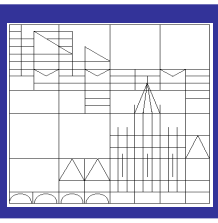
University of Konstanz
Department of Economics



Trust me, it is High Trust: On Trust and its Measurement

Christian Lukas and Peter Walgenbach

Working Paper Series
2010-09

http://www.wiwi.uni-konstanz.de/workingpaperseries

Trust me, it is high trust: On trust and its measurement

Christian Lukas[1] and Peter Walgenbach[2]

This version: October 2010

As the positive impact of trust on business success is widely undisputed, the question of how to measure trust naturally arises. Both expectations of trustworthy behavior and possible gains and losses from a trusting relationship influence the level of trust between two parties or individuals. This paper explores whether an exchange featuring (almost) equal expected gains and expected losses for a trusting individual is evidence for high trust or low trust; we argue that such an exchange tends to display low trust. A simple trust measure is suggested that can be applied both in experimental and analytical research.

Keywords: trust; trust measurement; trust as behavior; agency theory; behavioral game theory; investment game
JEL code: D6; M5

1 Introduction

Nowadays, it is widely accepted that trust is a crucial aspect of social and business life. Taking the latter as an example, trust has the potential to increase efficiency as it may foster teamwork and cooperation; at the same time it may substitute for (costly) formal control mechanisms. Trust research has therefore seen a substantial upsurge over the past several years.[3,4] To date, both a rich theoretical body of literature and a great deal of empirical literature on trust have emerged.[5] One objective of the latter is to quantify trust. Trust measurement can be carried out in different ways, for example, by counting the frequency of trusting behavior or by using surveys to measure trusting attitudes (Möllering (2006, p. 135f)). While the latter method is the preferred choice in field studies (see Dirks and Ferrin

[1] *Prof. Dr. Christian Lukas, Junior Professor of Business Administration and Accounting, University of Konstanz, Dept. of Economics, Box D 144, 78457 Konstanz, christian.lukas@uni-konstanz.de*

[2] *Prof. Dr. Peter Walgenbach, Chair of Organization, Leadership and Human Resource Management, School of Economics and Business Administration, Friedrich Schiller University of Jena, peter.walgenbach@uni-jena.de*

[3] Various journals published special issues on “trust”, e.g. Academy of Management Review (1998), Journal of Economic Behavior and Organization (2004), or Organization Science (2003).

[4] The phrase “investment game” yields about 1,000 hits in Google Scholar by title search for the period 2000-2009, while "trust game" as a synonym yields another 1,500 hits for that period. (Accessed March 13, 2009)

[5] See, for example, Möllering (2006); Nooteboom (2002); and Bachmann and Zaheer (2006); and the references therein. Additional empirical findings are presented in the meta-analyses by Dirks and Ferrin (2001, 2002) and the review by Cook and Cooper (2003).

(2001, 2002); Glaeser et al. (2000)), the former is often employed in experimental research (e.g., ; Burks et al. (2003); Deutsch (1958); Ferrin and Dirks (2003); Fetchenhauer and Dunning (2009); Kosfeld et al. (2005); Sally (1995)). Besides the frequency of trusting behavior, the extent of trusting behavior may serve as a measure for trust. This approach is applied in numerous experimental studies, for example in the investment game or trust game where the amount invested in a relationship provides the measure for trust (e.g., e.g., Altmann et al. (2008); Ashraf et al. (2006); Berg et al. (1995); Glaeser et al. (2000); Pillutla et al. (2003))[6,7]. In either case, i.e. frequency or extent of trusting behavior, it is an action that represents the trust proxy. A potential problem may be that perceptions (or intentions, depending on the point of view) do not matter for the proxy. However, trust is very often referred to (defined as) an "expectation concerning the behavior of others" (Nooteboom (2002, p.6)). Hence, perceptions should be reflected in a trust measure. Perceptions can be elicited by a survey. Usually, perceptions are tested for their predictive power, i.e. how well they explain observed behavior in the laboratory (e.g., Glaeser et al. (2000), Hill et al. (2009), Schweitzer et al. (2006), or Tjosvold (1985)). An alternative way to combine perceptions and actions would be to fuse them into a single trust measure. Such a measure could prove helpful especially in situations which are difficult to judge, e.g., when contracts, behaviors, and trust interact in long-term relationships[8], or when perceptions alone do not seem to possess much predictive power. It is the objective of this paper to propose a trust measure that is both conceptually founded and easy to apply.

Allow us to briefly illustrate the argument: Consider an employer who offers an employee a bonus of 10 units which is contingent on the employee's achieving a pre-specified performance target. The employee has an expectation about whether the employer keeps the promise or not. Next time, the employer offers a bonus of 15 and the employee's expectations may have changed, in part due to the previous (non-)fulfillment of the promise. If (a) actual achievement cannot be verified by a court and if (b) the employee cannot verify whether unforeseeable contingencies, e.g., financial distress or a worsening business outlook, may force the employer to break the promise, the bonus payment occurs at the discretion of the employer. In other words, the employee trusts the employer to fulfill the promise.

Comparing the two situations, the question – which of them shows more trust? – cannot be spontaneously answered in either case. If the employee's expectation remains unchanged, an increasing bonus at stake would suggest higher trust. Emanating from a change in expectation but a constant bonus, an increasing expectation would suggest higher trust again. It follows that the cases that are the most difficult to judge (and the most interesting) are those in which expectation and bonus do not jointly increase or decrease. They certainly exist in reality, because ever-increasing promises of bonuses will certainly not be in accord

[6] See Camerer (2003) for a review of trust games.
[7] Trust measurement in public good games may proceed in a very similar way. See Sally (1995) for a review.
[8] See Ferrin and Dirks (2003) or Tjosvold (1985) for interaction effects in static relationships.

with ever-increasing expectations of actually receiving a bonus.

This is why a measure that consistently quantifies trust is needed: it must incorporate both the expectation of trustworthy behavior and the stakes involved. Considering a transaction between a trusting individual (trustor) and a trusted individual (trustee), the trustor either incurs a loss contingent on untrustworthy behavior by the trustee or enjoys a gain contingent on trustworthy behavior by the trustee. However, as Friedland (1990) argues, trust should not lead to naivety. Therefore, the expected gain from a trusting transaction should never fall short of the expected loss. As the analysis will show, the answer to the following question provides the key to finding a consistent trust measure: Should a transaction featuring almost equal expected losses and expected gains for the trustor be labeled low trust or high trust? Stated differently, does naivety begin when trust reaches its maximum or minimum? The presence of potential losses in a transaction is evidence of the trustor's "willingness to be vulnerable" which represents a distinctive feature of a trust relationship (Mayer et al. (1995)). According to what we call the residual gain approach, which is based on research by Parkhe and Miller (2000), equality between expected losses and expected gains suggests high trust. In contrast, according to the avoided loss approach, as we will show, it would indicate low trust. We will argue in favor of the latter because if expectations of trustworthy behavior increase, expected losses decrease. Hence, the lower the ratio of expected losses over expected gains, the more trust is present. Based on this insight, we will suggest a simple trust measure that consistently quantifies trust. An obvious candidate for its application would be experimental research in game theory. Here the stakes are clearly identifiable (and controllable) and if they are combined with participants' perceptions the trust level can be determined. For example, consider two different "senders" in the investment game, i.e. two participants who decide on the amount invested in their respective relationship: Let both send an amount X but let them have different expectations as to how much of the *k*-fold amount of X the "recipients", i.e., their respective partners in the relationship, will return. Do both senders trust their respective recipient to the same extent because they sent the same amount? By equating the amount sent with the level of trust, the answer is yes; but based on this paper's approach, the answer is no.

The trust measure we suggest can also be applied to models in analytical trust research. Currently, trust is becoming an issue in analytical research (see, for example, Casadesus-Masanell (2004), Casadesus-Masanell and Spulber (2007), or Lukas and Schöndube (2010)). Relationships are modeled in a way that subjective expectations, actions, and contract parameters influence each other. As a consequence, changes in expectations affect the risks and rewards of a trusting relationship. In order to analyze whether the level of trust increases or decreases as a result of these interactions, one must be aware of the preconditions for trust and of consistent measures that correctly identify high trust or low trust. Otherwise, high trust transactions can be falsely referred to as low trust transactions, or vice versa.

In this paper, we proceed as follows. Section 2 reviews related literature and section 3 discusses preconditions for trust and concludes with requirements for a consistent trust measure. The following section investigates two different approaches to measure trust: one based on residual gains (4.1), the other based on avoided losses (subsection 4.2). In the final section, we discuss the results and implications of our study.

2 Related literature

A common but very general understanding of trust defines trust as an "expectation concerning the behavior of others".[9] Expectations are assumed to be somewhat diffuse or blurred; they do not account for all possible contingencies in a consistent manner.[10] Moreover, they may even disregard some contingencies. In this way, trust allows for complexity reduction (Luhmann (1989)). What will qualitatively matter for analytical and experimental work is that with trust increasing between contracting partners, higher-stake transactions become possible, which leads back to the notion that trust allows for a prediction of the behavior of contracting partners.[11] Allowing one party to predict behavior is not to be mistaken for *knowing* precisely what behavior the other party will display. If contracting partners know about each others' incentives to trust, they will rationally do so if incentives are sufficiently high. As such, "trust" would be calculative cooperation, or it could be labeled trust as prudence (James (2002)).[12] Such trust as prudence appears in (multi-period) agency models and, more generally, in the repeated games literature. Starting with agency theory, Casadesus-Masanell (2004), among others, defines trust as equilibrium behavior of rational individuals.[13] He analyzes a two-stage agency relationship. In the first stage, the agent develops dispositions, such as the desire to follow norms. Stage 2 features a standard agency problem that accounts for the agent's dispositions molded in stage 1. The principal knows these dispositions and is aware of the agent's commitment to the norms. Given that the agent feels pressure to obey rules and norms, the principal can offer lower incentives to induce a given action by the agent. Several features of a trust relationship show up here. The principal's action, i.e., the offering of a contract, does not restrict the agent's set of actions. Trustworthy behavior by the agent leaves the principal better off than does

[9] Nooteboom (2002), p.6

[10] There is an ongoing debate as to whether or not trust can be identified as a (subjective) probability since it entails risk. Dasgupta (1988), Kreps et al. (1982), Mayer et al. (1995), and Mui et al. (2002) are examples of advocates in this respect. Nooteboom (2002, p. 40) holds an opposing view on this matter; he objects to trust as a subjective probability because it can become one, implying certainty; but trust is related to *un*certainty. He prefers to speak of trust as an "expectation", leaving room for "residual uncertainty about agency and unforeseeable contingencies" (Nooteboom (2002), p. 41).

[11] In this vein, reputation is to be understood as a source of trust and trustworthiness. See Nooteboom (2002), p. 68f.

[12] Williamson (1993) calls calculative trust a contradiction in terms. See also Craswell (1993) for support and an elaboration of Williamson's (1993) arguments.

[13] See also Casadesus-Masanell and Spulber (2007).

untrustworthy behavior; in addition, in the event of untrustworthy behavior, the principal would be better off by offering a standard agency contract that does not account for dispositions.

Despite the presence of these features of trust, the principal and the agent maintain their ability to calculate completely and predict correctly the other party's behavior. Therefore, a genuine and essential feature of trust – the irreducibility of loss and uncertainty – is lacking.[14] Consequently, it appears inappropriate to label the principal's behavior – in this case – trust behavior.

Since trust relationships generally unfold in repeated, long-term interactions between individuals, one may analyze these interactions as repeated games. Most of the existing literature studies *reputation* in repeated games and a vast amount of literature exists on the topic.[15] However, a simple backward induction argument explains why reputation cannot be built up in finitely repeated games under perfect (or complete) information, which represents the standard result.[16] In contrast, as soon as individual players do not possess the same complete information, the development of reputation becomes possible and valuable. Early contributions by Kreps and Wilson (1982) and Kreps et al. (1982) demonstrate that a slight chance of incomplete information or bounded rationality already suffice to induce cooperation early on.[17] After one player behaves opportunistically for the first time, however, cooperation will never again resume.

Although the vast majority of game theoretic work in repeated games considers reputation as a means to induce and maintain cooperative behavior, there are some exceptions which explicitly refer to trust in that respect, e.g. Watson (1999; 2002) and Sobel (1985). Watson (1999; 2002) develops a model to determine equilibrium behavior in long-term relationships where partners are not entirely familiar with each other's type of cooperative behavior. He distinguishes between "high" and "low" types of cooperation. "High" types behave trustworthy and cooperate as long as their partners do so, but "low" types do "have an incentive to betray their partner's trust" (Watson (1999), p. 53) and behave in an untrustworthy manner. "Low" types may acquire a reputation for being of the "high" type if they mimic the latter's behavior. In equilibrium, then, partners initially engage in low-stake transactions but, as their reputation improves, increasingly engage in higher stake transactions later on. Eventually, the two types separate because the incentive to betray

[14] Uncertainty is to be thought of as uncertainty in a narrow sense such that no probabilities can be determined which is different from risk as random variation. See Möllering (2006), p. 7f.

[15] See, for example, Mailath and Samuelson (2006) and the references therein.

[16] Selten (1978) coined the term *chain-store paradox* for the corresponding backward induction argument. In infinitely repeated games (or repeated games with a stochastic end), opportunistic types can build up a reputation. Yet under imperfect monitoring it may not be possible to maintain a reputation for being non-opportunistic forever (Cripps et al. (2004)). Only if types change over time does reputation become valuable again (Phelan (2006); Wiseman (2008)).

[17] Andreoni and Miller (1993) find experimental evidence for the predictions made by Kreps et al. (1982).

their partner's trust becomes too attractive for "low" types and then they cannot mimic the "high" types ever again.

A similar pattern of "trusting" decisions is found by Sobel (1985) in a game between a sender of information – who exclusively knows the correct information – and a receiver of information who acts on the information received. Initially, the sender transmits information correctly but eventually cashes in on her/his reputation. Rising stakes in repeated transactions are very likely to be observable in trusting relationships. So in both Watson's (1999; 2002) and Sobel's (1985) argument this trust feature undoubtedly is present. For (at least) two reasons, however, one would not call the relationships analyzed in Watson (1999; 2002) and Sobel (1985) trusting ones. First, players are fully rational and therefore able to foresee the other players' strategies. Hence, "trust" is reduced to mere calculative cooperation. Second, players are able to verify decisions ex post, implying that irreducible uncertainty is absent from the model. It follows that once non-cooperative or untrustworthy behavior has been observed, a return to the cooperative or trust solution is precluded. To be precise, the latter represents an assumption (to sustain a cooperative equilibrium in the first place) rather than a consequence. At the same time, however, there is empirical evidence that trust can be rebuilt in such cases (Jonker et al. (2004); Schweitzer et al. (2006)).

In summary and in general, models with trust as equilibrium behavior of fully rational individuals will not be capable of capturing trust in multistage games (or relationships). Moreover, such models cannot explain trusting behavior in one-shot or end games. A step beyond rational calculation and full rationality is needed to do so. Then an expectation concerning the behavior of others still exists, but the crucial difference to trust as prudence is that uncertainty is prevalent and expectation does not result from mere rational calculations. Lukas and Schöndube (2010) model trust in an agency relationship by assuming that the agent is not able to foresee the principal's actions – contract offers and fulfillments – in a multi-period setting. In their model, the principal promises to pay a bonus to the agent contingent on the observation of sufficiently high performance by the agent. Since performance cannot be verified by a third party, the principal and agent rely on implicit contracting. The agent expects the principal to fulfill the implicit contract. However, financial shocks or other challenging situations may occur which prevent the bonus payment by the principal. These shocks cannot be foreseen or verified from the agent's perspective. Hence, the agent's expectation neither accounts for this contingency nor for the principal's actions. This is why the agent trusts the principal to pay the bonus. As the relationship evolves, higher stake transactions become possible because trust increases whenever the principal fulfills the implicit contract.[18]

In this case, the interaction of trust and incentives or stakes is evident. Experimental studies

[18] Hwang and Burgers (1997) present an analytical model of trust where the trust level remains constant after observing cooperation; while Boyle and Bonacich (1970), in contrast, develop a model where trust builds up rapidly but drops sharply after observing a violation of trust.

show that such interaction does indeed exist (Ferrin and Dirks (2003), Pillutla et al. (2003), Tjosvold (1985)).[19] A frequently used measure of trust in experimental game theory is the amount invested in a relationship (e.g., Altmann et al. (2008); Ashraf et al. (2006); Berg et al. (1995); Cox (2004); Glaeser et al. (2000); Pillutla et al. (2003); Schweitzer et al. (2006)). In the investment game or trust game, two players interact. One player (the "sender") receives an endowment and decides how much of it to send to the other player (the "receiver"). Any amount sent is (usually) tripled by the experimenter, and the receiver decides how much to send back to the sender. In the public goods game usually more than two players interact. Each of them receives an endowment and splits it into a contribution to a group project – the public good – and the amount to keep. The total contribution of all players is multiplied by a factor *k* by the experimenter. Each player derives a private benefit from the public goods which amounts to less than the *k*-fold total contribution. Hence, from an individual perspective, the contribution is not worthwhile because every unit invested yields less than one unit in return. However, if, for example, all players contribute their entire endowment private benefits derived from the public goods exceed initial endowments; an improvement if the total of private benefits constitutes the welfare measure.

At this point, it seems useful to clarify the understanding of trust in this paper. For this purpose, we refer to the following definition.[20]

Definition *Trust in persons or organizations means accepting that they may take advantage of a trustor although the trustor expects them not to do so.*

The trust definition is quite similar to Mayer et al.'s (1995) definition of trust as the "willingness to be vulnerable". It centers the expectation of trustworthy behavior, i.e., the behavior of the trusted individual (trustee) that will not take advantage of the trusting individual (trustor). That expectation in turn will (or will not) induce the trustor to display a certain behavior that will have a positive outcome if the expectation is met and negative consequences if not. In this sense, trust is identified as behavior (Deutsch (1958)).[21] The advantage of the approach lies in allowing straightforward application in analytical research and permitting experimental studies to test predictions.[22] Of course, given constant stakes, the higher the expectation of trustworthy behavior, the more trust is present. An equally

[19] Besides incentive structures other contextual factors are relevant, including competitive vs. cooperative organizations (Hill et al. (2009)); consensus vs. representative decision-making (Song 2008); country of origin (Holm and Danielson (2005)); reference points (Bohnet et al. (2010)) and even the administration of oxytocin (Kosfeld et al. (2005)).

[20] The definition is very much in line with Nooteboom (2002), p. 45.

[21] Operationalizing trust as behavior (Deutsch (1958); Coleman (1990)) is a common approach but differs from those defining it as a psychological state (Rousseau (1998)) or as a disposition (e.g., Rotter (1967)). These approaches share the view that trust involves (positive) expectation concerning the behavior of others. However, the latter two may use trust as an explanation for behavior while the first labels behavior as trust.

[22] The latter motivation already dates back to Deutsch (1958).

intuitive and correct statement would read that for any given level of expectation, the higher the stakes involved in a trusting relationship, the more trust is present again. Comparing cases where both expectation and stakes differ becomes difficult, however. One example is an employer who offers her employee a bonus of 10 units contingent on achieving a pre-specified performance target in a given year.[23] The employee holds an expectation about whether or not the employer will keeps her promise of, say, 0.9. Next time, the employer offers a bonus of 15 units the employee's expectation may have changed to 0.8, in part due to the previous (non-)fulfillment of the promise.[24] A second example would be two senders in the investment game who have different endowments but send the same amount.[25] Has trust increased, decreased, or remained unchanged? What can we say about the extent to which trust is present? To answer these questions, one must be able to quantify trust. A trust measurement must therefore include both the expectation of the partner's trustworthy behavior and stakes involved in the trusting relationship. For that purpose, the following section defines and formalizes expectation and stakes before possible trust measures will be presented and discussed in sections 4 and 5.

3 Requirements for a trust measure

Consider the following transaction. Let us assume A agrees to deliver a performance or service to B, while B promises to reward this with money or to provide some service in return. To speak of trust in that transaction, there must be uncertainty (see above). Therefore, suppose A does not know whether or not B will behave in a trustworthy manner, but A must provide a *Vorleistung* (Luhmann (1989)), a favor in advance, by giving the performance or providing the service. If A does so, this behavior can be labeled trust. Let L denote A's loss contingent on untrustworthy behavior by B. The loss could, for example, be the cost of providing the favor in advance or other foregone benefits because transacting with B does not allow A to transact with a third party. If B behaves in a trustworthy manner, A enjoys a gain G, the gain comprises B's reward or payment, or it could be the excess gain over the best alternative transaction available to A. Following Deutsch (1958), a necessary but not sufficient condition that A trusts B in a transaction is that A's loss contingent on untrustworthy behavior by B must be larger than A's gain contingent on trustworthy behavior by B:[26]

[23] Financial turmoil may deter the employer from paying the bonus. If the employee cannot verify its occurrence, the employee trusts the employer to pay the bonus.

[24] If individuals had stable values and consistent preferences and were always able to correctly and consciously calculate probabilities, identifying the level of expectation would suffice for the researcher to determine the existence of high or low trust. However, reality suggests something different and such a pure version of rational choice may not be appropriate. March (1994) elaborates on various ways of modifying assumptions of the rational choice model, e.g., unstable preferences or decision heuristics. The expectation could be thought of as the trustee's crude aggregate of assumed preferences and decision rules of the trustor.

[25] See Pillutla et al. (2003) for an investment game with varying endowments.

[26] See Coleman (1990), p. 99ff for the conceptualization of conditions (1) and (2) or Parkhe and Miller (2000).

$$R_1 = \frac{L}{G} > 1. \tag{1}$$

It is worth pointing out that (1) does *not* imply that A does not have trust in B if $\frac{L}{G} < 1$. A may well trust B without engaging in that transaction, but the point being made is that A's behavior in the transaction under consideration justifies the label trust only if (1) holds true.[27] In this sense, (1) represents a necessary but not a sufficient condition.

A second necessary but not sufficient condition to label behavior trusting behavior is that expected losses are lower than expected gains, that is

$$R_2 = \frac{L}{G} \cdot \frac{(1-p)}{p} < 1 , \tag{2}$$

where p denotes A's expectation of B acting non-opportunistically, meaning that (2) includes an attitudinal influence. Expectation p does not represent the outcome of a comprehensive evaluation of the transaction by a fully rational individual; expectation is rather blurred in the sense that A as a bounded rational individual is able to form an expectation but it may not correctly and consistently take into account all available information.[28,29]

Trust measurement based on (2) incorporates perceptions and not just observable actions. The separation of perception from action represents an indispensable ingredient to separate mere calculative cooperation from trust.[30] Condition (2) simply states trust should not lead to naivety (Friedland, 1990); a normative argument derived from it would state "if you expect a net loss from a transaction, do not engage in it". Given $R_1 > 1$, a minimum level of expectation $p = \underline{p} = \frac{1}{2}$ is required to make a trusting relationship feasible, so that

---

[27] Existence of a potential loss L for trust to matter has been demonstrated by Parks and Hulbert (1995). According to their experimental results, when trusting individuals are *not* exposed to a potential loss if trusted individuals do not cooperate, high- and low-trust individuals cooperate at the same rate.

[28] In this sense, the expectation formation resembles the cognitive bias in evaluating partners labeled knowledge-based trust (Yamagishi and Yamagishi (1994)). Yet if expectation forms in the process of repeated interaction – Yamagishi and Yamagishi (1994) call it general trust –, it may still be prone to bias or inconsistency given bounded rational individuals.

The expectation formation may also take into account the trustee's incentives to honor trust which Hardin (2001; 2003) calls trust as encapsulated interest. Since this paper does not need to impose restrictions on how expectation is formed, the analysis accommodates these different processes and conclusions apply accordingly.

[29] A possible safeguard, i.e. a threat of sanctions B will incur from behaving in an untrustworthy manner, is not explicitly included here. Such a safeguard would (positively) affect A's expectation p suggesting higher trust. However, increasing the safeguard enough would eventually force cooperative or trustworthy behavior. As Nooteboom (2002, p. 85) states: "To force trust is like forcing spontaneity: if it worked, it would not be genius." See Williamson (1993) or Eberl (2004) for a similar argument.

[30] See Möllering (2006), p. 41.

$$R_2\left(p = \underline{p} - \varepsilon\right) = 1, \varepsilon \in R^+, \varepsilon \to 0, \tag{3}$$

i.e., for an expectation marginally lower than the minimum level of expectation, expected loss equals expected gain. This lower bound $\underline{p}$ ensures that the A's decision represents a trusting choice, i.e. given $p > \frac{1}{2}$, the expectation of trustworthy behavior exceeds the one of untrustworthy behavior.[31] Bohnet and Zeckhauser (2004) present evidence that this lower bound $\underline{p}$ is indeed higher in trust situations than in risk situations.

Coleman (1990) argues against (1) as a necessary precondition for trust. In his view, it suffices to require (2) which is less restrictive than (1) because one can label transactions as trusting behavior that feature higher potential gains than losses $\left(\frac{L}{G} < 1\right)$ but display low expectation so that $R_2 < 1$. Taking the argument to its limit, if gains are sufficiently high, expectations of non-opportunistic behavior can approach zero and condition (2) still holds – the case of a confidence man. One may have serious reservations to calling engaging in such a transaction trusting behavior; presumably, the label gambling choice describes the behavior better.[32] Therefore, both conditions, (1) and (2), should hold. However, a condition $R_1 > 0$ reconciles both views to a limited extent and could justifiably be labeled baseline condition for trust. (In this case, a potential loss may be incurred.) Both R_1 and R_2 can be interpreted as indicators of trust.

The preceding arguments lead to the following requirements that a meaningful trust measurement or the trust situation under consideration should have:

1) Presence of loss, $R_1 > 0$ (or, more restrictive, $R_1 > 1$).
2) Absence of naivety, $R_2 < 1$.
3) Trust increases in perceived expectation of non-opportunistic behavior.

In the following, we discuss two possible ways of trust measurement: one based on what we call the residual gain approach and the other based on the avoided loss approach. The crucial difference between the two relates to the key question as to whether a state of almost equal expected loss and expected gain from an exchange ($R_2 \cong 1$) should be labeled low trust or high trust. The residual gain approach presumes such an exchange to be evidence of high trust, while the avoided loss approach presumes the opposite. We argue that the residual gain approach contains an inconsistency and that the measure based on the avoided loss approach will better allow for a consistent trust measurement.

[31] See Ripperger (1998), p. 87ff.
[32] See Ripperger (1998), p. 87f.

4 Two possible trust measures

Based on the arguments from the preceding section, the preconditions for trust can be fused into a simple ratio that measures trust. Here, two different ratios are discussed. While variant 1 does not consistently indicate high or low trust environments, variant 2 does just this.

4.1 Variant 1: Residual gain approach

The first candidate for a simple trust measure is suggested by Parkhe and Miller (2000).[33] They first interpret $(1-R_2)$ as the standardized spread of net expected gains, which corresponds to intuition given $(1-R_2)=\frac{pG-(1-p)L}{pG}$, and arrive at the trust measure defined as follows[34]:

$$R_3=\frac{R_1}{(1-R_2)}=\frac{R_1}{1-\left(\frac{1-p}{p}\right)R_1}. \tag{4}$$

If R_3 increases, more trust is present. To exemplify the intuitive idea behind (4), assume A is willing to accept a high potential loss compared to a possible gain. In this case, R_1 would be fairly high, indicating high trust. For any given expectation p, the higher R_1 is, the more trust exists. With R_1 rising, the expected loss approaches the expected gain, i.e., $R_2 \to 1$, so that the standardized spread of net expected gains decreases, i.e., $(1-R_2)\to 0$. Again, it indicates high trust. Since the spread $(1-R_2)\to 0$ can be labeled the normalized residual of expected gains over expected losses, we call the measurement approach in (4) the residual gain approach.

How does the ratio R_3 change in A's expectation of non-exploitative behavior p? Some algebra yields $(dR_13)/dp=-[R_11/((p-(1-p)R_11)]^12<0$. For any *fixed* loss/gain ratio R_1 trust will be lower the higher p is. Keeping both the potential loss and gain constant, A shows less trust in B the higher her/his expectation of non-opportunistic behavior is. For

[33] Though the dialogue section in the Academy of Management Review is not peer-reviewed, the proposal stands.

[34] Cf. Parkhe and Miller (2000), p. 11

example, a loss/gain ratio $R_1 = 2$ obtained from $L = 4, G = 2$ translates into lower trust given $p = 0.8$, $R_3(p = 0.8) = \frac{2}{1 - \frac{1 - 0.8}{0.8} \cdot 2} = 4.0$, than given $p = 0.9$, $R_3(p = 0.9) = \frac{2}{1 - \frac{1 - 0.9}{0.9} \cdot 2} = 2.6$. If a person has higher expectations of trustworthy behavior but accepts only the loss/gain ratio, trust is lower. Since intuition suggests that trust increases if positive expectations rise, one may wonder why we obtain the opposite outcome here. One crucial factor is that the potential loss does not depend on expectation, i.e., $L(p) = L$. Given the main effect of trust – one's willingness to accept more vulnerability in an exchange if trust increases[35] –, the potential loss should depend on p. For that purpose, let $L(p) = p$, so that higher loss is accepted if the subjective expectation of non-exploitative behavior p increases. Given condition (1), $G \leq 1$ must hold then. If p is sufficiently high, $R_1 > 1$. Plugging in $L(p) = p$ and G into (1), (2), and (4), one can calculate the following derivatives:

$$\frac{d}{dp} R_1(p) = \frac{1}{G} > 0 \quad ; \tag{5}$$

$$\frac{d}{dp} R_2(p) = -\frac{1}{G} < 0 \quad ; \tag{6}$$

$$\frac{d}{dp} R_3(p) = \frac{(G - 1)}{[G - (1 - p)]^2} < 0. \tag{7}$$

Strikingly, if A's expectation p increases, which in this case is indicative of more trust, higher trust would be correctly indicated only by R_1 while R_2 and R_3 suggest the opposite. Inspection of (7) shows that trust increases only if condition (1) does *not* hold.

A possible explanation for the implausible finding in (7) is the absence of any positive effect of trust on gains from the exchange. If trust directly results in higher gains or if it provides conditions that makes higher gains possible, G should depend on p.[36] Hence, let us assume trust positively affects gains. Now loss *and* gain are affected by A's expectation p.[37] Specifically, suppose expectation affects L and G as follows:

$$L^{mod} = p + kp \tag{8}$$

$$G^{mod} = p(1 - p) + kp \tag{9}$$

where $k, k \geq 0$, is a measure of A's ability to possess gains from the exchange. For example,

[35] See Mayer et al. (1995), and Dirks and Ferrin (2001).

[36] The direct effect is referred to as the main effect of trust, the indirect effect as the moderating effect of trust. See Dirks and Ferrin (2001).

[37] An early contribution that relates stakes and expectation is Boyle and Bonacich (1970).

A and B negotiate a bonus which is contingent on the success of A's effort. If A has high trust in B, he will exert more effort, thus increasing the value of the trusting relationship. That value is divided between A and B, and A's share could be the bonus. The bonus payment, however, depends on B's decision to pay, that is, to act non-opportunistically. It follows that if B behaves in a trustworthy manner and pays the bonus, A's potential gain increases, whereas the potential loss decreases if B does not. So with higher expectation p, loss and gain increase and condition (1) is met at any level of p.
After inserting (8) and (9) into (1), (2), and (4), some algebra yields

$$R_1(L^{mod}, G^{mod}) > 1 \tag{10}$$

$$R_2(L^{mod}, G^{mod}) < 1 \Leftrightarrow p > 1 + k - \sqrt{k + k^2}. \tag{11}$$

Figure 1 exemplifies how trust conditions and trust measure behave when p increases, where parameter k is set $k = 1$.

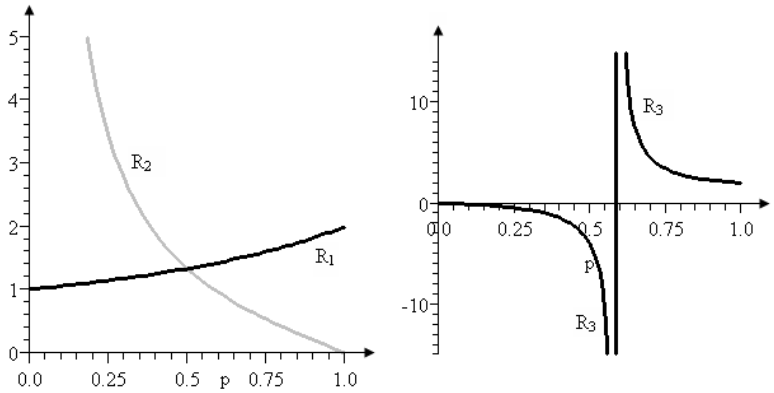


Figure 1: Trust conditions R_1 and R_2 (left), and trust measure R_3 (right) using residual gain approach

According to (11), a minimum expectation $p = 0.59$ is required if $k = 1$ to make a trusting relationship feasible (as expected loss then falls short of expected gain).[38] While R_1 indicates higher trust if p increases, both R_2 and R_3 do not:

$$\frac{dR_1(L^{mod}, G^{mod})}{dp} = \frac{k+1}{(1-p+k)^2} > 0 \tag{12}$$

$$\frac{dR_2(L^{mod}, G^{mod})}{dp} = -\frac{(k+1)(k+1+p^2-2p)}{[p(1-p+k)]^2} < 0 \tag{13}$$

$$\frac{dR_3(L^{mod}, G^{mod})}{dp} = -\frac{(k+1)(k+1-p^2)}{(k-2p-2kp+p^2+1)^2} < 0 \tag{14}$$

[38] The threshold value obtains by plugging in $k = 1$ into $R_2 = \frac{1-p}{p} \cdot \frac{(p+kp)}{[p(1-p)+kp]} = 1$ which leads to $p = 1 + k - \sqrt{k + k^2} = 0.59$.

Compare the derivatives (12)-(14) with their counterparts (5)-(7) where only the potential loss depends on expectation p : The results match exactly. While R_1 correctly suggests higher trust in both settings, both R_2 and R_3 fail to indicate higher trust in either setting. To remedy the inconsistent findings, a different approach to measure trust is needed.

4.2 Variant 2: Avoided loss approach

Two basic requirements for a useful trust approach are (1) to correctly identify high- or low-trust exchanges, and (2) to have consistent trust measures (or a hierarchy of measures). We argue that a combination of the indicators $R_1 > 1$ and $R_2 < 1$ such that

$$R_3 = \frac{(R_1)^2}{R_2} = \frac{L}{G} \cdot \frac{p}{1-p} \tag{15}$$

provides a consistent basis for trust measurement. The main difference to the residual gain approach is the idea that a small standardized spread of net expected gains, $(1 - R_2) = \frac{pG - (1-p)L}{pG}$, indicates *low* trust. Similarly, a decreasing ratio R_3 of expected loss over expected gain is suggestive of higher trust. The following example clarifies the properties of the trust measure (15). Let potential loss and gain be defined as in the previous section in (8) and (9).[39] Set $k = 1$ again. Then, if expectation of trustworthy behavior amounts to $p = 0.5$, trust measure R_3 obtains as $R_3(p = 0.5) = 1$; for $p = 0.7$ one obtains $R_3(p = 0.7) = \frac{14}{3}$ and $p = 0.9$ leads to $R_3(p = 0.9) = 18$. Figure 2 presents the corresponding plot for the trust measure.

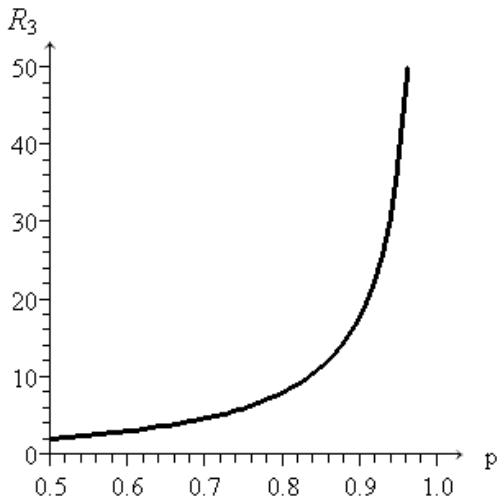


Figure 2: Trust measure R_3 as a function of expectation p of trustworthy behavior

[39] Obviously, if L and G do not depend on expectation p , the measure R_3 as defined in (15) monotonically increases in p .

If A trusts B in an exchange, s/he is willing to accept a higher loss compared to the possible gain (see condition (1)), but the fact that s/he trusts reduces the expected loss. When non-exploitative behavior is not expected at all, p approaches 1 and the expected loss $(1-p)\cdot L$ is reduced to 0 – as does R_2. Consequently, a decreasing ratio R_2 is consistent with increasing trust.[40] The trust measure R_2 in (15) relates expected avoided losses from non-opportunistic behavior to foregone gains from untrustworthy behavior; therefore, we label the trust measurement approach in (15) avoided loss approach. (Note that even if one adopts Coleman's (1990) view of not requiring $R_1 > 1$, the trust measure in (15) could be applied.)

To strengthen the argument, we elaborate on the subtle (and by no means apparent) inconsistency of the residual gain approach underlying trust measurement and its implications for trust research. The residual gain approach assumes that an increasing ratio of expected loss over expected gain indicates higher trust. If this were true, it remains questionable why an upper limit for R_2 is needed (see condition (2)). With R_2 increasing in expectation of non-opportunistic behavior, equality of expected loss and expected gain, $R_2 = 1$, marks the first exchange that is not acceptable; any exchange showing $R_2 > 1$ is not feasible because trust should not lead to naivety (Friedland, 1990). It follows that the upper limit is indispensable and both approaches consequently require one. However, this upper limit implies a *minimum* expectation $\underline{p}$ of non-opportunistic behavior (see condition (3)). From there on, with expectations rising, the relation between expected loss and expected gain must improve, which unambiguously brings about a decreasing indicator R_2. In the formalization of the residual gain approach, R_2 is supposed to increase in the expectation. As a result, one would obtain either $R_2(p \to 1) > 1$ or $R_2\left(p = \underline{p} - \varepsilon\right) < 1$, a contradiction to (2) or (3), or R_2 would be invariant to p , which implies it does not indicate trust which it should. In sum, while the residual gain approach and the avoided loss approach suggested in this paper coincide with respect to preconditions or indicators for trust ($R_1 > 1; R_2 < 1$) and a range of expectations $p \in [\underline{p}, 1)$ that allow for trusting relations, they substantially differ regarding the impact of expectation p on the trust indicator R_2. In our approach, increasing trust is suggested by a decreasing indicator, while it is suggested by an increasing indicator in the residual gain approach. The analysis shows that the trust indicator R_2 unambiguously decreases, whether or not L or G depend on p , and more importantly, regardless of the approach taken. The trust measure R_3 that builds on R_2 then fails to show high trust correctly. In addition, the inconsistency inherent in parameter R_2 remains. Our approach offers a remedy for the implausible findings and generates an increasing measure of trust. (It can readily be verified that $R_3 = \frac{(R_1)^2}{R_2}$ is increasing in

[40] A ratio $R_3 = R_1(1 - R_2)$ would ceteris paribus be decreasing in L and hence cannot provide a consistent basis for trust measurement either.

expectation p .)

5 Discussion and implications

The reader may now ask, why is the point being made rather generally and not specifically about an individual measure of trust? The interpretation of the ratio of expected loss over expected gain measured by the ratio R_2 is pivotal for trust measurement in general. This particularly holds true for changes in the ratio as a consequence of changes in subjective expectations. Any trust measure that builds upon the ratio with the interpretation of the residual gain approach (see section 3.1) is prone to (a) inconsistent results and to (b) incorrect findings with respect to the extent of trust present in an exchange. The analysis in this paper demonstrates the possibility of the occurrence of (a) and (b) in the residual gain approach. For this purpose, it employs a model that uses a very simple though specific measure in order to make this point clear.

It is important to point out that an objectively correct indication of high or low trust confirming the corresponding hypotheses under the residual gain approach does not refute our point. For example, in a static one-shot measurement of trust based on the measure in section 3.1, the residual gain could correctly identify high trust *but* inconsistencies could emerge in the (quite general) individual ratios used to measure trust (see (5)-(7)). Therefore, the avoided loss approach would yield a consistent set of indicators or measures, whereas the residual gain approach would not.

In summary, the analysis in this paper points out that thorough care should be taken when trust measures are designed and/or tested. In this paper, we have demonstrated that trust measurement based on the residual gain approach – to understand a small standardized spread of net expected gains from an exchange as high trust – could suggest implausible values for trust. At the same time, we suggest a different approach to trust – the avoided loss approach – that remedies these implausible findings, meaning it helps researchers better understand a small standardized spread of net expected gains from an exchange as *low* trust. The key point is that changes in subjective expectations which affect the risk/reward ratio of a trusting relationship, such as via implicit incentives or contracts, are reflected in our trust measure. Hence, analytical research could benefit from it by, e.g., determining the responsiveness of trust through comparative statics of contract changes; or experimental research could differentiate between participants who show the same behavior (by investing the same amount in a relationship) but have different expectations or endowments. So both analytical and empirical research must be aware of the distinct interaction of trust and performance. Otherwise, trust measures could indicate that trust is high where it is not and vice versa. However, our argument is a conceptual one that has not been validated in empirical studies. Thus, what is needed is an empirical analysis of when

and why individuals perceive trust as high (or low) and whether their assessments correspond to our conceptual argument.

The implications of our research, however, are apparent. The results of trust measurements are used for a number of different purposes: in order to identify how trust can be built up or how it will be destroyed, how trust is related to cooperation and teamwork, how trust improves employee relations, and many other purposes. The eventual goal, of course, is to answer the question of how trust can be a helpful means of improving a company's efficiency. If the measurement of trust in experimental or analytical research indicates high trust where low trust is present or vice versa, as our analysis of the residual gain approach suggests, recommendations for business decisions based on that research may be counterproductive or will be at least not as effective as they could be given correct identification of high (low) trust.

References

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

Altmann, S., T. Dohmen and M. Wibral (2008). Do the reciprocal trust less? *Economics Letters* 99: 454-457.

Andreoni, J. and J.H. Miller (1993). Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. *The Economic Journal* 103(418): 570-585.

Bachmann, R. and A. Zaheer, eds. (2006). *Handbook of Trust Research*. Cheltenham, Edward Elgar.

Berg, J., J. Dickhaut, K. McCabe (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10(1): 122-142.

Bohnet, I. and R. Zeckhauser (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization* 55(4): 467-484.

Bohnet, I., B. Herrmann, R. Zeckhauser (2010). Trust and the Reference Points for Trustworthiness in Gulf and Western Countries. *Quarterly Journal of Economics* 125(2): 811-828.

Boyle, R. and P. Bonacich (1970). The Development of Trust and Mistrust in Mixed-Motive Games. *Sociometry* 33(2): 123-139.

Burks, S. V., J.P. Carpenter, E. Verhoogen (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization* 51: 195-216.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, Princeton University Press.

Casadesus-Masanell, R. (2004). Trust in Agency. *Journal of Economics & Management Strategy* 13(3): 375-404.

Casadesus-Masanell, R. and D.F. Spulber (2007). Agency Revisited. Working Paper of the Harvard Business School and Northwestern University.

Coleman, J.S. (1990). *Foundations of Social Theory*. Belknap Press: Cambridge.

Cook, K.S. and R.M. Cooper (2003). Experimental Studies of Cooperation, Trust and Social Exchange. *Trust, Reciprocity and Gains from Association: Interdisciplinary Lessons from Experimental Research*. E. Ostrom and J. Walker (eds.). New York, The Russell Sage Foundation Publications: 277-333.

Craswell, R. (1993). On the uses of "trust": Comment on Williamson, "Calculativeness, trust, and economic organization. *Journal of Law and Economics* 36: 487-500.

Cripps, M.W., G.J. Mailath, et al. (2004). Imperfect Monitoring and Impermanent Reputations. *Econometrica* 72(2): 407-432.

Dasgupta, P. (1988). *Trust as Commodity*. In: Gambetta, D. (ed.). Trust: Making and Breaking of Cooperative Relations. Oxford: Blackwell, 49-72.

Deutsch, M. (1958). Trust and suspicion. *The Journal of Conflict Resolution*, 2(4): 265-279.

Dirks, K.D. and D.L. Ferrin (2001). The Role of Trust in Organizational Design. *Organization Science*, 12(4): 450-467.

Dirks, K.T. and D.L. Ferrin (2002). Trust in leadership: meta-analytic findings and implications for research and practice. *Journal of Applied Psychology*, 87(4): 611-628.

Eberl, P. (2004). The Development of Trust and Implications for Organizational Design: A

Game- and Attribution-Theoretical Framework. *Schmalenbach Business Review* 56: 258-273.

Fetchenhauer, D. and D. Dunning (2009). Do people trust too much or too little? *Journal of Economic Psychology* 30(3): 263-276.

Friedland, N. (1990). Attribution of Control as a Determinant of Cooperation in Exchange Interactions. *Journal of Applied Social Psychology*, 20(4): p. 303-320.

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

Hardin, R. (2001). Conceptions and Explanations of Trust. In: Cook, K.S. (ed.). *Trust in society*. New York, Russell Sage Foundation: 3-39.

Hardin, R. (2003). Gaming Trust. In: E. Ostrom and J. Walker (eds). *Trust & Reciprocity: Interdisciplinary Lessons from Experimental Research*. New York, Russell Sage Foundation: 80-102.

Hill, N.S., K.M. Bartol, P.E. Tesluk, G.A. Langa (2009). Organizational context and face-to-face interaction: Influences on the development of trust and collaborative behaviors in computer-mediated groups. *Organizational Behavior and Human Decision Processes* 108(2): 187-201.

Holm, H.J. and A. Danielson (2005). Tropic Trust Versus Nordic Trust: Experimental Evidence From Tanzania And Sweden. *The Economic Journal* 115(503): 505-532.

Hwang, P. and W.P. Burgers (1997). Properties of Trust: An Analytical View. *Organizational Behavior and Human Decision Processes* 69(1): 67-73.

James Jr., H.S. (2002). The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior & Organization* 47(3): 291-307.

Jonker, C.M., J.J.P. Schalken, J. Theeuwes, J. Treur (2004). *Human Experiments in Trust Dynamics*. Lecture Notes in Computer Science (Springer Berlin/Heidelberg) 2995: 206-220.

Kosfeld, M., M. Heinrichs, P.J. Zak, U. Fischbacher, E. Fehr (2005). Oxytocin increases trust in humans. *Nature* (435): 673-676.

Kreps, D.M. and R. Wilson (1982). Reputation and imperfect information. *Journal of Economic Theory* 27(2): 253-279.

Kreps, D.M., P. Milgrom, J. Roberts, R. Wilson (1982). Rational Cooperation in the Finitely

Repeated Prisoners' Dilemma. *Journal of Economic Theory* 27, 245-252.

Luhmann, N. (1989). *Vertrauen. Ein Mechanismus zur Reduktion sozialer Komplexität*, 3rd print, Enge: Stuttgart.

Lukas, C. and J.R. Schöndube (2010). Trust and adaptive learning in implicit contracts. To appear in: *Review of Managerial Science*.

Mailath, G.J. and L. Samuelson (2006). *Repeated Games and Reputations: Long-Run Relationships*. Oxford: Oxford University Press.

March, J.G. (1994). *A Primer on Decision Making: How Decisions Happen*. New York, The Free Press.

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

Möllering, G. (2006). *Trust: Reason, Routine, Reflexivity*. Amsterdam, Elsevier.

Mui, L., M. Mohtashemi, A. Halberstadt (2002). A Computational Model of Trust and Reputation for E-Businesses. Proceedings of the 35rd Hawaii International Conference on System Sciences, Vol. 7.

Nooteboom, B., 2002. *Trust: Forms, Foundations, Functions, Failures and Figures*. Cheltenham: Edward Elgar.

Parkhe, A. and S.R. Miller (2000). The structure of optimal trust: A comment and some extensions. *The Academy of Management Review*, 25(1): 10-11.

Parks, C. D. and L. G. Hulbert (1995). High and Low Trusters' Responses To Fear in a Payoff Matrix. *The Journal of Conflict Resolution* 39(4): 718-730.

Phelan, C. (2006). Public trust and government betrayal. *Journal of Economic Theory* 130: 27-43.

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

Ripperger, T. (1998). *Ökonomik des Vertrauens*. Tübingen, Mohr-Siebeck.

Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality* 35: 651-665.

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

Sally, D. (1995). Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Rationality and Society* 7(1): 58-92.

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

Selten, R. (1978). The chain store paradox. *Theory and Decision* 9(2): 127-159.

Sobel, J. (1985). A Theory of Credibility. *The Review of Economic Studies* 52(4): 557-573.

Tjosvold, D. (1985). Dynamics within Participation: An Experimental Investigation. *Group Organization Management* 10(3): 260-277.

Watson, J. (1999). Starting Small and Renegotiation. *Journal of Economic Theory* 85(1): 52-90.

Watson, J. (2002). Starting Small and Commitment. *Games and Economic Behavior* 38(1): 176-199.

Wicks, A.C., S.L. Berman, T.M. Jones (1999). The structure of optimal trust: Moral and strategic implications. *The Academy of Management Review* 24(1): 99-116.

Williamson, O.E. (1993). Calculativeness, Trust, and Economic Organization. *Journal of Law and Economics* 36(1): 453-486.

Wiseman, T. (2008). Reputation and impermanent types. *Games and Economic Behavior* 62: 190-210.

Yamagishi, T. and M. Yamagishi (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion* 18(2): 129-166.