Rationality and Society

http://rss.sagepub.com

The Dynamics of Contracts and Generalized Trustworthiness

Brent Simpson and Kimmo Eriksson

Rationality and Society 2009; 21; 59
DOI: 10.1177/1043463108099348

The online version of this article can be found at:
http://rss.sagepub.com/cgi/content/abstract/21/1/59

Published by:
SAGE
http://www.sagepublications.com

THE DYNAMICS OF CONTRACTS AND GENERALIZED TRUSTWORTHINESS

Brent Simpson and Kimmo Eriksson

ABSTRACT

Generalized trust, or trust in strangers, has been traced to a wide range of societal benefits. But generalized trust is not sustainable in the absence of widespread generalized trustworthiness, that is, the tendency for strangers to honor trust extended to them. While there has been much work on the origins and consequences of generalized trust, surprisingly little research has addressed the antecedents of generalized trustworthiness. We argue that generalized trustworthiness is negatively affected by prior exposure to a ubiquitous extrinsic motivator of trustworthy behavior, contracts. Specifically, drawing on classic social psychological research on the overjustification effect, we argue that actors previously constrained by contracts will attribute their own 'trustworthy' behavior in those interactions to the contract itself. According to overjustification arguments, this misattribution should lead to a decrease in intrinsic motivations to act trustworthily in interactions where the actor is not constrained by the contract. Results of a new experiment support this argument.

KEY WORDS • contracts • exchange • overjustification • trust • trustworthiness

Introduction

Trust is essential to social life (Rotter 1980), but it is by no means guaranteed. By definition, trust involves the relinquishing of control over one's own welfare to another person who may have an interest in abusing that trust. *Generalized trust*, or trust in strangers, may be especially problematic, due to the absence of relational constraints on the trustee's behavior. Yet, as many researchers have noted, it is exactly this type of trust that leads to broad-scale societal benefits like political and civic engagement (Brehm and Rahn 1997; Sullivan and Transue 1999),

Rationality and Society Copyright © 2009 Sage Publications. Los Angeles, London, New Delhi, Singapore and Washington DC, Vol 21 (1): 59–80.
http://rss.sagepub.com DOI: 10.1177/1043463108099348

economic development (Knack and Keefer 1997), and social order (Putnam 2000). The varied benefits of generalized trust have led to an explosion of research on the topic (e.g. Alesina and La Ferrara 2002; Brehm and Rahn 1997; Cook 2003; Delhey and Newton 2003; Yamagishi and Yamagishi 1994).

But generalized trust is only feasible if there is a high level of *generalized trustworthiness*, the tendency for an actor to honor trust extended by a stranger. As Hardin (2000, 18) puts it,

> It is commonly supposed that widespread trust is, loosely speaking, a public or collective good, especially in political life but also more generally in society and in the economy … This supposition cannot be generally correct. Rather, *generalized trustworthiness* would be collectively beneficial and then correctly acting on the trustworthiness of others would be beneficial not only to the truster of the moment but also more generally to the society.

Interestingly, despite the upsurge of research on generalized trust (e.g. Putnam 2000; Stolle 1998; Yamagishi and Yamagishi 1994), surprisingly little research has been directed at understanding its foundation, *generalized trustworthiness*. One goal of the present research is to begin to bring generalized trustworthiness into the literature.

What leads to trustworthiness toward strangers? Two general classes of explanations are possible (for reviews, see Kollock 1998; Mulder et al. 2006). The first views actors as *intrinsically motivated* to act in a trustworthy way (e.g. Van Lange 1999). The second addresses *extrinsic motivations* for trustworthy behavior (e.g. Horne 2004).

Perhaps the most common extrinsic motivator of trustworthiness is the use of formal contracts (see Malhotra and Murnighan 2002). For centuries, actors have used contracts to mitigate trust and malfeasance concerns in order to reap the benefits of cooperation. Of course, contracts are not used all the time, and for obvious reasons: not only can drawing up contracts be time-consuming and costly; they can also be difficult to enforce.

But contracts that are both readily available and easily enforceable can be powerful means to establishing trust and trustworthiness, potentially even creating positive ‘ downstream effects’ by leading to a habit or setting a norm of trustworthy behavior in future interactions. However, this conclusion ignores a potential byproduct of even the most carefully deployed contracts: as explained in detail below, they can damage intrinsic motivators of trust and trustworthiness (Bohnet et al. 2001; Malhotra and Murnighan 2002; see also Fehr and Rockenbach 2003; Mulder et al. 2006).

Given the extent to which both corporate and individual parties depend on the use of contracts, we believe it is critical to understand

their potential long-term effects. Building on previous work, we ask: what happens when actors historically constrained by (voluntarily imposed) contracts interact outside these constraints? We draw on both classic and contemporary social psychological research on extrinsic and intrinsic motivators to argue that the use of contracts damages generalized trustworthiness.

In contrast to previous work, we show that the detrimental effects of contracts do not require that actors make attributions about others' trustworthiness (as in Malhotra and Murnighan 2002). Nor do the effects require us to assume certain levels of heterogeneity of trustworthy types in the population, or that actors have information on others' trustworthiness (as in Bohnet et al. 2001). Instead, we argue that contracts can damage trustworthiness through a simple self-perception (Bem 1972) process. Thus, our work demonstrates how micro-processes (self-perception and the presence of contracts in dyadic interactions) can impact important macro-social outcomes (generalized trustworthiness and, by extension, generalized trust).

The remainder of this paper is organized as follows: we begin with a brief overview of problems of trust and trustworthiness, focusing on how formal mechanisms like contracts circumvent these problems. We then turn to the literature on intrinsic and extrinsic motivations, and discuss closely related work on the effects of contracts on various issues related to trust and trustworthiness. Building directly on this work, we outline an argument about the dynamic effects of contracts on intrinsic motivations. We then introduce a simple new experimental study to test our key prediction. The results support the argument. We conclude with a discussion of alternative explanations of the results, as well as some implications and directions for future work.

Why Study Trust and Trustworthiness?

Social and economic settings giving rise to questions of trust have two features in common (Rousseau et al. 1998: 395). First, there is some level of interdependence, such that the goals or interests of an actor cannot be realized without relying on another. Second, there exists uncertainty. As Dasgupta (1988) put its, uncertainty creates an opportunity for a test of trust (see also Kollock 1994). Trust, then, is defined as 'a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another' (Rousseau et al. 1998: 395). Trustworthiness, in turn, is defined as positive intentions or

behavior, i.e. one acts trustworthily if one does not abuse trust even when there exists an incentive to do so (e.g. see Hardin 2002). Generalized trust and generalized trustworthiness apply to trust in and trustworthiness toward strangers.

As noted earlier, much theory and research focuses on generalized trust. But generalized trust is not viable without generalized trustworthiness (Hardin 2000). This is because, if actors continually abuse trust (i.e. if untrustworthiness prevails), generalized trust does not pay. The question thus becomes: what leads to generalized trustworthiness?

Explaining Generalized Trustworthiness

To put the question into perspective, consider the distinction between generalized trustworthiness and *personalized trustworthiness*. Rational choice scholars who have assumed that actors are egoistic have had little trouble explaining trustworthiness in repeated exchanges. (See, for instance, Coleman's (1988) discussion of Jewish diamond merchants, and Kollock's (1994) discussion of raw rubber markets in Thailand.) Assuming the future casts a sufficiently long shadow, there may be a strong incentive for those who would otherwise (i.e. in the absence of a shadow of the future) renege on agreements to act in a personalized trustworthy way.

Explaining the antecedents of generalized trustworthiness (or explaining variation in generalized trustworthiness across contexts) presents researchers with a very different theoretical problem. This is because, by definition, one-shot interactions cannot be explained by a shadow of the future, or affective attachments based on prior interactions. In one-shot interactions between strangers, the standard prediction from rational egoist models is that trustees will abuse trust at any opportunity. Further, trusters will anticipate this abuse. Thus, no trust will be extended. (In game-theoretic parlance, not trusting is the subgame perfect Nash equilibrium.) This prediction is problematic because much research shows that trust is often extended and, when trust is extended, it is often honored. That is, strangers often engage in both trusting and trustworthy behaviors in one-shot interactions. The question is why?

Because much trustworthy behavior happens in the absence of formal institutions such as contracts, we can assume that a substantial number of actors are intrinsically motivated to be trustworthy toward strangers at least some of the time. (By extension, because trusting behavior happens in the absence of formal institutions, we can assume that a substantial number of trusters anticipate this trustworthiness.) Yet

the widespread use of formal institutions (such as contracts and sanctioning systems) that create extrinsic incentives shows us that intrinsic motivators may not always be sufficient to motivate generalized trust and trustworthiness.

Can Contracts Damage Trustworthiness?

As noted earlier, contracts are arguably the most ubiquitous means for creating extrinsic motivations for 'trust' and 'trustworthy' behavior.[1] But because interactants typically cannot foresee all possible contingencies, most contracts are incomplete (Bohnet et al. 2001; Rousseau 1995) and, as a result, difficult to enforce.

It seems to follow from the above that complete (and easily enforceable) contracts can provide a substitute for generalized trustworthiness. But classic social psychological work on the *overjustification effect* (e.g. Deci 1971; Greene et al. 1976) suggests that these formal institutions may backfire, leading to a reduction in intrinsic trust and trustworthiness over time.[2] Overjustification occurs when an actor who is intrinsically motivated to perform a given act receives a reward for performing it and/or a punishment for not performing it. Through a self-perception process (Bem 1972), the actor attributes his or her action to the conspicuous extrinsic motivator (reward or punishment). As a result, his or her intrinsic motivation to perform the act decreases. Overjustification effects have been documented for a wide array of phenomena (for a review, see Deci et al. 1999).

An overjustification approach shows why contracts can produce unintended effects. At the most basic level, the *use* of contracts can vary over time, either within given relations (exchange partners may use contracts for one type exchange but not another), or across relations (a person may be constrained by a contract in exchanges with one partner, but not others). Recent empirical work shows that these types of dynamics can have important effects on trust and trustworthiness. For instance, results from an experimental study by Bohnet et al. (2001) suggests that the level of trust and trustworthiness in a given interaction depends on the previous history of contract types, such that contracts with lower historical enforceability predict higher current trustworthiness in subsequent interactions with different partners. Similarly, Malhotra and Murnighan (2002) showed that the removal of contracts in a given relationship can have deleterious effects on *trust* in that particular relationship.

We are interested specifically in how the use of voluntary (and complete) contracts in a given relation affects behavior outside that relation, specifically generalized trustworthiness. We argue that contracts have important dynamic effects that occur through a much simpler process than has been demonstrated in previous work. Specifically, in contrast to Bohnet et al. (2001), we show that the negative effects of contracts do not depend on having heterogeneity in the level of trustworthiness in the population. Most importantly, the negative effects of contracts need not depend on actors having information about others' prior behaviors (as in both Bohnet et al. 2001, and Malhotra and Murnighan 2002). Instead, we predict that a *self-perception process* can lead to reduced generalized trustworthiness among actors voluntarily constrained by contracts in prior interactions.[3]

More generally, we extend previous work in a number of ways. First, while the primary focus of most of the laboratory work has been on the detrimental impact of contracts (Malhotra and Murnighan 2002) or other types of structural solutions (Mulder et al. 2006) on *trust*, we address the effects of structural solutions on trustworthiness, i.e. honoring extended trust. Moreover, whereas most previous research has focused on the effects of structural solutions on behavior vis-à-vis *a given interaction partner,* we focus on the impact of contracts on *generalized trustworthiness*, or trustworthiness toward new, one-shot, interaction partners. As noted earlier, generalized trustworthiness poses a very different type of problem. Thus, there is no reason to expect that arguments developed to explain personalized trustworthiness will apply to the problem of generalized trustworthiness, and there are a number of reasons to expect that they will not.

Our application of the overjustification effect to the problem of generalized trustworthiness is straightforward. As noted earlier, a wide range of studies show that, in the absence of structural solutions or extrinsic motivators, trustworthiness is much higher than would be expected from rational egoist models. Following others (Fehr and Fischbacher 2005), we interpret this result as evidence of substantial intrinsic motivators of trustworthiness. These intrinsic motivations could include altruism (Batson and Shaw 1991), other-regarding emotions (Frank 1988), fairness concerns (Rabin 1993), or a range of other motivations (for a review, see Dovidio et al. 2006).

Whatever the source of these intrinsic motivators of trustworthiness, we argue that extrinsic motivations for 'trustworthy' behavior will erode them. Specifically, we expect that actors who voluntarily enter into binding contracts that require 'fair' or 'trustworthy' behaviors will make

external attributions for their trustworthiness. (Again, we are not claiming that living up to a contract is trustworthy behavior. Instead, the contract provides a substitute for trustworthiness. See endnote 1.) Then, once the contract is removed, these external attributions will lead to untrustworthy behavior. That is, actors will attribute their own 'trustworthy' behavior to the contracts. When the perceived basis for their trustworthy behavior is no longer present, they will act in a less trustworthy way than they would have had their behaviors never been subject to external controls. Thus, we predict negative downstream effects of contracts on generalized trustworthiness:

Hypothesis 1a

> Contracts decrease generalized trustworthiness. That is, actors governed by contracts in previous interactions will be less trustworthy in subsequent interactions not governed by contracts than will actors not previously governed by contracts.

We test Hypothesis 1a against an alternative argument: As suggested earlier, contracts may establish 'norms or 'habits' of trustworthy behavior that will be carried on to subsequent interactions. According to this line of reasoning, repeated exposure to contracts will provide evidence that fair behavior can benefit both trusters and trustees. As a result, the effect of contracts may lead to sustained or positive, rather than negative, downstream effects:

Hypothesis 1b

> Contracts establish habits of trustworthy behavior that persist in subsequent interactions with different partners not governed by contracts. That is, actors governed by contracts in previous interactions will be at least as trustworthy in subsequent interactions not governed by contracts as actors not previously governed by contracts.

The experiment outlined in the section to follow offers a test of these competing predictions.

Design

Participants were recruited from introductory classrooms at a large Southeastern University using the opportunity to earn money as an incentive. A total of 68 students (40 females) participated. There was a single between-subjects factor: whether or not binding contracts governed early interactions.[4]

Settings and Procedures

Participants were scheduled in groups of ten to twelve. Upon entering the laboratory, each participant was escorted to a private subject room. After completing consent forms, they were given instructions (see below), which ensured them that they would not see other participants at any point during or after the study, and that participants would be identified only via letters (e.g. 'person A'). Although participants were told that they would interact with other persons in surrounding subject stations and adjoining rooms, in reality, the choices of others were simulated.

Trustworthiness Measure

After reading and signing consent forms, participants were given instructions (see Appendix). The experimental instructions began by stating that there would be two roles in the study: *Investor* (truster) and *Receiver* (trustee). (At no point during the study did the instructions use loaded terms such as 'trust,' 'truster,' 'trustworthiness.') The instructions stated further that the participant had been randomly assigned to the role of 'receiver.' In reality, all participants acted as receivers. The instructions then proceeded to explain the 'investment scenarios' in which the participant would be involved.

Our measure of trustworthiness is based on the investment game developed by Berg et al. (1995). Since its introduction, the procedure has become one of the most widely used behavioral measures of trust and trustworthiness. The game involves two players, an Investor (truster) and a Receiver (trustee). In our implementation of the game, the Investor (always a fictitious other) was given $10. The Investor ostensibly had to decide whether to invest the $10 in the Receiver (always a participant), or keep the $10. (Because the ostensible investor's decision was binary, it resembles the investor's decision in the trust game (Dasgupta 1988). The investment game is structurally similar to the trust game but both the investor's [truster's] and receiver's [trustee's] choices are continuous rather than binary.)

If the ostensible other did not invest, he or she would have kept the $10 and the Receiver (participant) would have earned nothing. If the Investor did invest (which, for this study, was always the case), the $10 was tripled. Thus, the participant received $30. The participant then had to decide how much of the $30, if any, to return to the Investor. The instructions stated that, following an investment, the participant could return any amount, from $0 to $30. Any amount not returned was the

participant's payoff for that particular investment scenario. Unlike the initial investment, the amount returned was not subject to a multiplier. Thus, for example, imagine that following an investment, the participant decided to return $12. In this case, the ostensible Investor would have earned $12 and the Receiver (participant) would have kept the remaining $18, or $30–$12.

The instructions informed participants that they would make decisions in several investment scenarios, always as a Receiver. We emphasized that each decision would be made with a *different* Investor, and underscored this point by assigning unique participant labels to each Investor with whom the participant ostensibly interacted. (As explained below, participants were told that their payment would be determined by their actions, and the actions of the person with whom they were paired for one of the investment-scenarios.) Following the instructions, participants completed a 'quiz,' designed to ensure that they understood key aspects of the procedures (e.g. that they would never be paired with the same other more than once). Thereafter, research assistants addressed any misunderstanding or confusion. Once the research assistant was certain the participant fully understood the procedure, he or she presented the participant with the materials for the first investment scenario.

Contracts and Control Conditions

The experimental instructions differed according to whether initial investment scenarios were governed by binding contracts. Specifically, instructions for participants in the contracts condition stated that, prior to investing, the Investor for a given decision scenario would be allowed to propose a 'contract' to the participant. A contract was an agreement by the Investor to invest the $10 if the participant agreed to return a specified amount of the resultant $30 back to the Investor. (The return amount was specified by the Investor in the proposed contract.)

Note that, from the participant's perspective, contracts were voluntary for both the participant and the investor who ostensibly proposed the contract. Furthermore, proposed contracts were non-negotiable and binding: receivers could either accept the terms of the contract (i.e. return the amount required by the proposed contract and keep the remaining amount), or decline the contract. If the participant declined the contract, no investment took place. Thus, the ostensible Investor kept the $10 and the participant received nothing for that investment scenario. Contracts lasted for only one investment scenario. Ostensible

Investors in the contracts condition proposed contracts in each of the first two investment scenarios.

We note that because declining a contract always results in an outcome of zero and accepting a contract always results in a positive outcome, a participant who seeks to maximize his or her earnings will always accept a contract. Of course, our situation does not correspond to all real-world situations in which contracts are proposed by one party to another who has an incentive to act in an untrustworthy way. But our goal is not to capture or model the variety of real-world situations. Perhaps more importantly, we believe that the incentive structure that confronts a participant to whom a contract is proposed has a number of parallels in the real world: a person who faces the option of not completing a profitable transaction (because his or her potential exchange partner will not otherwise agree to the transaction) versus completing the transaction with a contract will, *ceteris paribus*, 'rationally' agree to the contract.

We reasoned that, by first presenting a participant in the contract condition with the structure of incentives in the investment game and *then* presenting him or her with the contract, the participant would realize that the investment game posed a trust/trustworthiness problem, and that the contract provided a 'solution' to the problem.

Removal of Contracts

Although the participants were not told the exact number of investment scenarios for which they would make a decision, all participants were involved in a sequence of three investment scenarios. Following the second investment scenario, participants in the contracts condition received instructions stating that the rules would change slightly.[5] Thereafter, for the third investment scenario, participants in the contracts condition made decisions without the possibility of contracts. Thus, as explained more fully below, participants in the two conditions faced identical problems in the third (which was, unbeknownst to participants, the final) investment scenario.

Dependent Measure

Our dependent measure is trustworthiness, measured by the amount returned in the investment game (from 0 to 30), in the third investment scenario. As explained earlier, we expect that trustworthiness will be lower in the contracts condition than the control condition. In

order to develop a fair test of this hypothesis, we need to ensure that the only difference between the control and contract conditions is the presence of the contract proposed by the Investor in the earlier investment scenarios. The question thus becomes how much should the contract stipulate the Receiver (participant) return to the Investor?

Contract Content

To answer this question, we needed to anticipate the average amount that would be returned in the absence of contracts. Previous research (e.g. Glaeser et al. 2000) suggests that trustees return, on average, half the resulting endowment. Thus, we started the study by setting the return amounts requested in the contracts at $15 (out of $30).[6] As we amassed data for the two conditions, we tracked return amounts for the control condition to make sure they did, in fact, average $15. As small deviations from $15 began to emerge in the control condition, we 'yoked' return amounts from that condition to use as inputs (return amount requests) for the contracts condition. (Thus, while the majority of the contracts proposed a $15 return, we also included other values based on return amounts from the control condition.) As a result, as explained below, we have nearly identical average return amounts in the first two investment scenarios of the two conditions. Thus, we should be able to attribute any differences in return amounts in the final investment scenarios of the two conditions to the presence of contracts in prior interactions of the contracts condition.

Payment and Debriefing

Following related work (e.g. Malhotra and Murnighan 2002), the instructions explained that, at the end of the study, one scenario would be picked randomly and that the participant would be paid according to his or her actions (and the actions of the investor) for that scenario. The instructions emphasized that, because the participant's pay could be determined by any given scenario, it was important that they consider each decision very carefully. At the end of the study, all participants were paid $15 (the amount each would be paid if they agreed to the contract when contracts were permitted). Thereafter, the research assistant explained the study in detail and assessed each participant's suspicion using a funnel debriefing procedure. The entire procedure took approximately 45 minutes.[7]

Results

Suspicion Checks and Descriptive Statistics

Three participants (two in the contracts condition and one in the control condition) expressed some doubts about whether Investors were real. Because their suspicions were relatively mild, data from these three participants are included in the results reported below. Importantly, results from analyses with these participants excluded are virtually indistinguishable from those we report. (Results from analyses that exclude suspicious participants are available upon request from the first author.)

As mentioned above, only one participant turned down a contract in one of the first two investment scenarios of the contracts condition. This participant was offered a contract with a return request of $20 in the first and second investment scenarios and refused the contract for the second decision scenario. Because we are interested in the effects of contracts on behavior, we exclude data from this participant. (Note, however, that our substantive conclusions are virtually identical if we include this participant's data.) Table 1 gives average return amounts for each of the three decisions for the remaining participants.

As can be seen in Table 1, the manipulation of return amounts was successful: average return amounts for the first two investment scenarios (i.e. when contracts were present in the contracts condition) are very similar across the two conditions. The question is what happens when we remove binding contracts? More specifically, is generalized trustworthiness lower for actors previously subject to contracts (as suggested by Hypothesis 1a) or do contracts establish 'habits' of trustworthy behavior that persist in subsequent interactions with different partners (Hypothesis 1b)?

Table 1 shows that the average return amount in the third investment scenario is lower in the contract condition ($11.37) than in the control condition ($15.54). Thus, on the surface, these descriptive statistics seem to support our primary hypothesis (Hypothesis 1a). We now turn to statistical analyses designed to assess the effects of contracts on trustworthiness, net of relevant controls.

Analytic Methods and Control Variables

As stated earlier, our dependent variable is trustworthiness (amount returned) in the final (third) round. Our main predictor variable is condition (contracts versus control). We control for the average amount returned in the first two rounds. Findings from the literature on gender

Table 1. Average return amounts (out of 30) in contracts (n = 43) and control (n = 24) conditions (Standard deviations are in parentheses).

Item	*Investment scenario 1*	*Investment scenario 2*	*Investment scenario 3*
Contracts	15.79	15.79	11.37
	(1.73)	(1.73)	(5.18)
Control	15.63	16.04	15.54
	(6.41)	(6.11)	(7.50)

Table 2. Unstandardized coefficients from the regression of return amount in final investment scenario on condition (contracts versus control) and previous return amounts (average return amounts for decisions 1 and 2).

Independent variable	*Coef. (S.E.)*
Contract (coded 1)	–4.131(1.331)*
Previous Return Amounts	.824(.172)**

Note: *$p < 0.005$; **$p < 0.001$. All tests are two-tailed.

and cooperation (Buchan et al. *Forthcoming*; Kuwabara 2006; Simpson 2003) sometimes reveal gender differences in phenomena like trust, trustworthiness, and cooperation. But preliminary analyses showed no effect of gender on decisions. Thus, we exclude gender from further consideration.[8]

The results of our main analysis are given in Table 2. First, note that the control variable (the average amount returned in the first two rounds) is highly significant, $p < 0.001$. This effect is not surprising: It primarily stems from the fact that participants in the control condition who returned more in earlier rounds also returned more in the last rounds. For instance, looking only at decisions in the control conditions, the average return amount variable strongly predicts the return amount in the final decision, $p < 0.001$. Note, however, that we do not observe the same positive effects of contracted return amounts on trustworthiness. For the contract conditions, contract values in investment scenarios 1 and 2 do not predict the return amount in the final decision, $p = 0.36$.

Most importantly for our purposes, net of these effects of earlier decisions, condition significantly impacts generalized trustworthiness, $p < 0.005$. That is, participants in the contracts condition returned significantly less in interactions in which they were no longer bound by contracts than did participants in the control condition, who were never

bound by contracts. This supports our main hypothesis (Hypothesis 1a) about the negative effects of contracts on generalized trustworthiness. The findings do not support the competing hypothesis, i.e. that contracts establish habits or norms, such that trustworthiness remains high in interactions subsequent to the contract.

Discussion

The results of the experiment just outlined support our argument linking prior exposure to contracts to a decrease in generalized trustworthiness. In contrast to previous work (Bohnet et al. 2001; Malhotra and Murnighan 2002), participants in our study were not given information about others' prior behaviors. Thus, our findings could not have resulted from attributions (correct or not) about others' previous levels of trust or trustworthiness. Instead, following work on the overjustification effect (Deci et al. 1999) we have argued that the effects occur through a *self-attribution* process. Of course, such self-attributions can only be demonstrated indirectly: we did not directly measure whether participants in the contracts condition attributed their behavior to the extrinsic motivator and thus gave lower amounts when these contracts were removed. The reason is that actors are unaware of the negative effects of extrinsic motivators on their own (or others') behavior (see, e.g., Gneezy and Rustichini 2000). This is precisely the reason that extrinsic motivators such as contracts have unintended effects.

Limitations and Alternative Explanations

Because of the difficulty of directly demonstrating the cognitive and attribution processes assumed in overjustification arguments, studies designed to demonstrate overjustification effects are often subject to alternative explanations (cf. Carton 1996; Deci et al. 1999). This section reviews potential alternative explanations for the findings presented above. We show that these alternative explanations do not provide as good an explanation of the results as the overjustification account.[9]

First, we address whether it is possible that the contract 'primed' an economic exchange frame, whereas the control condition primed a social exchange or reciprocity frame. That is, people tend to think of contracts as governing economic agreements. Did the contract condition create more of an 'economic decision' frame than the control condition (and, as a result, lead to lower levels of trustworthiness

once contracts were removed)? We first note that, to the extent that contracts do generate economic decision frames, this effect would be worth demonstrating and future research should address this point. But more importantly, we want to emphasize that *both* conditions entailed terminology that would have likely generated an exchange frame, at least among those participants who would have been susceptible to such a frame. This is because both the contract and control conditions used terms such as 'investor,' 'investment decisions,' etc. Thus, we think that there are strong reasons to suspect that the tendency for participants to view the decision-scenarios in terms of an exchange frame would have been relatively constant across conditions. That said, future research might more explicitly manipulate economic vs. social exchange framings.

At a more basic level, is it possible that the rule change implemented after the second decision scenario (in the contract condition) might have led to behavioral differences across conditions? For instance, did the rule change lead to demand effects, such that participants in the contract condition might have suspected that we were interested in the effects of contract removal and, as a result, became 'less trustworthy'? At least two things make this alternative explanation unlikely. First, in the post-experiment debriefing sessions no participant (even those who, as discussed above, were mildly suspicious about the existence of other participants) expressed any indication that he or she knew or suspected the study was about the effect of contract removal.

Additionally, there is no reason (to our knowledge) to suspect that demand effects would have generated the effects we observed. As noted in the previous section, prior research (Gneezy and Rustichini 2000) shows that participants tend to be unaware of the overjustification effect (either as it affects their own or others' behavior). Thus, if anything, we might be more likely to expect the *opposite* pattern if demand effects were driving the results (along the lines of the effect suggested by Hypothesis 1b). In any case, although we believe this 'rule change explanation' is unlikely to account for the results, this is a very important consideration, because 'rule change' is endemic to the problem we are considering.

Summing up, while we believe the results provide important initial support for our overjustification account of the effects of contracts on generalized trustworthiness, more research is clearly needed. For now, we turn to some implications of the arguments and findings presented above.

Implications

There are a number of implications of our work for the development of trust and trustworthiness. For instance, as noted by Bohnet et al. (2001), the use of contracts may be the source of cross-national differences in trusting and trustworthy behavior. To name one example, compared to Americans, Japanese tend to be less trusting of strangers (e.g. Yamagishi and Yamagishi 1994). Yamagishi and his colleagues trace these differences to the greater prevalence of informal monitoring and sanctioning systems that result from Japanese social networks and organizations. These informal institutions create extrinsic motivations for trust and trustworthy behavior but, arguably, inhibit the development of trust and trustworthiness when actors operate outside the monitoring and sanctioning system. It remains to be demonstrated whether these effects occur through a self-perception process, as suggested by the overjustification argument presented above. In any case, an important question for future research is whether the levels of generalized trust and generalized trustworthiness in a given society can be traced to variation in the use of contracts and other extrinsic motivators.

Conclusion

This paper addresses whether the use of contracts can create unintended 'downstream' effects on generalized trustworthiness. As noted earlier, this question is important because generalized trustworthiness is necessary for the development of generalized trust and previous work shows that groups and societies high in generalized trust benefit in a number of ways. Yet very little research has addressed the origins of generalized trustworthiness. This paper explores one pathway through which generalized trustworthiness is affected by prior exposure to contracts.

The results of our simple experiment provide initial support for the argument linking the use of contracts to a subsequent reduction in generalized trustworthiness. In so doing, it echoes findings from previous work on the unintended byproducts of 'top down' solutions to problems of trust and cooperation (e.g. Bohnet et al. 2001; Fehr and Rockenbach 2003; Malhotra and Murnighan 2002; Mulder et al. 2006). Extrinsic motivators can overcome important hurdles to collectively beneficial outcomes. But there now exists much evidence (including the findings presented in this paper) that extrinsic motivators may backfire. An important goal for future work is thus to better understand how actors

who employ or enforce extrinsic motivators might anticipate and thus prevent these negative byproducts.

Appendix: General Instructions for Investments and Contracts

We are interested in how people make decisions in social situations under conditions of limited information. Thus, you will be given only limited information about the other participants in today's study. Similarly, they will be given only limited information about you.

As explained in more detail below, there are two types of roles in today's study – *Investor* and *Receiver. You have been randomly assigned to the role of Receiver*. As a Receiver, you will make several choices in investment-scenarios (explained below). *For each of the investment-scenarios, you will be paired with a different participant (Investor)* located in a different room. *You will never be paired with the same Investor for more than one investment scenario.*

The basic instructions for the scenarios are as follows (if, at any point, you have questions, please feel free to ask one of the research assistants): For each investment-scenario, the Investor will be given a coupon worth $10. The Investor has two options: keep the $10 for himself/herself, or 'invest' it with you. If the Investor invests the $10 with you, it will be tripled. Thus, you will receive $30. You will decide how much of that $30 (if any) you wish to return to the Investor. You can send *any* amount back: from nothing ($0) to everything ($30).

Thus, the amount of money that you and the Investor receive for a given investment-scenario depends on two factors: (1) whether or not Investor invests the $10 with you, and (2) if he or she invests, how much of the $30 you return.

- If Investor does not invest, you receive nothing for that investment-scenario and the Investor receives $10.
- If the Investor does invest (and thus the $10 becomes $30 and is then passed on to you), the amount each of you will receive depends on how much you decide to return. To give a few examples:
- If you returned $10 to the Investor, the Investor would receive $10 and you would receive $20.
- If you returned $15 to the Investor, the Investor would receive $15 and you would receive $15.
- If you returned $20 to the Investor, the Investor would receive $20 and you would receive $10.

{**Instructions for the Control Condition Continued**: There will be no communication between you and the Investor before you make your decisions.}

{**Instructions for the Contracts Condition Continued**: Prior to making his or her decision, the Investor will be able to propose a 'contract' to you. A contract is an agreement between you and the Investor. A contract states that the Investor will invest the $10 with you if you send a specified amount (to be determined by the contract) back to the Investor. Note that contracts are non-negotiable and binding. Either you accept and abide by the terms of the agreement (i.e. you send back the amount the contract requires), or you do not agree to the contract and you earn nothing for that investment-scenario. Note that contracts last for one investment scenario only.

Contracts will work as follows. At the beginning of the investment-scenario, the Investor will be given the opportunity to offer you a contract. A contract states that the Investor will invest the $10 coupon with you if you agree to send back the amount requested (by the Investor) in the contract. For instance, a contract from an Investor might read: "I will invest the $10 coupon with you, if you send back half ($15) of the resulting $30." If you sign the contract and send it back to the Investor, the transfer is automatic and the scenario is complete.}

{Both Conditions Continued}

A few more important things to note.

1) After an investment-scenario is complete, *you will not interact with the Investor from that scenario at any other point during or after the study*. You will be paired with a different participant for *every* investment-scenario. That is, you will not be paired with any other participant more than once in today's study.

{Contract Condition Only}

2) You have been assigned a unique Participant ID. If and when you sign contracts, do not use your name. To maintain confidentiality, use only your Participant ID.
3) You will make decisions in a number of investment-scenarios. Your total payment for today's study will depend on your decision (and the decision of the Investor) in <u>*one*</u> of these investment scenarios. Exactly which of the investment-scenarios you will be paid for will be determined randomly. Because your pay (and the pay of those with whom you are paired) may be determined by your decision in

any given investment-scenario, it is very important that you consider each scenario very carefully.

4) Finally, for some participants, the rules may change after a given number of investment-scenarios. If you are one of these participants, you will receive follow-up instructions later in the study.

Acknowledgments

This research was supported by grants from the National Science Foundation (SES-0240802 and SES-0551895), the Swedish Research Council, and the CULTAPTATION project (European Commission contract FP6-2004-NESTPATH-043434). We thank Nick Berigan, Alex Hirschfeld, Kyle Irwin, Tucker McGrimmon, and Rachael Russell for research assistance.

NOTES

1. We recognize that living up to a contract is not 'trust' or 'trustworthiness.' Instead, (complete) contracts serve as substitutes for these things. Our use of these terms in this context is for brevity.
2. There are a number of theoretical variants of the overjustification effect that are not important for our current purposes. For a discussion, see Deci et al. (1999).
3. Our theoretical focus and empirical work differ from these previous studies in other important ways as well. For instance, Malhotra and Murnighan (2002) study changes in trust within relations (i.e. specific trust), whereas we study the impact of trustworthiness across relations (i.e. generalized trustworthiness). Furthermore, Bohnet et al.'s (2001) primary question is how various types of contracts impact trust and trustworthiness. Meanwhile, we are simply interested in the presence of complete contracts versus no contracts on generalized trustworthiness. In addition, whereas Bohnet et al. impose some type of contract on *all* interactions, we address the impact of voluntarily entered contractual agreements. Finally, in contrast to both these studies, our work does not assume that actors have any information about others' prior behaviors.
4. All procedures were approved by the University Institutional Review Board. Note also that we ran a third condition designed to measure the impact of non-binding notes on trustworthy behavior. Because data from this condition will be used in a separate write-up on 'cheap talk,' we do not discuss them in this manuscript.
5. To decrease the chances that the rule change would be a surprise to participants in these conditions, the instructions read by all participants at the beginning of the study (Appendix) stated that the rules may change later in the study.
6. Besides being the modal return amount in a number of previous studies, there are additional advantages to setting the return amount in contracts at ½ the resulting

endowment. For instance, the amount is likely to be seen by Receivers as fairer than other 'obvious' amounts (e.g. $20). As a result, the amount should be less likely to create a backlash against contracts after they are removed. One-half therefore provides a more conservative test of our hypotheses than alternative values (e.g. $20).

7. We have nearly double the number of participants in the contract ($N = 44$), compared to the control ($N = 24$) condition. This happened for two reasons. First, as explained above, midway into data collection, we needed to yoke information (return amounts for decisions one and two) from the control condition. This allowed us to equate the means across the first two decision scenarios of the treatment and control conditions. Second, and more importantly, we assumed a significant number of participants would reject contracts, which would have reduced the number of usable data points in the contracts condition relative to the control. Thus, we assigned a greater number of participants to the contracts than the control condition. But, as explained below, only one participant (out of 44) rejected a contract. Thus, running more participants in the contract versus control condition turned out to be an unnecessary precaution and led to a larger-than-expected cell size for that condition.
8. More detailed analyses that include controls for gender are available upon request. We also conducted analyses that included an interaction term (contract x average return amounts for decisions 1 and 2). However, given that participants in the contract condition did not select their return amounts, whereas those in the control condition did, this interaction term is not theoretically meaningful. In any case, we do discuss how earlier return amounts affect return amounts in the third decision scenario for the control condition, but not for the contracts condition.
9. We thank an anonymous reviewer, whose careful reading of an earlier version of this manuscript led us to consider the following alternative explanations for our findings.

REFERENCES

Alesina, Alberto and Eliana La Ferrara. 2002. 'Who Trusts Others?' *Journal of Public Economics* 85: 207–34.

Batson, C. Daniel and Laura L. Shaw. 1991. 'Evidence for Altruism: Toward a Pluralism of Prosocial Motives.' *Psychological Inquiry* 2: 107–22.

Bem, Daryl J. 1972. 'Self Perception Theory.' *Advances in Experimental Social Psychology* 6: 1–62.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. 'Trust, Reciprocity, and Social History.' *Games and Economic Behavior* 10: 122–42.

Bohnet, Iris, Bruno S. Frey, and Steffen Huck. 2001. 'More Order With Less Law: On Contract Enforcement, Trust, and Crowding.' *American Political Science Review* 95: 131–44.

Brehm, John and Wendy Rahn. 1997. 'Individual-Level Evidence for the Causes and Consequences of Social Capital.' *American Journal of Political Science* 41: 999–1023.

Buchan, Nancy R., Rachel T. A. Croson, and Sara Solnick. Forthcoming. 'Trust and Gender: An Examination of Behavior, Biases, and Beliefs in the Investment Game.' *Journal of Economic Behavior and Organization.*

Carton, John. 1996. 'The Differential Effects of Tangible Rewards and Praise on Intrinsic Motivation: A Comparison of Cognitive Evaluation Theory and Operant Theory.' *Behavior Analyst* 19: 237–55.

Coleman, James. 1988. 'Social Capital in the Creation of Human Capital.' *American Journal of Sociology* 94: S95–S120.

Cook, Karen S. 2003. *Trust and Distrust in Society*. New York: Russell Sage.

Dasgupta, Partha. 1988. 'Trust as a Commodity.' In *Trust: Making and Breaking Cooperative Relations*, pp. 49–72, ed. D. Gambetta. New York: Blackwell.

Deci, Edward L. 1971. 'Effects of Externally Mediated Rewards on Intrinsic Motivation.' *Journal of Personality and Social Psychology* 18: 105–15.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. 'A Meta-Analytic Review of Experiments Examining Effects of Extrinsic Rewards on Intrinsic Motivation.' *Psychological Bulletin* 125: 627–68.

Delhey, Jan and Kenneth Newton. 2003. 'Who Trusts? The Origins of Trust in Seven Societies.' *European Societies* 5: 93–137.

Dovidio, John F., Jane Allyn Piliavin, David A. Schroeder, and Louis A. Penner. 2006. *The Social Psychology of Prosocial Behavior.* London: Lawrence Erlbaum Associates.

Fehr, Ernst and Bettina Rockenbach. 2003. 'Detrimental Effects of Sanctions on Human Altruism.' *Nature* 422: 137–40.

Fehr, Ernst and Urs Fischbacher. 2005. 'Human Altruism – Proximate Patterns and Evolutionary Origins.' *Analyse und Kritik* 27: 7–47.

Frank, Robert H. 1988. *Passions Within Reason*: *The Strategic Role of the Emotions*. New York: W. W. Norton & Company.

Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. 2000. 'Measuring Trust.' *The Quarterly Journal of Economics* 115: 811–46.

Gneezy, Uri and Aldo Rustichini. 2000. 'Pay Enough Or Don't Pay At All.' *The Quarterly Journal of Economics* 115: 791–810.

Green, David, Betty Sternberg, and Mark R. Lepper. 1976. 'Overjustification in a Token Economy.' *Journal of Personality and Social Psychology* 34: 1219–34.

Hardin, Russell. 2000. 'Trust in Society.' In *Competition and Structure* (pp. 17–46), ed. G. Galeotti, P. Salmon, and R. Wintrobe. New York: Cambridge University Press.

Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.

Horne, Christine. 2004. 'Collective Benefits, Exchange Interests, and Norm Enforcement.' *Social Forces* 82: 1037–62.

Knack, Stephen and Philip Keefer. 1997. 'Does Social Capital have an Economic Payoff? A Cross-Country Investigation.' *The Quarterly Journal of Economics* 112: 1251–88.

Kollock, Peter. 1994. 'The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust.' *American Journal of Sociology* 100: 313–45.

Kollock, Peter. 1998. 'Social Dilemmas: The Anatomy of Cooperation.' *Annual Review of Sociology* 24: 183–214.

Kuwabara, Ko. 2006. 'Nothing to Fear But Fear Itself: Fear of Fear, Fear of Greed, and Gender Effects in Two-Person Asymmetric Social Dilemmas.' *Social Forces* 84: 1257–91.

Malhotra, Deepak and J. Keith Murnighan. 2002. 'The Effects of Contracts on Interpersonal Trust.' *Administrative Science Quarterly* 47: 534–59.

Mulder, Laetitia, Eric Van Dijk, David De Cremer, and H. A. M. Wilke. 2006. 'Undermining Trust and Cooperation: The Paradox of Sanctioning Systems.' *Journal of Experimental Social Psychology* 42: 147–62.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.

Rabin, Matthew. 1993. 'Incorporating Fairness into Game Theory and Economics.' *American Economic Review* 83: 1281–1302.

Rotter, Julian B. 1980. 'Interpersonal Trust, Trustworthiness, and Gullibility.' *American Psychologist* 35: 1–17.

Rousseau, Denise M. 1995. *Psychological Contracts in Organizations.* New York: Sage Publications.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt and Colin Camerer. 1998. 'Not So Different After All: A Cross-Discipline View of Trust.' *Academy of Management Review* 23: 393–404.

Simpson, Brent. 2003. 'Sex, Fear, and Greed: A Social Dilemma Analysis of Gender and Cooperation.' *Social Forces* 82: 35–52.

Stolle, Dietlind. 1998. 'Bowling Together, Bowling Alone: The Development of Generalized Trust in Voluntary Associations.' *Political Psychology* 19: 497–525.

Sullivan, Michael J.L. and John E. Transue. 1999. 'The Psychological Underpinnings of Democracy: A Selective Review of Research on Political Tolerance, Interpersonal Trust, and Social Capital.' *Annual Review of Psychology* 625–50.

Van Lange, Paul A.M. 1999. 'The Pursuit of Joint Outcomes and Equality in Outcomes: An Integrative Model of Social Value Orientation.' *Journal of Personality and Social Psychology* 77: 337–49.

Yamagishi, Toshio and Midori Yamagishi. 1994. 'Trust and Commitment in the United States and Japan.' *Motivation and Emotion* 18: 129–66.

BRENT SIMPSON is associate professor of sociology at the University of South Carolina, and is currently a Wenner-Gren Fellow and a visiting researcher at the Center for the Study of Cultural Evolution at Stockholm University. Most of his research focuses on the antecedents and consequences of prosocial behavior.

ADDRESS: Department of Sociology, University of South Carolina, Columbia, SC 29208, USA [email: bts@sc.edu]

KIMMO ERIKSSON is professor of applied mathematics at Mälardalen University and the Center for the Study of Cultural Evolution at Stockholm University. His area of research is cultural and behavioral dynamics, investigated through mathematical modeling and laboratory experiments.

ADDRESS: Department of Mathematics and Physics, Mälardalen University, SE-721 23, Västerås, Sweden [email: kimmo.eriksson@mdh.se].