Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

SOEPpapers
on Multidisciplinary Panel Data Research

154

Ernst Fehr

On the Economics and Biology of Trust

Berlin, December 2008

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

On the Economics and Biology of Trust[*]

Ernst Fehr
University of Zurich - Institute for Empirical Research in Economics (IEW);
Massachusetts Institute of Technology (MIT) - Department of Economics;
Centre for Economic Policy Research (CEPR);
CESifo (Center for Economic Studies and Ifo Institute for Economic Research);
Institute for the Study of Labor (IZA)

ABSTRACT

In recent years, many social scientists have claimed that trust plays an important role in economic and social transactions. Despite its proposed importance, the measurement and the definition of trust seem to be not fully settled, and the identification of the exact role of trust in economic interactions has proven to be elusive. It is still not clear whether trust is just an epiphenomenon of good institutions or whether it plays an independent causal role capable of shaping important aggregate economic outcomes. In this paper, I rely on a behavioral definition of trust that enables us to relate it to economic primitives such as preferences and beliefs. I review strong biological and behavioral evidence indicating that trusting is not just a special case of risk-taking, but based on important forms of social preferences such as betrayal aversion. Behaviorally defined trust also opens the door for understanding national and ethnic trust differences in terms of differences in preferences and beliefs, and it suggests ways to examine and interpret a causal role of trust.

Contact:
Ernst Fehr
Institute for Empirical Research in Economics
University of Zurich
Blümlisalpstrasse 10
CH-8006 Zürich
Switzerland
E-mail: efehr@iew.uzh.ch

---

1. Introduction

Trust plays a role in almost all human relationships. It permeates friendship relations, family relations, and economic relations. People rely on the support of their friends, children trust their parents, and sellers trust their buyers to pay the bill. Thus, intuitively speaking, a social scientist has good reason to be interested in "trust" as a concept. Trust also seems particularly important in economic exchanges because it seems obvious that the absence of trust among the trading partners severely hampers market transactions.

During the last decade, there has been a surge of empirical research on trust. The development of experimental tools for measuring trust (Berg, Dickhaut and McCabe 1995, Camerer and Weigelt 1988, Fehr, Kirchsteiger and Riedl 1993) and its determinants (Bohnet and Zeckhauser 2004; Bohnet et al. 2008; Eckel and Wilson (2004); Schechter (2006), Houser, Schunk and Winter 2008), as well as the availability of survey measures of trust in international panel data sets greatly facilitated research on trust. They enabled both the analysis of the impact of institutions on trust (e.g., Bohnet and Huck 2004; Brown et al. 2004) and cross national comparisons of trust (Buchan, Croson and Dawes 2002, LaPorta, et al. 1997, Naef, et al. 2008). In addition, aggregate measures of trust at the country level have been related to important economic variables such as GDP growth, inflation, or the volume of trade between countries; several papers suggest that trust may be an important determinant of these variables. LaPorta and coauthors (LaPorta, et al. 1997) show that a larger share of trusting people is negatively correlated with inflation rates and positively correlated with GDP growth across countries. Knack and Keefer (Knack and Keefer 1997) report positive correlations between a measure of trust and a country's average annual GDP growth rate between 1980 and 1992. Zak and Knack (Zak and Knack 2001) document similar correlations. More recently, Guiso, Sapienza and Zingales (Guiso, Sapienza and Zingales 2009) show that higher bilateral trust between two countries is associated with more trade between the countries. In addition, this effect is stronger for more trust intensive goods. The same authors (Guiso, Sapienza and Zingales 2004, 2008) also provide microeconomic evidence on the role of trust in financial markets. In their 2008 paper, for example, they document that less trusting individuals are less likely to buy stock, and when they do so, they buy less of it. The authors suggest that lack of individual trust in the stock market could partly explain the "participation puzzle", that is, why so few people take advantage of the existence of a stock market.

1

Despite the wave of interesting and intriguing results on the role of trust, many unanswered questions remain. There is no consensus on the proper definition of trust and it is unclear in some of the literature how trust can be related to more fundamental economic primitives such as preferences and beliefs. Both the proper measurement of trust and the relationship between different measures of trust also remain open issues. And, above all, it is unclear whether trust plays an independent and sustainable causal role for important economic outcomes such as the trading volume, the gains from trade, or the overall welfare of social groups.

In this paper, I use a behavioral definition of trust that is based on Coleman (Coleman 1990). This definition is tightly connected with economic primitives such as preferences and beliefs. I document the recent accumulation of strong evidence – neurobiological (Baumgartner, et al. 2008, Kosfeld, et al. 2005), genetic (Reuter, et al. 2009) and behavioral (Bohnet, I., et al. 2008, Bohnet, I and Zeckhauser 2004, Cox 2004, Hong and Bohnet 2007) – that trusting cannot be captured by beliefs about other people's trustworthiness and risk preferences alone, but that social preferences play a key role in trusting behavior. Betrayal aversion as documented in the work of Bohnet and coauthors seems to play a particularly important role in trusting behavior. Betrayal aversion indicates an important departure from how economists have viewed decision-making under risk in the past because it suggests a fundamental distinction between risk constituted by asocial factors and that based on interpersonal interactions. Intuitively speaking, people are more willing to take risk when facing a given probability of bad luck than to trust when facing an identical probability of being cheated. Betrayal aversion is thus a major additional inhibitor of trusting behavior that adds importance to property rights and contract enforcement institutions – regardless of whether they are legally or informally constituted.

The behavioral definition of trust provides a neat organizing principle for the very reason that it enables the researcher to provide an account of trusting behavior in terms of beliefs about others' trustworthiness, and in terms of risk preferences and social preferences. I applied this definition to examine the determinants of trust as measured in surveys. The results indicate that survey measures of trust are predicted by risk preferences and measures of betrayal aversion as much as behavioral measures of trust that are derived from the trust game. This finding suggests that survey measures do not just capture beliefs about people's trustworthiness but are also influenced by their preferences. In addition, I show that preference measures do not affect beliefs about others' trustworthiness in a trust game. Thus,

if one needs a clean measure of beliefs in trustworthiness, one should ask respondents directly about the trustworthiness of trustees in the trust game.

The paper also includes a short section which indicates that the proposed concept of trust helps us understand differences in trust across countries and ethnicities. Here I rely on the work of Naef and coauthors (Naef, et al. 2008), documenting a substantial trust gap between the US and Germany which can be fully explained by differences in the preferences and beliefs in the two populations.

Finally, I demonstrate that the belief component of trust is a notoriously endogenous variable affected by the prevailing informal institutions. This fact provides the starting point for the discussion about the causal effects of trust. I argue that although it seems possible that optimistic beliefs about other people's trustworthiness have an independent role in causing different long-term outcomes – relative to pessimistic initial beliefs – there is yet no fully convincing evidence proving this claim. In this context, I discuss some of the papers which examined the causal role of trust in field data. In my view, economists still lack instrumental variables for trust that support causality claims beyond doubt. Thus, the most important open questions in research on trust seems to be the lack of fully convincing evidence supporting the notion that changes in trust cause sustainable changes in important economic variables. But in my view, it is exactly this piece of evidence that seems to be needed to justify social scientists massive interest in trust.

2. What is trust, and how can we measure it?

An individual (let's call her the trustor or investor) trusts if she voluntarily places resources at the disposal of another party (the trustee) without any legal commitment from the latter. In addition, the act of trust is associated with an expectation that the act will pay off in terms of the investor's goals. In particular, if the trustee is trustworthy the investor is better off than if trust were not placed, whereas if the trustee is not trustworthy the investor is worse off then if trust were not placed.

This definition of trust is based on Coleman (1990) and defines trust as a behavior. Trust defined in this way can be neatly examined using experimental one-shot games played anonymously, such as the gift exchange game (Fehr, Kirchsteiger and Riedl 1993), discussed in more detail in Section 7, or the trust game (Berg, Joyce, Dickhaut and McCabe 1995). Take, for example, the binary version of the trust game Bohnet and Zeckhauser (2004) use, which will play a role in the next section. In this game, both players have an initial

endowment of $10 and the investor only has two choices – to send her endowment to the trustee or to keep it. If the investor sends the money, it is doubled so that the trustee receives $20, giving him an overall amount of $30. The trustee also has only two choices – to send back $15 or to send back only $8. Thus, this game is capable of capturing behavior that meets our definition of trust. If the investor sends his endowment to an anonymous trustee she voluntarily places resources at the trustee's disposal without any real commitment from the latter. And if the trustee is trustworthy, the act of trust increases the investor's payoff from $10 to $15, while if the trustee is greedy (i.e., not trustworthy) the investor earns less than the initial endowment.

This game also illustrates why a behavioral definition of trust captures the essence of trust, which consists of the investor's willingness to make herself vulnerable to others' actions (Hong and Bohnet 2007), whereas a purely belief-based definition of trust misses this point. Suppose, for example, that two people, A and B, differ in their beliefs about the trustee's trustworthiness. In particular, A believes that the trustee will pay back $15 with probability 0.3 while B believes that this will occur with probability 0.4. Assume further that A makes herself vulnerable by sending her endowment, while B keeps her endowment. If the essence of trust consists in the willingness to make oneself vulnerable to others' actions it makes sense to say that A trusts while B does not trust. However, a purely belief-based definition of trust would express the opposite. Thus, our intuitive notion of trust is one that is associated with the act of trusting.

The above argument obviously does not mean that beliefs play no role in trust. If trust is a behavior involving trusting acts, then it is shaped by our beliefs about others' trustworthiness as well as our willingness to accept the risks involved in trusting acts. Thus, people's risk preferences should play a role in trusting behavior – a proposition that has received empirical support in recent papers (Karlan 2005, Schechter 2006).[2]

Although trust is best captured by behavioral measures of trust, no such measures are available in many cases, and researchers have to rely on survey measures of trust. The most frequently used measure is taken from the American General Social Survey (GSS), which has measured trust annually since 1972, and the World Values survey (WVS) which has been widely used to measure cross-cultural differences in trust. Both surveys capture trust using the following question: *Generally speaking, would you say that most people can be trusted or*

---

[2] We should add here that not all studies have found a link between risk preferences and trusting behavior (Eckel and Wilson (2004), Houser, Schunk and Winter (2008)). We will comment on these findings in the next section.

that you can't be too careful in dealing with people? The survey respondents can answer in a binary way to this question by agreeing either with "*Most people can be trusted*" or with "*Can't be too careful*".

Note that the question is very close to asking people about their behavioral inclinations ("can't be too careful"), and it seems quite likely that when people answer this question, they consult either their own experiences and behaviors in the past or introspect how they would behave in situations involving a social risk. Therefore, it seems likely that the answer to the GSS and WVS question is not only shaped by people's beliefs about others' trustworthiness, but also by their own preferences towards taking social risks. In fact, one problem with this trust measure is that it seems perfectly possible to agree with *both* answer categories for the very reason that beliefs and preferences both influence trusting behavior. A recent study (Miller and Mitamura 2003) pointed out that a risk averse or cautious person may share the view that "most people can be trusted" but that at the same time prudence or risk aversion may induce the person to say "Can't be too careful" because the person is unwilling to accept small probability risks that have large payoff consequences. Miller and Mitamura (2003) show that this "confound" is highly problematic when this question measures trust across cultures.

To rule out having reasonable people agree with both answer categories, Miller and Mitamura propose "one dimensional" questions that directly distinguish between trust and distrust such as "Do you think that most people can be trusted?" with answer categories on a 7-point Likert Scale from "not at all" (coded as 1) and "complete trust" (coded as 7). Similarly, the German Socio-Economic Panel (SOEP) asks questions such as "In general, one can trust people" with four answer categories (agree fully, agree somewhat, disagree somewhat, disagree fully). In view of the problems inherent in the GSS question, these new measures of trust are likely to be better.

One potential drawback of behavioral trust measures taken from the trust game is that the investor may send money for purely altruistic reasons (Cox 2004). These transfers might not be "trusting", although they place resources at the disposal of another party without any real commitment because the transfers are not associated with an expectation of a back transfer that renders the investor better off. There is little information about the strength of

altruistic motives behind first mover transfers in the trust game[3], but it does not seem all that plausible that a first mover who knows nothing about the anonymous second mover except that he has the *same* monetary endowment has altruistic feelings towards the unknown second mover.[4] Nevertheless, controlling for altruistic motives seems advisable (see section 4) because they might affect investors' behavior. Controlling for altruism makes sense even if one only uses survey measures of trust because if survey respondents derive their answers from introspection into their own likely behaviors in situations requiring trust, then one cannot rule out that altruism also affects survey measures of trust. If altruism affects behavioral measures of trust then it also seems likely it will affect survey measures of trust. In Section 4, I will examine this conjecture in more depth but before I can do this, a more complete description of the different types of preferences that affect trusting is needed.

3. Is trusting just a special case of risk taking?

From the viewpoint of the standard economic model, it is tempting to model situations involving trust just like situations involving risk or ambiguity. For example, a rational and self-interested investor in the binary trust game of Bohnet and Zeckhauser (2004) forms probability beliefs about the trustee's actions and chooses "trust" if the expected utility of the trusting act is larger than the expected utility of "distrust". From the perspective of the standard model, the source of the risk is completely inessential, i.e., it does not matter whether the risk is constituted through the uncertain behavior by the trustee, through a random mechanism that determines the trustee's behavior, or some other source of randomness. However, as I will show below, there is now strong neurobiological as well as behavioral evidence indicating that this view is untenable.

---

[3] Cox (2004) argues that one can derive a pure measure of behavioral trust by taking the difference between the first mover transfer in a trust game and what the person voluntarily gives in an otherwise identical dictator game. It is not implausible, however, that the trust game puts subjects in a very different mental frame compared to the dictator game. While the trust game is likely to trigger a social exchange frame ("If I trust you and you are trustworthy we are both better off") the dictator game may trigger a helping or generosity frame absent in the trust game. Therefore, the difference between the transfers in the two games may understate behavioral trust.

[4] There is evidence that a special form of altruistic preferences – the surplus-maximization motive – plays a role in certain circumstances (Charness and Rabin 2002, Engelmann and Strobel 2004, Fehr, Naef and Schmidt 2006). This motive may directly induce first movers to send money in the trust game because the transfer is usually doubled or tripled. However, it is easy to remove the impact of this motive on investors' behavior by implementing a game in which the first-mover can only make a lump-sum transfer while the second-mover's transfer increases the total surplus available to the two players. The gift exchange game, described in more detail later, has exactly these features. It rules out that the investor's behavior is driven by his surplus-maximization motive. While it may still be the case that the investor anticipates the trustee's surplus-maximization motive, this does not pose a problem because rational trust should be based on a full assessment of the trustee's motives. Thus, the gift exchange game may be a better measure of trust than the trust game because it rules out the investors' surplus-maximization motive.

3.1. Neurobiological Evidence

Kosfeld et al. (2005) conducted a version of the trust game with two groups of subjects – one group inhaled a spray containing the uniquely mammalian neuropeptide oxytocin, while the other group inhaled a placebo spray. The rationale for this experiment originates in evidence indicating that oxytocin plays a key role in certain prosocial approach behaviors in non-human mammals. For example, in prairie voles, one of the important animal "models" in neuropeptide research, oxytocin and vasopressin (a closely related neuropeptide) seem crucial in pair bonding and monogamy(Insel and Young 2001). In their influential survey Insel and Young (2001) write:

> *"Indeed, all the major behavioral aspects of monogamy can be facilitated in the prairie vole by central injections of either oxytocin or vasopressin even in voles that do not have the opportunity to mate. Conversely, these behaviors are inhibited by either oxytocin or vasopressin antagonists given to prairie voles just before mating. … Thus, in monogamous prairie voles oxytocin and vasopressin seem to be necessary and sufficient for pair bond formation".*

The formation of a long-term pair bond – sometimes tantamount to marriage in humans – can be viewed as a decision that involves substantial relation-specific investments and is therefore subject to risks because these investments are of little value outside the relationship. Thus, these neuropeptides might make animals – and people – more willing to take such social risks.

Another example of oxytocin's role in prosocial approach behavior comes from rats. In rats, as in many mammals, females have a natural avoidance of neonates, in particular of neonate odors. The onset of maternal behavior thus requires overcoming these natural tendencies to reject the newborn. Interestingly, lesions that inhibit olfactory processing in female rats facilitate the onset of maternal care (Fleming and Rosenblatt 1974), as does oxytocin. The neuropeptide is released centrally during parturition and is thought to decrease the firing rate of cells in the bulb, which then decreases olfactory processing. These results led researchers to hypothesize that oxytocin may be important for the transition from avoidance to approach of the young (Insel and Young 2001). Note that the willingness to provide maternal care involves high costs and risks because it makes the female animal more vulnerable to predators.

Based on the animal literature, Kosfeld et al. (2005) hypothesized that oxytocin might cause humans to exhibit more behavioral trust as measured in the trust game. In principle, it could have also been possible that the neuropeptide causes higher backtransfers in the trust game. For this reason, both investors and trustees were randomly assigned to a placebo and an oxytocin group. The subjects received placebo or oxytocin with a nasal spray because it is known that neuropeptides gain access to the brain by this method (Heinrichs and Domes 2008). Each investor participated in four independent trust games with four different trustees. Feedback about the trustees' behavior was only given at the end of the fourth game. The investors could invest 0, 4, 8 or 12 money units in the trustee. The amount sent was tripled and added to the trustees' endowment of 12. Then the trustee could send back whatever he wanted to. Figure 1a shows that the investors in the oxytocin group indeed sent more money to the trustees. The percentage of subjects who show maximal trust in all four games is, in particular, considerably higher in the oxytocin group.

This result raises several questions. First, did oxytocin increase subjects' trust by creating more optimistic beliefs? Second, did it make subjects generally more prosocial, that is, did it increase trust as well as the trustees' trustworthiness, or did it only increase trust? Figure 1b shows that oxytocin had no effect on investors' beliefs. In fact, for any given transfer level, investors' average belief about the trustees' back transfer is higher in the placebo condition, although the difference is not significant.

If beliefs remain unaffected, it must be the case that oxytocin influences behavior by affecting subjects' preferences. However, oxytocin did not make the trustees more willing to send back money, which rules out that it makes people indiscriminately more prosocial (see Figure 1c). Therefore, the higher level of trust the investors exhibit is unlikely to be the result of a generally higher level of niceness or generosity because this should also have raised the back-transfers.

Insert Figures 1a – 1c here

But since risk preferences are predicted to shape trusting behavior, oxytocin also might have made people more risk seeking. To check this possibility, Kosfeld et al. conducted risk experiments that were identical to the trust experiment except for the fact that there was no

trustee. Instead, subject's faced a computer which mimicked the trustees' behavior in trust games played previously. The subjects in the risk experiment had exactly the same investment opportunities as in the trust game and they received vague (ambiguous) information about the likely behavior of the computer, information that was intended to mimic the vagueness of investors' information in the trust game. The results of the risk game showed that oxytocin did not affect the investors' behavior, suggesting that the neuropeptide affects some other preference component than risk or ambiguity preferences. A second risk experiment reinforces this conclusion. After the end of the risk game isomorphic to the trust game, subjects made 12 decisions involving binary lotteries that varied the probability of winning and the prizes in case of winning or losing. Kosfeld et al. found no significant difference between the placebo and the oxytocin group in any of the 12 lotteries.

These results suggest that neurophysiological mechanisms affect subjects' preferences in the trust game, but that these preferences are distinct from those towards risk or ambiguity. In other words, trust taking is not just a special case of risk taking or decision making under ambiguity. There must thus be some preference components that the concept of risk or ambiguity preferences cannot capture. One natural candidate for the missing preference type is social preferences because some categories of social preferences imply a special aversion towards trusting. Inequity aversion (Fehr and Schmidt 1999), for example, implies that the inequality the investor experiences if he faces an untrustworthy trustee reduces his utility beyond that merely associated with the reduction in income. Likewise, theories of reciprocity (Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006) imply that subjects derive extra disutility from non-reciprocated trust. Thus, these theories of social preferences imply a kind of special aversion against being a sucker or being exploited by untrustworthy partners. Kosfeld et al., therefore, tentatively conclude that oxytocin affects the neurophysiological mechanisms that underlie these kinds of social preferences.

The interpretation that trust taking involves neurophysiological mechanisms (or neural activity) beyond those needed to account for risk and ambiguity preferences received further support in a follow-up study (Baumgartner, et al. 2008). This study examines the differences in the neural circuitry involved in trust and (isomorphic) risk taking in subjects with placebo and oxytocin. In contrast to Kosfeld et al., the new study has a within subject design and focuses on subjects' trust and risk taking *after* they experienced relatively low back transfers in the trust game or low investment success in the risk game. The study shows that oxytocin selectively deactivates neural circuitry known to be involved in fear processing in the trust

game (relative to the risk game). More specifically, oxytocin deactivates neural activity in the amygdala and midbrain regions in the trust game relative to the activation levels observed in the risk game. In addition, this deactivation in fear circuitry is associated with a differential behavioral response to the meager feedback information in the trust and risk game. While the subjects in the risk game reduce their investments after they experienced limited investment success, subjects in the trust game (who received exactly the same feedback regarding the payoffs associated with their investments) keep their investment at pre-feedback levels. Thus, as in Kosfeld et al. (2005), oxytocin selectively affects trust taking, but we can also infer from the new study that this behavioral effect is associated with selective deactivation in fear circuitry. This study, therefore, supports the notion that social risk taking (i.e., trusting) involves neural mechanisms that go beyond what is needed for taking asocial risks.

Reuter et al. (2009) provide another fascinating piece of neurobiological evidence. These authors examined behavior in a trust and a risk game. They show that subjects who have a particular variant of the oxytocin receptor gene exhibit more trust taking than those who exhibit the alternative variant of the gene. In addition, subjects with the trust enhancing gene variant do not behave differently in the risk game, nor do they display more trustworthiness, suggesting that the gene variant does not generally increase prosocial behavior. The results of Reuter et al. suggest that the distinction between social and asocial risk taking is indeed fundamental, as it even seems to be encoded in our genes. These results are also consistent with recent findings in a twin study (Cesarini, et al. 2008) that suggest that trust is partially heritable.

3.2. Behavioral Evidence

The conclusion that trust taking differs significantly from behavior towards non-socially constituted risks also receives strong support from recent papers by Bohnet and Zeckhauser (2004) and Bohnet et al. (2008). The main tool in their studies is the binary trust game described in section 2. But instead of simply allowing the investors to make a decision to trust or distrust, they elicit the investors' minimum acceptance probability (MAP), which provides a much more complete picture of their behavior in the trust game. The MAP is the minimum percentage of trustees who need to choose the trustworthy action to induce the investor to trust. Thus, if the investor chooses a MAP of 0.7 but the actual share of trustworthy trustees turns out to be less than 0.7, the chosen MAP implies that the investor will not trust. If, instead, the actual share of trustworthy trustees exceeds 0.7, the chosen MAP implies that the

investor trusts. Note that telling the truth in this game is a weakly dominant strategy for the investors because they cannot gain anything from misrepresenting the MAP relative to their true MAP, while misrepresentation can result in losses because they do not know the actual share of trustworthy trustees. If, for example, a subject's true MAP is 0.4 while he or she states a MAP of 0.6, his or her expected utility is reduced if the actual share of trustworthy people turns out to be 0.5.

To control for risk aversion, Bohnet and coauthors also collect data from a non-interactive decision problem with exactly the same material payoffs as in the trust game. The only difference is that there is no trustee in the decision problem who makes a decision and earns payoffs. Instead, the investor simply faces the choice between a sure payoff of 10 or a lottery $L = \{15$ with probability p and 8 with $(1-p)\}$. Similar to the trust game, the investors have to state a MAP, which in the decision problem is just the minimum probability with which the payoff of 15 has to occur in order for the subject to be willing to choose the lottery. Since a risk neutral investor will state a MAP of 2/7, the deviation of the actually stated MAP from 2/7 provides information about investors' risk preferences.[5] In particular, if subjects' MAP in the decision problem (MAP^{DP}) is larger than 2/7, the subject is risk averse.

If subjects' MAP in the trust game (MAP^{TG}) is higher than MAP^{DP}, we can infer that there are trust inhibiting preference components that cannot be captured by risk preferences alone. Bohnet et al. (2008) conducted these experiments in 6 different countries – Brazil, Oman, China, Switzerland, Turkey, and the US. The average MAP^{DP} in all countries is significantly larger than 2/7, indicating that risk aversion plays an important role in the decision problem. However, we also observe in all countries that the average MAP^{TG} is higher than the average MAP^{DP}, suggesting that risk preferences cannot tell the whole story in the trust game.

To gain deeper insights into the nature of the additional preference component that inhibits trusting behavior, the authors also conducted a so-called risky dictator game. This game is identical to the trust game, except that a random mechanism, rather than the trustee him or herself, determines the trustee's choice. Thus, there is still a trustee in this game that receives the payoffs but the trustee cannot make a decision; instead a random mechanism makes the "choice". This means that outcome-based social preferences can still play a role in

---

[5] The expected utility of L is $15p + 8(1 – p)$, which is equal to 10 if $p = 2/7$. Subjects did not know the actual probability p with which the outcome 15 was chosen. Due to this uncertainty, stating the true MAP in the decision problem is incentive compatible.

the risky dictator game but betrayal aversion, i.e., the pure disutility of experiencing or anticipating non-reciprocated trust, is ruled out because the trustee has no opportunity to reciprocate in this game. The difference between MAP^{TG} and the MAP in the risky dictator game (MAP^{RDG}) is, therefore, a measure of betrayal aversion. Bohnet et al. find that betrayal aversion exists in all 6 countries and is – averaging across the six countries – as important as risk aversion. These results thus provide strong evidence that not only risk aversion, but betrayal aversion as well, constrain behavior towards inter-personal risks.[6]

If betrayal aversion is important for trust, then risk preferences alone might have little explanatory power for elucidating trusting behavior in the trust game. If researchers cannot control for betrayal aversion, regressions of first mover behavior in the trust game on measures of risk preferences suffer from a lot of noise and omitted variable bias, possibly preventing significant results in smaller samples. This may explain why some researchers have not found an impact of risk preferences on trust (Eckel and Wilson 2004, Houser, Schunk and Winter 2008). In fact, the evidence in Houser et al. can be taken as indirect support for betrayal aversion because the authors find that their measure of risk preferences has excellent predictive power for behavior in risk games but not in trust games.

The existence of betrayal aversion is likely to be very important for institutional design questions because some of the most important risks that people face in their economic activities are socially constituted, such as the risk of being cheated by the trading partner or the risk of expropriation by politicians or corrupt civil servants. Thus, the existence of betrayal aversion is likely to be a potent inhibitor of trade and economic activity which renders the design and implementation of efficient legal enforcement institutions all the more important.

---

[6] Currently it is an open question whether betrayal aversion is just a variant of preferences for reciprocity or whether additional motivational forces that go beyond reciprocity preferences are needed to explain betrayal aversion. However, as the material allocations in the risky dictator game and the trust game are identical, the behavioral evidence implies that people care about how an allocation is reached (i.e., whether bad luck or selfish human behavior is the source of a potentially bad outcome). Bohnet et al. (2008) find no evidence for the role of outcome-based social preferences for behavioral trust (perhaps because the effects of inequality aversion and surplus maximization cancel each other in the aggregate) but Hong and Bohnet (2007) show that inequity aversion is an inhibitor of trust in lower status groups of the US population – women, minorities, young adults and non-Protestants – while higher status groups – men, Caucasians, middle-aged people and Protestants – display betrayal aversion. Thus, as in Bellemare and Kröger (2008), the relative importance of outcome and intention-based social preferences seems to vary across socio-economic groups.

4. Do risk and social preferences predict survey trust?

If betrayal and risk aversion are determinants of trust, they should have predictive power in trust regressions. I examine this question with data from the German SOEP which contains survey measures of trust, betrayal aversion, and risk preferences.[7] Trust is measured with three questions: (i) "In general, one can trust people", (ii) "Nowadays, you can't rely on anybody" and (iii) "When dealing with strangers, it is better to be cautious before trusting them". In all three cases the answer categories are "disagree strongly", "disagree somewhat", "agree somewhat" and "agree strongly". I code the answers to questions (i) – (iii) in such a way that a higher number always represents higher trust.

Betrayal aversion means that people dislike non-reciprocated trust. It is plausible that people who experience particularly high disutility from non-reciprocated trust have a high willingness to punish non-reciprocating players. The reverse also makes sense: people with a strong preference for negative reciprocity (i.e., a preference for punishing non-reciprocal behavior) are – ceteris paribus – more likely to feel betrayed in case of non-reciprocated trust. Measures of negative reciprocity should therefore be good proxies for betrayal aversion. The German SOEP provides such a measure based on two questions taken from the Reciprocity questionnaire (Perugini, et al. 2003): (i) "If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the costs" and (ii) If someone offends me, I will also offend him/her". People answer these questions on a Likert Scale from 1-7.

The SOEP also contains an experimentally validated measure of risk preferences which is based on the question: "Are you, generally speaking, a person who is fully prepared to take risk, or do you try to avoid taking risks?" The respondents can answer this question on an 11 point Likert Scale ranging from 0 (very risk averse) to 10 (very risk seeking). A recent paper shows (Dohmen, et al. 2005) that this measure of risk preferences is a good predictor of behavior in risk taking experiments.

Finally, because altruism may affect measures of trust, I also include a measure of altruism or generosity as an explanatory variable. The SOEP has a question that provides information regarding the frequency with which people volunteer for clubs or social services: "How do you spend your free time? Please indicate how often you engage on average in each

---

[7] This section relies heavily on the idea's developed in Naef et al. (2008). In particular, the measures of betrayal aversion and altruism have been developed in this paper. However, Naef et al. (2008) use these variables to explain the behavioral trust gap between the US and Germany while the purpose of this section is to show that survey trust is also affected by risk preferences and social preferences.

of the following activities?" The menu of possible activities ranges from meeting friends, relatives or neighbors to watching TV or videos, to volunteering in clubs and social services. Answer categories include "never, seldom, monthly, weekly or daily". I include a dummy into the regression that takes the value of zero if the respondent never volunteers and a value of 1 otherwise. The volunteering variable may not only indicate a subject's altruism, but his or her sociability as well. I thus also include a direct "sociability" variable, as I am interested in the altruistic component of volunteering: subjects who meet "friends, neighbors or relatives" at least daily or weekly are considered sociable because of their frequent "out-of-family" social contacts. This control for sociability increases the chances that the volunteering measure picks up other-regarding concerns.

Table 1 shows how these preference measures affect the answers to each of the trust questions and to the average answer across all three questions. The table indicates robustly that risk preferences, betrayal aversion, and altruistic concern expressed through volunteering significantly affect trust. Regardless of which trust measure I use, subjects who exhibit high risk aversion trust less compared to those who show intermediate risk aversion, and the latter trust less than those who indicated a very low level of risk aversion. Likewise, subjects with high betrayal aversion trust less than those with intermediate levels, who in turn trust less than those with low levels of betrayal aversion. It is also interesting that the coefficients on betrayal aversion are even larger than those on risk aversion, suggesting that it is of substantial importance in trust. Controlling for sociability, subjects who never volunteer are also less likely to trust others – suggesting that other-regarding concerns affect trust measures.

I introduce demographic and socio-economic controls in Table 2. This somewhat reduces the coefficients of our preference measures, but most of them remain significant. Our preference measures, in particular, continue to robustly affect our average trust measure ("trust index"), probably because the trust index provides a less noisy measure of trust than the individual measures.

Insert Tables 1 and 2 about here

5. Disentangling behavioral trust and beliefs about trustworthiness

If we want to understand trust there is a need to distinguish between beliefs in others' trustworthiness and trusting behavior. In the trust game this distinction is straightforward. The investor's transfer constitutes the behavioral measure and by eliciting the investors' belief about the trustee's trustworthiness we have a separate belief measure. However, how should we interpret survey measures of trust? Are they a behavioral measure of trust or do they just represent people's beliefs about others' trustworthiness? A priori, it seems plausible to assume that the answer to the statement "In general, one can trust people" elicits people's beliefs about others' trustworthiness. However, the results of Table 1 and Table 2 call this into question. The fact that risk and social preferences significantly influence survey measures of trust is consistent with the idea that people derive their answers to trust questions from introspecting on their own likely behaviors in situations requiring trust. Therefore, the survey measures of trust that are currently used are also likely to reflect a composite of preferences and beliefs.

This does of course not mean that survey measures are better behavioral measures than the observed trusting behavior itself, i.e., observation of investors' behavior in the trust game. Observing actual trusting behaviors towards unknown strangers will probably always provide a more convincing measure than answers to a question.[8] In addition, participation in the trust game requires that the game is explained in detail so that subjects know all contingencies. Such detail is typically not available when simple questions are posed, implying that less is known about the subjects' implicit assumptions. In addition, the trust game offers the possibility of a clean separation between the act of trusting and the assessment of the trustee's trustworthiness because the investor in the trust game is put into a specific situation that asks for two well-defined responses – an action response and a belief response. In contrast, the belief component of trust and the preference component are inextricably interwoven in the currently used survey questions.

---

[8] Naef and Schupp (2008) found that questionnaire measures of trust – even those not subject to the same criticism as the GSS measure – seem to be affected by social desirability bias, while behavioral measures of trust are not affected. The German SOEP survey takes advantage of the fact that a representative subsample of the German population participated in a trust game for several years. The same people who participated in the experiment also answered survey questions on trust and questions designed to measure social desirability bias. The advantage of the behavioral trust measure could be due to the fact that the usual experimental economics care was applied in creating an unobtrusive situation for the participants in the trust game. In particular, while the individual SOEP interviewers did not know the SOEP participants' decisions in the trust game, no interviewer-participant anonymity existed with regard to the answers to the survey questions.

However, if beliefs about others' trustworthiness were equally affected by preference measures as the investor's back transfer or survey measures of trust, one could doubt whether the trust game enables a clean separation between behavioral trust and expected trustworthiness. Therefore, we examine next whether subjects' beliefs in the trust game are predicted by their risk and social preferences.

Fortunately, Naef et al. (2008) collected a large data set that provides a nationally representative behavioral trust measure for the US and Germany and an equally representative measure of beliefs in the trustees' trustworthiness. In total, more than 1400 investors participated in this experiment. I present the impact of risk preferences and social preferences on the beliefs of the German and the US population in the trust game in Table 3. Naef et al. (2008) asked the investors about their beliefs about the trustees' back transfers at investments of 0, 5 and 10.[9] The table shows the results of regressions with the expected back transfers as the dependent variable. The same measures of risk preferences and social preferences are used as in Table 1 and 2. Table 3 shows that the preference measures have basically no impact on the investors' beliefs in the trust game – the coefficients for risk and betrayal aversion in all regressions have no significant impact on beliefs. Thus, in contrast to the questionnaire measures of trust, which are likely to be driven by an unknown mixture of belief and preference components, preferences do not affect the belief measure from the behavioral trust experiment. This finding is good news for a rational choice approach towards trust because such an approach relies on the fundamental distinction between beliefs and preferences. According to this approach, preferences and beliefs drive trusting behavior, but preferences should not affect beliefs, otherwise they cannot be considered rational. Of course, the absence of a preference influence on beliefs is not sufficient for the existence of rational beliefs, but it is an important necessary condition.

Insert Table 3 about here

---

[9] The subjects in Naef et al (2008) played the following modified trust game. Both players were endowed with €10 (in Germany) or $10 (in the US). Each subject could transfer between 0 and 10 money units to the "partner". The investor's transfer *and* the trustee's back-transfer was doubled. These modifications simplified the game which was important in view of the constraints imposed on nationally representative games that are embedded in a survey.

While risk and social preferences do not affect beliefs, they influence investors' actions. Naef et al. (2008) show that the preference measures used in Table 1 -3 significantly affect investors' trust in the trust game. In particular, people who display high betrayal aversion send less money to the trustee than those who display medium betrayal aversion, while the latter send less than those who display little betrayal aversion. The same qualitative order of effects holds for risk aversion. In addition, the subjects' beliefs about the trustworthiness of their anonymous trustee determine the amount sent. These results reinforce our claim that risk preferences, social preferences, and beliefs shape trust simultaneously. Taken together, the trust game thus enables us to achieve what we want. It provides a suitable behavioral measure of trust and it provides the opportunity to elicit a clean – preference-free – measure of beliefs about the trustworthiness of anonymous strangers.

6. Explaining national and ethnic trust differences

To what extent does the framework laid out in the previous sections help us understand national and ethnic trust differences? Or put differently, do differences in beliefs and preferences across nations and ethnic groups drive national and ethnic trust differences? Or are other factors, such as the socio-economic differences between nations and ethnic groups, the main drivers of differences in trusting behavior?

Naef et al. (2008) applied the preferences and beliefs framework to their representative trust data. They found that Americans display a much higher trust level in the trust game compared to the German population. In addition, they observed that African Americans show a much lower level of trust than Caucasian Americans. The question then is whether the preferences and beliefs framework is capable of explaining a substantial part of these trust differences. For this purpose, Naef et al. examined the distribution of risk and social preference in the US and in Germany on the basis of the preference measures described in Section 4 above. The key findings can be summarized as follows: (i) The US population is significantly less risk averse and less betrayal averse than the German population. (ii) The US population is also significantly more altruistic in terms of the volunteering measure described above. (iii) Finally, the US population also has significantly more optimistic beliefs about the trustworthiness of the trustees[10]. Therefore, if risk preferences, social preferences, and beliefs about the anonymous partner's trustworthiness affect trusting behavior, the preferences and

---

[10] Note that US investors faced trustees from the US and German investors faced trustees from Germany. Thus, the beliefs about the back transfers reflect beliefs about the average trustworthiness of the respective populations.

belief differences across the two countries should also explain the trust gap across countries. And they do. Naef et al. show that differences in preferences and beliefs can explain *the entire* trust gap between Germany and the US. Interestingly, the importance of preferences and beliefs shows up even if the authors control for a host of socio-economic variables. If the preferences and belief measures and the socio-economic measures are simultaneously used as explanatory variables, the preferences and beliefs variables still explain roughly 90 percent of the trust gap, while socio-economic variables explain the rest.

The approach in Naef et al. can also be applied to explain the trust gap between African and Caucasian Americans. It turns out that African Americans have significantly more betrayal aversion than Caucasian Americans, while there are no differences in terms of risk aversion. However, African Americans are much less optimistic about the trustees' behavior. Taken together, ethnic differences in betrayal aversion and beliefs about trustworthiness account for roughly two-thirds of the black-white trust gap.

7. Trust as an endogenous variable – the trust shaping role of informal institutions

It is standard practice in economics to assume that preferences are given exogenously, while the prevailing equilibrium determines beliefs endogenously. In order to justify this view, it is not necessary to assume that preferences are completely fixed because they may change over long periods of time. All that is needed is that preferences are stable relative to the economic problem under investigation. Beliefs, in contrast, are likely to be more malleable so that they change more quickly in response variations in the prevailing conditions.

If we apply the standard view to problems involving trust, we must conclude that trust is a partly endogenous and a partly exogenous variable, endogenous to the extent that people's beliefs shape it and exogenous in the scope that their preferences influence it. The likely endogeneity of trust poses a major problem for field studies investigating the causal impact of trust on economic variables such as investment or trading volume, gross national product, etc. I will deal with this question in the next section. In this section I demonstrate the endogeneity of trust with the help of the experiments that examined the impact of reputation formation opportunities on markets with moral hazard problems (Brown, Falk and Fehr 2004).[11]

---

[11] There are several experimental papers which document the endogeneity of trust and the role of institutions in trust formation (see, e.g., Bohnet and Baytelman 2007; Bohnet and Huck 2004; Huck, Lünser and Tyran 2007).

Subjects in these experiments are randomly allocated to two treatments – a one-shot gift exchange treatment and a gift exchange treatment with an opportunity for reputation formation. The basic gift exchange design has the following features. Some subjects are in the role of an employer, others in the role of workers. There are more workers than employers, and each employer can only employ one worker. Employers offer employment contracts in a one-sided continuous auction; an offer stipulates a wage offer and a requested effort level for the current period. At the beginning of each of 15 periods, the employers make offers and the workers can accept them. Once accepted, they have to choose an effort level $e \in \{1,2, \ldots, 10\}$. The employer's wage offer as accepted in the contract is legally binding, while workers can still choose any effort level they like, i.e., they are not legally obliged to choose the requested effort level. This set-up mimics a situation where only the employer and the worker can observe the actual effort, but third parties cannot verify it. Thus, because the employer cannot condition the current wage on the current effort level, effort needs to be enforced endogenously – either by social preferences alone (in the one-shot treatment) or by a combination of social preferences and reputation incentives (in the reputation treatment).

In the reputation treatment, workers and employers have an identification number which remains fixed throughout the whole 15 periods of the experiment. Thus, if a worker, say W7, performs well, the employer can again make a wage offer in the next period to W7, or he can fire the worker by not making him a new offer. Fixed identification numbers thus imply that workers can acquire a reputation for high performance. This reputation is "bilateral", however, because only the current employer knows his worker's effort level while other employers do not have this information. Thus, employers in the reputation treatment can condition their current wage offer on the worker's past performance, which provides an incentive for workers to provide effort which is not available in the one-shot treatment. In the latter, the identification numbers are randomly reassigned every period so that no individual reputation can be acquired and employers cannot condition their new offers on past performance.

Trust in the gift exchange game is measured by the employer's wage offer because this offer becomes legally binding if accepted, while the worker is free to choose any feasible effort level. One can thus measure the extent to which behavioral trust differs by observing wage levels across treatments. In addition, because Brown et al. elicited employers' beliefs about the effort level they expect, I have a separate measure of beliefs in workers' trustworthiness. Finally, I also have a measure of the workers' actual level of trustworthiness

because I can observe the actual effort level. Note also, that there is no reason ex-ante, i.e., before period 1 of the experiment, to believe that there are differences in preferences or prior beliefs across the two treatments because the subjects were randomly assigned to the treatments. Thus, any emerging differences in behaviors and beliefs across treatments must be the result of the absence or presence of the reputation formation opportunity.

The results displayed in Figure 2a – 2c show the impressive consequence of the reputation opportunity. If workers can acquire a good reputation, some of them do so and employers' punish shirking by firing "bad" workers, while "good" workers are rewarded with high-wage offers in the future.[12] The figures show that the reputation treatment simultaneously increases the expected trustworthiness as measured by the expected average effort (Figure 2a), behavioral trust measured by average wages (Figure 2b), and actual trustworthiness measured by actual average effort (Figure 2c). It is also remarkable how strongly these three variables are correlated over time. Behavioral trust, believed trustworthiness, and actual trustworthiness all increase over time in the reputation treatment, until they reach a peak in period 13 and decline thereafter. In contrast, all three variables in the one-shot treatment decline initially until they stabilize at a baseline level.

Insert Figures 2a – 2c here

Because of randomization to treatments, there can be no doubt that the informal reputation formation institution is the ultimate driver of behavioral trust, beliefs about trustworthiness and actual trustworthiness. However, imagine an outside observer who does not know that the treatment difference is due to a difference in informal institutions. Such an observer may erroneously conclude that the higher trust in the reputation treatment is the ultimate cause for higher effort and higher output.

---

[12] Brown, Falk and Fehr (2004) show that these behaviors arise from an interaction between fairness preferences and reputation incentives. Note that because it is common knowledge that the experiment lasts 15 periods, backward induction rules out the possibility of reputation formation if all subjects are purely selfish. But if a heterogeneous mix of fair and selfish types exists, equilibria with reputation formation, in which the selfish types mimic the fair types in all periods but the last one, become possible.

8. Causal effects of trust?

A researcher interested in examining the causal impact of trust on economic outcomes such as investment, trading volume, gains from trade, or economic growth is in the position of an observer who typically lacks information about the details of informal institutions and norms. Such institutions and norms can depend on seemingly minor details such as people's spatial mobility or the existence of a citizen registry that enables principals to find out where they can find agents. However, because informal institutions are likely to shape trust strongly, and because of the largely unknown interactions between formal and informal institutions, assessing the causal role of trust in field data becomes extremely difficult.

Trust measures are endogenous, regardless of how one measures them. Both behavioral measures of trust taken from a nationally representative trust game and survey measures of trust are equally subject to the endogeneity problem because subjects' beliefs about the prevailing trustworthiness affect both measures. These beliefs, in turn are likely to be affected by people's experiences about others' trustworthiness, which in turn are very likely to be influenced by formal and informal institutions.

The most common strategy for solving such endogeneity problems is using instrumental variables. In the seminal paper by Knack and Keefer (Knack and Keefer 1997), for example, the authors instrument "trust" in a two-stage least squares approach to explain average annual growth rates in GDP across 29 countries. They use a variable that indicates ethno-linguistic homogeneity in a country (the percentage of a country's population belonging to the largest ethno-linguistic group) and the number of law students in 1963 as a percentage of all postsecondary students as instruments. It is plausible that a more homogeneous population exhibits more trust; likewise, a higher percentage of law students may indicate more need for lawyers because of problems with legal institutions. Thus, both variables pass the easy part of the test for valid instruments – they are correlated with trust as measured in the World Values survey. However, it is also easy to construct arguments why both variables violate the second condition for a valid instrument, i.e., why they are probably correlated with the error term in the growth rate regression: a population that is more ethno-linguistically homogenous is likely to be associated with denser social networks that simplify communication and trade. In addition, it may be associated with a higher degree of impartiality in the legal enforcement of property rights and contracts because discrimination across groups plays less of a role. All these factors may thus plausibly have a direct impact on growth, invalidating this instrument.

Likewise, if a higher percentage of law students indicates problems in the legal enforcement of property rights and contracts and the absence of effective social norms, this variable is likely to have a direct impact on growth. It is also noteworthy that the use of lagged variables does not solve the problem because the lack of contract enforcement and beneficial informal social norms is a long term phenomenon likely to be associated with large hysteresis effects. Thus bad legal enforcement in 1963 is likely to be correlated with bad legal enforcement decades later.[13] Taken together, these objections against the instruments suffice in questioning the claimed causal role of trust for GDP growth rates. It seems equally possible that the trust variable in the growth equation just captures the consequences of formal and informal economic institutions that are not controlled for by the other available control variables.

It is not my aim here to claim that no good instruments for trust can be found or that trust cannot have a causal impact on economic outcomes, but with regard to the instruments I have seen some doubt remains. Typically, the instruments used are correlated with measures of trust, but there are also plausible arguments supporting a direct impact on the economic outcome of interest, i.e., they are not exogenous to the error term beyond doubt. A recent study by Guiso, Sapienza, Zingales (2009) has perhaps gone farthest in achieving true exogeneity. These authors provide fascinating evidence on the role of mutual trust in trade across countries. They use "common religion" as an instrument for mutual trust because trading volume and mutual trust are likely to be positively correlated. This instrument seems to have the advantage that it is exogenous to current trading volumes because the commonality of religion was typically determined hundreds of years earlier. However, it is not only important for the validity of an instrument whether it is exogenous to the model or to the outcome variable of interest. Instead, we must validate the hypothesis that common religion has no direct influence on trade, an assumption that is unlikely to be true. Common religion not only influences trust, but does many other things as well, because it is probably associated with more frequent interactions between the two countries, compared to cases with different religions, and this may well have a direct impact on trade. For this reason, the authors introduce a second instrument – somatic similarity between populations – and apply a Hausman test for overidentification. Under the *assumption* that one of the two instruments is valid (i.e., exogenous to the error term) this test does not reject the null hypothesis that both instruments are valid. Thus, we cannot reject the null hypothesis but we also don't know

---

[13] Knack and Keefer also use distance to the Equator as an instrument for trust. However, distance to the equator is also correlated with a lot of climate variables (such as temperature) which may have a direct impact on work morale, effort costs and, therefore, GDP growth.

whether it is true. In addition, because we already have expressed doubt about "common religion" as a valid instrument, the question is whether the instrument "somatic similarity" is exogenous to the error term (so that the assumption of the test is met). In my view, this could be true but even for such a sophisticated instrument one can raise plausible doubts. It is, for example, well known that people display more altruism towards ingroup members (Yamagishi, Jin and Miller 1998), and they are less likely to violate social norms if the victim of a norm violation is an ingroup member (Bernhard, Fischbacher and Fehr 2006, Goette, Huffman and Meier 2006). Under the plausible assumption that somatic similarity is associated with ingroup status, people will tend to be more altruistic towards others who are more similar, and they tend to have a lower propensity to violate a norm if a similar other is the victim of the norm violation. If this argument is true, then somatic similarity between two countries may have a direct effect on trade because trade also depends on the party's willingness to honor social norms of contract enforcement.

The difficulty in finding valid field instrumental variables suggests that laboratory experiments can be useful because they provide valid instruments more easily. In fact, a set-up where one treatment group is induced to have a low level of trust while the subjects in the other treatment group are induced to have a high level of trust provides the best instrument. One can then observe the impact of the exogenous manipulation of trust by randomly assigning subjects to the two treatments. However, I know of no experiment that has done this.[14]

9. Possible mechanisms behind a causal role of trust

When studying the causal role of trust for economic outcomes, considering the potential mechanisms behind causality is useful. One obvious mechanism is that a population that is less betrayal and risk averse is more likely to invest when facing a given level of uncertainty, and more likely to trade for a given level of imperfect contract enforcement. Likewise, less risk and betrayal aversion on the part of principals is likely to lead to less waste in terms of monitoring and policing agents. Thus, all else being equal, it seems plausible that less risk and

---

[14] In a trivial sense, the trust game shows that trust has an economic impact because the investor's transfer is multiplied by a number bigger than one. Thus, higher trust is associated with a higher surplus. However, this is the trivial result of the multiplication. The real question is whether higher trust *causes* subsequent behaviors that justify the initially higher level of trust, and whether this process leads to *sustainable* changes in economic outcomes relative to a situation with an *initial* low trust level. We are currently experimenting with this possibility but have no results yet.

betrayal aversion, and the higher level of trust that is associated with it, lead to more investment and trade and lower transaction costs in terms of monitoring.

However, preferences are probably not easily malleable, meaning that economic policy cannot easily shape them. Nor is it clear whether implementing policies that deliberately aim at shaping people's preferences is desirable. The other component of trust, however, namely beliefs about other' trustworthiness, may be more easily affected. This raises the question whether it is possible to show that trust exerts a causal impact on economic outcomes via the belief component of trust, i.e., by raising people's beliefs about others' trustworthiness.

The literature on gift exchange games (Brandts and Charness 2004, Charness 2004, Charness, Frechette and Kagel 2004, Fehr, Kirchsteiger and Riedl 1993) and the more recent papers on the counterproductive effects of sanctions and other measures that constrain agents' shirking (Bohnet, I., Frey and Huck 2001, Falk and Kosfeld 2006, Fehr and Falk 2002, Fehr and List 2004) suggest that beliefs about others' trustworthiness may be self-reinforcing. Principals in gift exchange games who trust more in terms of paying higher up-front wages induce a higher effort level on average. The literature mentioned above on the counterproductive effects of incentives also suggests that beliefs about trustworthiness can be self-reinforcing. For example, in Fehr and List (2004), subjects (CEOs and students) in the role of an investor have the opportunity of combining a transfer in a modified trust game with a small credible sanction if the back-transfer does not meet the amount requested. However, the investor can also voluntarily refrain from using the sanctioning threat. Surprisingly, the back-transfers are significantly lower if the investor uses the threat, perhaps because the threat is interpreted as a hostile act to which trustees respond with lower back-transfers. A similar result has been observed by Falk and Kosfeld (2006), who show that restrictions on the agents' action space can be detrimental for their average performance – although they limit shirking of selfish agents – because they induce agents with social preference (those who are in principal trustworthy) to increase shirking.

This empirical evidence suggests that trust can be self-reinforcing.[15] Principals or investors who have pessimistic beliefs exhibit a low level of trust and subsequently experience a low level of trustworthiness; those with optimistic beliefs display a high level of trust and subsequently experience a high level of trustworthiness. In fact, the principals in

---

[15] Reuben, Sapienza, Zingales (2008) show in a recent paper that trustees who know that their investor expects a low level of trustworthiness will send back less. Thus, mistrust tends to be self-fulfilling and high expectations tend to elicit higher back transfers. Unfortunately, the authors do not report a correlation between behavioral trust and expectations.

Falk and Kosfeld behave exactly in this way. They experienced what they expected, so that they had no reason to change their expectations.

Although the literature above suggests that high trust *can be* self-reinforcing, it does not show that it *is* self-reinforcing. To show the self-reinforcing nature of trust, multiple periods are necessary in order to see whether the principals' behavior in the high (low) trust condition causes levels of trustworthiness that justify and maintain the initial level of high (low) trust. I do not know any evidence supporting this claim; in fact, if anything, there is evidence that casts doubt on this claim. In a recent paper (Bartling, Fehr and Schmidt 2008), the ideas in Falk and Kosfeld (2006) have been applied to a *repeated* one-shot gift exchange game that lasts for 15 periods: in each period, the principals make a contract offer consisting of a wage and a requested effort level. The wage offer is a legally binding commitment by the principal, but effort is not third party verifiable and hence the agent is free to choose any available effort level. However, the principal can restrict the agent's available effort levels from $e \in \{1,2, \ldots, 10\}$ to $e \in \{3,4, \ldots, 10\}$. The restriction of the agent's effort choices can be viewed as a reduced form representation of an incentive and monitoring system that enforces a minimal effort of 3. For convenience, contract offers associated with a restriction on effort choices are called control contracts, while contracts that do not restrict the agent are called trust contracts.

The question then is whether the principals offer generous trust contracts with relatively high wages that appeal to the agents' reciprocity and trustworthiness or whether they offer control contracts with relatively low wages. It turns out that the principals offer trust contracts in about 50% of the cases during the first few periods, but that they converge towards low wage control contracts over time; less than 20% of the contracts were trust contracts towards the end of the experiment. We have here a case where an initial high level of trust was not sustainable because the lack of agents' trustworthiness did not support the initial high level of trust.[16],[17]

---

[16] Note that the principal's revenue function for trust contracts was 5*e* (*e* denotes effort) while for control contracts it was only 4*e*, compatible with the notion that monitoring and controls impose a burden on the agents and lower their productivity. In the context of our discussion, this feature makes the decline of trust contracts over time even more remarkable: although trust contracts were more efficient in terms of "physical" productivity, the principals largely preferred control contracts.

[17] A similar finding was reported in Fehr and Zehnder (2008), which examined the viability of a one-shot credit market where no third party enforced the repayment of debt. Initially, roughly 80% of the lenders offered credit contracts but their willingness to offer contracts declined steadily to less than 20% due to the low repayment rate. Thus, like in the previous experiment, the initial high level of trust was not sustainable, and behavioral trust levels gradually declined.

These findings suggest that proving that exogenously manipulated trust causes different long-term outcomes may be very hard. Of course, in both above-mentioned experiments, the initial high level of trust and its subsequent decline to levels matching the prevailing level of trustworthiness, temporarily caused higher levels of trade and higher gains from trade. But this was only a temporary phenomenon and trust ultimately converged towards rather low levels.

Currently, I see only one way in which the belief component of trust may have causal long-term effects. If high trust equilibria and low trust equilibria exist, initial variations in trust across groups may have long-term effects because an initially high level of trust acts as an equilibrium selection device. However, I know no study that provides support for this argument.

10. Conclusions

I selectively reviewed and assessed recent research on trust in this paper. My view is based on a behavioral definition of trust. This definition highlights the importance of risk preferences, social preferences, and beliefs about other people's trustworthiness for trusting behavior. From an economic viewpoint, in fact, all individual behaviors – including trusting behavior – should be captured by preferences and beliefs, and there is little reason why other "determinants" of trust – such as religiousness, for example – should play an independent role that does not work through its effect on preferences and beliefs. However, as our measures of trust, preferences, and beliefs are almost always imperfect, even demonstrating that measures of beliefs and preferences are indeed significant predictors of trust would be useful.

I have shown that risk preferences and social preferences are indeed predictors of survey trust. This complements work by Bohnet and coauthors (2004, 2008) and Naef et al. (2008), which show that these types of preferences play a key role in behavioral trust as measured in the trust game. Thus, both trust measured by survey methods and trust measured in the trust game seems to be based on risk and social preferences. This economic approach towards trust also receives particular vindication by the results of Naef et al. (2008), who show that measures of beliefs, risk, and social preferences explain almost the whole trust gap between the US and Germany.

We also reviewed experimental research that unambiguously shows that trust, expected trustworthiness, and actual trustworthiness are endogenous variables shaped by formal or

informal institutions. This endogeneity problem is a major obstacle that inhibits the provision of fully convincing evidence for a causal role of trust. Despite recent progress in finding sophisticated instrumental variables for trust, it is in my view still plausible that trust is simply an epiphenomenon of the institutional environment, i.e. it may not cause lasting effects on important economic outcomes such as the volume of trade, investment, or reliance on monitoring technologies. However, the economic approach towards trust also provides suggestions how a causal role of trust could be established experimentally. The most convincing strategy seems to be to induce optimistic or pessimistic beliefs about other people's trustworthiness exogenously and observe whether this leads to interesting and lasting changes in behaviors and economic outcomes.

References

Bartling, B., E. Fehr, and K. M. Schmidt (2008). "Reputation and Contract Design," *Working Paper, Institute for Empirical Research in Economics, University of Zurich.*

Baumgartner, T., M. Heinrichs, A. Vonlanthen, U. Fischbacher, and E. Fehr (2008). "Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans," *Neuron*, 58, 639-650.

Bellemare, C., S. Kröger, and A. Van Soest (2008). "Preferences, Intentions, and Expectations: A Large-Scale Experiment with a Representative Subject Pool " *IZA Working paper No. 3022.*

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

Bernhard, H., U. Fischbacher, and E. Fehr (2006). "Parochial Altruism in Humans," *Nature*, 442, 912-915.

Bohnet, I., and Y. Baytelman (2007). "Institutions and Trust - Implications for Preferences, Beliefs and Behavior," *Rationality and Society*, 19, 99-135.

Bohnet, I., B. S. Frey, and S. Huck (2001). "More Order with Less Law: On Contract Enforcement, Trust, and Crowding," *American Political Science Review*, 95, 131-144.

Bohnet, I., F. Greig, B. Herrmann, and R. Zeckhauser (2008). "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States," *American Economic Review*, 98, 294-310.

Bohnet, I., and S. Huck (2004). "Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change," *American Economic Review*, 94, 362-366.

Bohnet, I., and R. Zeckhauser (2004). "Trust, Risk and Betrayal," *Journal of Economic Behavior & Organization*, 55, 467-484.

Brandts, J., and G. Charness (2004). "Do Labour Market Conditions Affect Gift Exchange? Some Experimental Evidence," *Economic Journal*, 114 497, 684-708.

Brown, M., A. Falk, and E. Fehr (2004). "Relational Contracts and the Nature of Market Interactions," *Econometrica*, 72, 747-780.

Buchan, N. R., R. T. A. Croson, and R. M. Dawes (2002). "Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange," *American Journal of Sociology*, 108, 168-206.

Camerer, C., and K. Weigelt (1988). "Experimental Tests of a Sequential Equilibrium Reputation Model," *Econometrica*, 56, 1-36.

Cesarini, D., C. T. Dawes, J. H. Fowler, M. Johannesson, P. Lichtenstein, and B. Wallace (2008). "Heritability of Cooperative Behavior in the Trust Game," *Proceedings of the National Academy of Sciences of the United States of America*, 105, 3721-3726.

Charness, G. (2004). "Attribution and Reciprocity in an Experimental Labor Market," *Journal of Labor Economics*, 22 3, 665-88.

Charness, G., G. R. Frechette, and J. H. Kagel (2004). "How Robust Is Laboratory Gift Exchange?," *Experimental Economics*, 7 2, 189-205.

Charness, G., and M. Rabin (2002). "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117, 817-69.

Coleman, J. (1990). *Foundations of Social Theory*. The Belknap Press of Harvard University Press.

Cox, J. C. (2004). "How to Identify Trust and Reciprocity," *Games and Economic Behavior*, 46 2, 260-81.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005). "Individual Risk Attitudes - Evidence from a Large, Representative, Experimentally-Validated Survey," *IZA Working Paper No. 1730.*

Dufwenberg, M., and G. Kirchsteiger (2004). "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47, 268-298.

Eckel, C., and R. Wilson (2004). "Is Trust a Risky Decision?," *Journal of Economic Behavior and Organization* 55, 447-465.

Engelmann, D., and M. Strobel (2004). "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, 94, 857-869.

Falk, A., and U. Fischbacher (2006). "A Theory of Reciprocity," *Games and Economic Behavior*, 54, 293-315.

Falk, A., and M. Kosfeld (2006). "The Hidden Costs of Control," *American Economic Review*, 96, 1611-1630.

Fehr, E., and A. Falk (2002). "Psychological Foundations of Incentives," *European Economic Review*, 46, 687-724.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993). "Does Fairness Prevent Market Clearing - an Experimental Investigation," *Quarterly Journal of Economics*, 108, 437-459.

Fehr, E., and J. A. List (2004). "The Hidden Costs and Returns of Incentives--Trust and Trustworthiness among Ceos," *Journal of the European Economic Association*, 2 5, 743-71.

Fehr, E., M. Naef, and K. M. Schmidt (2006). "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment," *American Economic Review*, 96, 1912-1917.

Fehr, E., and K. M. Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817-868.

Fehr, E., and C. Zehnder (2008). "Reputation and Credit Market Formation: How Endogenous and Legal Enforcement of Credit Contracts Interact," *Working Paper, Institute for Empirical Research in Economics, University of Zurich*.

Fleming, A. S., and J. S. Rosenblatt (1974). "Olfactory Regulation of Maternal Behavior in Rats: Ii. Effects of Peripherally Induced Anosmia and Lesions of the Lateral Olfactory Tract in Pup-Induced Virgins," *Journal of Comparative Physiological Psychology*, 86, 233-246.

Goette, L., D. Huffman, and S. Meier (2006). "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups," *American Economic Review*, 96 2, 212-16.

Guiso, L., P. Sapienza, and L. Zingales (2004). "The Role of Social Capital in Financial Development," *American Economic Review*, 94, 526-556.

— (2008). "Trusting the Stock Market," *Journal of Finance*, forthcoming.

— (2009). "Cultural Biases in Economic Exchange?," *Quarterly Journal of Economics*, forthcoming.

Heinrichs, M., and G. Domes (2008). "Neuropeptides and Social Behavior: Effects of Oxytocin and Vasopressin in Humans," *Progress in Brain Research*, 170.

Hong, K., and I. Bohnet (2007). "Status and Distrust: The Relevance of Inequality and Betrayal Aversion," *Journal of Economic Psychology*, 28, 197-213.

Houser, D., D. Schunk, and J. Winter (2008). "Distinguishing Trust from Risk: An Anatomy of the Investment Game," *Working Paper, Institute for Empirical Research in Economics, University of Zurich*.

Huck, S., G. K. Lünser, and J. R. Tyran (2007). "Competition Fosters Trust," *Working Paper No. 235, University College London*.

Insel, T. R., and L. J. Young (2001). "The Neurobiology of Attachment," *Nat Rev Neurosci*, 2, 129-36.

Karlan, D. (2005). "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions," *American Economic Review*, 95, 1688-99.

Knack, S., and P. Keefer (1997). "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *Quarterly Journal of Economics*, 112, 1251-1288.

Kosfeld, M., M. Heinrichs, P. J. Zak, U. Fischbacher, and E. Fehr (2005). "Oxytocin Increases Trust in Humans," *Nature*, 435, 673-676.

Laporta, R., F. Lopezdesilanes, A. Shleifer, and R. W. Vishny (1997). "Trust in Large Organizations," *American Economic Review*, 87, 333-338.

Miller, A. S., and T. Mitamura (2003). "Are Surveys on Trust Trustworthy?," *Social Psychology Quarterly*, 66, 62-70.

Naef, M., E. Fehr, U. Fischbacher, J. Schupp, and G. Wagner (2008). "Decomposing Trust: Explaining National and Ethnical Trust Differences," *Working Paper, Institute for Empirical Research in Economics, University of Zurich*.

Naef, M., and J. Schupp (2008). "Measuring Trust: Experiments and Surveys in Contrast and Combination," *Working Paper, Royal Holloway College, London*.

Perugini, M., M. Callucci, F. Presaghi, and A. P. Ercolani (2003). "The Personal Norm of Reciprocity," *European Journal of Personality*, 17, 251-283.

Reuben, E., P. Sapienza, and L. Zingales (2008). " Is Mistrust Self-Fulfilling?," *Working Paper, Kellog School of Management, Northwestern University*.

Reuter, M., C. Montag, S. Altmann, F. Bendlow, C. Elger, P. Kirsch, B. Weber, and A. Falk (2009). "Genetically Determined Differences in Human Trust Behavior: The Role of the Oxytocin Receptor Gene," *Working Paper*.

Schechter, L. (2006). "Trust, Trustworthiness, and Risk in Rural Paraguay," *Experimental Economics*, 9, 173-173.

Yamagishi, T., N. Jin, and A. S. Miller (1998). "In-Group Bias and Culture of Collectivism," *Asian Journal of Social Psychology*, 1, 315-328.

Zak, P. J., and S. Knack (2001). "Trust and Growth," *Economic Journal*, 111, 295-321.

TABLE 1. The role of risk preferences and social preferences for survey trust

Dependent variable	Trust index		In general, one can trust people		Nowadays you can't rely on …		When dealing with strangers…	
Risk aversion: high	-0.16	***	-0.09	***	-0.11	***	-0.17	***
(Base: low)	(0.02)		(0.02)		(0.02)		(0.02)	
Risk aversion: medium	-0.07	***	-0.04	**	-0.05	***	-0.07	***
	(0.02)		(0.02)		(0.02)		(0.02)	
Betrayal aversion: high	-0.36	***	-0.33	***	-0.32	***	-0.16	***
	(0.03)		(0.03)		(0.03)		(0.03)	
Betrayal aversion: medium	-0.11	***	-0.11	***	-0.13	***	-0.00	
(Base: low)	(0.02)		(0.02)		(0.02)		(0.02)	
Dummy for volunteering	0.25	***	0.18	***	0.17	***	0.22	***
	(0.02)		(0.02)		(0.02)		(0.02)	
Dummy for sociability	0.15	***	0.12	***	0.16	***	0.04	***
	(0.01)		(0.01)		(0.01)		(0.01)	
Constant	0.05	**	0.03		0.04	*	0.03	
	(0.02)		(0.02)		(0.02)		(0.02)	
Observations	18732		18732		18732		18732	
Adjusted-R2	0.04		0.02		0.03		0.02	

OLS-Regression of trust on measures of risk aversion, betrayal aversion, and altruism (robust standard error in parenthesis). We also performed ordered probit regressions which are not reported. They yield the same conclusions, i.e., risk aversion, betrayal aversion and volunteering affect our trust measures significantly. All trust measures are taken from the German Socio-Economic Panel. They are standardized so that they have mean zero and a standard deviation of 1, implying that the regression coefficients provide information about how many standard deviations trust changes if the independent variable changes by one unit. The trust index takes averages of the answers to all three trust questions. We partition the risk and betrayal aversion measure into three categories: high aversion, medium aversion and low aversion. Risk aversion is "high" if respondents circle 0-3 on the Likert Scale, "medium" if they circle 4-6 and "high" if subjects choose 7-10. Betrayal aversion is "high" if respondents circle on average 6-7 on the Likert Scale for negative reciprocity, it is medium for 3-5 and we classify a subject as "low" betrayal averse if he chooses 1-2 on the Likert Scale. The volunteering variable is a dummy that takes a value of zero if subjects never volunteer and a value of 1 otherwise. This regression does not control for demographic and socio-economic influences on trust, but it controls for "sociability" which is a dummy variable with value 1 if the respondent meets at least daily or weekly with "friends, relatives or neighbors". * p<0.10, ** p<0.05, *** p<0.010

TABLE 2. The role of preferences and socio-economic factors for survey trust

Dependent variable (std)	Trust index		In general, one can trust people		Nowadays you can't rely on …		When dealing with strangers ….	
Risk aversion: high	-0.09	***	-0.07	***	-0.04	**	-0.09	***
(Base: low)	(0.02)		(0.02)		(0.02)		(0.02)	
Risk aversion: medium	-0.04	**	-0.03	*	-0.03		-0.03	*
	(0.02)		(0.02)		(0.02)		(0.02)	
Betrayal aversion: high	-0.30	***	-0.28	***	-0.27	***	-0.13	***
(Base: low)	(0.03)		(0.03)		(0.03)		(0.03)	
Betrayal aversion: medium	-0.10	***	-0.10	***	-0.13	***	-0.00	
	(0.01)		(0.02)		(0.02)		(0.02)	
Dummy for volunteering	0.18	***	0.12	***	0.11	***	0.17	***
	(0.02)		(0.02)		(0.02)		(0.02)	
Dummy for sociability	0.16	***	0.15	***	0.16	***	0.05	***
	(0.01)		(0.02)		(0.02)		(0.02)	
Dummy of being a female	0.01		0.03	*	0.03	*	-0.03	*
	(0.02)		(0.02)		(0.02)		(0.02)	
Age: 18 - 30 (Base: 31 - 50)	0.07	***	0.01		0.13	***	0.01	
	(0.02)		(0.02)		(0.02)		(0.02)	
Age: 51 - 92	0.06	***	0.10	***	0.04	*	-0.00	
	(0.02)		(0.02)		(0.02)		(0.02)	
Household income: 2nd quartile	0.09	***	0.09	***	0.09	***	0.03	
(Base: 1st quartile)	(0.02)		(0.02)		(0.02)		(0.02)	
Household income: 3rd quartile	0.18	***	0.17	***	0.17	***	0.06	***
	(0.02)		(0.03)		(0.02)		(0.02)	
Household income: 4th quartile	0.37	***	0.30	***	0.32	***	0.21	***
	(0.03)		(0.03)		(0.03)		(0.03)	
Household size	-0.02	***	-0.01	*	-0.01		-0.02	***
	(0.01)		(0.01)		(0.01)		(0.01)	
Head of household	0.04	**	0.03		0.05	***	0.02	
	(0.02)		(0.02)		(0.02)		(0.02)	
Lives in a 1- to 2-family house	0.05	***	0.05	***	0.05	***	0.01	
	(0.02)		(0.02)		(0.02)		(0.02)	
Education: high school	0.06	***	0.03		0.07	***	0.02	
(Base: less than high school)	(0.02)		(0.02)		(0.02)		(0.02)	
Education: more than high school	0.30	***	0.16	***	0.29	***	0.23	***
	(0.03)		(0.03)		(0.03)		(0.03)	
Employment status: unemployed	-0.17	***	-0.17	***	-0.15	***	-0.07	**
(Base: Employed)	(0.03)		(0.03)		(0.03)		(0.03)	
Employment status: self employed	0.02		-0.04		-0.03		0.13	***
	(0.03)		(0.03)		(0.03)		(0.03)	
Employment status: retired	-0.02		0.05	**	-0.00		-0.08	***
	(0.02)		(0.02)		(0.02)		(0.02)	
Employment status: non labor market	0.07	***	0.06	**	0.06	**	0.04	
	(0.02)		(0.03)		(0.02)		(0.02)	
Constant	-0.29	***	-0.26	***	-0.31	***	-0.08	*
	(0.04)		(0.05)		(0.04)		(0.04)	
Observations	18509		18509		18509		18509	
Adjusted-R2	0.08		0.04		0.06		0.04	

OLS-Regression of trust on measures of risk aversion, betrayal aversion, and altruism (robust standard error in parenthesis). We also performed ordered probit regressions which are not reported. They yield the same conclusions, i.e., risk aversion, betrayal aversion and volunteering affect our trust measures significantly. All trust measures are taken from the German Socio-Economic Panel. They are standardized so that they have mean zero and a standard deviation of 1. The trust index takes averages of the answers to all three trust questions. Our preference measures displayed in Table 2 are explained in more detail in Table 1. * $p<0.10$, ** $p<0.05$, *** $p<0.010$

TABLE 3. The impact of risk preferences and social preferences on investors' beliefs about the trustees' back-transfers in the trust game

Dependent Variable	Expected back-transfer if investor transfers five		Expected back-transfer if investor transfers ten		Average expected back-transfer	
Dummy of being a U.S. resident	0.32	***	1.04	***	0.34	***
	(0.11)		(0.21)		(0.10)	
Risk aversion: high	-0.14		-0.31		-0.16	
(Base: low)	(0.13)		(0.25)		(0.13)	
Risk aversion: medium	0.03		-0.26		-0.05	
	(0.11)		(0.22)		(0.11)	
Betrayal aversion: high	0.06		0.00		-0.18	
(Base: low)	(0.21)		(0.40)		(0.19)	
Betrayal aversion: medium	0.05		0.11		0.00	
	(0.10)		(0.19)		(0.10)	
Dummy for volunteering	-0.04		0.42	**	0.01	
	(0.10)		(0.20)		(0.10)	
Dummy for sociability	0.11		0.30		0.10	
	(0.10)		(0.20)		(0.10)	
Constant	4.08	***	6.30	***	3.99	***
	(0.15)		(0.28)		(0.14)	
Observations	1423		1423		1423	
Adjusted-R2	0.01		0.04		0.01	

OLS-Regression of US and German investors' beliefs about the trustees' back-transfers on measures of risk aversion, betrayal aversion and altruism (robust standard error in parenthesis). We also performed ordered probit regressions which are not reported. They yield the same conclusions, i.e., the preference measures have no significant impact on subjects' beliefs. The data are taken from Naef et al. (2008) who conducted nationally representative trust games in Germany and the US. The regressions are based on the pooled data. * p<0.10, ** p<0.05, *** p<0.010.

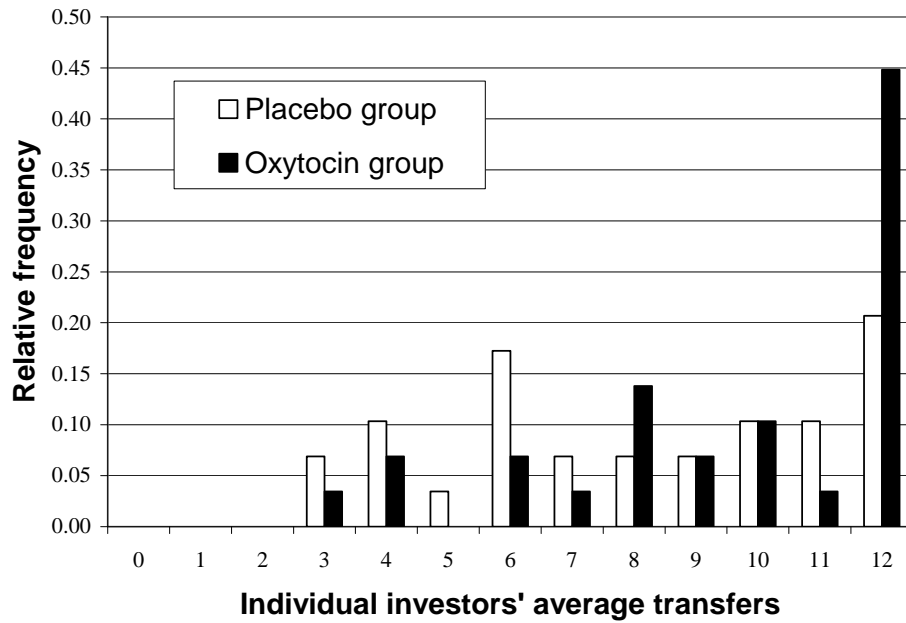Figure 1a: The distribution of trust in the oxytocin and the placebo group

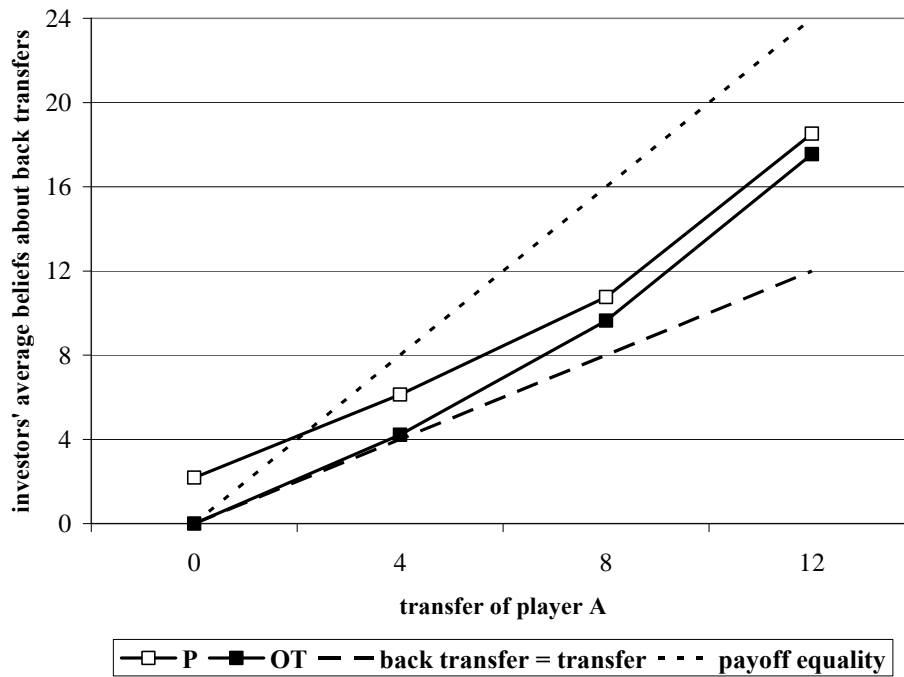Figure 1b: Investors' beliefs about back transfers in the oxytocin and the placebo group

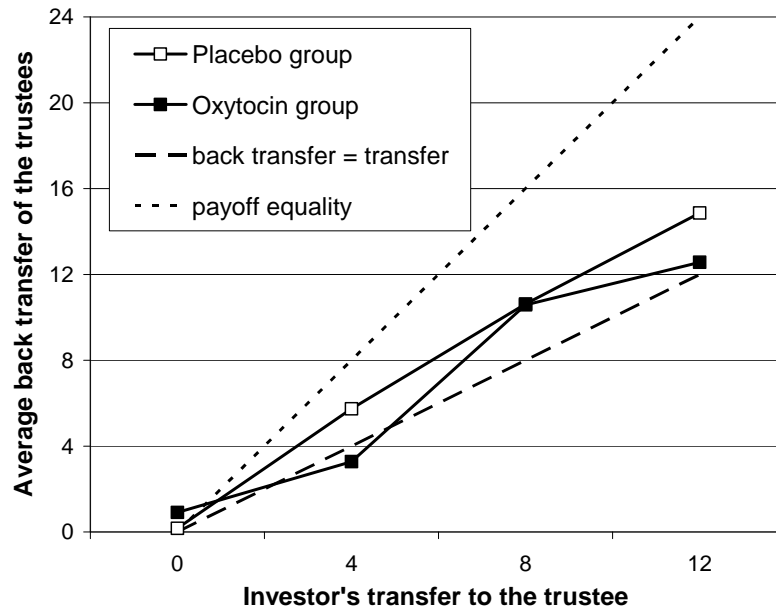Figure 1c: Trustees' backtransfers in the oxytocin and the placebo group

Figure 2a: Employers' beliefs in workers' trustworthiness (measured in terms of expected effort)
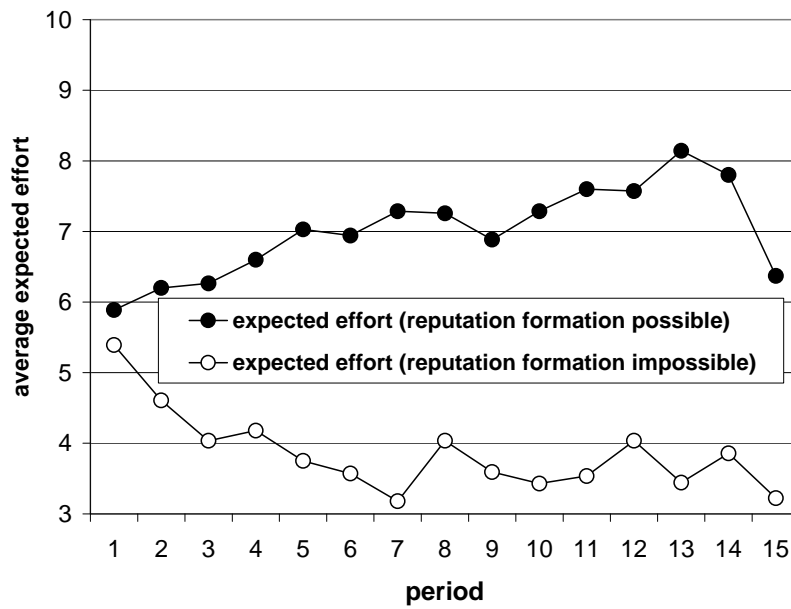

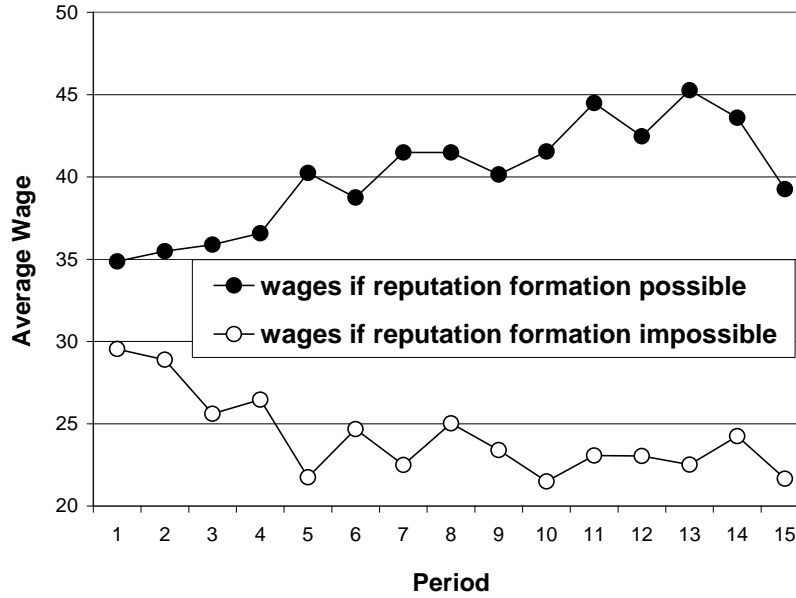Figure 2b: Employers' behavioral trust measured in terms of actual wages

Figure 2c: Workers' trustworthiness measured in terms of actual effort