Psychological Foundations of Trust

Jeffry A. Simpson

University of Minnesota, Twin Cities Campus

ABSTRACT—*Trust lies at the foundation of nearly all major theories of interpersonal relationships. Despite its great theoretical importance, a limited amount of research has examined how and why trust develops, is maintained, and occasionally unravels in relationships. Following a brief overview of theoretical and empirical milestones in the interpersonal-trust literature, an integrative process model of trust in dyadic relationships is presented.*

KEYWORDS—*trust; interdependence; strain tests; felt security*

Trust: "confidence that [one] will find what is desired [from another] rather than what is feared." (Deutsch, 1973, p.148)

Trust involves the juxtaposition of people's loftiest hopes and aspirations with their deepest worries and fears. It may be the single most important ingredient for the development and maintenance of happy, well-functioning relationships. Several major theories, including attachment theory (Bowlby, 1969) and Erikson's (1963) theory of psychosocial development, are built on the premise that higher levels of trust in relationships early in life lay the psychological foundation for happier and better-functioning relationships in adulthood. Tooby and Cosmides (1996) claim that trust-relevant emission and detection mechanisms should have evolved in humans, given the importance of gauging accurately the intentions of others.

Considering the centrality of trust in relationships across the lifespan, one might expect the topic would have received widespread theoretical and empirical attention. Surprisingly, it has not. Although there have been significant pockets of theory (e.g., Holmes & Rempel, 1989) and research (e.g., Mikulincer, 1998; Rempel, Holmes, & Zanna, 1985) on the subject, relatively little is known about how and why interpersonal trust develops, is maintained, and unravels when betrayed.

Address correspondence to Jeffry A. Simpson, Department of Psychology, University of Minnesota, Minneapolis, MN 55455-0344; e-mail: simps108@umn.edu.

Why has trust received such limited attention? To begin with, trust is a complex, multidimensional construct, making it difficult to operationalize, measure, and interpret. Second, trust can be construed in different ways, and it might have varying importance at different stages of relationship development. Third, trust emerges and changes in situations that are difficult to observe and study, such as in "strain test" situations (Holmes, 1981). In strain-test situations, one individual is highly outcome dependent on his or her partner, but the actions that would promote the individual's own interests differ from those that would benefit the partner. For example, if Chris desperately needs Susan's help to complete an important task and Susan willingly helps despite the fact that doing so impedes what she really wants or needs to accomplish, Susan has "passed" a strain test and, accordingly, Chris should trust her more.

In this article, I first discuss major theoretical and empirical accounts of interpersonal trust from which four basic principles can be distilled. I then describe a new process model of dyadic trust that integrates these principles.

THEORETICAL AND EMPIRICAL CONCEPTUALIZATIONS OF TRUST

Historically, there have been two main approaches to conceptualizing interpersonal trust. The earliest work adopted a dispositional (person-centered) view. According to this perspective, trust entails general beliefs and attitudes about the degree to which other people are likely to be reliable, cooperative, or helpful in experimental game situations (Deutsch, 1973) or in daily-life contexts (Rotter, 1971). Beginning in the early 1980s, conceptualizations and measures of trust started to focus on specific partners and relationships (Holmes & Rempel, 1989; Rempel et al., 1985). According to the dyadic (interpersonal) perspective, trust is a psychological state or orientation of an actor (the truster) toward a specific partner (the trustee) with whom the actor is in some way interdependent (that is, the truster needs the trustee's cooperation to attain valued outcomes or resources). What makes trust particularly difficult to study is that it involves three components (e.g., "*I* trust *you* to do X"; Hardin, 2003). Thus, trust is a

function of properties of the self (I), the specific partner (you), and the specific goal in a current situation (to do X).

Kramer and Carnevale (2001) argue that trust involves a set of beliefs and expectations that a partner's actions will be beneficial to one's long-term self-interest, especially in situations in which the partner must be counted on to provide unique benefits or valuable outcomes. Trust-relevant situations typically activate two cognitive processes: (a) feelings of vulnerability; and (b) expectations of how the partner is likely to behave across time, particularly in strain-test situations. When the partner promotes the individual's best interests rather than his or her own, both parties should experience heightened trust. Trust is also likely to be higher in a relationship when (a) each member's self-interested outcomes match those that are best for their partner or the relationship, or (b) both members believe that their partner will act on what is best for the relationship even when the members' personal self-interests diverge.

Kelley et al. (2003) claim that trust can be assessed in certain interpersonal situations. Trust situations involve the configuration of high interdependence (such that the actions of each partner strongly impact the other), a blend of rules for coordination and exchange that sustain interdependence, and moderately corresponding interests (see Kelley et al., 2003, for further details). One prototypical trust situation is depicted in Figure 1. In this situation, trust should be facilitated when partners in a relationship repeatedly make A1/B1 (i.e., mutually

Partner A's Choices

		A1	**A2**
B1		+20 / +20	+10 / −10
B2		−10 / +10	0 / 0

Partner B's Choices

Fig. 1. The payoffs (outcomes) in Kelley et al.'s (2003) trust situation. Payoffs above the diagonal in each cell are for Partner A; those below the diagonal are for Partner B. If both partners choose option 1 (e.g., they work together on a difficult but important task, reflected in the A1/B1 cell), each partner receives 20 units of benefit because the task gets done and partners enjoy each other's company. If both partners select option 2 (neither works on the task, represented by the A2/B2 cell), neither partner receives benefits because nothing gets accomplished. If partner A chooses option 2 (not to work on the task: A2) whereas partner B chooses option 1 (works solo on the task: B1), partner A benefits by 10 units because progress is made on the task, but partner B experiences a net loss of 10 units because he or she is saddled with all the work. The reverse pattern exists when partner A chooses A1 (to work on the task) and partner B chooses B2 (to not work on the task). Trust situations have three special properties: First, they involve collective rationality in that cooperative behavior by both partners (A1/B1) always yields better outcomes than when partners fail to cooperate (A1/B2 or A2/B1). Second, the best outcome always occurs when both partners make the cooperative choice (A1/B1). Third, cooperative choices are risky, because if one's partner decides to make a noncooperative choice, the cooperative choice generates the worst possible outcomes because one has been exploited.

beneficial) decisions that yield maximum rewards for both individuals.[1]

Most previous research on interpersonal trust has been guided by dispositional or interpersonal perspectives. Dispositionally oriented work has revealed that individuals who are more insecurely attached, have lower self-esteem, or have more poorly differentiated self-concepts (i.e., self-concepts that are less diversified, more imbalanced, and poorly tied together) trust their relationship partners less (see Simpson, 2007, for a review). Interpersonally oriented work has confirmed that trust is typically higher when individuals believe their partners are more committed to the relationship and have more benevolent intentions and motivations. It is also higher when partners regularly display prorelationship transformations of motivation (that is, turn initial gut-level negative reactions to caustic partner behaviors into constructive responses that benefit the relationship), which then generate self-sacrificial or accommodative behaviors. Research testing Holmes and Rempel's (1989) dyadic model of trust has also indicated that the development of trust involves a process of uncertainty reduction as individuals move from having confidence in their partner's general predictability to having confidence in their prorelationship values, motives, goals, and intentions (Holmes & Rempel, 1989).

Based on a recent review of the interpersonal trust literature (Simpson, 2007), four core principles of interpersonal trust stand out.

First, individuals gauge the degree to which they can trust their partners by observing whether partners display proper transformation of motivation in *trust-diagnostic* situations (that is, in trust or strain-test situations in which partners make decisions that go against their own personal self-interest and support the best interests of the individual or the relationship).

Second, trust-diagnostic situations often occur naturally and unintentionally during the ebb and flow of everyday life. Depending on situational circumstances, however, individuals may enter, transform, or occasionally create trust-diagnostic situations to test whether their current level of trust in a partner is warranted.

Third, individual differences in attachment orientations, self-esteem, or self-differentiation (i.e., working models of self and others as relationship partners) should affect the growth or decline of trust over time in relationships. People who are more securely attached, have higher self-esteem, or have more differentiated self-concepts should be more likely to experience trust as well as increases in trust in relationships across time.

Fourth, neither the level nor the trajectory of trust in relationships can be fully understood without considering the dis-

---

[1]In contrast to Prisoner's Dilemma games in which decision making is independent of the partner, partners in established relationships often engage in joint decision making in which separate (totally independent) action may not be realistic. See Kelley et al. (2003) for how game outcomes involving established partners can be interpreted.

positions and actions of *both* relationship partners, especially in trust-diagnostic situations.

A DYADIC MODEL OF TRUST

A model of how these core principles may be linked together in dyadic social interactions is shown in Figure 2. The Dyadic Model of Trust in Relationships (Simpson, 2007) contains both normative (typical) and individual-difference components. The normative components are depicted in the five boxes (constructs) in the middle of the figure. The individual-difference components reflect the relevant dispositions of each relationship partner (e.g., attachment orientations, self-esteem, self-differentiation) and their connections to each normative construct. Feedback loops from the final normative construct in the model (each partner's degree of felt security following an interaction) to the construct that launches future trust-relevant interactions (each partner's decision regarding whether or not to enter the next trust-relevant situation) are not shown, but are presumed to exist. According to the model, each individual's perceptions of

his or her own and the partner's standing on each construct are necessary to explain and understand what happens for each partner further along in the model.

The model assumes that information about the relevant dispositions of *both* partners is essential to understanding and explaining the growth of trust—or lack thereof—in a relationship across many interactions. The dispositional tendencies discussed above should motivate or enable individuals to enter, transform, and occasionally create social interactions that enhance trust over time. Two types of situations should give individuals a particularly good opportunity to gauge the level of trust warranted in a partner or relationship: (a) *trust situations* (Kelley et al., 2003), in which partners can repeatedly make or fail to make A1/B1 (i.e., mutually beneficial) decisions (see Fig. 1); and (b) strain-test situations (Holmes, 1981), in which partners can demonstrate or fail to demonstrate their willingness to make personal sacrifices for the good of the partner or relationship. Before trust-diagnostic situations can be entered, transformed, or created, however, one or both partners must have enough confidence to take the interpersonal risks necessary to confirm or
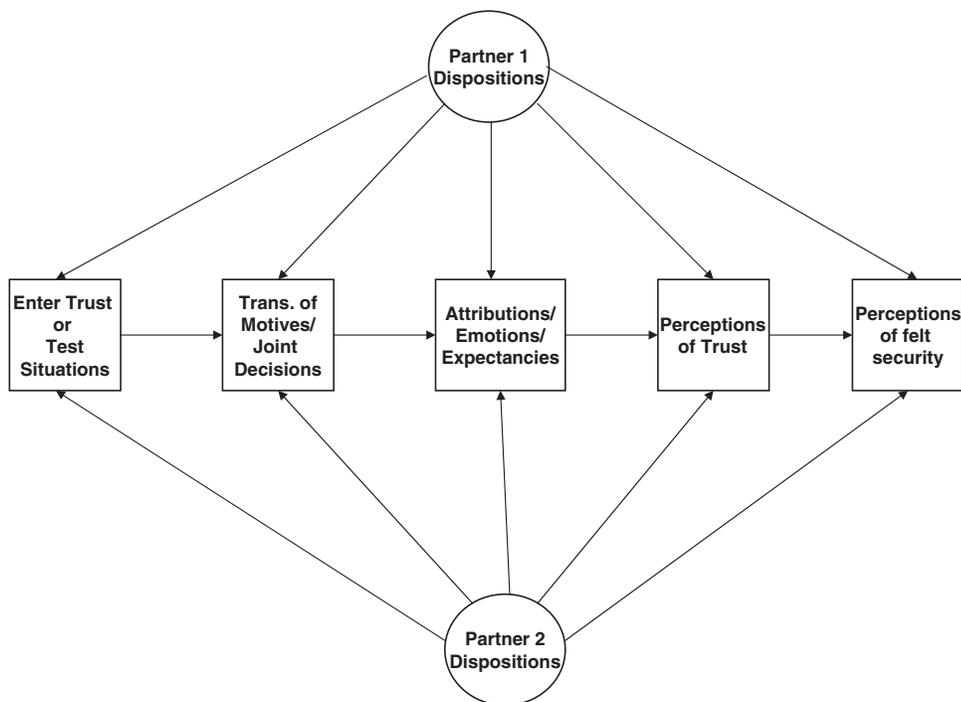

Fig. 2. The Dyadic Model of Trust in Relationships (Simpson, 2007). Individuals who have certain dispositions (positive working models) should be more likely to enter, transform, or occasionally create trust-diagnostic situations in their relationships. When encountered, such situations should provide good opportunities for these individuals to experience relationship-enhancing transformations of motivation, which should increase their willingness to enter mutually beneficial decisions (promoting the goals of the partner and/or the relationship over the self). If both partners make mutually beneficial decisions, this should generate positive patterns of attributions, emotions, and future expectancies, which in turn should enhance perceptions of trust and felt security, at least temporarily. The working models of each partner in the relationship are likely to affect outcomes at each stage of the model. Feedback loops from the final construct in the model (each partner's perceived degree of felt security) to the construct that launches future trust-relevant interactions (each partner's decision whether or not to enter, transform, or create the next trust-relevant situation) are not depicted, but are presumed to exist.

reaffirm that the partner *can* be trusted. People who have more positive working models should be more inclined to take these risks and leaps of faith.

Once in trust-diagnostic situations, individuals who display the transformation of motivation necessary to make decisions that benefit the partner or relationship at some cost to the self should experience greater trust and felt security. Partners who have more positive working models ought to display partner- or relationship-based transformations more often and more extensively. Consequently, they should be more motivated and capable of steering trust-relevant social interactions toward mutually beneficial decisions. Once such decisions have been made, working models should influence how individuals interpret the amount of transformation that the self and the partner have undergone. Those who have more positive working models may grant themselves and their partners "fuller credit" for each partner's willingness to prioritize partner or relationship outcomes over egocentric interests, whereas the reverse should be true of those who harbor more negative models (Murray, Holmes, & Collins, 2006).

This process should then trigger benevolent attributions of one's own and the partner's relationship motives. For example, after Susan makes a major personal sacrifice to help Chris complete his important task, Chris should infer that he needs and values Susan and the relationship and that she genuinely cares about him. Such attributions may result in more constructive problem solving, more adaptive emotion regulation, and/or more optimistic expectations about future trust-diagnostic interactions. Patterns of attribution, emotion regulation, and situation-specific expectancies are presented in a single box in Figure 2 because the temporal order of these processes may depend on specific aspects of the interaction, the working models of each partner, or unique properties of the relationship. These positive outcomes, in turn, should increase perceptions of trust, which should elevate felt security at least temporarily. At this point, positive working models might further boost perceptions of trust and felt security. These perceptions should then set up the next trust-relevant interaction, determining whether or when the next trust situation or strain test is entered, transformed, or created.

In many instances, individuals enter, transform, or create trust or strain-test situations without forethought or deliberation. Deliberate attempts to create such situations should occur when important, unexpected, or suspicious events lead people to question whether they can truly trust their partners. Though potentially very diagnostic, premeditated tests probably are conducted rather infrequently.

Over time, individuals who repeatedly experience mutually beneficial outcomes with their partners should begin to perceive greater "added value" (see Fig. 1), especially in trust-diagnostic situations. These repeated outcomes may encourage individuals and their partners to engage in further relationship-sustaining or relationship-building acts (e.g., disparaging attractive alterna-

tive partners, perceiving the partner in an overly positive light) that might increase the likelihood of more mutually beneficial decisions and outcomes in the future. These effects should diminish, however, if one or both partners have negative working models, repeatedly decide not to enter into mutually beneficial agreements, or harbor negative attributions regarding their partner's overarching relationship motives.

This raises an interesting paradox. Early in relationships, greater transformation of motivation by both members should be a good barometer of the level of trust that is warranted in a partner or the relationship, especially when both individuals are equally dependent on each other for unique or important outcomes (Holmes & Rempel, 1989) or have similar levels of vulnerability or commitment (Wieselquist, Rusbult, Foster, & Agnew, 1999). As relationships progress, however, most individuals incorporate (merge) their partners and relationships with their own self-concepts. This means that less transformation of motivation should take place in well-established relationships. Couples in long-term relationships, therefore, may occasionally need to identify new trust-diagnostic situations in which each individual's self-interests start out being discrepant from what might be best for his or her partner or the relationship. If partners jointly reaffirm that they are still willing to undergo large transformations when new trust-diagnostic situations arise, this should sustain or increase trust even in longstanding pairs.

FUTURE DIRECTIONS AND CONCLUSIONS

Our knowledge of how trust is generated, sustained, and compromised in relationships remains surprisingly limited. Future theory and research would profit from addressing the following issues. First, we need to gain a better understanding of how the dispositions and behaviors of both individuals in a relationship affect how they think, feel, and behave in trust-relevant situations. Second, research should explore how and why certain combinations of partner attributes promote or impede the development and maintenance of trust. For example, relationships in which one partner has much more power than the other could hinder the development of trust if the powerful person self-servingly takes advantage of the less powerful partner, yet this combination could generate high levels of trust if the high-power partner continually forgoes his or her best self-interest for the best interest of the low-power partner. Third, we need to determine whether the normative component of the Dyadic Model of Trust in Relationships holds with equal effect at different stages of relationship development. Fourth, we need to know more about how normative processes and individual differences reciprocally influence one another over time, particularly in trust-relevant situations. Fifth, we need to disentangle the constructs and processes depicted in the middle box of the model (attributions/emotions/expectancies). Sixth, we must learn more about the roles played by factors critical to the development of intimacy—such as feeling understood, validated, and cared for

by one's partner—in fostering trust and felt security (Reis & Shaver, 1988), especially when they occur in trust-relevant interactions. Finally, we need to know more about when and how perceptions of "added value" arise.

In conclusion, there are few constructs in psychology that are more central to interpersonal functioning and outcomes than trust. Indeed, many of the principles discussed in this article might apply to other social contexts and relationships, including coworker relationships, business negotiations, and perhaps even interactions between political groups. Whether the model and psychological processes discussed above extend beyond the realm of close relationships remains unknown. Nevertheless, for the science of relationships to advance, we must gain a deeper and more complete understanding of how trust emerges, changes, is maintained, and declines in different types of relationships across time.

Recommended Reading

Holmes, J.G., & Rempel, J.K. (1989). (See References)
Kelley, H.H., Holmes, J.G., Kerr, N.L., Reis, H.T., Rusbult, C.E., & Van Lange, P.A.M. (2003). (See References)
Ostrom, E., & Walker, J. (Eds.). (2003). *Trust and reciprocity: Interdisciplinary lessons from experimental research*. New York: Russell Sage Foundation.
Simpson, J.A. (2007). (See References)

REFERENCES

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. New York: Basic Books.
Deutsch, M. (1973). *The resolution of conflict*. New Haven, CT: Yale University Press.
Erikson, E. (1963). *Childhood and society*. New York: Norton.
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.
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: 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.
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.
Kramer, R.M., & Carnevale, P.J. (2001). Trust and intergroup negotiation. In R. Brown & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 431–450). Malden, MA: Blackwell Publishers.
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.
Murray, S.L., Holmes, J.G., & Collins, N.L. (2006). Optimizing assurance: The risk regulation system in relationships. *Psychological Bulletin*, *132*, 641–666.
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.
Rempel, J.K., Holmes, J.G., & Zanna, M.P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*, 95–112.
Rotter, J.B. (1971). Generalized expectancies of interpersonal trust. *American Psychologist*, *26*, 443–452.
Simpson, J.A. (2007). Foundations of interpersonal trust. In A.W. Kruglanski & E.T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 587–607). New York: Guilford.
Tooby, J., & Cosmides, L. (1996). Friendship and the banker's paradox: Other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy*, *88*, 119–143.
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.