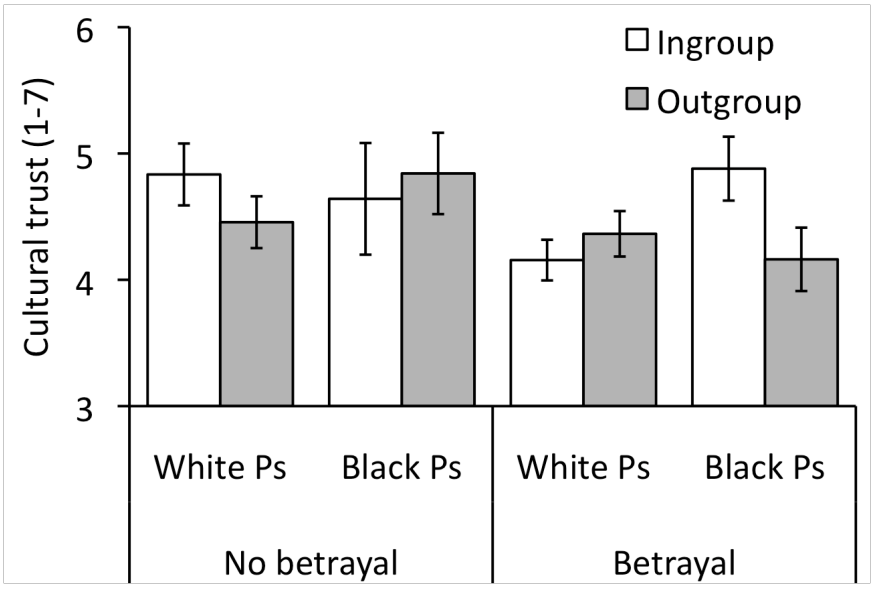


Figure 2. Mean (±1 *SE*) cultural trust by betrayal condition, target, and participant race, with means adjusted for participant gender, social desirability, pro-White affect, and pro-Black affect in Study 2a.

Discussion

As predicted, Black participants reported lower cultural trust levels when imagining a betrayal by a White friend as opposed to a Black friend. These results are consistent with the claim

that members of lower-status groups may be especially vigilant to cues that members of other higher-status (historical perpetrator) groups cannot be counted on to behave favorably toward them. Statistically significant group-based differences did not emerge for (overall) trust in the target, counter to predictions. Possibly the target's five trust-building behaviors provided such a strong signal of trustworthiness that an intergroup betrayal seemed inconsequential, except when participants considered an additional situational stressor (e.g., cultural differences) that could trigger trust violations. In support of the assertion that these findings extend traditional prejudice research, the group-based cultural trust gap for Blacks emerged when affect toward Whites and Blacks was held constant, suggesting that this phenomenon is not an outgrowth of basic interracial animus. Moreover, the absence of significant effects for liking is consistent with the claim that intergroup liking or harmony does not necessarily translate into intergroup trust between members of lower-status and higher-status groups, particularly if a member of a higher-status group has previously betrayed the trust of a lower-status-group member.

Study 2b: Virtual Interracial Trust Betrayals

Study 2b was designed as a replication of Study 2a, using only one (late) betrayal condition, as no differences emerged between the early-betrayal and late-betrayal conditions in Study 2a.

Method

The study used a 2 (behavior: no betrayal, betrayal) × 2 (target: ingroup, outgroup) × 2 (participant race: Black, White) design. White and Black participants were asked to imagine a series of six detailed interpersonal interactions with a fictive friend, then report the extent to which they trusted and liked this person. Random assignment determined which participants imagined a same- versus different-race target and whether the interactions included a trust violation or not.

Participants. Originally, 143 Black and White Princeton undergraduate Facebook.com users participated for partial course credit or $5. Analyses excluded 10 participants: 2 failed a manipulation

check, 6 reported suspicion that the study examined race, 1 reported distraction and disengagement, and 1 gave invariant responses ($SD = 0$) across many scales. The final sample ($N = 133$) comprised 58 Black and 75 White participants, including 92 women and 41 men, with a mean age of 19.7.

Materials and measures. All study materials and measures were identical to Study 2a, except as indicated, and were presented via Qualtrics.com.

Behaviors. The imagined interactions were revised very slightly for clarity (see Appendix D), modifying elements that a few Study 2a participants remarked upon as odd.

Trust, cultural trust, and liking. The trust, cultural trust, and liking scales proved reliable, respective αs = .92, .76, and .82.

Results

Participants' liking, trust, and cultural trust for Devon were each entered into a 2 (behavior: no betrayal, betrayal) × 2 (target: ingroup, outgroup) × 2 (participant race: Black, White) ANCOVA, with participant gender,[7] social desirability ($M = 1.40$), pro-White affect ($M = 64.22$), and pro-Black affect ($M = 65.19$) as covariates (see Table 4 for means and *SD*s by condition). For liking, the interaction of behavior, target, and participant race was not significant, $F(1, 121) < 1$, $\eta^2_p < .01$, as predicted. Participants reported liking the target more in the no-betrayal ($M = 5.15$) than the betrayal ($M = 4.49$) conditions, $F(1, 121) = 15.00$, $p < .001$, $\eta^2_p = .11$. A significant target by participant race interaction emerged, $F(1, 121) = 5.35$, $p = .022$, $\eta^2_p = .04$: Averaging across betrayal conditions, White participants reported liking outgroup targets ($M = 5.08$) more than ingroup targets ($M = 4.61$), $F(1, 121) = 4.34$, $p = .039$, $\eta^2_p = .04$, whereas Black participants liked outgroup ($M = 4.64$) and ingroup ($M = 4.96$) targets comparably, $F(1, 121) = 2.01$, $p > .10$, $\eta^2_p = .02$.

[7] No gender differences emerged for liking, trust, or cultural trust, Fs(1, 121) < 1, nor did gender qualify the behavior × target × participant race interaction for general or cultural trust, Fs(1, 114) < 1, or liking, $F(1, 114) = 2.17$, $p = .143$, $\eta^2_p = .02$. Collapsing across Studies 2a and 2b, a trend emerged for men to report more cultural trust than women, $F(1, 244) = 2.48$, $p = .116$, $\eta^2_p = .01$.

Table 4: *Mean (SD) Trust, Cultural Trust, Liking, and IOS by Target Behavior, Participant (P) Race, and Target Group in Study 2b*

Target behavior	P race	Target group	*n*	Trust	Cultural trust	Liking	IOS
No betrayal	White	Ingroup	19	5.06 (0.82)	4.43 (0.94)	4.94 (0.95)	4.05 (1.31)
		Outgroup	21	5.11 (0.84)	4.41 (0.96)	5.49 (0.98)	4.71 (1.34)
	Black	Ingroup	13	5.06 (0.84)	4.40 (0.96)	5.36 (0.97)	4.08 (1.33)
		Outgroup	16	5.11 (0.84)	4.45 (0.96)	4.83 (0.98)	4.14 (1.34)
Betrayal	White	Ingroup	18	3.72 (0.85)	3.99 (0.97)	4.28 (0.99)	3.24 (1.35)
		Outgroup	17	4.09 (0.85)	4.43 (0.98)	4.67 (0.99)	3.87 (1.36)
	Black	Ingroup	15	4.46 (0.85)	4.83 (0.97)	4.56 (0.98)	3.75 (1.35)
		Outgroup	14	4.13 (0.83)	4.07 (0.95)	4.45 (0.97)	3.54 (1.33)

Note. Adjusted means covary for participant gender, social desirability, pro-White affect, and pro-Black affect.

For trust, as in Study 2a, only the main effect of behavior emerged as significant: Participants trusted Devon more in the no-betrayal (M = 5.08) than betrayal (M = 4.10) conditions, $F(1, 121) = 44.40$, $p < .001$, $\eta^2_p = .27$. The predicted three-way interaction of behavior, target, and participant race failed to attain significance for trust, $F(1, 121) = 1.30$, $p = .257$, $\eta^2_p = .01$, though the means for Black participants in the betrayal condition (outgroup = 4.13 vs. ingroup = 4.46) were in the predicted direction. Cultural trust showed a marginal target-by-participant-race interaction, $F(1, 121) = 2.81$, $p = .097$, $\eta^2_p = .02$, which was qualified by the predicted three-way interaction of behavior, target, and participant race, $F(1, 121) = 3.33$, $p = .070$, $\eta^2_p = .03$. (No other effects approached significance.) To interpret this three-way interaction, follow-up ANCOVAs confirmed that the target by participant race simple interaction was significant in the betrayal condition, $F(1, 121) = 5.77$, $p = .018$, $\eta^2_p = .05$, but not in the no-betrayal condition, $F(1, 121) < 1$, $\eta^2_p < .01$. As hypothesized, White participants in the betrayal condition did not report significantly different cultural trust levels for ingroup (i.e., White) and outgroup (Black) targets, $F(1, 121) = 1.78$, , $p = .185$, $\eta^2_p = .02$; whereas Black participants in the betrayal condition trusted ingroup (Black) targets more than outgroup (White) targets, $F(1, 121) = 4.57$, $p = .034$, $\eta^2_p = .04$ (see Figure 3).

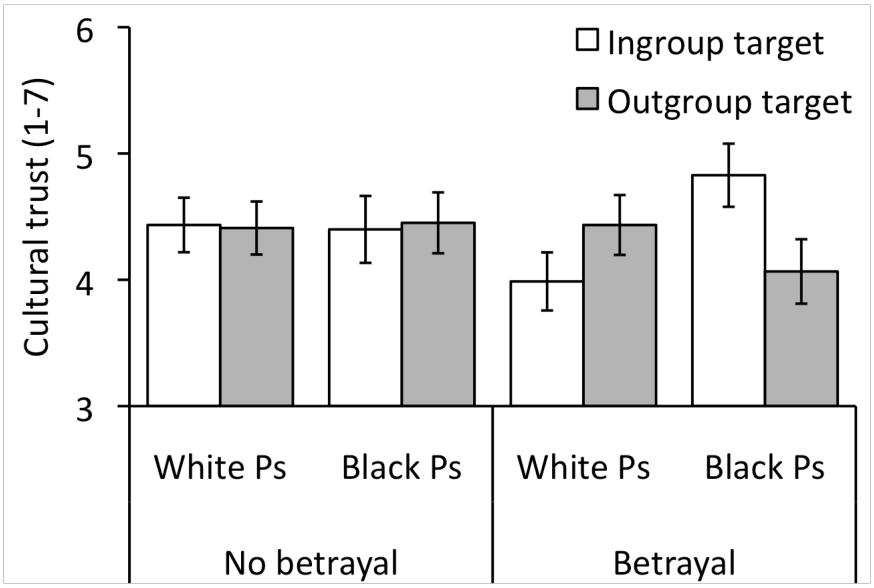


Figure 3. Mean (±1 *SE*) cultural trust by betrayal condition, target, and participant race in Study 2b, adjusted for participant gender, social desirability, pro-White affect, and pro-Black affect.

Combined analysis of Studies 2a and 2b. The data from the no-betrayal and late-betrayal conditions of both studies were pooled to test the results of closeness as a proposed moderator of interracial trust gaps with adequate statistical power (N = 257; see Green, 1991). Study (2a or 2b) was included as a covariate in all analyses, which were conducted using the GLM univariate procedure, with closeness as continuous predictor and social desirability, pro-Black affect, and pro-White affect as covariates. Closeness, operationalized as IOS immediately following the imagined sequence of interactions, significantly moderated the three-way interaction of behavior, target, and participant race for cultural trust, $F(1, 199) = 4.45$, $p = .036$, $\eta^2_p = .02$ (see Figure 4). For participants 1 *SD* above the mean on IOS, the three-way interaction of behavior, target, and participant race was not significant, $F(1, 199) = 0.00$, $\eta^2_p < .01$, whereas for participants 1 *SD* below the mean on IOS, this interaction was significant, $F(1, 199) = 6.97$, $p = .009$, $\eta^2_p = .03$. Tests of simple interactions 1 *SD* below the mean on IOS confirmed that the target by participant race interaction was significant in the betrayal condition, $F(1, 199) = 6.24$, $p = .013$, $\eta^2_p = .03$, but not in the no-betrayal condition, $F(1, 199) = 1.71$, $p = .193$, $\eta^2_p < .01$. Among participants low in IOS in

the betrayal condition, no difference emerged between ingroup and outgroup targets for cultural trust among Whites, $F(1, 199) = 1.54$, $p = .216$, $\eta^2_p < .01$, but ingroup targets elicited more cultural trust than outgroup targets did among Blacks, $F(1, 199) = 4.96$, $p = .027$, $\eta^2_p = .02$.

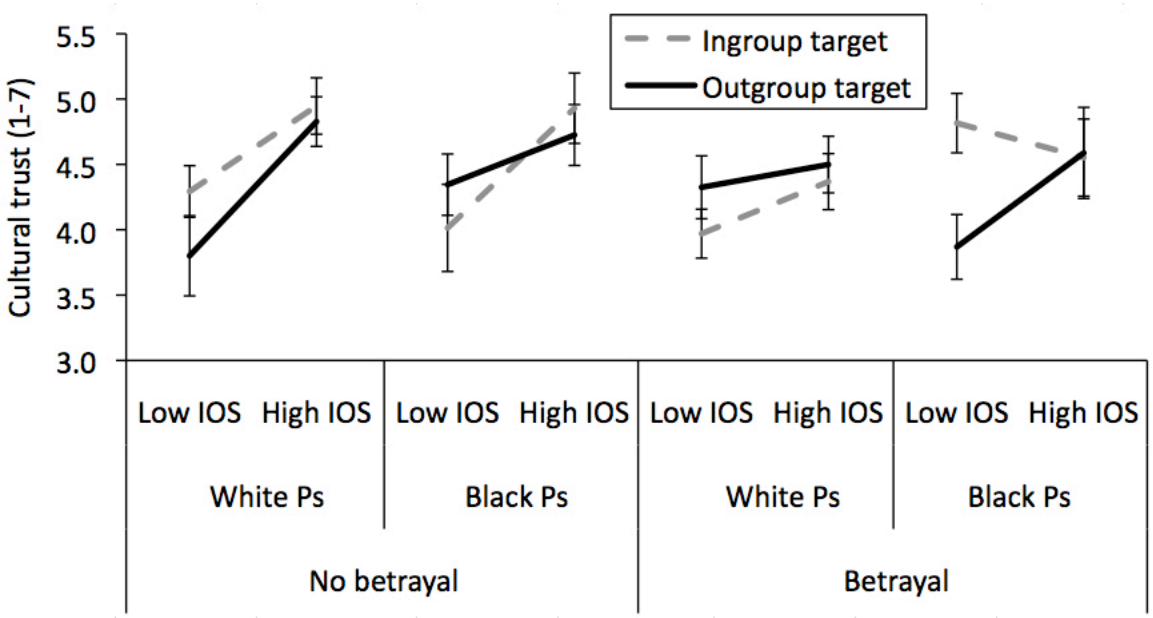


Figure 4. Marginal means (±1 *SE*) of cultural trust by betrayal condition, target, and participant race plotted at 1 *SD* below and above the mean for inclusion of other in the self (IOS) in Studies 2a and 2b, adjusting for study, participant gender, social desirability, pro-White affect, and pro-Black affect.

Discussion

Replicating Study 2a, Black participants in Study 2b reported lower cultural trust levels when imagining a betrayal by a White as opposed to a Black friend. Significant group-based differences again did not emerge for trust in the target, counter to predictions. The group-based cultural trust gap for Blacks emerged with intergroup affect held constant, suggesting that distrust need not derive from animus. Consistent with the claim that subjective closeness is a key determinant of trust and can attenuate intergroup trust deficits, IOS moderated the group-based cultural trust gap for Blacks who experienced a betrayal: Black participants who felt less close to the target reported lower cultural trust for White than Black targets, but Black participants who felt closer reported no gap.

CHAPTER 5: BUFFERING TRUST WITH INDUCED CLOSENESS

This chapter examines closeness or IOS, a factor that mediated and moderated race-based trust effects in Studies 1 and 2a/b, respectively. Emphasizing self-other merging as a determinant of trust aligns with theorizing about trust's distinct forms and bases in developing relationships (Lewicki et al., 2006). In this account, strangers initially use "calculus-based" trust that relies on incentives and deterrents, until "knowledge-based" trust emerges, which depends on understanding partners sufficiently to predict their behavior. Eventually, in select relationships people attain "identification-based" trust, in which self and partner preferences are merged and internalized. Similarly, theories that emphasize relationship-enhancing transformation of motivation in building trust (J. Simpson, 2007a, 2007b; Wieselquist et al., 1999) posit that the interests and motivations of individuals in close relationships merge over time.

In the context of close relationships, merging of self- and other-interests can take on especially profound meaning. Deepening trust fosters a sense of mutuality, an "experience of union, where the boundaries of people's own interests and their partners' interests become blurred" (Holmes & Rempel, 1989, p. 202). Theories of intimacy in close relationships also stress the importance of "habitual patterns by which partners respond to each other's needs and desires, and integrate them with their own needs and desires" (Reis & Shaver, 1988, p. 382). This blurring of boundaries between the interests, needs, and desires of self and other appears consistent with work on IOS. Individuals tend to include their partners and relationships more centrally in their own self-concept as (romantic and friendship) relationships progress (Aron, Aron, & Norman, 2001).

Closeness and Forgiveness

Several lines of work suggest that closeness or IOS holds promise for buffering trust against betrayals. Merging of self and other may lead individuals to construe departures from self-interest that benefit their partners as compatible with self-interest (see Agnew, Van Lange, Rusbult, &

Langston, 1998). Correlational evidence exists that closer friendships and romantic relationships endure longer into the future than less close relationships. This correlation emerges across studies using the Relationship Closeness Inventory (RCI; Berscheid, Snyder, & Omoto, 1989), subjective closeness items like those in Study 1, and the IOS scale (Aron et al., 1992; Berscheid et al., 1989). Individuals who report feeling more committed to their partners are more willing to forgive partner betrayals, whether these betrayals were imagined in the lab, recalled, or tracked over time (Finkel et al., 2002). In an experimental study, participants who were primed to think about themselves as interconnected with, as opposed to independent from, their partners subsequently were more willing to forgive hypothetical betrayals and demonstrated fewer destructive response tendencies (i.e., exit and neglect; Finkel et al., 2002). Making more positive attributions for actual betrayals mediated committed participants' greater willingness to forgive betrayals, suggesting that inducing closeness may lead participants to reinterpret trust violations more charitably.

Closeness in Intergroup Contexts

Closeness and IOS can play a decisive role in intergroup contexts. Including the ingroup in the self corresponds to stronger ingroup identification (Tropp & Wright, 2001), and including an outgroup in the self-concept is theorized to decrease anti-outgroup negativity (e.g., Aron et al., 2004). Automatically associating collective characteristics of groups with the self may enable intergroup interactions to unfold more like same-group interactions via an expanded understanding of the ingroup (Gaertner & Dovidio, 2000; Pettigrew & Tropp, 2006; West, Pearson, Dovidio, Shelton, & Trail, 2009). Implicit and explicit measures of IOS and closeness indeed mediate the relationship between cognitive accessibility of intergroup friends and positive intergroup interaction expectations as well as adaptive hormonal responses (Page-Gould et al., 2010). Both Whites' and minorities' outgroup attitudes (including trust) are positively associated with knowing more ingroup members who have outgroup friends and with the perceived closeness of these friendships (Wright,

Aron, McLaughlin-Volpe, & Ropp, 1997). Merely knowing about other ingroup members' close outgroup friendships may increase outgroup trust.

Study 3: Trust Betrayals in a Prisoner's Dilemma Game

This study replicated and extended Studies 2a and 2b by assessing trust, cultural trust, and liking between dyad members in the context of a rigged mixed-motive game that included either a betrayal or consistently positive partner behavior. In an adaptation of the Lount et al. (2008) paradigm, participants completed an iterative prisoner's dilemma—supposedly with their partner—in which the partner ostensibly either defected on the first two trials then cooperated on the remaining trials or cooperated on all trials. Black participants were hypothesized to report lower levels of trust following trust violations by outgroup as opposed to ingroup partners. Participants in prior studies experienced a partner's repeated defection at the outset of a mixed-motive task as an unambiguously negative signal that substantially lowered their own trust and cooperation levels (Lount et al., 2008), so in this paradigm a race-based trust asymmetry was predicted for overall trust more so than cultural trust. Because trust and liking are understood to be theoretically and empirically distinct constructs, no systematic group-based liking differences were predicted.

Unlike Studies 2a and 2b, in which subjective interpersonal closeness was measured, Study 3 involved a direct experimental manipulation of closeness. Pairs of participants completed either IOS-inducing or control tasks together prior to the prisoner's dilemma game. If subjective closeness is a key determinant of trust that can attenuate intergroup trust deficits and facilitate trust repair, inducing closeness should reduce the group-based trust gap for Blacks who experienced a betrayal. Finally, this study expanded upon the previous findings by testing whether negative affect, specifically anxiety or negative other-directed affect, hinders trust repair following a betrayal.

Method

The study used a 2 (closeness: IOS induced, control) × 2 (behavior: no betrayal, betrayal) × 2 (participant race: Black, White) × 2 (dyad race: same-race, interracial) design. Random assignment determined which participants completed IOS-inducing or control tasks, experienced partner defection versus constant cooperation, and interacted with a same- versus different-race partner.

Participants. Originally, 160 Black and White Princeton undergraduates participated in same-gender dyads for partial course credit or $12. Analyses excluded 20 dyads: 5 who knew each other previously; 2 who experienced technical problems; and 13 dyads in which a participant suspected that the game might be rigged (n = 7) or that the partner might be a confederate (1), misperceived the other's race (1), failed a manipulation check (1), reported being very distracted and disengaged (1), defected on the first trial (1), or defected on all critical trials (1).[8] The final sample (N = 120) comprised 52 Black and 68 White participants, including 84 women and 36 men, with a mean age of 20.1; half participated in same-race dyads, half in interracial dyads.

Procedure. Participants came to the lab for an "Interpersonal Processes Study" that entailed "joint activities, discussion, and decision making" with another student. First, White and Black participants were seated together at a small table and photographed by the experimenter while holding placards marked "Participant A" and "Participant B." Participants then completed a series of three IOS-inducing or control tasks involving similarities, face drawing, and discussion together in dyads. In the similarities task, participants received a paper displaying a large Venn diagram and spent 5 minutes identifying and recording similarities in the overlapping portion of the two shapes labeled "A" and "B". Participants focused on similarities between each other in the IOS condition

[8] Excluding participants who defect on the first trial parallels prior work on trust betrayals in experimental games (Bottom, Gibson, Daniels, & Murnighan, 2002; Komorita & Mechling, 1967; Lount et al., 2008); such participants would be less likely to interpret a defection as a breach of trust. Present analyses also excluded one dyad with a participant who defected on all critical trials for being an extreme univariate outlier (more than 7 *SD*s below the mean level of cooperation).

or between two specific others (whom they nominated) in the control condition. Next, participants completed a face-drawing task in which they produced a fused image of two faces, intended to visually reflect merging of self and other (Aron, Eberhardt, Davies, Wright, & Bergsieker, 2006). Participants first folded a piece of paper in half then spent 90 sec drawing the right half of their own face then—keeping the paper folded—turned over the paper and spent 90 sec drawing the left half of another person's face: that of either their partner (IOS condition) or the person that they nominated in the previous task (control condition). At the end of this task, participants unfolded the papers to spend about 30 seconds viewing the entire "split-face representation" that they had created. In the third task, participants had 5 minutes to discuss four questions of their choosing from a list of eight "closeness-generating" (IOS condition) or "small-talk" (control condition) questions used in prior research (Aron et al., 1997), such as, respectively, "What do you value most in a friendship?" and "How often do you get your hair cut? Where do you go?" After the discussion task, participants indicated which questions they had discussed and then were led to separate rooms, where they filled out a brief questionnaire about the tasks and their closeness to their partner.

Next, participants completed a "joint decision making task"—an iterative prisoners' dilemma game ostensibly with their partner via networked computers; in fact, all "partner" responses were preprogrammed. All game instructions were presented and responses recorded via computer. Participants viewed a prisoner's dilemma payoff matrix (see Appendix E) and demonstrated their comprehension of it by correctly indicating the amount of their expected payoff in each of the four possible scenarios. To encourage participants to begin with cooperative choices, before the game participants read a "Tutorial on Cooperation" (see Appendix F) adapted from Murnighan (1991, pp. 13-27; see also Bottom et al., 2002; Lount et al., 2008) describing the benefits of cooperation in repeated prisoner's dilemmas. To ensure that participants took the task seriously, participants were informed at the start of the study and reminded right before the game that several participants,

selected at random, would receive the actual dollar amounts they earned in the task, in addition to the standard study compensation. Following the Lount et al. (2008) paradigm, participants in the betrayal condition experienced two sequential defections by their programmed "partners" on two initial trials. In both the betrayal and control conditions, partner programmed responses were consistently cooperative for the remaining 20 trials (i.e., 10 trials following the defection plus 10 trials following an endgame announcement), irrespective of the participant's responses. To make the game more engaging and immersive for participants, the number of cooperative trials was reduced from 30 (Lount et al., 2008) to 20, and photos of the participants (and partners) were embedded in the task. After each trial, participants saw an outcome screen with photos of them and of the partner with each person's choice ("X" or "Y") and earnings ($0, $3, $7, or $10) displayed below the corresponding photo (see Appendix G).

After the prisoner's dilemma game, participants rated their affective experience during the interaction (including negative other-directed affect) and the extent to which they trusted and liked this person. Last, participants completed manipulation checks about their partner's race and behavior, demographics, questions about their engagement level, and background measures of social desirability and general affect toward social groups, then were thanked, debriefed, and compensated.

Materials and measures. The pre- and post-game questionnaires were administered via the Qualtrics.com website; the prisoner's dilemma game was conducted using DirectRT 2010 software. The prisoner's dilemma game used embedded idiographic stimuli: color digital photos of each participant cropped to 428 × 320 pixels.

Payoff matrix. The payoff matrix presented to participants (see Appendix E) awarded players $7 each for mutual cooperation, $3 each for mutual defection, or $10 for the defector and $0 for the cooperator if their choices differed. The Lount et al. (2008) payoff matrix used the respective payoffs of $24, $12, $30, and $6, combined with the same provision that some participants would

receive their actual earnings from 5 randomly selected trials. That matrix was adapted for purposes of the present study to fit a budget constraint of payoffs ranging from $0 to $10, and to more closely conform to the optimal prisoner's dilemma specifications provided by Kelley and Thibaut (1978), namely, a matrix with a non-correspondence index of -0.80. In this optimal matrix, temptation (i.e., payoff for sole defection minus mutual cooperation), risk (i.e., payoff for sole defection minus joint defection), and gain (i.e., payoff for mutual cooperation minus mutual defection) are equivalent, with a ratio of 1:1:1. For a $0-$10 budget constraint, these payoffs would be $6.67, $3.33, $10, and $0; in the present study, these amounts were rounded to the nearest dollar to increase perceptual fluency, yielding a non-correspondence index of -.72 and a temptation-risk-reward ratio of 1:1:1.33. In contrast, the Lount et al. (2008) matrix had a non-correspondence index of -.60 and a 1:1:2 ratio, which represents a slight increase in the convergence of participants' interests and in the reward for mutual actions (cooperation or defection) relative to discrepant actions (this component, termed "mutual behavioral control," should be set to zero in prisoner's dilemmas; Kelley & Thibaut, 1978).

Task evaluations. Participants rated the extent to which they found the three tasks *interesting* (α = .56), *difficult* (α = .42), and *enjoyable* (α = .60) on a scale from 1 (*Not at all*) to 7 (*Extremely*).

Closeness. The IOS scale (Aron et al., 1992) assessed closeness, with these instructions: "select the pair of circles below that you feel best represents how close you are to your partner."

Affect. Participants indicated the extent to which they "felt the following emotions during the final join decision-making task" using the response options "*Not at all*," "*A little*," "*Moderately*," "*Quite a lot*," and "*Very much*," respectively scored 1-5. Four types of affect were assessed: positive (*content*, *happy*, *friendly*, *satisfied*; α = .86), anxious (*nervous*, *tense*, *uneasy*, *anxious*, *uncomfortable*; α = .86), negative self-directed (*annoyed at myself*, *disappointed with myself*, *angry at myself*; α = .79), and negative other-directed (*hostile*, *resentful*, *angry at others*, *upset at others*, *irritated with others*; α = .88).

Trust, cultural trust, and liking. Participants indicted how much trust (α = .92), cultural trust (α = .73), and liking (α = .84) they felt toward their partner using the scales in Table 1.

Manipulation checks, engagement items, and suspicion probes. After all dependent measures, manipulation check questions asked participants to indicate Devon's race and most negative behavior (if any). To assess engagement, two questions asked participants to report how distracted they felt during the study and how seriously they took the study, with instructions encouraging candor and emphasizing that their responses would not affect compensation. The next questions assessed participants' understanding of the study purpose and whether they were suspicious about any aspects of the study. A verbal funnel debriefing probed for suspicion.

Background measures. The social desirability and general affect toward Whites/European Americans and Blacks/African Americans measures from Studies 2a and 2b were assessed but not included as covariates in the reported analysis, as they did not affect the main dependent variables.

Results

Because the participants were nested in 60 dyads (both interracial and same-race), dyadic mixed-model analyses were used to control for non-independence and other dyad-level effects, as in Study 1. All predictors in the model were effects coded: participant race (-1 = minority, 1 = White), dyad race (-1 = interracial, 1 = same-race), closeness (-1 = control, 1 = IOS), and betrayal (-1 = no betrayal, 1 = betrayal). No gender differences emerged for the dependent measures, nor did gender significantly qualify any of the results reported.

Closeness induction. The first set of analyses confirmed that the closeness induction raised participants' self-reported closeness in the IOS condition (M = 3.39, SD = 1.35) relative to the control condition (M = 2.85, SD = 1.35), $t(58) = 2.03$, $p = .047$. Notably, participants rated the IOS-inducing and control tasks as comparably interesting (respective Ms = 5.02 & 4.68, SDs = 0.85

& 0.95), $t(58) = 1.88$, $p = .065$, difficult (respective Ms = 3.43 & 3.15, SDs = 1.10 & 0.97), $t(58) = 1.42$, $p = .161$, and enjoyable (respective Ms = 4.75 & 4.65, SDs = 0.97 & 1.03), $t(58) < 1$.

Cooperation. For purposes of analysis, participants' choices on the prisoner's dilemma game trials were scored 0 for defection and 1 for cooperation, and then average cooperation levels were computed, overall and for five bins of trials: two early-game bins, a mid-game bin, and two end-game bins. In the Lount et al. (2008) paradigm, only data from participants who cooperate on the initial trial should be analyzed, meaning that responses on the very first trial were invariant (as noted previously, the one dyad in which a participant initially defected was excluded from all analyses). Thus, early-game bin 1 comprised the second and third trials, and bin 2 comprised the fourth and fifth trials for each participant. End-game bins 4 and 5 respectively comprised the penultimate 2 trials and final 2 trials. All remaining intervening trials were included in the mid-game bin. (The mid-game bin contained 11 trials in the no-betrayal condition and 13 trials in the betrayal condition: Although all participants experienced 20 trials with ostensibly cooperative partner choices, those in the betrayal condition experienced an additional 2 betrayal trials at the start.) The early- and end-game bins are of particular interest because defection is likeliest immediately after a defection or prior to the end of an iterated prisoner's dilemma game (Lount et al., 2008).

Analysis of cooperation levels revealed a robust effect of ostensible partner betrayal, which appeared clearly in the early-game trials, faded in the mid-game trials, and vanished by the end-game trials. The main effect for betrayal condition was significant for overall cooperation, $t(48) = 6.50$, $p < .001$: Participants whose partners ostensibly betrayed them cooperated 89% of the time on average, whereas those with consistently cooperative partners cooperated 99% of the time. The effect of partner betrayal on cooperation was significant in bins 1, 2, and 3, respective ts(48) = 8.53, 6.20, and 2.91: Cooperation rates by bin were 0.49, 0.62, and 0.96 in the betrayal condition and 0.99, 0.99, 1.00 in the no-betrayal condition, respectively. No partner-betrayal effects emerged for bins 4 and 5,

respective *t*s(48) = 0.49 and 1.18, both *p*s > .20: Cooperation rates by bin were 0.99 and 0.93 in the betrayal condition and 1.00 and 0.97 in the no-betrayal condition, respectively. Contrary to predictions, no significant effects on cooperation—overall or in any bin of trials—emerged for closeness induction, participant race, or dyad race, all *p*s > .15.

Trust, cultural trust, and liking. Trust, cultural trust, and liking scores were submitted to mixed-model analysis (see Table 5 for means and *SD*s by condition). Participants whose partners ostensibly defected (vs. cooperated) reported less trust, $t(48) = 3.26$, $p = .002$, but equivalent cultural trust, $t(48) < 1$, and liking, $t(48) = 1.41$, $p = .166$. No other significant main effects emerged.

Table 5: *Mean (SD) Trust, Cultural Trust, and Liking by Closeness Condition, Partner Behavior, Participant (P) Race, and Dyad Race in Study 3*

Closeness condition	Partner behavior	P race	Dyad race	*n*	Trust	Cultural trust	Liking
Control	No betrayal	White	Same-race	4	5.15 (0.56)	3.88 (0.32)	5.79 (0.70)
			Interracial	8	4.76 (0.55)	3.56 (0.93)	5.48 (0.34)
		Black	Same-race	6	4.70 (0.93)	5.13 (0.77)	5.67 (0.94)
			Interracial	8	5.63 (0.98)	4.94 (0.98)	6.02 (0.65)
	Betrayal	White	Same-race	10	4.71 (0.53)	4.35 (1.23)	5.65 (0.74)
			Interracial	7	4.39 (0.55)	4.18 (0.43)	5.10 (0.93)
		Black	Same-race	4	4.94 (0.51)	4.31 (0.13)	6.17 (0.69)
			Interracial	7	3.56 (1.21)	3.50 (1.27)	4.57 (1.19)
IOS	No betrayal	White	Same-race	14	4.99 (0.93)	4.25 (1.07)	5.40 (1.03)
			Interracial	9	5.48 (0.50)	4.75 (0.56)	5.98 (0.56)
		Black	Same-race	6	5.07 (0.78)	4.67 (0.56)	5.67 (0.64)
			Interracial	9	4.92 (0.86)	4.53 (1.39)	5.67 (0.68)
	Betrayal	White	Same-race	10	4.62 (0.68)	4.28 (0.67)	5.45 (0.67)
			Interracial	6	4.81 (0.82)	4.67 (1.33)	5.86 (0.83)
		Black	Same-race	6	4.71 (0.68)	4.67 (0.61)	5.25 (1.06)
			Interracial	6	4.69 (1.35)	4.29 (1.34)	5.61 (0.99)

The hypothesized effect of interest is the four-way interaction of closeness induction, betrayal, participant race, and dyad race, which proved significant for trust, $t(98.1) = 2.22$, $p = .029$, but not cultural trust, $t(99.8) < 1$, or liking, $t(88.4) = 1.74$, $p = .086$. Further probing revealed that the closeness induction × betrayal × dyad race interaction effect on trust was significant among Black participants, $t(68.8) = 2.55$, $p = .013$, but not White participants, $t(74.2) < 1$. Among Black

participants, the betrayal × dyad race interaction effect on trust was significant in the control condition, $t(67.0) = 3.32$, $p = .001$, but not the IOS condition, $t(70.7) < 1$. Black participants who did the control tasks reported slightly higher trust levels in interracial and same-race dyads after their partner ostensibly cooperated, $t(69.2) = 2.05$, $p = .044$, but less trust in interracial than same-race dyads after a betrayal, $t(65.5) = 2.62$, $p = .011$ (see Figure 5).

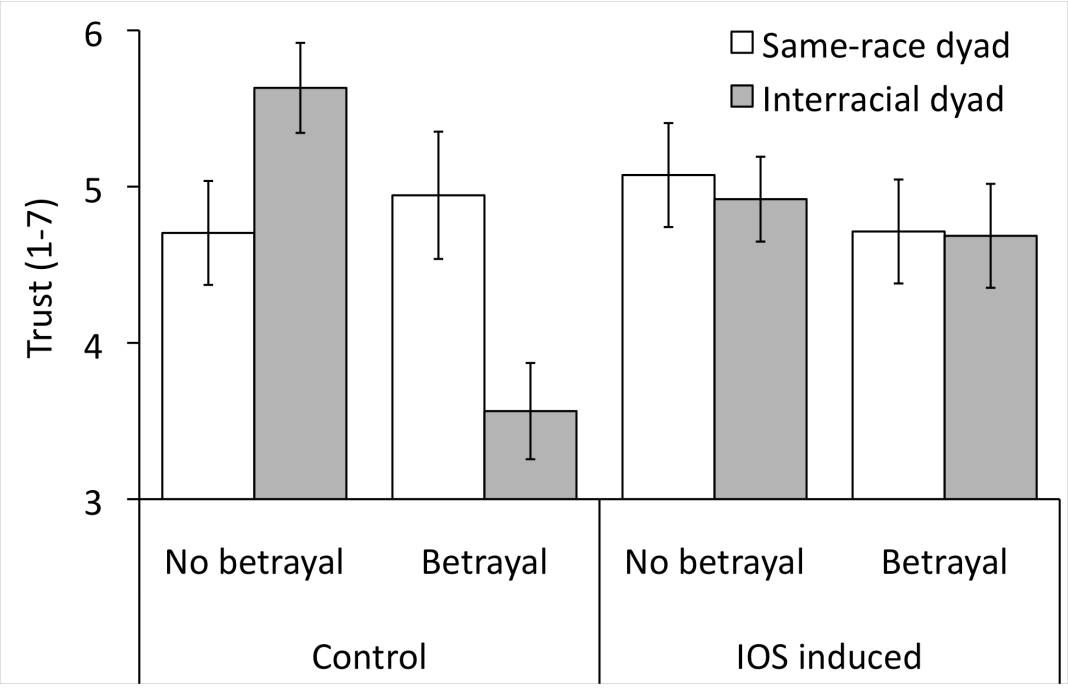


Figure 5. Mean (±1 *SE*) trust among Black participants by closeness condition, partner behavior, and dyad race in Study 3.

To further clarify the significant closeness induction × betrayal × dyad race interaction observed for Black participants' trust in their partners, additional tests of simple interactions and simple effects were conducted. Among Black participants, the closeness induction × betrayal interaction effect for trust was non-significant in same-race dyads, $t(48) < 1$, but significant in interracial dyads, $t(101.1) = 3.04$, $p = .003$. Black participants in interracial dyads reported trusting their White partners less following a betrayal (vs. cooperative partner behavior) in the control condition, $t(101.1) = 4.89$, $p < .001$, but this trust gap disappeared in the IOS condition, $t(101.1) < 1$.

Affect. Prior to analysis, univariate outliers on each of the affect measures were trimmed (not deleted) so as to deviate no more than 2.5 *SD* from the mean. This correction reduced positive skew for the anxiety, self-directed negativity, and (especially) other-directed negativity measures, as well as mild negative skew for positive affect, reducing the statistical leverage of extreme cases. Affect scores were submitted to mixed-model analysis (see Table 6 for means and *SD*s by condition). Participants whose partners ostensibly betrayed them (vs. consistently cooperated) reported feeling less positive affect, $t(48) = 3.57$, $p = .001$, more negative self-directed affect, $t(48) = 3.38$, $p = .001$, more negative other-directed affect, $t(48) = 5.78$, $p < .001$, and marginally more anxiety, $t(48) = 1.83$, $p = .074$. Participants in the IOS (vs. control) condition reported more positive affect, $t(48) = 2.30$, $p = .026$; no other main effects proved significant. No significant interaction effects emerged for positive affect or negative self-directed affect, so they are not discussed further.

Table 6: *Mean (SD) Affect by Closeness Condition, Partner Behavior, Participant (P) Race, and Dyad Race in Study 3*

Closeness condition	Partner behavior	P race	Dyad race	Positive affect	Negative self-directed affect	Anxiety	Negative other-directed affect
Control	No betrayal	White	Same-race	3.88 (0.25)	1.00 (0.00)	1.55 (0.68)	1.00 (0.00)
			Interracial	2.78 (0.57)	1.13 (0.35)	1.48 (0.52)	1.00 (0.00)
		Black	Same-race	3.79 (0.62)	1.00 (0.00)	1.70 (0.79)	1.00 (0.00)
			Interracial	3.63 (0.92)	1.04 (0.12)	1.45 (0.49)	1.00 (0.00)
	Betrayal	White	Same-race	3.13 (0.67)	1.23 (0.42)	1.70 (0.30)	1.34 (0.43)
			Interracial	3.14 (0.86)	1.14 (0.26)	2.00 (0.53)	1.58 (0.49)
		Black	Same-race	2.88 (0.66)	1.08 (0.17)	1.25 (0.10)	1.00 (0.00)
			Interracial	2.68 (0.99)	1.33 (0.38)	2.20 (0.90)	1.60 (0.53)
IOS	No betrayal	White	Same-race	3.68 (0.85)	1.08 (0.29)	2.03 (0.69)	1.04 (0.12)
			Interracial	3.97 (0.57)	1.00 (0.00)	1.31 (0.41)	1.04 (0.13)
		Black	Same-race	3.67 (0.93)	1.11 (0.17)	1.37 (0.51)	1.00 (0.00)
			Interracial	3.78 (0.96)	1.15 (0.34)	1.60 (0.50)	1.02 (0.07)
	Betrayal	White	Same-race	3.45 (0.89)	1.31 (0.45)	1.74 (0.83)	1.31 (0.47)
			Interracial	3.38 (0.52)	1.22 (0.27)	1.53 (0.45)	1.33 (0.41)
		Black	Same-race	3.13 (0.68)	1.51 (0.48)	1.93 (0.47)	1.38 (0.54)
			Interracial	3.42 (1.34)	1.11 (0.17)	1.53 (0.56)	1.13 (0.16)

The primary effect of interest is the interaction of closeness induction, betrayal, and dyad race, especially for Black participants. The three-way closeness induction × betrayal × dyad race interaction was significant for anxiety, $t(48) = 2.20$, $p = .032$, and was qualified by the hypothesized four-way closeness induction × betrayal × dyad race × participant race interaction, $t(102.7) = 2.22$, $p = .029$. Further probing revealed that the closeness induction × betrayal × dyad race interaction effect on anxiety was significant among Black participants, $t(86.5) = 3.00$, $p = .004$, but not White participants, $t(93.4) < 1$. Among Black participants, the betrayal × dyad race interaction effect on anxiety was significant in the control condition, $t(83.9) = 2.71$, $p = .008$, but not the IOS condition, $t(89.2) = 1.50$, $p = .137$. Black participants who did the control tasks reported comparably low anxiety levels in interracial and same-race dyads after their partner ostensibly cooperated, $t(87.1) < 1$, but more anxiety in interracial than same-race dyads after a betrayal, $t(81.4) = 2.84$, $p = .006$.

A very similar pattern of results emerged for negative other-directed affect. The closeness induction × betrayal × dyad race interaction proved significant for negative other-directed affect, $t(48) = 2.45$, $p = .018$. The hypothesized four-way closeness induction × betrayal × dyad race × participant race interaction failed to attain significance, $t(103.5) = 1.39$, $p = .166$, but analysis of simple interactions confirmed the *a priori* prediction that for negative other-directed affect, like anxiety, the closeness induction × betrayal × dyad race interaction was significant only among Black participants, $t(76.7) = 2.59$, $p = .012$, not White participants, $t(83.3) < 1$. Among Black participants, the betrayal × dyad race interaction effect on negative other-directed affect was significant in the control condition, $t(74.5) = 2.45$, $p = .017$, but not the IOS condition, $t(79.2) = 1.17$, $p = .245$. Black participants who did the control tasks reported comparably low negative other-directed affect levels in interracial and same-race dyads after their partner ostensibly cooperated, $t(77.3) < 1$, but more negative other-directed affect in interracial than same-race dyads after a betrayal, $t(72.4) = 3.24$, $p = .002$ (see Figure 6).

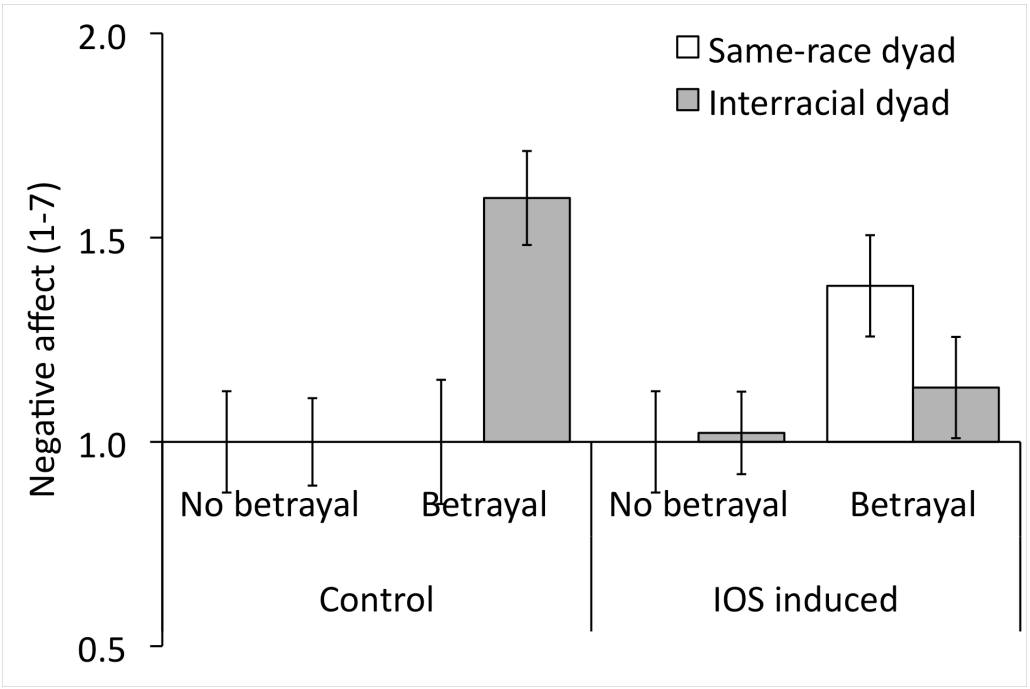


Figure 6. Mean (±1 *SE*) negative other-directed affect among Black participants by closeness condition, partner behavior, and dyad race in Study 3.

Additional simple interactions and simple effects tests probed the significant closeness induction × betrayal × dyad race interaction evident among Black participants for both anxiety and negative other-directed affect, yielding results largely convergent with the trust findings. For Black participants in interracial dyads, the closeness induction × betrayal interaction proved marginally significant for anxiety, $t(100.5) = 1.89$, $p = .061$, and significant for negative other-directed affect, $t(104.0) = 2.17$, $p = .032$. Following a betrayal (vs. cooperative partner behavior) in the control condition, Black participants in interracial dyads reported more anxiety, $t(100.5) = 2.48$, $p = .015$, and negative other-directed affect, $t(104.0) = 3.81$, $p < .001$; those in the IOS condition, however, reported comparable levels of anxiety, $t(100.5) < 1$, and negative other-directed affect, $t(104.0) < 1$. Among Black participants in same-race dyads, the closeness induction × betrayal interaction was non-significant for negative other-directed affect, $t(48) = 1.52$, $p = 0.134$, but significant for anxiety, $t(48) = 2.35$, $p = .023$. Simple effects tests revealed that Black participants in same-race dyads did not

significantly differ in their anxiety levels following a betrayal versus cooperation in either the control, $t(48) = 1.39$, $p = .170$, or IOS condition, $t(48) = 1.96$, $p = .056$.

Mediation. Subsequent mediation analyses tested whether affect might partially account for Black participants' lower levels of trust in their White (vs. Black) partners after completing control (vs. IOS-inducing) tasks together and experiencing a simulated betrayal (vs. cooperation). The meditational procedures outlined by Baron and Kenny (1986) were extended by using the *INDIRECT* macro (Preacher & Hayes, 2008) for SPSS with 1000 bias-corrected bootstrap resamples. Bootstrapping gives a more robust test of the indirect path through a proposed mediator while maintaining reasonable control of Type I error.[9]

Mediation analyses focused on anxiety and negative other-directed affect because, as noted previously, the key effects of interest — the closeness induction × betrayal × dyad race and closeness induction × betrayal × dyad race × participant race interactions — were not significant for positive affect and negative self-directed affect, ruling them out as potential mediators. Regressing trust on anxiety revealed a significant relationship between these two variables, $t(112.2) = 2.06$, $p = .042$, but this relationship became non-significant in a simultaneous regression with the other manipulated variables (closeness induction, betrayal, dyad race, participant race, and their interactions) as predictors, $t(91.3) < 1$. Thus, anxiety did not significantly mediate this race-based trust gap. A comparable regression confirmed that negative other-directed affect significantly predicted trust

[9] The commonly used Sobel (1982) *Z* test is based on the standard normal distribution, yet products of path coefficients typically have a skewed distribution. Bootstrapping, a nonparametric resampling procedure, does not impose a normality assumption and instead empirically approximates the sampling distribution of the indirect effect, testing whether its 95% confidence interval includes zero. Relative to the *Z* test, bias-corrected bootstrapping provides more powerful tests of indirect effects in both single-level (MacKinnon, Lockwood, & Williams, 2004) and multi-level models (Pituch, Stapleton & Kang, 2006). The INDIRECT macro randomly resamples individual cases, not dyads, which is acceptable for multi-level data only when intraclass correlations approach zero (see Pituch et al., 2006). Tests of nonindependence between dyad members were not significant for either trust or negative other-directed affect, respective partial ICCs (controlling manipulated variables) = 0.09 and -0.09, *p*s = .505 and .528.

both alone, $t(114.1) = 4.07$, $p < .001$, and when controlling for effects of the manipulated variables, $t(98.9) = 2.04$, $p = .044$. Among Black participants who did the control tasks then experienced an ostensible partner betrayal, the path from dyad race (-1 = interracial, +1 = same-race) to trust was significant at first, $t(65.5) = 2.62$, $p = .011$, but dropped to marginal significance when including negative other-directed affect as a mediator, $t(68.9) = 1.95$, $p = .055$. Although the indirect path from dyad race to trust through negative other-directed affect was marginally significant according to the Sobel test, $Z = 1.73$, $p = .084$, its 95% C.I. [0.01, 0.44] excluded zero, indicating a significant mediating role of negative other-directed affect (see Figure 7).

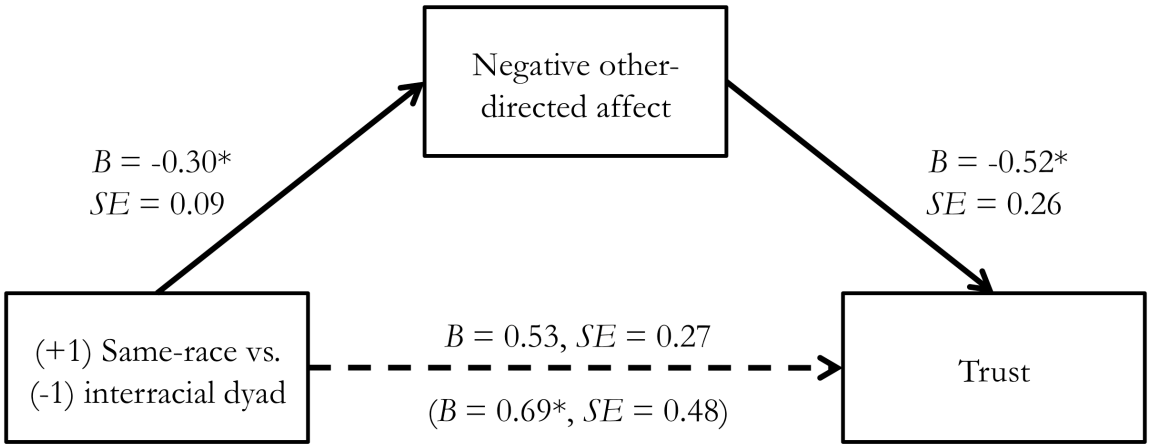


Figure 7. Model of negative other-directed affect mediating dyad race effect on trust among Black participants in the control condition after a partner's betrayal in Study 3. $^{*}p < .05$

Discussion

Study 3 extended the prior studies by experimentally manipulating both betrayal and closeness, using established methods for instantiating betrayal (partner defection in a mixed-motive game) and inducing closeness (the "fast friends" procedure). The absence of race- or closeness-based effects for the behavioral cooperation data may appear surprising, and the lack of a significant betrayal effect on endgame cooperation contrasts with the findings of Lount et al. (2008). The very high levels of cooperation—over 90% on the final trial and over 95% on the other 9 endgame

trials—in this study exceed those observed previously, suggesting a ceiling effect. In prior work, participants who experienced an immediate partner defection cooperated 69% and 61% of the time (in Studies 1 and 2 of Lount et al., 2008, respectively), whereas those in the control condition cooperated 88% and 83% of the time. This present study differed from the original Lount et al. (2008) experiments primarily in that participants interacted with each other one-on-one and face-to-face for 15 min. prior to the game, read a shorter tutorial on cooperation (see Appendix F for both versions), and encountered a payoff matrix in which the absolute dollar amounts were somewhat lower but also less correspondent. Because the changes to the cooperation tutorial and payoff matrix were relatively modest and involved countervailing factors (e.g., lower payoffs may reduce the appeal of defecting but less correspondent payoffs incentivize defection), it seems most plausible that the interaction prior to the game increased cooperation. Engaging in a live social interaction with another undergraduate before completing a mixed-motive game together may heighten normative pressures toward cooperation. Indeed, friends tend to display higher-than-usual levels of cooperation in mixed-motive games (Oskamp & Perlman, 1966). The race- and closeness-based null effects for cooperation appear inconclusive in light of the overall ceiling effect.

In a conceptual replication of Studies 2a and 2b, Black participants in the control condition reported trusting White study partners less than Black partners following a betrayal. Notably, in Study 3 the race-based trust discrepancy emerged for overall trust, not cultural trust, which showed no significant race-based effects. In contrast to Studies 2a and 2b, in which the target person displayed numerous clearly trustworthy behaviors (e.g., involving helping, support, responsiveness, and self-sacrifice) that enabled high trust regardless of race even after a betrayal, partners' ostensible initial defection in Study 3 gave a signal of untrustworthiness strong enough that their partner's cooperation on subsequent trials evidently did not suffice to restore basic trust for Black participants with White partners in the control condition. In the IOS condition, by contrast, Black participants

who experienced a partner betrayal reported comparable levels of trust in White versus Black partners (as well as significantly more trust than Black participants whose White partners ostensibly betrayed them in the control condition). This finding implies that inducing interpersonal closeness—even between strangers—can attenuate race-based trust gaps in the aftermath of betrayals or other negative interpersonal interactions. Again, the absence of significant effects for liking is consistent with the claim that intergroup trust, more so than liking or harmony, is fragile and damaged by negative interpersonal interactions, particularly if a member of a higher-status group has previously betrayed the trust of someone in a lower-status, historically victimized group.

Finally, Study 3 provides some insight into one mechanism by which trust may be damaged and not subsequently repaired following an interpersonal betrayal. Both anxiety and negative other-directed affect increased for Black participants following a betrayal by a White (vs. Black) partner, and negative other-directed affect partially mediated the race-based discrepancy in trust. For individuals who belong to a group that has direct past and present experience with discrimination, betrayal by a high-status individual (vs. an ingroup member) evokes more anger, hostility, and resentment, which endure after the partner's subsequent cooperation and may hamper trust repair.

CHAPTER 6: GENERAL DISCUSSION

Three unifying themes emerge from Studies 1 through 3. First, consistent with sociological, economic, and political science reports of racial asymmetries in generalized trust and group-based interracial trust, Black participants in each study reported at least somewhat more interpersonal trust in an ingroup versus an outgroup individual. In contrast, White participants displayed a comparable trust bias favoring ingroup over outgroup individuals only when initially describing a real-world friendship (Study 1, Time 1). This race-based asymmetry in trust extended to outgroup versus ingroup individuals comes with a critically important qualification: Black participants did not report more trust for Black than White individuals in all instances, as work on Black-White gaps in generalized and group-based trust might imply. On the contrary, Black participants, much like their White counterparts, reported comparable levels of interpersonal trust in Black and White individuals during a later stage of friendship development (Study 1, Time 2), when imagining interactions with a friend who displayed uniformly trustworthy behaviors (Studies 2a & 2b), and when experimentally led to experience feelings of closeness and friendship to a study partner who then displayed either cooperative or betraying behavior (Study 3). These results converge with prior findings that Black participants playing a mixed-motive behavioral trust game with Whites were not significantly less trusting, and in fact tended to be more trustworthy, relative to Whites (B. Simpson et al., 2007). In sum, although this research documents race-based interpersonal trust asymmetries between Blacks and Whites in interracial and same-race interactions that contain betrayals (particularly in Studies 2a, 2b, & 3), such asymmetries are not necessarily invariant or inevitable across all intergroup contexts.

The second major theme of this work involves the tight coupling between betrayals and trust, but not liking, within established, imagined, and experimentally induced interracial friendships. These studies number among relatively few others (e.g., Reis et al., 2010; Righetti & Finkenauer, 2011; Tam et al., 2009) that consistently distinguish trust from liking (or evaluation) in an effort to

isolate trust-related phenomena from effects of interpersonal interaction or relationship overall quality. In addition, the present work focuses on betrayals, operationalized variously as a friend "failing to follow through" (Study 1), deceptively breaking a promise for selfish gain (Studies 2a & 2b), or defecting in the prisoner's dilemma game (Study 3), rather than negative interaction incidents more generally. In matters of trust, betrayals differ from other kinds of negative incidents that researchers have simulated in intergroup settings (e.g., rude or cold behaviors, racist comments; Greenberg & Pyszczynski, 1985; Kawakami, Dunn, Karmali, & Dovidio, 2009; Murphy, Richeson, Shelton, Rheinschmidt, & Bergsieker, 2012; Operario & Fiske, 2001). Little work on intergroup interactions examines betrayals, though one study found that low-power group members who receive fewer resources than expected from members of a high-power group tend to retaliate with aggression (Hawi, Saguy, Dovidio, & Tropp, 2012). Thus, the present studies make a novel contribution to the existing trust literature by demonstrating that trust, more so than liking, (a) predicted the likelihood than individuals would (not) experience frequent friend betrayal behaviors in the following year (Study 1) and (b) decreased after simulated betrayals (Studies 2a, 2b, & 3).

A third noteworthy theme relates to the role of interpersonal closeness or IOS in regulating the interpersonal trust that Whites and Blacks extend to ingroup and outgroup individuals. Across all studies, higher levels of closeness reduced race-based disparities in trust. Black and White participants reported comparable trust levels for same- and different-race friends toward whom they felt equivalent interpersonal closeness in established friendships (Study 1), high interpersonal closeness in imagined friendships (Studies 2a & 2b), and induced closeness in experimentally created friendships (Study 3). These results are encouraging because they suggest a potential mechanism for buffering trust against betrayals in interracial contexts, enabling interracial relationships to flourish. These findings converge with recent work underscoring intergroup benefits of including outgroup individuals in the self (Cwir, 2011) and maintaining intergroup friendships (Davies et al., 2011).

Strengths and Limitations

In addition to these empirical contributions, these studies display several methodological strengths. First, this multimethod program of research examines same-race and interracial trust in three relatively distinctive contexts—established friendships, imagined friendships, and experimentally induced friendships—using both correlational and experimental approaches. For instance, this work investigates a key antecedent of trust—subjective interpersonal closeness or IOS—by variously measuring and manipulating it. In contrast to most existing studies of intergroup friendships (for a review, see Davies et al., 2011), Study 1 adopts a relational approach by incorporating the perspectives of both individuals involved in a given friendship (see Shelton & Richeson, 2006b). This study also addresses the need voiced by other researchers (Turner & Feddes, 2011) for studies of friendship over time that extend beyond a few weeks or a single semester. All three studies assess perceptions of minority participants, in some instances (Studies 2a, 2b, & 3) placing emphasis on their potentially distinctive perspectives and experiences. Despite the difficulties inherent in simulating a symmetric interpersonal betrayal between naïve participants in the course of an interaction, no confederates were used, allowing interpersonal processes in which individuals mutually influence one another to unfold more naturally.

Notwithstanding the aforementioned strengths, the present work has several noteworthy limitations. For example, although Studies 2a, 2b, and 3 produced very similar race-based patterns of results for trust, the specific trust measures that detected significant effects varied across studies, with the hypothesized effects emerging only for cultural trust in Studies 2a and 2b, and for overall trust in Study 3. This discrepancy would be readily interpretable if the trust violation in Studies 2a and 2b (see Appendix D) contained racial or cultural content, but this episode was intended to reflect a straightforward, race-neutral betrayal. In contrast to the prisoner's dilemma game's two binary alternatives—purely selfish defection versus mutually beneficial cooperation—in Study 3, the

betrayal scenario in Studies 2a and 2b afforded substantially more latitude for interpretation. Given this attributional ambiguity, perhaps participants invoked differing racial or cultural norms related to punctuality or spontaneity when interpreting the scenario.

The evidence for race-based asymmetries in trust also varies across studies. Although Studies 2a, 2b, and 3 show a consistent pattern of Black participants (but not White participants) trusting ingroup individuals more than outgroup individuals following a trust violation, no significant race-based asymmetries emerge between Black and White participants describing same-race and interracial friendships in Study 1. Differing theoretical assumptions influence whether it seems more surprising that Whites did not show a pro-ingroup trust bias in Studies 2a, 2b, and 3, or that Blacks did not display a significantly larger pro-ingroup trust bias than Whites in Study 1. The experimental studies were designed to contain a high ratio of positive to negative behaviors even in the betrayal conditions (5:1 scenarios in Studies 2a & 2b; 20:2 trials in Study 3), so White participants' tendency not to report greater trust in White than Black partners may stem from the overall positivity of these encounters, coupled with lack of vigilance to negative incidents. (Conceivably, political correctness could have led White participants to inflate their reported trust in Blacks, but the inclusion of social desirability as a covariate should render this explanation less probable.) Regarding the Study 1 findings of a symmetric pro-ingroup trust for Black and White participants—mean trust levels were descriptively but not significantly lower for Blacks than Whites in interracial friendships—it bears noting that (a) there may be a lower bound on how little one trusts another person before he or she ceases to be considered a friend, such that this person would not be nominated as friend for study purposes, and that (b) Black students at a predominantly White institution likely have daily contact with more White students than vice versa, such that they may have a broader array of potential outgroup individuals to nominate. Low numbers of Black participants (*n*s = 48 & 35 at Times 1 & 2) also reduced this study's power to reliably detect lower trust levels for Blacks relative to Whites.

A third limitation stems from the failure to observe converging behavioral evidence of trust in Study 3. (Although Study 1 provides some indication of behavioral consequences of trust in the significant association of initial trust with friend betrayal behaviors reported a year later, the latter measure was assessed through self-report and the correlational design of Study 1 precludes strong causal claims.) As discussed previously, the null effects for race- and closeness-related variables as predictors of behavior in Study 3 may well be attributable to the cooperative norm created by an extended and enjoyable social interaction immediately preceding the prisoner's dilemma game. The prisoner's dilemma game, rather than another mixed-motive dilemma (e.g., the trust game; Berg et al., 1995), was incorporated in Study 3 because the prisoner's dilemma game proved effective in prior trust research (Lount et al., 2008), because it preserves symmetry in roles and simultaneity in decisions for both members of a dyad—making it more suitable for planned extensions of this work—and because the trust game's use of incremental rather than binary outcomes led participants to become suspicious in prior research using a preprogrammed set of partner responses (Rotella, Richeson, & Chiao, 2011). That said, the binary nature of participants' alternatives on each prisoner's dilemma trial, namely, between obvious selfishness or cooperation, may have reduced its sensitivity to detect subtle variations in behavioral trust in the present research.

Implications and Future Directions

Assessing the contribution of this work entails considering whether these effects generalize beyond White and Black American college students to other populations, which depends upon the mechanism driving the interracial trust asymmetries. In addition, different conceptualizations of interpersonal closeness or IOS may carry different implications for intergroup trust. Future work should also examine the consequences of trust betrayals and resulting decreases in trust on subsequent intergroup interaction dynamics as they unfold over time.

Generalizability

The experiences and perspectives of White and Black Americans diverge in numerous respects (see Shelton & Richeson, 2006b), so trust asymmetries between them may be multiply or over-determined. However, these findings converge with other lines of work on trust asymmetries using populations, suggesting that more general processes may be at play. Two possible mechanisms include salience of past victimization and status differences. At an individual level, past criminal or economic victimization experiences lead individuals to be less trusting of novel others (Effron & Miller, 2011; Ross et al., 2002); likewise, recalling past group victimization may foster wariness of outgroup members. Studies involving Jewish or politically conservative participants demonstrated that perceptions that one's social group is currently or has been victimized in society influences intergroup trust (Rotella et al., 2011). Following a high or low perceived group victimhood prime, participants played an economic trust game ostensibly with other ingroup or outgroup individuals. In all studies higher perceived group victimhood induced participants to extend more behavioral trust (i.e., monetary allocations) to ingroup than outgroup others, an effect driven by enhanced ingroup trust. Black participants in Studies 2a, 2b, and 3 may have responded more negatively to betrayals by White than Black partners due to recalling Whites' history of exploiting Blacks.

Some controversy surrounds the relationship between socioeconomic status and tendency to trust others (e.g., see Piff, Kraus, Côté, Cheng, & Keltner, 2010; but also Ross et al., 2001; Smith, 2010; Uslaner, 2002). Recent work, however, shows a strong relationship between experimentally induced high status and higher behavioral (economic) trust in others (Lount & Pettit, 2012). Parallel effects emerged for self-reported status, and increased belief that others have positive intentions toward one's self accounted for the relationship between higher status and greater trust. Given that poor Black individuals trust high-status partners less than fellow low-status individuals (Hall, 2008), these studies suggest that status gaps may give rise to the asymmetries observed in the present work.

Closeness Considerations

Although the current results provide encouraging evidence that subjective closeness or IOS may improve intergroup relations, several caveats may apply. For example, the direction of the connection may matter: When Whites contemplate how a Black individual is like them, as opposed to how they are like him or her, they subsequently report less empathy for that person's group (Nadolny et al., 2012). If interconnectedness with an outgroup other involves merely projecting the self onto the other, effectively assimilating the him or her into the superordinate group, this process clashes with the established preferences of minorities for expressing a dual identity and having their subgroup membership valued by majority group members (see Dovidio et al., 2008). Notably, the IOS scale, which is used in all of the present studies, permits a range of interpretations, and prior work has demonstrated that lay construals of it are bidirectional. Participants more often see the scale as a measure of connection than mere similarity, and they perceive elements of the other expanding to include the self, as well as the self expanding to include the other (Aron et al., 1992). Future work should address the directionality of perceived closeness as a contributing factor to intergroup trust, for instance by assessing individuals' perceptions of integration between own and others' interests and motives.

Dyadic Dynamics

Leading models of trust in interpersonal relationships agree that disruptions to trust can cause a downward spiral of interpersonal phenomena (Holmes & Rempel, 1989; Murray & Holmes, 2009; J. Simpson, 2007b; Wieselquist et al., 1999). In the context of interracial interactions, after minorities have experienced a betrayal by a White person, they may shift to a state of heightened vigilance for additional cues of untrustworthiness (see Campbell et al., 2010). Efforts to continually monitor the behavior of the White person may prove cognitively taxing (see Shelton & Richeson, 2006b), and the negative other-directed affect aroused by betrayals may undermine their desire to

continue engaging in interactions with this person. In the language of the Murray and Holmes (2009), Black participants may withhold commitment or even withdraw from the relationship altogether. Particularly if the White person who initiated the transgression did so unintentionally or attempted to restore goodwill by subsequent favorable actions (as in Studies 2a, 2b, & 3), he or she might be confused by the Black person's continued distrust.

Despite widespread assumptions about the adaptiveness of trust (see Ross et al., 2001), important exceptions may exist. Vigilance to subtle cues of prejudice, even if these signals take the form of relatively minor interpersonal betrayals, may protect minorities from becoming vulnerable to potentially harmful discrimination. That said, because interpersonal interactions are fraught with "noise," namely discrepancies between the outcomes one person intends to produce and those that the other person experiences (Van Lange et al. 2002), construing benign discrepancies as intentional trust betrayals may unnecessarily damage relationships (Rusbult & Van Lange, 2003), especially if one or both parties is primed to detect negative behaviors. In addition, because feeling betrayed by another person fosters reluctance to trust other individuals in the future (Effron & Miller, 2011), negative noise could have cascading negative effects that extend to other outgroup individuals. Future work should examine whether inducing closeness enables both Black and White individuals to more accurately decipher each other's intentions and maintain trust and cooperation in the face of disruption during real-world interactions unfolding over time.

REFERENCES

Abelson, R. P., Dasgupta, N., Park, J., & Banaji, M. R. (1998). Perceptions of the collective other. *Personality and Social Psychology Review, 2*, 243-250.

Aboud, F. E., Mendelson, M. J., & Purdy, K. T. (2003). Cross-race peer relations and friendship quality. *International Journal of Behavioral Development, 27*, 165 – 173.

Agnew, C. R., Van Lange, P. A. M., Rusbult, C. E., & Langston, C. A. (1998). Cognitive interdependence: Commitment and the mental representation of close relationships. *Journal of Personality and Social Psychology, 74*, 939–54.

Alesina, A., & La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics, 85*, 207–234.

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Aron, A., & Aron, E. N. (1986). *Love as the expansion of self: Understanding attraction and satisfaction*. New York, NY: Hemisphere.

Aron, A., Aron. E. N., & Norman, C. (2001). The self expansion model of motivation and cognition in close relationships and beyond. In M. Clark & G. Fletcher (Eds.), *Blackwell handbook of social psychology: Interpersonal processes* (pp. 478 – 501). Malden, MA: Blackwell.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*, 596-612.

Aron, A., Eberhardt, J. L., Davies, K., Wright, S. C., & Bergsieker, H. B. (2006). *A social-psychological intervention to improve community relations with police: Initial results.* Manuscript in preparation.

Aron, A. McLaughlin-Volpe, T., Mashek, D., Lewandowski, G., Wright, S. C., Aron, E. N. (2004). Including others in the self. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 15, pp. 101-132). Hove, UK: Psychology Press.

Aron, A., Melinat, E., Aron, E. N., Vallone, R., & Bator, R. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin, 23,* 363 – 377.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

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

Berscheid, E., Snyder, M., & Omoto, A. M., (1989). The Relationship Closeness Inventory: Assessing the closeness of interpersonal relationships. *Journal of Personality and Social Psychology, 57*, 792-807.

Bourdieu, P. (1983). Forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York, NY: Greenwood.

Bottom, W. P., Gibson, K., Daniels, S., & Murnighan, J. K. (2002). When talk is not cheap: Substantive penance and expressions of intent in the reestablishment of cooperation. *Organizational Science, 13*, 497-513.

Brewer, M. B. (1997). The social psychology of intergroup relations: Can research inform practice? *Journal of Social Issues, 53*, 197-221.

Brewer, M. B. (2001). *Ingroup identification and intergroup contact: When does ingroup love become outgroup hate*? In R. Ashmore, L. Jussim, & D. Wilder (Eds.). Social identity, intergroup conflict, and conflict reduction (pp. 2-41). New York: Oxford University Press.

Brown, E., & Ferris, J. M. (2007). Social capital and philanthropy: An analysis of the impact of social capital on individual giving and volunteering. *Nonprofit and Voluntary Sector Quarterly, 36*, 85-99.

Burns, J. (2006). Racial stereotypes, stigma and trust in post-apartheid South Africa. *Economic Modelling, 23*, 805-821.

Campbell, L., Simpson, J. A., Boldry, J., & Kashy, D. A. (2005). Perceptions of conflict and support in romantic relationships: The role of attachment anxiety. *Journal of Personality and Social Psychology, 88,* 510–531.

Campbell, L., Simpson, J. A., Boldry, J. G., & Rubin, H. (2010). Trust, variability in relationship evaluations, and relationship processes, *Journal of Personality and Social Psychology, 99*, 14-31.

Cehajic, S., Brown, R., & Castano, E. (2008). Forgive and forget? Antecedents and consequences of intergroup forgiveness in Bosnia and Herzegovina. *Political Psychology, 29*, 351-367.

Clark, M. S., Graham, S. M., Williams, E., & Lemay, E. P. (2008). Understanding relational focus of attention may help us understand relational phenomena. In J. P. Forgas & J. Fitness (Eds.) *Social relationships: Cognitive, affective, and motivational Processes* (pp. 131-146). New York: Psychology Press.

Collins, N. L. (1996). Working models of attachment: Implications for explanation, emotion, and behavior. *Journal of Personality and Social Psychology, 71,* 810–832.

Collins, N. L., & Feeney, B. C. (2000). A safe haven: An attachment theory perspective on support seeking and caregiving in intimate relationships. *Journal of Personality and Social Psychology, 78*, 1053-1073.

Collins, N. L., & Feeney, B. C. (2004). Working models of attachment shape perceptions of social support: Evidence from experimental and observational studies. *Journal of Personality and Social Psychology, 87*, 363–383.

Conley, T. D., Moors, A. C., Ziegler, A. & Feltner, M. R. (2011). Trust and satisfaction in adult child–mother (and other) relationships. *Basic and Applied Social Psychology, 33*, 239-254.

Cook, K. S., & Cooper, R. M. (2003). Experimental studies of cooperation, trust, and social exchange. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 209-244). New York: Russell Sage Foundation.

Cook, K. S., Hardin, R., & Levi, M. (2005). *Cooperation without trust?* New York: Russell Sage Foundation.

Cottrell, C. A., Neuberg, S. L., & Li, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology, 92,* 208–231.

Crawford, M. (1977). What is a friend? *New Society, 20*, 116-117.

Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology, 60*, 218-228.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Counseling Psychology, 24*, 349-354.

Cwir, D. (2011). *The power of social connections: Feelings of connectedness result in sharing goals, emotions, and intergroup empathy*. Unpublished doctoral dissertation. University of Waterloo, Waterloo, Canada.

Czopp, A. M. (2007). When is a compliment not a compliment? Evaluating expressions of positive stereotypes. *Journal of Experimental Social Psychology, 44*, 413-420.

Davies, K. M. (2009). *Identifying key themes in cross-group friendship formation.* Unpublished doctoral dissertation. Stony Brook University, Stony Brook, NY.

Davies, K., Tropp, L. R., Aron, A., Pettigrew, T. F., & Wright, S. C. (2011). Cross-group friendships and intergroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 15*, 332-351.

Davis, J. A., Smith, T. W., Marsden, P. V. (2006). *General social surveys, 1972–2006*. Chicago: National Opinion Research Center, producer. Storrs, CT: Roper Center for Public Opinion Research, distributor.

de Dreu, C. K. W. (2010). Social value orientation moderates ingroup love but not outgroup hate in competitive intergroup conflict. *Group Processes and Intergroup Relations, 13*, 701-713.

Derlega, V. J., Metts, S., Petronio, S., & Margulis, S. T. (1993). Self-*disclosure*. Newbury Park, CA: Sage Publications.

Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution, 2*, 265-279.

Deutsch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. New Haven, CT: Yale University Press.

Dovidio, J. F. (2001). On the nature of contemporary prejudice: The third wave. *Journal of Social Issues, 57*, 829-849.

Dovidio, J. F. (2008). *Racial bias and interracial interaction: Anxiety and intergroup inferences*. Invited address at Princeton University, Princeton, NJ.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 1 – 51). San Diego, CA: Academic Press.

Dovidio, J. F., Gaertner, S. L., John, M.-S., Halabi, S., Saguy, T., Pearson, A. R., et al. (2008). Majority and minority perspectives in intergroup relations: The role of contact, group representations, threat, and trust in intergroup conflict and reconciliation. In A. Nadler, T. E. Malloy & J. D. Fischer (Eds.), *The social psychology of intergroup reconciliation* (pp. 227-253). Oxford: Oxford University Press.

Dovidio, J. F., Gaertner, S. L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology, 8*, 88-102.

Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology, 70,* 1327–1343.

Dyer, J. H., & Chu, W. (2003). The role of trustworthiness in reducing transaction costs and improving performance: Empirical evidence from the United States, Japan, and Korea. *Organization Science, 14*, 57-68.

Eckel, C. C., & Wilson, R. K. (2003). The human face of game theory: Trust and reciprocity in sequential games. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 245-274). New York: Russell Sage Foundation.

Effron, D. A., & Miller, D. T. (2011). Reducing exposure to trust-related risks to avoid self-blame. *Personality and Social Psychology Bulletin, 37*, 181-192.

Ensari, N., & Miller, N. (2002). The out-group must not be so bad after all: The effects of disclosure, typicality, and salience on intergroup bias. *Journal of Personality and Social Psychology, 83*, 313–329.

Fehr, B. (1996). Friendship processes. Thousand Oaks, CA: Sage.

Ferrin, D. L., Bligh, M. C., & Kohles, J. C. (2007). Can I trust you to trust me? A theory of trust, monitoring, and cooperation in interpersonal and intergroup relationships. *Group and Organization Management, 32*, 465-499.

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology, 82*, 956-974.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurement, 53*, 417-424.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist, 48*, 621-628.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social perception: Warmth and competence. *Trends in Cognitive Science, 11,* 77-83.

Frey, F. E., & Tropp, L. R. (2006). Being seen as individuals versus as group members: Extending research on metaperception to intergroup contexts. *Personality and Social Psychology Review, 10*, 265-280.

Gable, S. L., & Reis, H. T. (2001). Appetitive and aversive social interaction. In J. Harvey & A. Wenzel (Eds.), *Close romantic relationships: Maintenance and enhancement* (pp. 169–194). Mahwah, NJ: Erlbaum.

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The Common Ingroup Identity Model.* Philadelphia, PA: The Psychology Press.

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

Gordon, H. S., Street, Jr., R. L., Sharf, B. F., Kelly, P. A., & Souchek, J. (2006). Racial differences in trust and lung cancer patients' perceptions of physician communication. *Journal of Clinical Oncology, 24*, 904-909.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*, 499-510.

Greenberg, J., & Pyszczynski, T. (1985). The effect of an overheard ethnic slur on evaluations of the target: How to spread a social disease. *Journal of Experimental Social Psychology, 21*, 61-72.

Gulati, R., & Sytch, M. (2008). Does familiarity breed trust? Revisiting the antecedents of trust. *Managerial and Decision Economics, 29*, 165-190.

Hall, C. C. (2008). *Decisions under poverty: A behavioral perspective on the decision making of the poor.* Unpublished doctoral dissertation. Princeton University, Princeton, NJ.

Hallinan, M. T., & Williams, R. A. (1987). The stability of students' interracial friendships. *American Sociological Review, 52*, 653-664.

Hardin, R. (2002). *Trust and trustworthiness.* New York: Russell Sage Foundation.

Hardin, R. (2003). Gaming trust. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research.* (pp. 80-101). New York: Russell Sage Foundation.

Hawi, D., Saguy, T., Dovidio, J. F., & Tropp, L. R. (2012). *When positive expectations are not met: How power dynamics affect disadvantaged group members' reactions following different forms of intergroup contact.* Poster presented at the 12th Society for Personality and Social Psychology Meeting in San Diego, CA.

Hewstone, M., Kenworthy, J. B., Cairns, E., Tausch, N., Hughes, J., Tam, T., et al. (2008). Stepping stones to reconciliation in Northern Ireland: Intergroup contact, forgiveness, and trust. In A. Nadler, T. E. Malloy & J. D. Fischer (Eds.), *The social psychology of intergroup reconciliation* (pp. 199-226). Oxford: Oxford University Press.

Holmes, J. G. (1981). The exchange process in close relationships: Microbehavior and macromotives. In M. J. Lerner & S. C. Lerner (Eds.), *The justice motive in social behavior* (pp. 261-284). New York, NY: Plenum.

Holmes, J. G., & Rempel, J. K. (1989). Trust in close relationships. In C. Hendrick (Ed.), *Close relationships* (pp. 187-220). Newbury Park, CA: Sage.

Insko, C. A., & Schopler, J. (1998). Differential distrust of groups and individuals. In C. Sedikides, J. Schopler & C. A. Insko (Eds.), *Intergroup cognition and intergroup behavior: Applied social research* (pp. 75-107). Hillsdale, NJ: Erlbaum.

Insko, C. A., Schopler, J., Hoyle, R. H., Dardis, G. J., & Graetz, K. A. (1990). Individual-group discontinuity as a function of fear and greed. *Journal of Personality and Social Psychology, 58*, 95-111.

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

Jost, J. T., & Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology, 36*, 209–232.

Katz, D., & Braly, K. (1933). Racial stereotypes in one hundred college students. *Journal of Abnormal and Social Psychology*, 28, 280-290.

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science, 323*, 276-278.

Kelley, H. H. (1979). *Personal relationships: Their structures and processes.* Hillsdale, NJ: Erlbaum.

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An atlas of interpersonal situations*. New York: Cambridge University Press.

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations*. New York: Wiley.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.

Kiecolt, K. J, Morris, E., & Toussaint, J. (2006). *Race and social trust*. Paper presented at American Sociological Association, Montreal, Canada.

Klik, K. A., & Butz, D. A. (2012). *The role of openness in interracial and same-race interactions*. Poster presented at the 12th Society for Personality and Social Psychology Meeting in San Diego, CA.

Koleva, S. P. (2012). *Birds of a moral feather: The role of morality in romantic attraction and relationship satisfaction*. Poster presented at the 12th Society for Personality and Social Psychology Meeting in San Diego, CA.

Komorita, S. S., & Mechling, J. (1967). Betrayal and reconciliation in a two-person game. Journal of Personality and Social Psychology, 6, 349-353.

Kramer, R. M. (2004). Collective paranoia: Distrust between social groups. In R. Hardin (Ed.), *Distrust* (pp. 136-166). New York: Russell Sage Foundation.

Kramer, R. M., & Carnevale., P. J. (2001). Trust and intergroup negotiation. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 431-450). Malden, MA: Blackwell.

Larzelere, R. E., & Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family, 42*, 595-604.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability in the positive evaluation of in-groups). *Journal of Personality and Social Psychology, 93*, 234-249.

Lemay, E. P., Clark, M. S., & Feeney, B. C. (2007). Projection of responsiveness to needs and the construction of satisfying communal relationships. *Journal of Personality and Social Psychology, 92*, 834-853.

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

Li, P. P. (2008). Toward a geocentric framework of trust: An application to organizational trust. *Management and Organization Review, 4*, 413–439.

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

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

Mabry, J. B., & Kiecolt, K. J. (2005). Anger in Black and White: Race, alienation, and anger. *Journal of Health and Social Behavior, 46*, 85-101.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99-128.

Marshall, M. J., & Stolle, D. (2004). Race and the city: Neighbourhood context and the development of generalized trust. *Political Behaviour 26*, 125–153.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*, 24-59.

McClintock, C. G. (1972). Social motivation—a set of propositions. *Behavioral Science, 17,* 438-454.

McEvily, B., & Tortoriello, M. (2011). Measuring trust in organisational research: Review and recommendations. *Journal of Trust Research, 1*, 23-63.

McLaughlin-Volpe, T. (2004). Understanding stigma from the perspective of the self-expansion model. In S. Levin & C. van Laar (Eds.), *Stigma and group inequality: Social psychological perspectives* (pp. 213-234). Mahwah, NJ: Erlbaum.

Mendelson, M. J., & Aboud, F. E. (1999). Measuring friendship quality in late adolescents and young adults: McGill Friendship Questionnaires. *Canadian Journal of Behavioural Science, 31*, 130-132.

Mikulincer, M. (1998). Attachment working models and the sense of trust: An exploration of interaction goals and affect regulation. *Journal of Personality and Social Psychology, 74*, 1209–1224.

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

Monteith, M. J., & Spicer, C. V. (2000). Contents and correlates of Whites' and Blacks' racial attitudes. *Journal of Experimental Social Psychology, 36*, 125-154.

Murnighan, J. K. (1991). *The dynamics of bargaining games*. Englewood Cliffs, NJ: Prentice Hall.

Murphy, M. C., Richeson, J. A., Shelton, J. N., Rheinschmidt, M. L., & Bergsieker, H. B. (2012). *Cognitive costs of contemporary prejudice*. Manuscript under review.

Murray, S., & Holmes, J. (2009). The architecture of interdependent minds: A motivation-management theory of mutual responsiveness. *Psychological Review, 116*, 908-928.

Nadler, A., & Liviatan, I. (2006). Intergroup reconciliation: Effects of adversary's expressions of empathy, responsibility, and recipients' trust. *Personality and Social Psychology Bulletin, 32*, 459-470.

Nadolny, D. A., Cwir, D., & Spencer, S. J. (2012). *Social connection: A catalyst for cultural diversity or for color blindness?* Poster presented at the 12th Society for Personality and Social Psychology Meeting in San Diego, CA.

Norton, M. I., & Sommers, S. R. (2011). Whites see racism as a zero-sum game that they are now losing. *Perspectives on Psychological Science, 6*, 215-218.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*, 11087-11092.

Operario, D., & Fiske, S. T. (2001). Ethnic identity moderates perceptions of prejudice: Judgments of personal versus group discrimination and subtle versus blatant bias. *Personality and Social Psychology Bulletin, 27,* 550–561.

Oskamp, S., & Perlman, D. (1966). Effect of friendship and disliking on cooperation in a mixed-motive game. *Journal of Conflict Resolution, 10*, 221-226.

Page-Gould, E., Mendoza-Denton, R., Alegre, J. M., & Siy, J. O. (2010). Understanding the impact of cross-group friendship on interactions with novel outgroup members. *Journal of Personality and Social Psychology, 98*, 775 - 793.

Paolini, S., Hewstone, M., Cairns, E., & Voci, A. (2004). Effects of direct and indirect cross-group friendships on judgments of Catholics and Protestants in Northern Ireland: The mediating role of an anxiety-reduction mechanism. *Personality and Social Psychology Bulletin, 30*, 770-786.

Pearson, A. R., West, T. V., & Dovidio, J. F. (2008). *Barriers to contact: The perils of intergroup interaction at minimal acquaintance*. Talk presented at 116th Annual Convention of the American Psychological Association in Boston, MA.

Pearson, A. R., West, T. V., Dovidio, J. F., Powers, S. R., Buck, R., & Henning, R. (2008). The fragility of intergroup relations: Divergent effects of delayed audio-visual feedback in intergroup and intragroup interactions. *Psychological Science, 19*, 1272–1279.

Pettigrew, T. F. (1997). Generalized intergroup contact effects on prejudice. *Personality and Social Psychology Bulletin, 23,* 173-185.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65-85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751–783.

Petty, R. E., & Mirels, H. L. (1981). Intimacy and scarcity of self-disclosure: Effects on interpersonal attraction for males and females. *Personality and Social Psychology Bulletin, 7*, 493-503.

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology, 99*, 771-784.

Pituch, K. A., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research, 41*, 367-400.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*, 741–763.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891.

Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York, NY: Simon & Schuster.

Putnam, R. D. (2007). E Pluribus Unum: diversity and community in the twenty-first century: The 2006 Johan Skytte prize lecture. *Scandinavian Political Studies, 30*, 137–74.

Reis, H. T. (2006). Implications of attachment theory for research on intimacy. In M. Mikulincer & G. S. Goodman (Eds.), *Dynamics of romantic love: Attachment, caregiving, and sex* (pp. 383–403). New York, NY: Guilford.

Reis, H. T. (2007). Steps toward the ripening of relationship science. *Personal Relationships, 14*, 1-23.

Reis, H. T., & Shaver, P. R. (1988). Intimacy as an interpersonal process. In S. W. Duck (Ed.), *Handbook of personal relationships* (pp. 367–389). Chichester, England: Wiley.

Reis, H. T., Smith, S. M., Carmichael, C. L., Caprariello, P.A., Tsai, F.-F., Rodrigues, A., & Maniaci, M. R. (2010). Are you happy for me? How sharing positive events with others provides personal and interpersonal benefits. *Journal of Personality and Social Psychology, 99*, 311-329.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*, 95–112.

Rempel, J. K., Ross, M., & Holmes, J. G. (2001). Trust and communicated attributions in close relationships. *Journal of Personality and Social Psychology, 81*, 57-64.
Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: The role of perceived self-control in interpersonal trust. *Journal of Personality and Social Psychology, 100*, 874-886.
Ross, C. E., Mirowsky, J., & Pribesh, S. (2001). Powerlessness and the amplification of threat: Neighborhood disadvantage, disorder, and mistrust. *American Sociological Review, 66*, 568-591.
Ross, C. E., Mirowsky, J., & Pribesh, S. (2002). Disadvantage, disorder, and urban mistrust. *City & Community, 1*, 59-82.
Rotella, K. N., Richeson, J. A., Chiao, J. Y. (2011). Effects of perceived victimhood on intergroup trust and loyalty. Poster presented at the 12th Annual Meeting of the Society for Personality and Social Psychology in San Antonio, TX.
Rothbart, M., & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology, 50,* 131-142.
Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*, 651-665.
Rotter, J. B. (1971). Generalized expectancies of interpersonal trust. *American Psychologist, 26*, 443–452.
Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35*, 1-7.
Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*, 393-404.
Rubin, Z. (1973). *Liking and loving: An invitation to social psychology*. New York: Holt, Rinehart & Winston.
Rusbult, C. E., & Van Lange, P. A. M. (2003). Interdependence, interaction, and relationships. *Annual Review of Psychology, 54,* 351–375.
Saguy, T., Tausch, N., Dovidio, J. F., & Pratto, F. (2009). The irony of harmony: Intergroup contact can produce false expectations for equality. *Psychological Science*, *20*, 114-121.
Salvatore, J., & Shelton, N. (2007). Cognitive costs of exposure to racial prejudice. *Psychological Science, 18*, 810-815.
Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918-24.
Sapadin, L. A. (1988). Friendship and gender: Perspectives of professional men and women. *Journal of Social and Personal Relationships, 5,* 387-403.
Sawyer, P. J., Major, B., & Mendes, W. B. (2008). *Attributions for nonprejudiced action shape psychological and physiological responses to evaluative intergroup situations*. Poster presented at the 9th Annual Meeting of the Society for Personality and Social Psychology in Albuquerque, NM.
Shallcross, S. L., & Simpson, J. A. (2012, January 16). Trust and responsiveness in strain-test situations: A dyadic perspective. *Journal of Personality and Social Psychology*. Advance online publication. Doi: 10.1037/a0026829
Shelton, J. N. (2000). A reconceptualization of how we study issues of racial prejudice. *Personality and Social Psychology Review, 4*, 374-390.
Shelton, J. N. (2003). Interpersonal concerns in social encounters between majority and minority group members. *Group Processes and Intergroup Relations, 6*, 171-185.
Shelton, J. N., & Bergsieker, H. B. (2008). *Thinking about race (or not): Implications for intergroup friendships*. Manuscript under revision.
Shelton, J. N., & Richeson, J. A. (2005). Intergroup contact and pluralistic ignorance. *Journal of Personality and Social Psychology, 88*, 91-107.

Shelton, J. N., & Richeson, J. A. (2006a). Ethnic minorities' racial attitudes and contact experiences with White people. *Cultural Diversity and Ethnic Minority Psychology, 12*, 149-164.
Shelton, J. N., & Richeson, J. A. (2006b). Interracial interactions: A relational approach. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 38, pp. 121 – 181). New York: Academic Press.
Shelton, J. N., Richeson, J. A., & Salvatore, J. (2005). Expecting to be the target of prejudice. Implications for interethnic interactions. *Personality and Social Psychology Bulletin, 31*, 1189-1202.
Shelton, J. N., Trail, T. E., West, T. V., & Bergsieker, H. B. (2010). From strangers to friends: The interpersonal process model of intimacy in developing interracial friendships. *Journal of Social and Personal Relationships, 27*, 71-90.
Sidanius, J., Levin, S., & Pratto, F. (1998). Hierarchical group relations. Institutional terror, and the dynamics of the criminal justice system. In J. Eberhardt & S. T. Fiske (Eds.), Confronting racism: The problem and the response (pp. 136–165). Thousand Oaks, CA: Sage.
Sidanius, J., Pratto, F., & Mitchell, M. (1994). Ingroup identification, social dominance orientation, and differential intergroup social allocation. *Journal of Social Psychology, 134*, 151-167.
Simpson, B., McGrimmon, T., & Irwin, K. (2007). Are Blacks really less trusting than Whites? Revisiting the race and trust question. Social Forces, 86, 525-552.
Simpson, J. A. (2007a). Foundations of interpersonal trust. In A. W. Kruglanski & E. T. Higgins, (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 587-607). New York: Guilford.
Simpson, J. A. (2007b). Psychological foundations of trust. *Current Directions in Psychological Science, 16*, 264-268.
Simpson, J. A., Rholes, W. S., & Nelligan, J. (1992). Support-seeking and support-giving within couples in an anxiety-provoking situation: The role of attachment styles. *Journal of Personality and Social Psychology, 62*, 434-446.
Simpson, J. A., Rholes, W. S., & Phillips, D. (1996). Conflict in close relationships: An attachment perspective. *Journal of Personality and Social Psychology, 71,* 899–914.
Smith, S. S. (2010). Race and trust. *Annual Review of Sociology, 36*, 453-475.
Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290-312). Washington, DC: American Sociological Association.
Soss, J., Langbein, L., & Metelko, A. R. (2003). Why do White Americans support the death penalty? *Journal of Politics, 65*, 397-421.
Steinmetz, S. (Ed.). (1993). *The Random House Webster's dictionary*. New York, NY: Ballantine Books.
Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*, 157-175.
Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*, 96-102.
Tam, T., Hewstone, M., Cairns, E., Tausch, N., Maio, G., & Kenworthy, J. (2007). The impact of intergroup emotions on forgiveness. *Group Processes and Intergroup Relations, 10*, 119-136.
Tam, T., Hewstone, M., Kenworthy, J. B., & Cairns, E. (2009). Postconflict reconciliation: Intergroup forgiveness and implicit biases in Northern Ireland. *Journal of Social Issues, 64*, 303-320.
Tausch, N., Tam, T., Hewstone, M., Kenworthy, J., & Cairns, E. (2007). Individual-level and group-level mediators of contact effects in Northern Ireland: The moderating role of social identification. *British Journal of Social Psychology, 46*, 541-556.
Terrell, F., & Terrell, S. (1981). An inventory to measure cultural mistrust among Blacks. *The Western Journal of Black Studies, 5*, 180-185.
Tesch, S. A., & Martin, R. R. (1983). Friendship concepts of young adults in two age groups. *The Journal of Psychology, 115*, 7-12.

Tropp, L. R. (2008). The role of trust in intergroup contact: Its significance and implications for improving relations between groups. In U. Wagner, L. R. Tropp, G. Finchilescu & C. Tredoux (Eds.), *Improving intergroup relations: Building on the legacy of Thomas F. Pettigrew* (pp. 91-106). Malden, MA: Blackwell Publishing.

Tropp, L. R., & Pettigrew, T. F. (2005). Relationships between intergroup contact and prejudice among minority and majority status groups. *Psychological Science, 16*, 951-957.

Tropp, L. R., & Wright, S. C. (2001). Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin, 27*, 585-600.

Turner, R. N., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology, 93*, 369-388.

Turner, R. N., & Feddes, A. R. (2011). How intergroup friendship works: A longitudinal study of friendship effects on outgroup attitudes. *European Journal of Social Psychology, 41*, 914–923.

Turner, R. N., West, K., & Christie, Z. (in press). Outgroup trust, intergroup anxiety, and outgroup attitude as mediators of the effect of imagined intergroup contact on intergroup behavioural tendencies. *Journal of Applied Social Psychology*.

Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. New York: Russell Sage Foundation.

Uslaner, E. M. (2002). *The moral foundations of trust*. Cambridge, UK: Cambridge University Press.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly, 42*, 35–67.

Vorauer, J. D. (2006). An information search model of evaluative concerns in intergroup interaction. *Psychological Review, 113*, 862-886.

Webb, W. M., & Worchel, P. (1986). Trust and distrust. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 213-228). Chicago: Nelson-Hall.

West, T. V., Pearson, A. R., Dovidio, J. F., Shelton, J. N., & Trail, T. E. (2009). Superordinate identity and intergroup roommate friendship development. *Journal of Experimental Social Psychology, 45*, 1266-1272.

West, T. V., Popp, D., & Kenny, D. A. (2008). A guide for the estimation of gender and sexual orientation effects in dyadic data: An Actor-Partner Independence Model approach. *Personality and Social Psychology Bulletin, 34*, 321–336.

Wieselquist, J., Rusbult, C. E., Foster, C. A., & Agnew, C. R. (1999). Commitment, pro-relationship behavior, and trust in close relationships. *Journal of Personality and Social Psychology, 77*, 942–966.

Worchel, S., Cooper, J., & Goethals, G. R. (1991). *Understanding social psychology*. Pacific Grove, CA: Brooks/Cole.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology, 73*, 73-90.

Wright, S. C., Aron, A., & Tropp, L. R. (2002). Including others (and groups) in the self: Self-expansion and intergroup relations. In J. P. Forgas & K. D. Williams (Eds.), *The social self: Cognitive, interpersonal, and intergroup perspectives* (pp. 343–363). New York, NY: Psychology Press.

Wrightsman, L. S. (1964). Measurement of philosophies of human nature. *Psychological Reports, 14*, 743-751.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 129-166.

Yovetich, N. A., & Rusbult, C. E. (1994). Accommodative behavior in close relationships: Exploring transformation of motivation. *Journal of Experimental Psychology, 30*, 138-164.
Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science, 9*, 141-159.
Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151-175.

Appendix A

Descriptive Statistics and Raw Bivariate Correlations Within and Between Time Points for Variables (Study 1)

Variables	N	*ICC*	Mean	*SD*	DR	FL	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	Time 2				Time 2		Correlations within time points																	
1. Trust	138	.50*	5.61	1.06	.02	.33*	—	.64*	.67*	.56*	.61*	.65*	.74*	.63*	.51*	.39*	.48*	.60*	.64*	—	—	—	—	—
2. Cultural trust	138	.45*	5.39	1.03	-.10	.26*	.70*	—	.50*	.39*	.48*	.45*	.49*	.47*	.30*	.19*	.31*	.31*	.48*	—	—	—	—	—
3. Liking	136	.44*	5.79	1.21	-.10	.36*	.80*	.62*	—	.36*	.59*	.46*	.45*	.67*	.34*	.13	.29*	.40*	.48*	—	—	—	—	—
4. Closeness	138	.72*	4.68	1.49	.12	.33*	.65*	.57*	.58*	—	.63*	.60*	.64*	.47*	.76*	.61*	.71*	.64*	.78*	—	—	—	—	—
5. Stimulating companionship	136	.46*	5.85	1.23	.03	.38*	.70*	.60*	.83*	.69*	—	.64*	.62*	.49*	.56*	.44*	.54*	.55*	.66*	—	—	—	—	—
6. Emotional security	136	.37*	5.64	1.38	.07	.38*	.74*	.54*	.71*	.67*	.76*	—	.61*	.53*	.63*	.49*	.58*	.59*	.64*	—	—	—	—	—
7. Reliable alliance	136	.56*	5.80	1.20	.06	.38*	.80*	.61*	.76*	.73*	.75*	.76*	—	.50*	.56*	.45*	.48*	.58*	.66*	—	—	—	—	—
8. Self-expression	134	.40*	5.71	1.16	-.02	.27*	.68*	.50*	.70*	.61*	.67*	.67*	.67*	—	.54*	.22*	.38*	.43*	.62*	—	—	—	—	—
9. Self-disclosure	134	.64*	5.12	1.57	.15	.35*	.61*	.47*	.59*	.79*	.69*	.75*	.68*	.68*	—	.70*	.69*	.61*	.69*	—	—	—	—	—
10. Perceived other disclosure	132	.56*	5.10	1.61	.29*	.20*	.55*	.43*	.46*	.75*	.55*	.61*	.60*	.54*	.77*	—	.68*	.54*	.50*	—	—	—	—	—
11. Help/support	134	.42*	5.01	1.53	.13	.24*	.52*	.41*	.52*	.78*	.64*	.69*	.60*	.55*	.79*	.74*	—	.69*	.60*	—	—	—	—	—
12. Perceived help/support	132	.56*	5.18	1.64	.17	.39*	.73*	.51*	.62*	.77*	.67*	.77*	.75*	.65*	.77*	.71*	.73*	—	.64*	—	—	—	—	—
13. Perceived understanding	132	.47*	5.34	1.37	.08	.39*	.77*	.60*	.78*	.75*	.76*	.77*	.77*	.79*	.73*	.64*	.66*	.79*	—	—	—	—	—	—
14. Contact quantity (absolute)	136	.76*	5.03	1.77	.19*	.16	.45*	.53*	.38*	.73*	.49*	.43*	.51*	.35*	.57*	.51*	.60*	.59*	.53*	—	—	—	—	—
15. Contact quantity (relative)	136	.47*	2.96	1.25	.23*	.21*	.46*	.43*	.38*	.70*	.48*	.48*	.51*	.47*	.61*	.56*	.58*	.59*	.51*	.78*	—	—	—	—
16. Contact quality	136	.29*	3.87	0.90	.12	.22*	.62*	.50*	.52*	.67*	.50*	.49*	.64*	.55*	.56*	.53*	.50*	.63*	.58*	.62*	.63*	—	—	—
17. Trust-building behaviors	135	.42*	3.06	1.04	.11	.15	.51*	.45*	.39*	.73*	.51*	.62*	.61*	.46*	.70*	.67*	.63*	.68*	.61*	.61*	.55*	.46*	—	—
18. Trust-betraying behaviors	135	.29*	1.75	0.70	.13	-.11	-.41*	-.20*	-.44*	.05	-.20*	-.21*	-.20*	-.25*	-.03	-.01	.04	-.08	-.22*	.22*	.14	-.18*	.13	—
	Time 1				Time 1		Correlations between time points																	
1. Trust	198	-.19†	5.79	0.89	.21*	.18*	**.60***	.42*	.38*	.39*	.31*	.49*	.47*	.35*	.38*	.33*	.36*	.48*	.50*	.32*	.27*	.41*	.39*	-.17
2. Cultural trust	198	.06	5.39	0.97	.11	.09	.48*	**.54***	.35*	.29*	.29*	.32*	.32*	.29*	.25*	.19*	.26*	.32*	.39*	.26*	.19*	.28*	.14	-.10
3. Liking	198	.09	5.98	0.96	.08	.01	.43*	.37*	**.49***	.31*	.36*	.36*	.42*	.35*	.28*	.16	.26*	.32*	.42*	.24*	.21*	.26*	.26*	-.12
4. Closeness	198	.66*	4.91	1.17	.29*	.39*	.33*	.33*	.23*	**.61***	.40*	.48*	.41*	.29*	.55*	.48*	.55*	.49*	.47*	.45*	.39*	.32*	.54*	.23*
5. Stimulating companionship	198	.29*	6.12	0.85	.23*	.20*	.43*	.45*	.39*	.55*	**.58***	.51*	.48*	.36*	.54*	.45*	.55*	.51*	.52*	.43*	.35*	.35*	.50*	.09
6. Emotional security	198	.23*	5.53	1.28	.15*	.20*	.37*	.27*	.28*	.43*	.37*	**.60***	.43*	.36*	.51*	.45*	.51*	.51*	.49*	.30*	.31*	.21*	.49*	.05
7. Reliable alliance	198	.42*	6.01	0.98	.33*	.35*	.46*	.39*	.28*	.50*	.31*	.50*	**.49***	.37*	.49*	.43*	.45*	.58*	.55*	.35*	.26*	.36*	.48*	.04
8. Self-expression	198	-.07	5.97	0.88	.10	.10	.40*	.30*	.39*	.40*	.35*	.37*	.47*	**.47***	.40*	.26*	.34*	.36*	.46*	.26*	.26*	.29*	.34*	-.06
9. Self-disclosure	198	.54*	5.26	1.28	.34*	.24*	.32*	.23*	.23*	.50*	.34*	.47*	.42*	.30*	**.59***	.50*	.52*	.50*	.41*	.37*	.34*	.27*	.51*	.14
10. Perceived other disclosure	198	.42*	5.37	1.13	.25*	.16*	.22*	.18*	.09	.37*	.24*	.42*	.26*	.12	.43*	**.52***	.48*	.38*	.25*	.22*	.19*	.15	.39*	.19*
11. Help/support	198	.48*	5.43	1.39	.27*	.09	.17*	.19*	.12	.42*	.31*	.46*	.30*	.15	.45*	.45*	**.60***	.44*	.30*	.38*	.33*	.21*	.50*	.23*
12. Perceived help/support	198	.42*	5.58	1.40	.26*	.16*	.30*	.19*	.19*	.42*	.30*	.44*	.35*	.26*	.46*	.44*	.47*	**.53***	.35*	.32*	.31*	.32*	.50*	.17
13. Perceived understanding	198	.40*	5.35	1.12	.22*	.31*	.44*	.45*	.39*	.56*	.50*	.58*	.52*	.46*	.54*	.49*	.58*	.58*	**.60***	.41*	.37*	.38*	.50*	.08

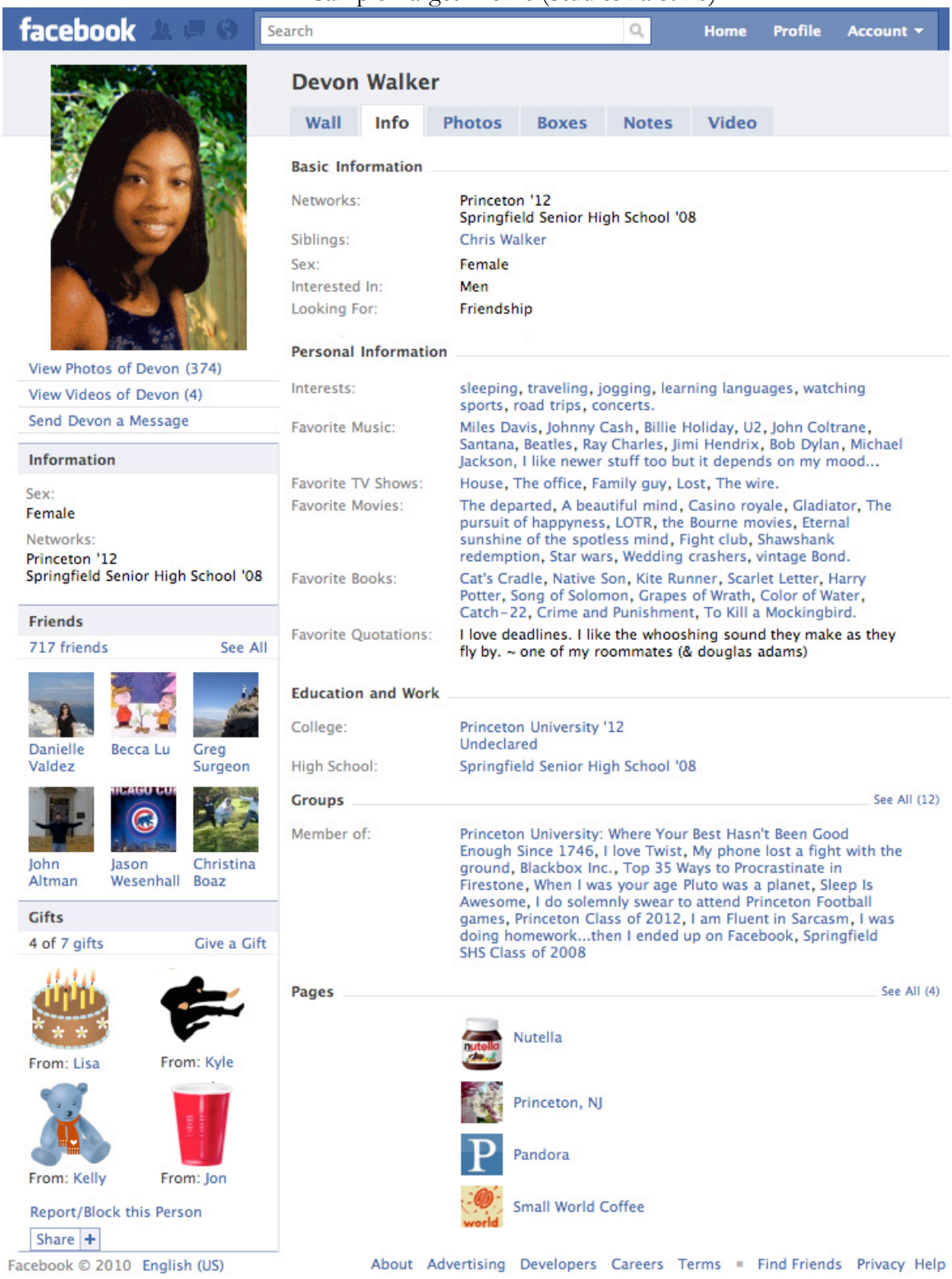
Note. DR = dyad race (effects coded: -1 interracial, 1 same-race), FL = friendship length. For correlations within time points, Time 1 data appear above the diagonal, with Time 2 data below. For correlations between time points, Time 1 variables (1-13) run along the left side and Time 2 variables (1-18) run along the top. Boldface indicates correlations for a given variable across time points. $^{*}p < .05$

Appendix B

Target Photos (Studies 2a & 2b)

	Female		Male	
Target race	Photo 1	Photo 2	Photo 1	Photo 2
Black				
White				

Appendix C
Sample Target Profile (Studies 2a & 2b)



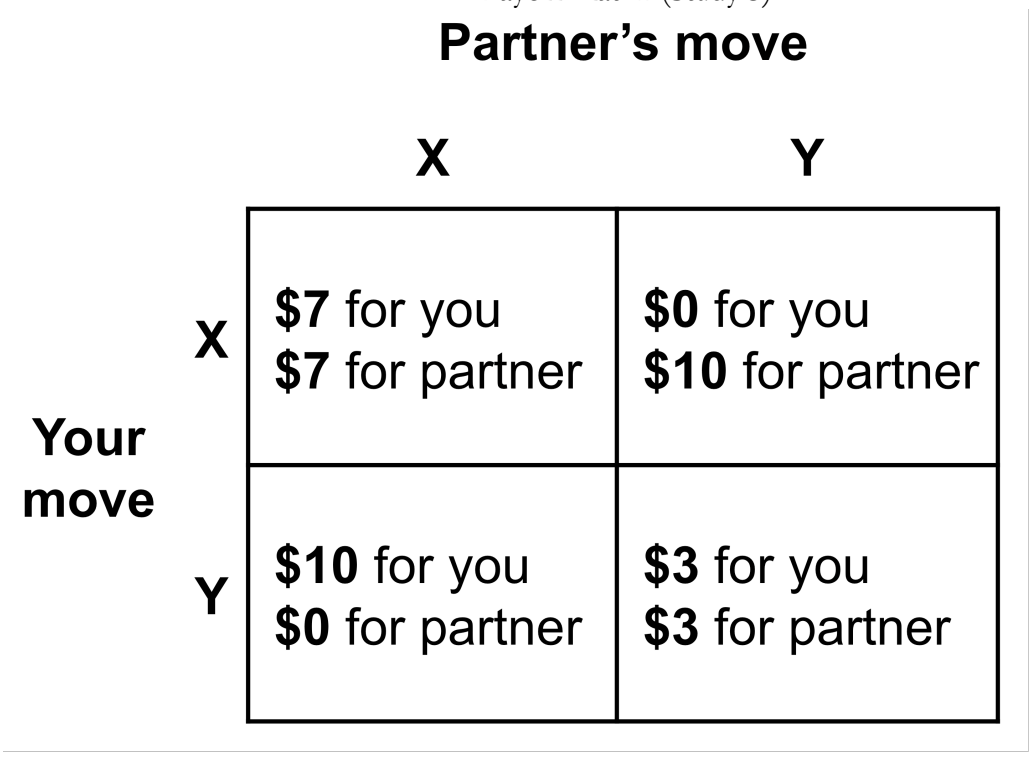
Appendix D

Imagined Interpersonal Interactions (Studies 2a & 2b)

Imagined interactions by type	Order by condition: Early	Late	Never
Trust-betraying			
One day after class, you and Devon decide that you should get together that night to go over a problem set that is due after the weekend. Devon promised to call you to let you know when to meet up. Instead, Devon sends you a last-minute text message saying that night won't work. The next day, you see on Facebook that Devon has been tagged in several photos from a party that happened the same night you had been planning to go over the problem set together.	1	4	—
Trust-building			
You and Devon attended a lecture in which you were sneezing and coughing up a storm and feeling awful. Devon left right after class, but then posted on your wall that night: "Hey hope you're doing ok and [not dying of/haven't caught] the flu! If you're still not feeling well on Wednesday you can totally have my class notes. Feel better!"	4	1	1
You mention to Devon that you are really unhappy with one of your professors, whose lectures you find to be very unclear, and you are getting worried about the midterm you will have the following week. Devon mentions that a close friend had taken the class and found a website that was really helpful. That night, Devon posts on your Facebook wall and gives you the link to the website.	2	2	2
You noticed Devon was on Facebook one evening when you were both procrastinating, so you initiated a Facebook chat. You ended up chatting until really late, and it felt like Devon shared thoughts on just about everything (family issues, fitting in, dating drama, academic worries, etc.). You learned that you and Devon shared several similar struggles.	3	3	3
You posted a Facebook status update stressing out about your paper (due tomorrow) that was still in really bad shape. Devon commented on your status, [mentioning a coffee run and asking if you wanted anything. Devon also offered to help proofread your paper/saying "I feel your pain – I'm way behind on work this week!" Also, Devon mentioned going on a late-night coffee run soon and asked if you wanted anything].	5	5	5
You mention to Devon that your aunt is coming to Princeton to visit you, and that you have no idea where to go for dinner. A few hours later, Devon posts on your Facebook wall to suggest two restaurants that were recommended by friends, and gives you the links to their websites so that you can check their menus.	6	6	6
Trust-neutral			
One day, Devon [is late to/looks really tired in] class. Later that evening, Devon posts on your Facebook wall: "Hey I [missed/spaced out during] the first five minutes of class today. When is our first paper due again?"	—	—	4

Note. Minor wording variants in across Studies 2a and 2b are noted in brackets, in order.

Appendix E
Payoff Matrix (Study 3)



Appendix F
Tutorial on Cooperation (Study 3)

Current version (Study 3): To maximize your overall earnings, recall that they depend on your partner's actions and that you will complete several rounds together. If you both choose X and each get $7, this choice may signal mutual willingness to pass up the highest payoff to get the next-best option (which is still good relative to the others). This cooperative signal suggests that additional rounds of jointly choosing X and earning $7 payoffs will benefit both of you. If you both choose Y and each get $3, this negative signal suggests that you were each unwilling to risk choosing X or hoped to get $10 at the other's expense. Starting with Y may make it difficult to get to the mutually beneficial state of repeatedly earning $7 each. If one of you chooses X and the other chooses Y, the first person earns $10 (to the other's $0) in that round but may earn less overall as the other may be more likely to choose Y in subsequent rounds.

Lount et al. (2008) version: You will soon be presented with some information about a possible interaction you might have with another person. Some of that information will consist of a set of payoffs to you and to them. Not surprisingly, the choices you make and the choices that they make will influence what each of your outcomes will be. Your outcomes will be:

	P	Q
A	you get $24; they get $24	you get $6; they get $30
B	you get $30; they get $6	you get $12; they get $12

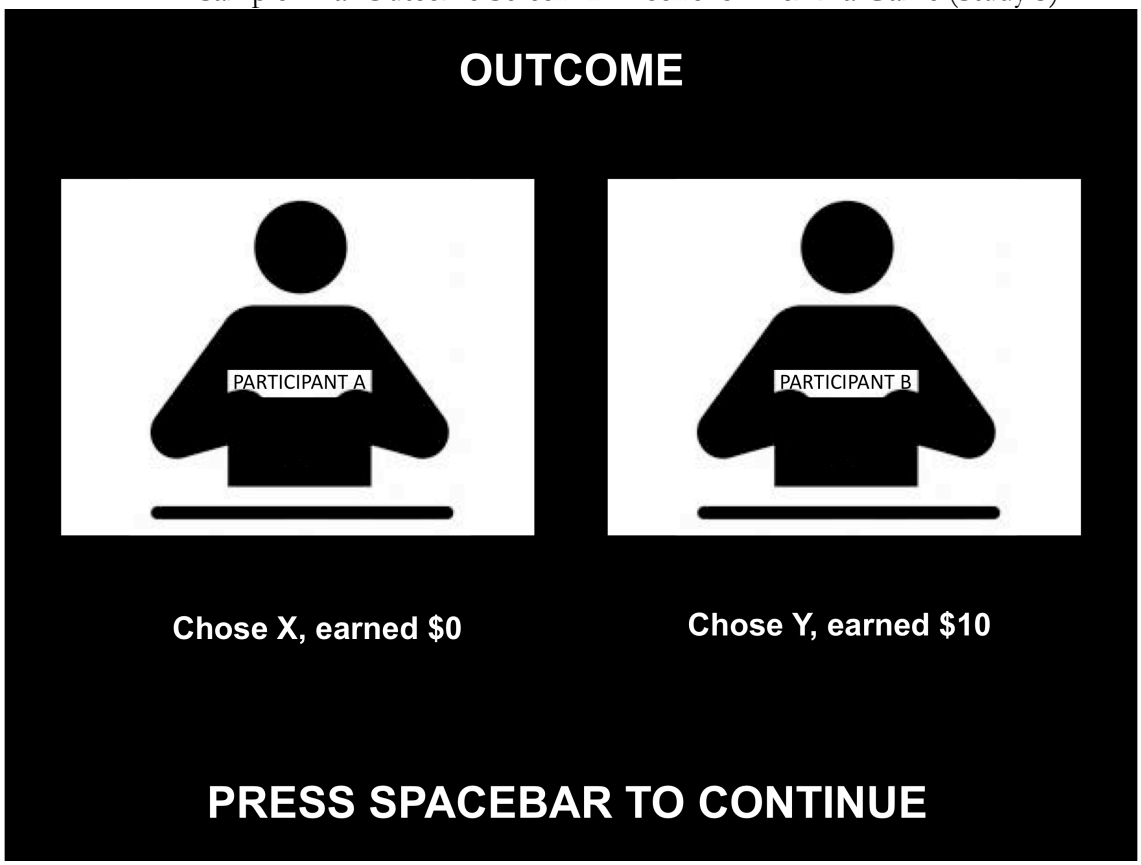
Your best outcome here is $30; your worst outcome here is $6. The payoffs are symmetric, so their best is also $30 and their worst outcome is also $6. You and the other person will make several choices, one after the other. After each choice, you will know what your outcomes are for that round (i.e., which of the four outcomes). This way you will know, in advance, what both of you have chosen before you make your next choice.

Now imagine that you chose A and they chose P: you would each get $24. You could both interpret your choices as signals that you are both willing to resist the temptation of your highest outcomes so that you can achieve the next best outcome (which is still quite good, in comparison to the other possibilities). This would be a very cooperative signal that also suggests that additional rounds of the same choices, and their $24 payoffs, make sense for both of you. (This exemplifies thinking about this interaction from your perspective and theirs.)

Why are these such important signals? Well, consider what would happen if you chose A and they chose Q, with you getting $6 and them getting $30. They might very well be delighted with this outcome, but what would you do on the next round? Most people in your shoes shift to B for their next choice. If they chose Q again (highly likely in this situation) you would both get $12 - decidedly worse that the $24 you could both have by playing A and P. You would also have a hard time shifting to the $24 each outcome, which is better for both of you (and increasingly better as your choices continue and you participate in this interaction repeatedly).

The same thing might happen if your roles were reversed, i.e., if you chose B and they chose P. Thus, although choosing A is risky (you could get $6), it makes sense in the long run - and in your upcoming interaction, there will be repeated choices by both you and the other person, making the long run really important, increasing the value of each cooperative $24 outcome, over and over again. (Note that if, on the first round, the two of you chose B and Q, yielding $12 for each of you, this would be a negative signal, suggesting that each of you were not willing to take the risk to be cooperative or that each of you was hoping to get $30 at the other person's expense. Starting out with B and Q choices makes it harder to get to the mutually beneficial, cooperative outcomes of repeatedly getting $24 each.) All this is by way of saying that, yes, cooperation is risky, but that, once it is achieved, it is very much worth repeating.

Appendix G
Sample Trial Outcome Screen in Prisoner's Dilemma Game (Study 3)



Note. Actual outcome screens included color photographs of participants and their partners.