

In: S.R. Thye, E.J. Lawler, M.W. Macy and H.A. Walker (2001, eds.), *Advances in Group Processes* 18. Amsterdam: JAI, Elsevier Science, pages 129-160.

*Do you trust? Whom do you trust? When do you trust?*

Chris Snijders  
Utrecht University  
Heidelberglaan 1  
3584 CS Utrecht  
The Netherlands  
C.Snijders@fss.uu.nl

Gideon Keren  
Technical University of Eindhoven  
P.O.B. 513  
TEMA building 0.08  
5600 MB Eindhoven  
The Netherlands  
G.B.Keren@tm.tue.nl

Running head: Trust in single-shot encounters

## Abstract

We try to provide a broader view on the factors that influence the decision to trust and honor trust. Using the “Trust Game” as our experimental paradigm, we consider three classes of factors that may be related to trust issues. The first one considers individual differences with regard to the probability to trust others (and honor trust of others), or *disposition factors*. Which kinds of people are more likely to trust? Second, we examine who is more likely to be trusted (*anticipation factors*), focusing on the appearance of the person who is to be trusted. And third, we analyze the circumstances under which trust is more likely to evolve (*situation factors*). Trusting is easy if there is not much at stake, but if the stakes and the risk increase, then how does that affect the willingness to trust? In short, we consider the decision to trust to be dependent on who you are, on who it is that has to be trusted, and on the specific situation. Moreover, we analyze which of these three classes is more important, using a set of experiments designed to measure the impact of disposition, anticipation, and situation factors. The data suggest that disposition factors play a minor role; the differences between people with regard to their likelihood to trust are relatively small. Anticipation factors, operationalized by varying alter’s appearance, had a larger but somewhat paradoxical effect. Those with a trustworthy appearance are indeed trusted more easily, but they do not actually behave more trustworthy. By far the strongest influences were found among the situation factors. Both the payoffs and the structure of the game have a large impact on trust and honoring trust.

## INTRODUCTION

Suppose you are searching for a '69 Ford Mustang and come across a newspaper advertisement offering exactly the car you were searching for. Arriving at the address mentioned in the ad, you are confronted with a gray-haired guy, about 75 years old, called Sal, who tells you he is an ex-mechanic who patches up old cars. The polish on the shiny Mustang in the garage has the same roguish twinkle as the guy's eyes. The price he asks sounds reasonable assuming the car is in as good a shape as he claims ("It's just a hobby, my friend, I don't want to make a profit."). Actually, if you were sure the car was in a good shape as he claimed, you would certainly buy it. You realize, however, that Sal could just as easily be an ex-salesman as an ex-mechanic, and that the twinkle in his eye might not reflect his honest nature but the prospect of a large sum of money coming his way for a car that isn't worth half the price. Would you buy the car? Would you buy it if the price were lower? Would you buy the car if the guy were not a twinkling-eyed, gray-haired, 75-year-old, but a flashy 30-year-old yuppie? On what factors would you base your decision?

The basic issue of interpersonal trust is this: Ego has to decide whether or not to hand over control over the situation to Alter. Alter can choose to honor or betray the trust of Ego, has a certain incentive to betray, and Ego knows it. Many daily life situations, including the one mentioned above, have a similar structure that is well captured by the "Trust Game" (Dasgupta 1988, Kreps 1990) in Figure 1.

<< INSERT FIGURE 1 (TWO TRUST GAMES) ABOUT HERE >>

In the Trust Game, player 1 moves first and has to choose between two possible acts: moving left (i.e., no trust) thus ending the game (resulting in each player getting \$20), or moving right (i.e., trust) thus letting player 2 decide. Player 2, if getting his turn, must choose between moving left (i.e., abuse trust) and keeping \$100 for himself, or moving right (i.e., honor trust) entailing an even split. If player 1 would be sure that player 2 would split the \$100, he would certainly let player 2 decide. Player 1, however, is uncertain about player 2's choice, and if sufficiently unsure may want to settle for the 20 dollars. Figure 1 shows both this Trust Game, and the more general format in which  $P$  represents the payments both players receive when player 1 decides not to trust,  $R$  stands for payments to both players when player 1 trusts player 2 and player 2 reciprocates, and, finally,  $S$  represents what player 1 ends up with when player 1 has trusted player 2 and player 2 subsequently decides to abuse trust and receive  $T$ . The chosen symbols emphasize the resemblance with the Prisoner's Dilemma. As in the Prisoner's Dilemma, the relation  $S < P < R < T$  has to be satisfied.

The approach presented here is based on observing individuals' behavior in Trust Games, and in that respect differs from many studies that appraise trust by using one or another measurement scale. We measure a behavioral response instead of an attitude or a mental disposition. This implies that our pragmatic operational definition of "trusting" is "choosing to let player 2 decide in a Trust Game", and similarly "being trustworthy" is taken to be identical with "choosing the  $(R,R)$  payoff in a Trust Game". This definition fits much of the recent literature on trust (e.g., Mayer, Davis, and Schoorman, 1995; Gambetta, 1988; Rousseau, Sitkin, Burt, and Camerer, 1998). In a cross-discipline review

on the meaning of trust, Rousseau et al. (1998) conclude that there seems to be a consensus that trust is “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another”. In the experimental approach presented here, this intention is assessed with responses in Trust Games. This implies that we consider trust as a decision under risk, where the risk stems from the possibility of opportunistic behavior by alter. The trust is in the *intention*: ego trusts alter, who is capable of honoring trust but may choose not to. You trust in the intention of your friend to return the book he has borrowed. Another kind of trust is trust in *capability*: ego trusts alter to be able to perform a certain task and alter tries to deliver but may not be able to. You trust in the capability of the pilot to fly the airplane and you can in all likelihood assume that a breach of trust is not related to the pilot’s intentions (cf. Misztal, 1996, chapter 1). The current paper is solely devoted to the former interpretation of trust.

By combining the findings of several experiments employing the Trust Game, we try to provide a broader view on the factors that influence the decision to trust and honor trust. To this end, the experiments reported here are categorized into three classes: We first describe experiments that examine individual differences trying to identify which people are more likely to trust (*disposition factors*). We try to identify individuals or groups of individuals who are, in general, more likely to trust (or, more likely to honor trust). Moreover, we want to find out what the size of the difference between (groups of) individuals is. Do males trust more often than females? Or, do older people honor trust more often? Can these differences be quantified? Second, we examine who is more likely to be trusted (*anticipation factors*), focusing on the appearance of the person who is to be trusted. Do people differ with regard to their trustworthy appearance in the sense that people agree about who appears trustworthy and who doesn’t? How veridical are these perceptions on which judgments of trust are formed? Finally we analyze the circumstances under which trust is more likely to evolve (*situation factors*). Trusting is easy if there is not much at stake, but if the stakes and the risk increase, then how does that affect the willingness to trust? In short, whether you trust someone depends on who you are, on the one to be trusted, and on the specific situation. In terms of the Trust Game, whether you trust someone in a Trust Game depends on who you are, on with whom you are playing the Trust Game, and on what that Trust Game looks like. In the present chapter we examine the relative impact of these three classes of factors. As is apparent from the structure of the Trust Game, our investigations are limited to static two-person settings.

Naturally, interactions between elements from different categories are possible. For instance, some people may trust more under some conditions but not under others. Or similarly, under some circumstances some people may be trusted more than others. It is easy to come up with ways in which all three kinds of factors interact. Some people may have different perceptions of the same situation (disposition), which influences the way in which they look at the other person (anticipation) in certain Trust Games (situation). Though such interactions are not only possible but also probable, they are beyond the realm of this study; only main effects are considered.

An important advantage of using the Trust Game is that it explicitly disentangles two concepts that are sometimes treated as identical: *trustfulness* (willingness to trust) and *trustworthiness* (willingness to honor given trust). Trust is sometimes considered to

be a lubricant of social and economic life. Underlying this assertion is the assumption that trustfulness and trustworthiness come hand in hand, which need not be true. In fact, most scales measuring trust, like Rotter's (1967) Interpersonal Trust Scale, or the Dyadic Trust Scale by Larzelere and Houston (1980), actually measure only trustfulness. Finally, using the Trust Game enables us to manipulate the variables that form the basis for decisions on trust simultaneously.

The current literature suggests that all the three factors mentioned above play a role in the decision to trust (or mistrust). Regarding disposition, there is evidence that some individuals seem to be more prone to trust than others (Yamagishi, 1986). For instance, research on social dilemmas suggests that economists tend to be less cooperative (Frank, Gilovich, and Regan, 1993) and therefore perhaps less trustful and trustworthy. Trust also seems to depend on beliefs regarding the others' intentions and inclinations (Parks, Henager, and Scamahorn, 1996). Indeed, predicting and anticipating others' behavior is based on previous knowledge and experience with those others, or by physical appearance such as facial outlook and other behavioral signals (van Lange and Kuhlman, 1994). Finally, situational variables may also influence considerations of trust. For instance, there is substantial evidence showing that trustful behavior in interdependent situations, like public good games, strongly depends on the size of monetary incentives (Ledyard, 1995).

In essence, we address the explanatory power of the "dispositionalist view" as opposed to the "situationalist" view (Mischel, 1968), applied to the question of trust. Mischel has convincingly shown that cross-situation consistencies are low for several traits, like honesty and aggressiveness. Other research (e.g., Bem and Allen, 1974) challenged these findings and argued, among other things, that it may very well be that consistencies are stronger for specific domains. Trust could be one of them.

In what follows, we describe a series of experiments on trust, conducted in the past few years, and interpret the results in light of a broader literature. The first three sections address the dispositional, the anticipation, and the situational questions, respectively. The final section is an attempt to combine the different results in order to obtain an overall assessment of the determinants of trust.

## DO YOU TRUST?

The first approach to the question of trust is to examine whether some people are more likely to trust than others. Several studies examined which characteristics of people correlate with trustful behavior yet the empirical evidence is quite mixed. Some studies investigated sex differences and trust, but – possibly because trust is not consistently operationalized – findings are not unequivocal (compare, for instance, Aranoff and Tedeschi 1968, Rapoport et al. 1976, Dawes et al. 1977, Terrell and Barrett 1979, Wright and Sharp 1979, Heretick 1984, Lagace and Gassenheimer 1989, and Steel 1991). Given the inconsistent evidence, one may doubt whether any universal sex differences, as far as trust in others is concerned, do exist. Such doubts have been echoed by Orbell et al.'s (1994) review of sex (gender) differences in social dilemmas, and by Kagel and Roth's (1995) review of public goods experiments.

Other characteristics that correlate with trust have been less extensively studied. For instance, Larzelere (1984) found that religious (Christian) students were less trustful than secular students. Yamagishi and Yamagishi (1994) provide empirical evidence for

the claim that Americans, in general, are more likely to trust other people compared with Japanese. There is apparently a positive relation between trust and marital status (Larzelere and Huston 1980), a negative relation between trust and suspicion (Misra and Kalro 1979), and some complicated relations between trust, race, and socio-economic status (Switkin and Gynther 1974, Terrell and Barrett 1979, Steel 1991). Wheelless and Grotz (1977) report a relation between trust and self-disclosure.

Liebrand et al. (1986) introduced the “Might versus Morality” effect: some people evaluate behavior in terms of strong versus weak (“might”), while others evaluate behavior in terms of good and bad (“morality”). Those who emphasize might are more likely to behave non-cooperatively in trust issues (i.e., they are less trustful and less trustworthy), because they tend to behave in a way that maximizes their own profits without regard for the profit or loss of others. Likewise, those that emphasize morality are more likely to behave cooperatively in trust issues, because they tend to behave in a way that maximizes profit from the point of view of the collectivity. Persons who have been frequently exposed to the evaluation of behavioral alternatives in terms of might, like those who have studied game theory or economics, may therefore be less likely to trust others and less likely to honor the trust of others. The latter factor in particular appears to have a strong effect. Economists tend to trust significantly less often, as has been reported by Marwell and Ames (1981), Carter and Irons (1991), and Frank et al. (1993).

In the first set of experiments we report on (Snijders, 1996), participants (university students) played several Trust Games and some of the individual characteristics mentioned above were examined. Participants in the other study were respondents of a survey among households who played a Trust Game, controlling for various socio-economic background characteristics. The main results of these experiments are briefly described below.

#### Experiment 1: Effects of disposition in a student population

Participants (466, mostly students) in the first set of experiments played several (3, 5 or 6) one-shot Trust Games. Sessions were run under laboratory conditions with about 20 participants per session. In total, 36 different Trust Games were employed (average values of  $S$ ,  $P$ ,  $R$ ,  $T$  were 0, \$10, \$20, \$40 respectively). For each Trust Game, participants were asked to indicate what they would choose in the role of player 1 *and* what they would do in the role of player 2. Participants were told in advance that at the end of each session, two participants would be randomly determined as winners: a player 1 and a player 2. Subsequently, one Trust Game would be randomly chosen and the winners would be paid in accordance with their choices in that particular Trust Game. This procedure implies that participants knew that “the other player” in a Trust Game would in fact be a randomly determined other in the same session. In addition, participants filled out a questionnaire measuring several characteristics that we considered likely to correlate with trustful or trustworthy behavior. All participants received a flat fee (about 5 US\$). Further details are provided in Snijders (1996) and Snijders and Keren (1999).

The demographic information obtained from participants consisted of sex, age, field of study, whether they carried a donor codicil, were blood donors, had any knowledge about game theory, and a factor score of their high school graduation package. As a last measure of disposition, the “Ring” measure of social orientations

(Liebrand 1984) was included. This standard test measures the extent to which a person values payoffs to others (Liebrand and McClintock 1988). Usually, the test results are used to divide subjects in “pro-socials” (or “cooperators”), “individualists”, and “competitors”. Competitors are relatively scarce (about 15% of the subjects), so that individualists and competitors are often clustered and labeled “pro-self”. In general, the prediction is that pro-socials are more cooperative, more trusting, less cynical, more optimistic, and more caring about other than pro-selves. Table 1 summarizes the results of our analyses (Snijders 1996). Results are based on a repeated measures Probit analysis. The first column indicates the direction of the hypothesized effects. The numeric entries in the following two columns represent the effect on the probability to trust (honor trust). For instance, there is a mean difference of 0.10 in the probability to trust between subjects carrying a donor codicil and those who do not.

<< INSERT TABLE 1 (NET EFFECTS OF DISPOSITION) ABOUT HERE >>

Inspection of Table 1 suggests that males tend to be more trustful and less trustworthy, and participants who carry a donor codicil seem to be more trustful. Pro-socials are both more trustful and more trustworthy. Most of the other indicators were not statistically significant. These weak results regarding disposition emphasize the danger of assessing other persons’ trustworthiness through stereotyping or categorization (that is, by basing assessments on the group to which the person belongs to). Apparently, it is difficult to find a category of people who trust (or honor trust) significantly more than others, at least in a student population. Measuring a person’s social orientation provides the most useful information (hence, look for pro-socials, beware of pro-selves).

The above experiment should not be considered as conclusive evidence that differences in disposition to trust are scarce and small. Nevertheless, a literature scan suggests that only few individual-differences indicators have been unveiled, and even those usually show a weak effect.

A possible objection regarding this experiment is that for some indicators there is probably not enough variation in the subject population to be able to find relevant differences. For instance, if people become less trustful as they get older, we are not likely to find such an effect in a population of students with ages between 18 and 25 years. This lack of variation was the impetus for conducting a second experiment.

#### Experiment 2: Effects of socio-economic characteristics in a population at broad.

The data for this experiment were obtained as part of a large survey of Dutch households, with 1533 two-person households and 288 singles (Kalmijn, Bernasco, and Weesie, 1995). Every interview ended with presenting respondents with a booklet consisting of several socio- and psychometric scales, one or another version of the Prisoner’s Dilemma, and one Trust Game (Bruins and Weesie, 1996). About 70% (2283 out of 3354) booklets were returned, of which half are of interest for our analysis here depending on which characteristics are taken in to account (for instance, not all respondents were willing to state their incomes, or their political preference). Behavior of the trustor in a Trust Game was examined and correlated with several socio-economic

characteristics of the respondents. Admittedly, the choice of the socio-economic variables is somewhat arbitrary, and the analyses should be considered explorative. Four clusters were constructed labeled as 'General characteristics', 'View on life', 'Control over short term gratification', and 'Previous breach of trust'.

Included under 'General characteristics' were sex, education, household income and age. These characteristics are quite broad and permit different sorts of reasoning regarding their relation with trust. For instance, one could conjecture that as people grow older, they will trust more (because they care less about trust being abused), less (because their trust has been betrayed more often) or use better judgment (because they have learned when to trust and when not). Hence, these variables were simply added as controls

The cluster 'View on life' was assessed by kind of job, religion, political affiliation, and whether one lives in a large city or small village. The third group of indicators ('Control over short term gratification') was supposed to examine susceptibility to short-term incentives. Underlying this cluster is the supposition that the more likely one opts for immediate gratification (like smoking or drinking), the less trustful she will be because the inability to resist the (usually short-term) temptation to abuse trust. Finally, the category 'Previous breach of trust' consisted of indicators that could be conceived as major breaches of trust: having divorced parents, having a working mother (measured at age 14), the number of previous relationships, the number of previous cohabiting relations, and the number of prior marriages. Naturally, the effects of the different payoffs in the different Trust Games were controlled. The results of this study, based on a Probit analysis that takes clustering within households into account, are summarized in Table 2. The numeric entries represent the effect on the probability to trust. For instance, on average the difference in the probability of trust between Roman Catholics and others is 0.06. Roman Catholics trust less.

<< INSERT TABLE 2 ABOUT HERE >>

Inspection of Table 2 suggests that only few socio-economic characteristics are correlated with trust. Regarding trust, no evidence was found that responses obtained from highly educated youngsters differ from responses from a representative sample of inhabitants of the Netherlands. Two significant effects stand out. People who donate to charity are also the ones high on trust, and people whose mother worked when they themselves were fourteen years old are also more likely to trust. The latter is inconsistent with our original intuitions: The variable 'mother worked when respondent 14' was meant to serve as an admittedly extremely crude proxy for, say, "growing up in a cold nest".

Overall, the two experiments display a similar pattern. There is little evidence to suggest that some groups, or people with certain traits, tend to trust more often than others. Certainly our sample size warrants the expectation that even small differences between individuals would have tended to materialize. A related concern is that perhaps the Trust Game is not a very reliable measurement instrument, and that we therefore look at trust only through a haze of measurement error (cf. Thye 2000). A conclusive reply to

this argument can only be given based on analyses taking the measurement error explicitly into account. What we can say is that in the remainder of this chapter, other effects *do* materialize, in spite of possible measurement error, which provides some support for the assertion that the Trust Game has reasonable reliability as long as one assumes that measurement error does not vary systematically with the experimental conditions. The effects of disposition we find here, though relatively weak, seem to relate to pro-social behavior. Those who score high on a scale measuring pro-social behavior and those who donate to charity, trust more often. These results are compatible with Uslaner's (in press) claim that trust is related to a person's general world-view and not so much to a person's experience with issues regarding trust.

### WHOM DO YOU TRUST?

How important are the characteristics of a person you consider to trust? Certainly, the extent to which one is willing to trust someone else would depend on characteristics of that other. Trusting a brother or other family member is different from trusting a stranger. There is also evidence suggesting that one is more likely to trust someone who is in some sense similar to oneself. For instance, the Dutch trust the Dutch more than they trust other nationalities (Snijders, 1998). Similar results have been reported on cooperation between in-group members, where in fact "belonging to the in-group" can be manipulated quite easily (Tajfel and Turner, 1979; 1986), for instance by dividing a group of people in two separate groups, with each group wearing different hats.

Kinship relation (e.g., family, friends, in-group) may naturally affect the level of trust. Notwithstanding, in the current section we focus on trust based on *appearance*. Even without any information about a person, trust in daily life is often based on guts feelings and the belief (justified or not) that we are able to accurately "read" a person's face or appearance ("first impressions are often truest"). There are (at least) two reasons why appearance of others could guide our assessments regarding their trustworthiness.

First, it is known that people are particularly susceptible to what has been coined the *fundamental attribution error* or *correspondence bias*. There is an inclination to attribute observed behavior to an actor's disposition more than to situational circumstances, even when obvious situational incentives exist that are sufficient to explain that behavior (Ross, 1977). In other words, we tend to interpret behavior as a strong sign of character. Moreover, not only a person's behavior but also his or her appearance is often taken as displaying that person's character. The shape and expression of the face, body language, and clothing can be conceived as providing information about a person's nature, and people are more than willing to use such cues.

A second, and related, argument relies on what has been termed (*dispositional*) *category based expectancies* (Jones, 1990, page 79): the belief that members of a particular group (e.g., rock stars, people over 65, men with thick black eyebrows, etc) share similar dispositions. A glance at someone's appearance is often enough to relate that person to a group. A category-based expectancy is frequently formed, especially when no other information about the other person is available. Whether these conclusions about others, based on whatever clue, are warranted is of course an entirely different question.

Evidence in the literature suggesting that people's behavior towards other persons is affected by physical appearance is, to the best of our knowledge, mainly related to physical attractiveness. Examples abound in diverse areas of interest. For instance, physically attractive students are judged more favorably by their teachers on a number of dimensions including intelligence and various social skills (Ritts, Patterson, and Miles 1992); physical attractiveness of children is associated with their teacher's expectations about them (Clifford and Walster 1973); physically attractive persons with HIV are considered more responsible for their condition than others (Agnew, Thompson and Vaida 1994); physically attractive female-offenders are treated differently by male intake personnel (Rosenbaum and Chesney 1994); men are more likely to help physically attractive females (Harrel 1978), and personnel professionals are reported to have a (slight) bias in favor of physically attractive job candidates (Morrow et al 1990). The effect of appearance is not limited to physical attractiveness. Brownlow (1992) suggests that credibility is affected by facial appearance, and Lennon (1986) demonstrated an effect of clothing on first impressions of strangers. Given the empirical evidence, the purpose of experiment 3 was to directly examine possible relationships between trust and appearance.

### Experiment 3: Do people base their decision to trust on appearance?

Participants, in one of our recent experiments, were shown pictures of four different persons and were asked to assume that they had to play a Trust Game with one of these persons. They were asked to assume that they would play the role of player 1 and the persons displayed in the picture would play as player 2 (the trustee). Their task was to rank order the pictures such that the person with whom they would most like to play a Trust Game would be ranked first, and the person they would least like to play with would be ranked last. Lacking any other cues, it was assumed that participants would judge the pictures based on how trustful the persons' pictures look like.

The pictures (taken from an old student year-book) were chosen by the researchers and two assistants. Using our own subjective judgments, two pictures (T1 and T2) were chosen that seemed to reflect trustworthy appearances, and another two pictures (U1 and U2) that supposedly represented untrustworthy appearances. The pictures were displayed in several sequences so as to avert possible order effects.

The results were unequivocal: Subjects displayed a strong preference for the two pictures that were originally judged as reflecting trustworthy appearance (mean rank ordering for pictures T1 and T2 was 1.54 and 1.94 respectively, compared with 3.23 and 3.26 for pictures U1 and U2). More than 75% of the subjects ranked T1 and T2 over U1 and U2, and the difference in mean ranking between the former and the last pair was so large that it did not require any statistical test. These results are compatible with previous research showing that people tend to trust those who appear more friendly (Mulford et al., 1998).

Admittedly, although the results demonstrate the possible role of appearance in judgments of trust, it does not disclose which facets make the appearance of a person to be seen as more or less trustworthy. To answer this question, a research program is required that is beyond the scope of the present paper.

We consider two natural follow-up questions. First, will the level of agreement exhibited by our participants be maintained with pictures that are not chosen selectively,

as was the case in the previous experiment? Our experiment only demonstrated that it is possible to select faces on which people will exhibit high level of agreement regarding their trustworthiness, yet real life is messier. A second, perhaps more important, question is whether trustworthiness judgments, based on physical appearance, are accurate. The evidence is mixed, though empirical evidence seems to suggest that judging trustworthiness by physical appearance is often unwarranted.

Notwithstanding, there are some indications that appearance is not entirely an unreliable cue. For instance, some studies show that physically attractive people actually behave differently than others. They are reported to be more assertive (Campbell, Olson, and Kleim, 1990), and it seems that their attractiveness has some impact on achievement and psychological well-being (Umberson and Hughes 1987). Adams (1977), in a review article on physical attractiveness, concluded that a relation exists between appearance and internal behavioral processes. Note, however, that the fact that physically attractive persons behave differently than others does not necessarily entail that other people can correctly predict the behavior of such persons.

Another argument in favor of relying on cues based on appearance is that people are reasonably accurate in judging another person's (self-reported) degree of extraversion (Levesque and Kenny, 1993). As noted already by Cronbach (1955, 1958), research in interpersonal perception and *target accuracy* is problematic. Kenny (1994) presented an extensive discussion on methodological issues associated with this type of research and concluded that "social science is beginning to see that the view of the person perceiver as a biased and faulty perceiver is not complete" (1994, page 142).

Other empirical evidence casts doubts about the predictive power of expectations about the behavior of others. Dispositional category based expectancies may have some value in the sense that correlations exist between being a member of a certain group and the score on a trait (such as trustworthiness) or actual behavior (such as trustworthy behavior). However, these correlations are often small and people are poor assessors of correlations. They tend to overestimate the degree of covariation, especially when causal explanations along the lines of stereotypes can be made (Nisbett and Ross, 1980, chapter 5). Moreover, their use of correlational information is often inadequate in that they produce estimates that are too far away from the population base-rate, particularly when the relevant correlation is weak (Nisbett and Ross, 1980, chapter 7).

Given the inconclusive evidence from the existing literature, it was our aim to examine whether the perceived trustworthiness of other persons is a predictor for the probability of trustful behavior of the perceiver, and whether this perception is correlated with the actual behavior of the perceived target. As in the other studies, the Trust game was employed as a measure of actual behavior.

#### Experiment 4: Effect of appearance when perceived trustworthiness is not manipulated

In a laboratory experiment, 134 participants (mostly students) played a one-shot Trust Game with each of four other participants separately. The experiment consisted of 17 sessions with 8 participants each (occasionally 7). Participants were seated in two opposite rows of four. Each participant was asked to play a single (and the same) Trust Game with each of the four other participants he or she was facing. For each dyad, a participant was told to consider the Trust Game and choose what he or she would do as player 1 given that the other participant was player 2, *and* what he or she would do as

player 2 given that the other participant was player 1 (the so-called “strategy-method”). For each session, all participants were asked to decide on the basis of the same Trust Game. For purposes of generality, four different Trust Games were used across different sessions. After playing the Trust Games, participants filled out a questionnaire in which they were asked, among other things, to rate each of the 4 persons in the other row on some personality traits, including trustworthiness. Communication between participants was not permitted (other than just looking at the other person). To prevent participants from thinking that their ratings served as a consistency test for their responses in the Trust Games, they were told that the task was meant to examine the degree of consistency in the first impressions people have about other people (see Snijders and Keren, 1995 for details).

To assess whether people are indeed more likely to trust those of whom they say that they appear trustworthy, we compared the choices made by a participant in the role of player 1 with that participant’s rating of the trustworthiness of the other. Answers to other items of the questionnaire (age, sex, being a blood donor, donor codicil ownership, knowledge of game theory) were used as control variables. Values of the payoffs in the Trust Game were used as additional controls (employing dummy-variables). The data were submitted to a logistic regression analysis of subjects’ decisions to trust, and a summary of the results is portrayed in Table 3. The numeric entries represent the average effect on the probability to trust (honor trust).

<< INSERT TABLE 3 (NET EFFECTS OF PERCEIVED TRUSTWORTHINESS)  
ABOUT HERE >>

The analysis is intricate because of the dependencies in the data. Complications arise not only because each subject made eight decisions (four times on trustfulness, four times trustworthiness), but also because four subjects in a single session played with the same players 2. The difficulty of dealing with these interdependencies is precisely why, according to Kenny (1994), a lot of previous results should be interpreted with care. We refer the reader to Snijders (1996) for the details of the analysis that was used here.

Table 3 shows that the mean effect of perceived trustworthiness on the probability to trust equals 0.12, which needs some clarification. Trustworthiness was measured on a 20-point scale. The estimated effect of perceived trustworthiness equals 0.02 per unit on that scale. About 90% of the subjects assessed the difference between the most and the least trustworthy person they were facing as less than 13 units on the scale, with a mean of about a 6 unit difference. One could say that for the most extreme subjects, the effect of perceived trustworthiness equals 0.26 (13 times 0.02). On average, it is about 0.12 (6 times 0.02). The effect of perceived trustworthiness on actual trustworthiness cannot be statistically distinguished from zero. First impressions regarding the trustworthiness of alter are not an adequate predictor of the actual behavior of alter. This can be understood loosely as providing empirical evidence against Frank’s (1988) idea that tell-tale signs like appearance ease trust and cooperation between strangers. More precisely, we should say that appearance apparently does not tell the right tale.

Additionally, subjects did not agree at all about who appears trustworthy (inter-rater agreement kappa of 0.04). Real life is indeed messier than our Experiment 3 in the sense that in this standard student population there is no consensus about who looks trustworthy at all. It should be emphasized that the experiment does not disclose which cues are used to assess trustworthiness. As mentioned earlier, there is evidence in the literature for a mixture of potential cues. Clothing (Lennon 1986), physical beauty (Eagly 1991) and facial features (Berry and McArthur 1985), are probably the most prominent potential cues in the experimental design. Potential cues could also be based on (perceived) *similarity* rather than on generally acknowledged properties of the other individual. It is a well established research finding that people tend to like those people whom they perceive to be similar to themselves, and likewise that people assume that similar others will also like them (e.g., Newcomb 1961, Aronson and Worchel 1966, Byrne 1971, and Condon and Crano 1988). Similarity could possibly account for the fact that subjects disagreed about who looked trustworthy: they only assessed those subjects as trustworthy who, in some way, resembled themselves. One could hypothesize on the basis of similarity that males perceive other males to be more likely to honor their trust (“he is one of us guys”), and females perceive other females as more likely to honor their trust (“she is one of us girls”). Alternatively, if a male judges another male’s trustworthiness, he has the advantage of being similar to the one he judges. The other person is a male just like him, and therefore may better be able to judge how that other person is going to behave in a Trust Game. This suggests that player 1 may be more likely to rely on what he or she would do if he or she were in player 2’s shoes, if player 2 happens to be of the same sex. Additional analyses revealed that no such effects are supported by the data.

## WHEN DO YOU TRUST?

Besides possible disposition and physical appearance effects, it is quite likely that characteristics of the specific situation under which trust is called for, may also play a role. Here, a wide range of “situational variables” can be thought of. For instance, trust is more easily established when one expects a long-term relationship with the partner (Axelrod’s 1984 *shadow of the future*). It is also more readily established given a history of mutually trustful and trustworthy behavior with the same partner. The presence of a third party may also facilitate the formation of trust because the potential trust abuser fears for his reputation. One straightforward aspect of the situation, and one that we will focus on, is what is at stake. Lending someone \$ 5 requires a different level of trust than lending \$ 5,000. Indeed, the experimental literature provides ample support regarding the sensitivity of people to the size of what is at stake. Even in situations where a narrow interpretation of economic theory (people behave like rational egoists) predicts non-cooperative behavior independent of the payoffs, people are sensitive to changes in the payoffs (Rapoport, Guyer and Gordon, 1976; Steele and Tedeschi, 1967; Ledyard, 1995; Sally, 1995).

In the present section we investigate the claim that the degree of trustfulness and trustworthiness depend on what is at stake. We thus predict that the level of trust in the Trust Game would depend, among other things, on the values  $S$ ,  $P$ ,  $R$ , and  $T$ . The

question is, of course, how differences in these parameters can account for differences in trust. To this end we adopt what Kelly and Thibaut (1978) called the “effective decision matrix” according to which people behave as if they transform the given monetary payoffs into “effective payoffs”, where all attributes relevant to the decision are included. Several ways of creating an “effective Trust Game” are possible. For instance, one could argue that what is lacking from the model is the feeling of guilt that player 2 will have when he or she abuses the trust of player 1. This can be modeled as an effective Trust Game by assuming that if player 2 in a Trust Game abuses the trust given by player 1, player 2 values the payoff he then gets ( $T$ ) somewhat less. Figure 2 displays the effective Trust Game.

<< INSERT FIGURE 2 (EFFECTIVE TRUST GAME USING GUILT) HERE >>

In the guilt model, player 2 honors trust if and only if that player’s “guilt parameter”  $\gamma_2$  is larger than  $(T-R)/(T-S)$ . Following the model, player 2 will not abuse player 1’s trust due to her feelings of guilt for pocketing more cash than player 1 ( $T$  for player 2, as opposed to  $S$  for player 1). Specifically, this guilt feeling is amplified by the fact that it is exactly player 1’s behavior that enables player 2 to receive more money, in the first place. Other researchers have proposed the same ratio as being related to the percentage of cooperative choices in a Prisoner’s Dilemma (Harris, 1969; Rapoport and Chammah, 1965), though these ratios were inspired by arguments based on repeated rather than one-shot games. Here, the ratio is a natural consequence of the assumptions of the guilt model. Now suppose that indeed all persons in the role of player 2 bear their personal guilt parameter and behave accordingly. Then, as the index increases, the percentage of players 2 who possess a large enough guilt parameter to be willing to honor trust decreases. Players 1, in the guilt model, will trust only if the probability that they assign to player 2 honoring their trust is larger than  $(P-S)/(R-S)$ . The probability that player 2 honors trust depends on  $(T-R)/(T-S)$ , so the behavior of player 1 depends on both indices. For convenience, we label these ratios *risk* and *temptation* respectively.

The effective Trust Game in Figure 2 is just one out of many possible effective Trust Games, and it is tempting to debate which is the most appropriate way to model guilt (or regret, or altruism). However, that is not so much the issue here. What can be shown is that the same indices (risk and temptation) tend to pop up under different reasonable effective Trust Games. More precisely, under different “reasonable” effective Trust Games, risk is a major determinant of player 1’s behavior, whereas temptation affects the behavior of both player 1 and player 2. Though other indices can be derived, these are often highly correlated with risk and temptation (Snijders, 1996; Snijders and Keren, 1999). Note that this way of modeling Trust Games uses “psychological parameters”, like  $\gamma_2$  introduced in Figure 2, but generates hypotheses about the relation between the indices and behavior that do not necessitate actual measurement of these parameters. The main conjecture of the present section is that trustfulness (the behavior of player 1) varies with risk and temptation, and trustworthiness (the behavior of player 2) varies with temptation.

Reanalysis of Experiment 1: The effect of the monetary incentives in Trust Games.

The data obtained in a set of experiments described earlier (Experiment 1) was reanalyzed with a focus on the effects of the monetary incentives. Recall that participants in this set of experiments played several one-shot Trust Games. In total, 36 different Trust Games were employed (average values of  $S$ ,  $P$ ,  $R$ ,  $T$  were 0, \$10, \$20, \$40 respectively). The values of  $S$ ,  $P$ ,  $R$ , and  $T$  were chosen such that there was an adequate variation in the values of risk  $((P-S)/(R-S))$  and temptation  $((T-R)/(T-S))$ . The value of risk varied between 0.08 to 0.71, and temptation between 0.09 and 0.83 (both risk and temptation are contained between 0 and 1). For each Trust Game, participants were asked to indicate their choice in the role of player 1 *and* their choice in the role of player 2.

Table 4 summarizes the results of our analyses. Besides the payoff indices (risk and temptation), we also include the results of a separate analysis using the net effects of the separate payoffs. The numeric entries represent the effect on the probability to trust (honor trust) of an increase of 1 in the independent variable. For instance, on average the difference in the probability of trustfulness between a risk index of 0.2 and a risk index of 0.7 equals  $(0.7 - 0.2) * -0.85 = -0.425$ .

<< INSERT TABLE 4 (NET EFFECTS OF MONETARY PAYOFFS) ABOUT HERE  
>>

Table 4 indicates unequivocally, that the strongest predictors are risk for trustfulness and temptation for trustworthiness. Among games with low risk and high risk, we find an estimated difference in probability of trustfulness of 0.85. That is, risk can vary between 0 and 1, and the percentage of trustfulness then roughly varies between 0.95 for low risk and 0.10 for high risk. An even larger difference between games associated with low or high temptation is observed with regard to trustworthiness (0.94). These effects are substantially larger than the effects of disposition and anticipation found in the previous experiments. The data thus suggest that what matters most, given the factors we considered, is the situation (defined by the payoffs).

In some respect, the results of the above analysis are at odds with the standard dictates of game theory. Following a strict game theoretical interpretation, player 1's behavior should depend entirely on what player 2 will do. Since player 2's behavior is determined by the temptation to abuse trust, player 1's behavior should therefore heavily depend on player 2's temptation. Our empirical results, however, suggest that the extent to which player 1 will decide to trust is mainly determined by the risk involved. Specifically, it depends on what player 1 can gain as compared to the no trust situation. The temptation of the other seems to be of relatively minor importance (though still of the same order of magnitude as the strongest effects of disposition and anticipation in the previous experiments). Apparently, people are somewhat myopic with regard to trust: In considering whether to trust, they underscore the potential consequences to themselves and tend to put less emphasis on the likelihood that their trust will indeed be honored.

The separate effects of the payoffs  $S$ ,  $P$ ,  $R$ , and  $T$  are harder to interpret because, for instance,  $P$  is constrained between  $S$  and  $R$  so that the potential increase in the probability to trust, depends on the particular Trust Game being played. It should be noted that, contrary to our initial expectations, the probability of trustworthy behavior

does not depend on the payoff  $P$  (and also not on risk). The data show that not all people abuse trust, even if there is a strong monetary temptation to do so. The question is what kind of motives may induce player 2 not to give in to the temptation? One such motive could arise from the fact that player 2 appreciates that player 1 has actually gone out on a limb by trusting, and thus feels a sense of obligation to reciprocate. The larger the potential sacrifice of player 1, the larger the motivation for reciprocity (i.e., players 2 are more likely not to give in to temptation for larger values of  $P$ , or for larger values of risk). This, in turn, triggers the question what it is that people are reciprocating when they decide to honor the trust given to them. To answer this question we turn to another situational aspect besides the monetary payoffs, namely the structure of the game.

Why is trust being honored? One reason, as suggested above, is reciprocity. An alternative explanation is based on plain altruism. If a trustor is lucky enough to be confronted with a trustee who has genuine feelings for the trustor (or for trustors in general), trust will be honored. There is, however, an essential difference between genuine altruism and reciprocity. Reciprocity implies a certain obligation or debt, but this obligation is conditional on previous behavior of the trustor. It is this conditional aspect that discerns it from altruism. This is also why the judgment of the extent to which reciprocity has taken place in an experimental setting is complicated by a methodological problem. Just looking at what subjects do when confronted with one or several Trust Games does not suffice. In fact, several researchers (e.g, McCabe, Rassenti, and Smith 1996) have employed the concept of reciprocity to account for cooperative behavior in different types of games, without taking into account rival explanations based on altruism. Indeed, there is ample empirical evidence that considerations of fairness and decency cause people to act cooperatively even in games (like “dictator” or “ultimatum”) in which there is no place for reciprocity (e.g., Kagel and Roth, 1995, p. 270; Dufwenberg and Gneezy, 1996). Therefore, to infer reciprocity, it is essential to employ a control group where the possibility of reciprocity is precluded.

We propose two ways in which characteristics of the situation may influence the triggering of reciprocity, which we label *Bookkeeping* (of utility associated with gains and losses) and *Perceived Intentionality* (Keren and Snijders, 1999). The first is based on calculations for achieving a balance in terms of costs and benefits, the second is a subjective appraisal of the other’s willingness to arrive at mutual cooperation.

In the context of the Trust Game, the bookkeeping conjecture, as the name suggests, is based on sole considerations of “give and take”. It relates to three different factors: The amount of utility gained by the trustee (due the trust she was endowed with), the possible loss conceded by the trustor, and the initial relative advantage the trustor has given up. For further exposition, consider the Trust Games depicted in Figure 3. The two Trust Games differ only with respect to the payoffs associated with player 1 not trusting player 2. According to the bookkeeping conjecture there are three possible components that determine whether given trust will be reciprocated. We use subscripts to denote the payoffs to the two players, respectively. First, the difference  $(P_1 - S)$  embodies the possible loss player 1 incurs by trusting player 2, and in this respect reflects a potential sacrifice. In the second Trust Game in Figure 3 this difference is larger ( $40 - 0$  versus  $20 - 0$ ), so the potential sacrifice is larger. Second, the difference  $(R_2 - P_2)$ , or  $(T - P_2)$  if player 2 does not honor trust, represents the increment or possible gain in player 2’s utility as a consequence of player 1 trusting. Again, this difference is larger in the second

Trust Game: player 2 gains more if player 1 trusts in the second Trust Game. Finally, the difference ( $P_1 - P_2$ ) may also play a role. Specifically, if  $P_1 > P_2$ , which implies a relative advantage for player 1, then giving up this advantage can also be considered as part of the sacrifice of player 1. In other words, the condition  $P_1 > P_2$  may magnify the sacrifice of player 1 and trigger reciprocal behavior of player 2. Once more, player 1 sacrifices more when trusting in the second Trust Game. The bookkeeping conjecture is supported if players 2 would tend to cooperate more in games where ( $P_1 - S$ ) is large, ( $R_2 - P_2$ ) is large, and  $P_1 > P_2$ . In other words, one would expect more honoring of trust in the second Trust Game in Figure 3. This conjecture is tested by examining the difference in behavior of participants for different pairs of Trust Games, like the pair depicted in Figure 3.

<< INSERT FIGURE 3 (TWO TRUST GAMES) HERE >>

In contrast to the bookkeeping hypothesis, the perceived intentionality conjecture is independent of the specific payoffs or utilities at hand. For trust to be honored, the trustor has to exhibit a genuine intention for mutual cooperation. Perceived intentionality then refers to the extent to which the trustor provides the reciprocator with adequate cues of her cooperative intentions. Consider, for instance, the Trust Game depicted in the upper part of Figure 4 (portraying exactly the same game as in Figure 1).

<< INSERT FIGURE 4 (TRUST GAME, COIN-GAME + CONTROL) ABOUT HERE >>

Now imagine that the role of player 1 is eliminated and the game is reduced to player 2's decision between the right and the left option, as depicted in the lower part of Figure 4. Obviously, in this 'restricted' game there is no place for reciprocity since player 1 did not act at all and hence there is no action that can be reciprocated. A decision to choose the right option can only be motivated by considerations of altruism or fairness. Evidence in favor of the intentionality conjecture would be a difference in the proportion of players 2 who would choose cooperatively across the two games mentioned above (thus, more cooperation by player 2 in the Trust Game than in the restricted game). This conjecture was tested by comparing behavior of trustees in different Trust Games with behavior of trustees in the control condition

In addition, in order to examine whether different levels of intentionality play a role, the extent to which player 1 exhibited cooperative intentions was varied. For instance, suppose that the Trust Game is being played, but the rules of the game are such that player 1 has to make her choice by flipping a coin. In that case, to choose right entails that player 1 has made a sacrifice, but it was clearly not by free will. Since player 1's decision is not intentional, there is less room for expecting reciprocating behavior on player 2's part. We vary the intentionality by varying the rules regarding the basis on which player 1 has to decide (free will versus coin-flip).

Experiment 5: Empirical test of the bookkeeping and intentionality hypotheses .

In total, 879 participants, mostly students, played a single Trust Game, a single control condition, or a single “coin-flip Trust Game” (thus, all comparisons are based on between-subjects designs). As independent factors, we considered whether participants were paid in accordance with their choices or received a flat fee, and whether they participated in the experimental lab or were interviewed on campus. Five different base Trust Games were used. All participants were assigned the role of player 2. See Keren and Snijders (1999) for further details. A summary of the results is portrayed in Table 5. We focus on the results of testing the bookkeeping and intentionality hypothesis. The numeric entries in this table represent the average effect on the probability to honor trust. For instance, on average there is a difference of 0.25 in the probability of trustworthiness between players 2 in a Trust Game and players 2 in the control condition.

<< INSERT TABLE 5 (NET EFFECTS OF BOOKKEEPING AND ...) ABOUT HERE  
>>

The results are clearly in favor of the intentionality hypothesis. The bookkeeping hypothesis, which should lead to differences in trustworthiness because of the payoff differences, received no support for any of the three hypothesized effects (although combining all conditions does show a significant but small effect). In games where all payoff differences are relatively large simultaneously, trustworthy behavior is about 10% more likely. The strongest result is where we compare different kinds of Trust Games with their control condition. Honoring of trust is on average about 25% more likely in the Trust Game than in its control condition. These results can be interpreted as showing a main effect of intentionality, but not an interaction effect based on bookkeeping. What is important as a trustor is that you convey intentionality (as opposed to not conveying intentionality at all). The degree to which you show intentionality, as operationalized by payoff differences, is of minor importance. This result can be conceived as rendering support for the notion that interdependencies lead to trust and commitment, as argued by Lawler and Yoon (1993). Likewise, it provides some empirical support for Frank’s (1988) idea that signaling cooperative intent triggers cooperative behavior.

## CONCLUSION AND DISCUSSION

Some people are more likely to trust and to honor trust. The strongest evidence we found pointed to people who score high on the social orientation scale, and to those who donate to charity. In general, this supports the view that people with a more altruistic nature tend to trust and honor trust more often. The effect size for these variables is about 10 percentage points. Surprisingly, there was no evidence for trust differences with respect to sex, age, education, training in economics, and many other variables that one could expect to have an influence. As an interesting aside, especially the lack of a difference with regard to age and education may serve as an argument against the objections raised against the frequent use of a student population for experiments.

There are also trust differences connected to the appearance of the one to be trusted. It is possible to pick out photographs of people who, according to the majority of

the population, appear less trustworthy than others. When we tried to extrapolate these findings to interactions that resemble real life more adequately, the results were weaker. There is certainly a substantial correlation between what people say about the appearance of another player and how they themselves behave in a Trust Game with that player, thus supporting the view that appearance matters (with an effect size in the same order of magnitude as the effects of disposition: about 10 to 15 percentage points). However, whereas participants did agree about pre-selected photos, they did not agree about real life others at all. Moreover, their assessments of the other person's trustworthiness had no relation to the actual behavior of the other person in a Trust Game.

The most prominent effects on trust were found when we analyzed situational characteristics: the payoffs and, to a lesser extent, the structure of the game. Trustfulness correlates highly with the risk involved ( $(P-S)/(R-S)$  in the Trust Game) for the trustor and trustworthiness correlates even higher with the monetary temptation ( $(T-R)/(T-S)$  in the Trust Game). Effect sizes are large: about an 80 percentage point difference can be expected between situations with the most extreme risk and temptation values. The structure of the game also shows a relatively large effect. The plain fact that trust can be honored only when it is given, induces trustworthy behavior by the trustee (an effect size of about 25 percentage points). It appears that subjects reciprocate the intentionality that is displayed by the trustor. The strength of the intentions, however, is unrelated to differences in payoffs.

Though situational characteristics seem to exert the strongest influences on trust, it is worthwhile to note explicitly that this situationalist view need not be at odds with possible mechanisms on the dispositional level guiding these situational differences with regard to trust. That is, although we find strong effects of monetary payoffs, the differences we documented may still be caused by people's dispositions. For instance, people may honor trust more easily because of a feeling of guilt. In that sense, YY situational factors (but not by differences in people's dispositions). Given a certain distribution of the degree to which people are sensitive to feelings of guilt when they abuse trust, one can find differences in the percentage of trustworthy behavior between Trust Games. For games with a larger value of temptation, a smaller proportion of the people will experience feelings of guilt that are sufficiently strong to withstand the temptation, and the percentage of trustworthy behavior therefore decreases. Even though one finds a difference in temptation to be accountable for this decrease, the effect of the difference in trust then runs through differences with regard to, for instance, feelings of guilt.

It should be emphasized that the comparison between the effects associated with the three proposed categories is far from being completed. Effects of kinship, homophily, in-group favoritism, reputation, and culture are likely or sometimes even shown to exist, but were not part of the present study.

A most important omission of the research reported here is the role of time, an aspect that has been neglected in much of the research on trust. We considered single shot Trust Games between strangers, whereas one could rightfully argue that trust is typically something that grows slowly over time. In fact, one of the more puzzling issues is why it grows slowly, but seems to decrease rapidly once abused (Dasgupta, 1988; Gauthschi, 2000). Our single shot games are not suited to deal with such issues, though it seems logical that setups using repeated Trust Games could shed some light on the matter.

There is even a more compelling question at the heart of our approach, namely how well the intricate topic of trust is captured by the abstract games we employ to measure it. Some critical comments are in order.

Perhaps the most central issue is whether trust behavior can be conceived (and studied) in an exclusively rational framework. Does the construct of trust necessarily require consistency, given the structure of the situation and the preferences of the actors? We know that predictions based on a narrow economic interpretation of game theory (nobody trusts and nobody honors trust) are plainly wrong. In response, one may claim the need for more sophisticated models, for instance models that would more accurately capture actors' preferences or would better represent the affective aspects associated with trust. Alternatively, one may argue that an exclusive rational model cannot (in principle) capture all the subtleties underlying trust considerations on the simple grounds that people are not always rational, especially when trust is concerned. The Trust Game undoubtedly captures some primary aspects of trust yet it may not lend itself to some ("non-rational") particulars associated with messy real life situations. In addition, the specific structure of the trust game and the manner in which it is presented in controlled experiments may implicitly guide participants to behave consistently. The situation presented to participants is well structured, creating a "conversational logic" (Schwarz, 1996) that triggers neat and orderly behavior from subjects. Asking participants to play several Trust Games could be interpreted as what is referred to as 'demand characteristics'. Specifically, participants might conclude (consciously or unconsciously) that the experimenter's aim is to test whether their behavior is consistent across the games. By responding to such demands, participants may exhibit behavior that is consistent across payoffs, whereas in real-life this consistency would have been absent.

A related worry is that we cannot exclude that real-life trust is to a large extent based on "minor considerations" typically unsuited for treatment in systematic research: the rise of a brow, the bad taste of your coffee, or the weather being sunny. Statistical models can predict behavior in a single shot Trust Game reasonably well, not perfectly (Snijders & Keren, 1999). The unexplained part is still large, but what is worse is that this unexplained part might be, for the larger part, random noise. What does it mean when we find out that we will never explain more than 5% of the variance?

We feel we have only scratched the surface of the intricate subject of trust. Our story in the introduction about the mechanic selling the car paints a picture ("Can he be trusted? Should I do it? What is he saying between the lines?") that can only be partially captured in the abstract world of Trust Games, and even in that abstract world we only looked at part of the main effects. Still, were we forced to take a stand on the issue raised in the title of our contribution, then, based on the empirical evidence available to us at this moment, we would claim that the question is not so much who will trust, and also not whom will be trusted, but whether the *situation* induces trustful and trustworthy behavior.

## REFERENCES

- Adams, G.R. (1977) Physical attractiveness research: Toward a developmental social psychology of beauty. *Human development*, 20, 4: 217-239.
- Agnew, C.R. and Thompson, V.D. (1994) Causal inferences and responsibility attributions concerning an HIV-positive target: the double-edged sword of physical attractiveness. *Journal of social behavior and personality*, 9, 1: 181-190.
- Aranoff, D. and Tedeschi, J. (1968) Original stakes and behavior in the prisoner's dilemma game. *Psychonomic science*, 12: 79-80.
- Aronson, E. and Worchel, S. (1966) Similarity versus liking as determinants of interpersonal attractiveness. *Psychonomic science*, 5: 157-158.
- Axelrod, R. (1984) *The evolution of cooperation*. New York: Basic Books.
- Bem, D.J. and Allen, A. On predicting some of the people some of the time: the search for cross-situational consistencies in behavior. *Psychological review*, 81: 506-520.
- Berry, D.S. and McArthur, L.Z. (1985) Some components and consequences of a babyface. *Journal of personality and social psychology*, 48: 312-323.
- Brownlow, S. (1992) Seeing is believing: Facial appearance, credibility, and attitude change. *Journal of nonverbal behavior*, 16, 2:101-115.
- Bruins, J. and Weesie, J. (1996) HIN95Exp. A set of experiments conducted in connection with HIN95. Codebook of HIN95Exp. ISCORE paper, no. 77. Utrecht University.
- Byrne, D. (1971) *The attraction paradigm*. New York: Academic Press.
- Campbell, K.E., Olson, K.R., and Kleim, D.M. (1990) Physical attractiveness, locus of control, sex role, and conversational assertiveness. *Journal of social psychology*, 130, 2: 263-265.
- Carter, J. and Irons, M. (1991) Are economists different, and if so, why? *Journal of economic perspectives*, 5, 2: 171-177.
- Clifford, M.M. and Walster, E. (1973) The effect of physical attractiveness on teacher expectations. *Sociology of education* 46, 2: 248-258.
- Condon, J.W. and Crano, W.D. (1988) Inferred evaluation and the relation between attitude similarity and interpersonal attraction. *Journal of personality and social psychology*, 58: 644-663.

Cronbach, L.J. (1955) Processes affecting scores on “understanding of others” and “assumed similarity.” *Psychological bulletin*, 52: 177-193.

Cronbach, L.J. (1958) Proposals leading to analytic treatment of social perception scores. Pages 353-379 in Tagiuri, R. and Petrullo, L. (Eds.), *Person perception and interpersonal behavior*. Stanford, CA: Stanford University Press.

Dasgupta, P. (1988) Trust as a commodity. Pages 49-72 in D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations*. Oxford: Basil Blackwell.

Dawes, D. McTavish, J., and Shaklee, H. (1977) Behavior, communication, and assumptions about other people’s behavior in a commons dilemma situation. *Journal of personality and social psychology*, 35: 1-11.

Dufwenberg, M. and Gneezy, U. (1996) Efficiency and expectations in an experimental game. Working paper, CENTER for Economic Research, Tilburg University, The Netherlands.

Eagly, A.H., Ashmore, R.D., Makhijani, M.G., and Longo, L.C. (1991) What is beautiful is good, but ...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological bulletin*, 110: 109-128.

Frank, R. H., (1988) *Passions within reason. The strategic role of the emotions*. New York: Norton.

Frank, R.H., Gilovich, and Regan, D.T. (1993) Does studying economics inhibit cooperation? *Journal of economic perspectives*, 7, 2:159-171.

Gambetta, D. (1988) Can we trust trust? Pages 213-237 in Gambetta, D. (Ed.), *Trust: Making and braking cooperative relations*. Oxford: Basil Blackwell.

Gautschi, T. (2000) History effects in social dilemma situations. *Rationality and society*, 12: 131-162.

Harrel, W.A. (1978) Physical attractiveness, self-disclosure, and helping behavior. *Journal of social psychology*, 104, 1: 15-17.

Harris, R.J. (1969) A geometrical classification for 2x2 interval-symmetric games. *Behavioral Science*, 14: 138-146.

Heretick, D.M.L. (1984) Trust-Suspicion and gender differences in interpersonal functioning. *Journal of research in personality*, 18: 27-40.

Jones, E.E. (1990) *Interpersonal perception*. New York: W.H. Freeman and Company.

Kagel, J.H. and Roth, A.E. (Eds.) (1995) *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Kalmijn, M. (1996) Kalmijn, M., Bernasco, W., and Weesie, J. Households in the Netherlands 1995. Codebook of HIN95. ISCORE paper no. 67. Utrecht University.

Kelly, H.H. and Thibaut, J.W. (1978) Interpersonal relations: a theory of interdependence. New York: Wiley.

Kenny, D.A. (1994) Interpersonal perception. A social relations analysis. New York: The Guilford Press.

Keren, G. and Snijders, C. (1999) Trust and reciprocity. ISCORE paper no. 168. Utrecht University.

Kreps, D.M. (1990) Corporate culture and economic theory. Pages 90-143 in J.E. Alt and K.A. Shepsle (Eds.), Perspectives on positive political economy. Cambridge: Cambridge University Press.

Lagace, R.R. and Gassenheimer, J.B. (1989) A measure of global trust and suspicion: replication. Psychological reports, 65: 473-474.

van Lange, P.A.M. and Kuhlman, D.M. (1994) Social value orientations and impressions of partner's honesty and intelligence: a test of the might versus morality effect. Journal of personality and social psychology, 67: 126-141.

Larzelere, R.L. (1984) Dyadic trust and generalized trust of secular versus christian college students. Journal of psychology and theology, 12, 2: 119-124.

Larzelere, R.L. and Huston, T.L. (1980) The dyadic trust scale: toward understanding interpersonal trust in close relationships. Journal of marriage and the family, 595-604.

Lawler, E.J. and Yoon, J. (1993) Power and the emergence of commitment behavior in negotiated exchange. American Sociological Review, 58: 465-481.

Ledyard, J.O. (1995) Public goods: a survey of experimental research. Chapter 2 in Kagel, J.H. and Roth, A.E. (Eds.) The handbook of experimental economics. Princeton, NJ: Princeton University Press.

Lennon, S.J. (1986) Additivity of clothing clues in first impressions. Social behavior and personality, 14, 1: 15-21.

Levesque, M.J. and Kenny, D.A. (1993) Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. Journal of personality and social psychology, 65: 1178-1187.

Liebrand, W.B.G. (1984) The effect of social motives, communication and group size on behavior in an n-person multi-stage mixed-motive game. *European journal of social psychology*, 14: 239-264.

Liebrand, W.B.G. and McClintock, C.G. (1988) The Ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European journal of personality*, 2: 217-230.

Liebrand, W.B.G., Jansen, R.W.T.L., Rijken, V.M., and Suhre, C.J.M. (1986) Might over morality: social values and the perception of other players in experimental games. *Journal of experimental social psychology*, 22: 203-215.

Marwell, G. and Ames, R.E. (1981) Experiments on the provision of public goods. II. Provision points, stakes, experience, and the free-rider problem. *American Journal of Sociology*, 85, 4: 926-937.

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

McCabe, K.A., Rassenti, S.J. and Smith, V. (1996) Game theory and reciprocity in some extensive form experimental games. *Proceedings of the National Academy of Sciences U.S.A. (Economic Sciences)*, 93, 13421-13428.

Mischel, W. *Personality and assessment*. New York: Wiley.

Misra, S. and Kalro, A. (1979) Triangle effect and the connotative meaning of trust in prisoner's dilemma: a cross cultural study. *International journal of psychology*, 14, 4: 255-263.

Misztal, B.A. (1996) *Trust in modern societies. The search for the bases of social order*. Oxford: Basil Blackwell.

Morrow, P.C., McElroy, J.C., Stamper, B.G., and Wilson, M.A. (1990) The effects of physical attractiveness and other demographic characteristics on promotion decisions. *Journal of management*, 16, 4: 723-736.

Mulford, M., Orbell, J. et al. (1998) Physical attractiveness, opportunity, and success in everyday exchange. *American journal of sociology* 103, 6: 1565-1592.

Newcomb, T.M. (1961) *The acquaintance process*. New York: Holt, Rinehart, and Wilson.

Nisbett, R. and Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Orbell, J., Dawes, R.M., and Schwartz-Shea, P. (1994) Trust, social categories, and individuals; the case of gender. *Motivation and emotion* (special issue: Trust and distrust: psychological and social dimensions), 18, 2: 109-128.

Parks, C.T., Henager, R.E., and Scamahorn, S.D. (1996) Trust and reactions to messages of intent in social dilemmas. *Journal of Conflict Resolution*, 40, 1: 134-151.

Rapoport, A. and Chammah, A.M. (1965) *Prisoner's Dilemma*. Ann Arbor: University of Michigan Press.

Rapoport, A., Guyer, M.J., and Gordon, D.G. (1976) *The 2x2 game*. Ann Arbor: University of Michigan Press.

Ritts, V. Patterson, M.L., and Tubbs, M.E. (1992) Expectations, Impressions, and judgments of physically attractive students: a review. *Review of educational research*, 62, 4: 413-426.

Rosenbaum, J.L. and Chesney-Lind, M. (1994) Appearance and delinquency: A research note. *Crime and delinquency*, 40, 2: 250-261.

Ross, L. (1977) The intuitive psychologist and his shortcomings. Pages 174-221 in Berkowitz, L. (Ed.) *Advances in experimental social psychology* (Vol. 10). New York: Academic Press.

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

Rousseau, D.M., Sitkin, S.B., Burt, R.S., and Camerer, C. (1998) Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23 (3): 393-404.

Sally, D. (1995) Conversation and cooperation in social dilemmas. *Rationality and society*, 7, 1:58-92.

Schwarz, N. (1996) *Cognition and communication. Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Snijders, C. (1996) *Trust and commitments*. Amsterdam: Thela Thesis.

Snijders, C. (1998) *Inter-national trust and organizational cooperation*. ISCORE paper, no. 121, Utrecht University.

Snijders, C. and Keren, G. (1995) Trust is in the eye of the beholder: The effect of physical appearance on trust. ISCORE paper, no. 32. Utrecht University.

Snijders, C. and Keren, G. (1999) Determinants of Trust. In Budescu, D., Erev, I., and Zwick, R. (Eds.) Games and human behavior. Mahwah, NJ: Lawrence Erlbaum Associates.

Steel, J.L. (1991) Interpersonal correlates of trust and self-disclosure. Psychological reports, 68, 3, 2:1319-1320.

Steele, M.W. and Tedeschi, J.T. (1967) Matrix indices and strategy choices in mixed-motive games. Journal of experimental psychology: general, 115:131-154.

Switkin, L.R. and Gynther, M.D. (1974) Trust, activism, and interpersonal perception in black and white college students. Journal of social psychology, 94, 1: 153-154.

Terrel, F. and Barret, R.K. (1979) Interpersonal trust among college students as a function of race, sex, and socioeconomic class. Perceptual and motor skills, 48: 1194.

Thye, S. (2000) Reliability in experimental sociology. Social Forces.

Umberson, D. and Hughes, M. (1987) The impact of physical attractiveness on achievement and psychological well-being. Social psychology quarterly, 50, 3: 227-236.

Uslaner, E.M. (in press) The moral foundations of trust. Draft downloadable from <http://www.bsos.umd.edu/gvpt/uslaner>.

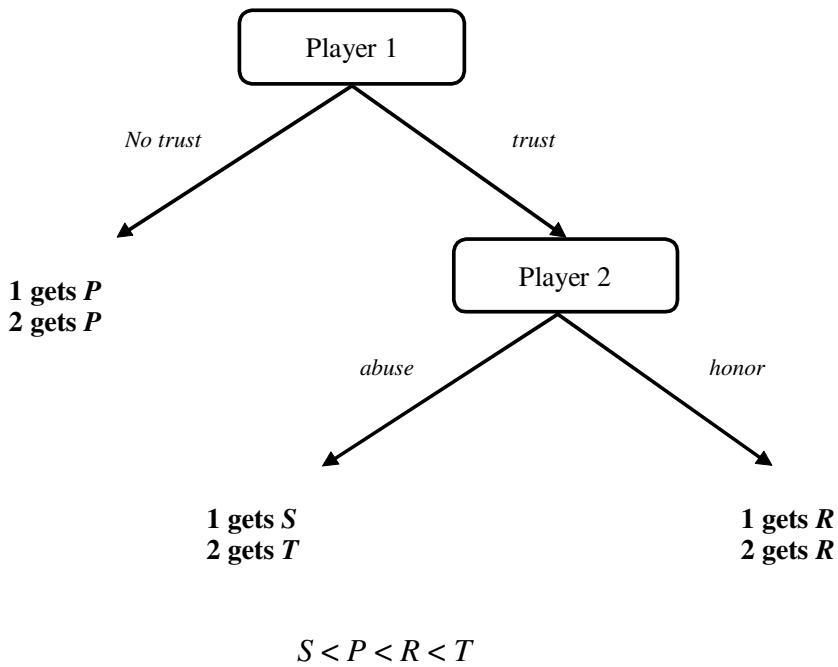
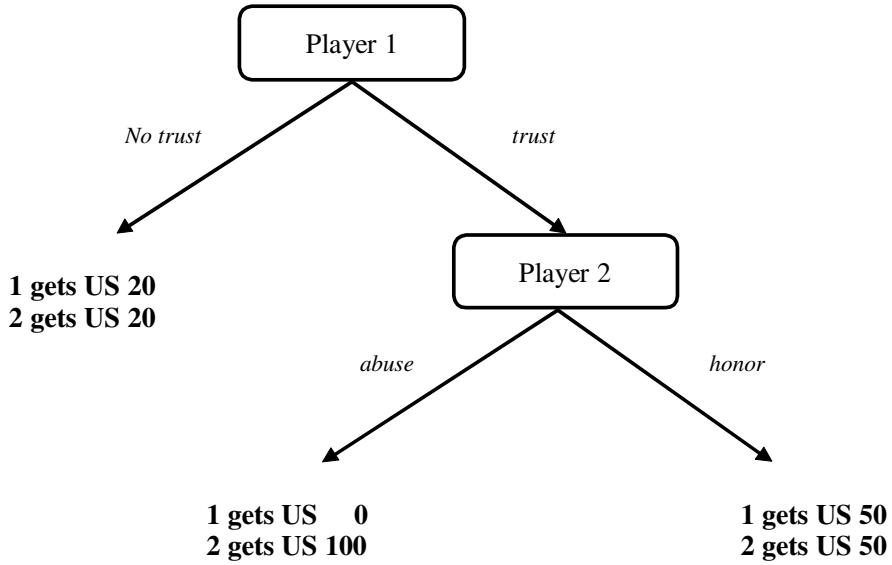
Wheless, L.R. and Grotz, J. (1977) The measurement of trust and its relationship to disclosure. Human communication research, 3, 3: 250-257.

Wright, T.L. and Sharp, E.G. (1979) Content and grammatical sex bias on the interpersonal trust scale and differential trust toward women and men. Journal of consulting and clinical psychology, 47, 1: 72-85.

Yamagishi, T. (1986) The provision of a sanctioning system as a public good. Journal of Personality and Social Psychology, 51, 1: 110-116.

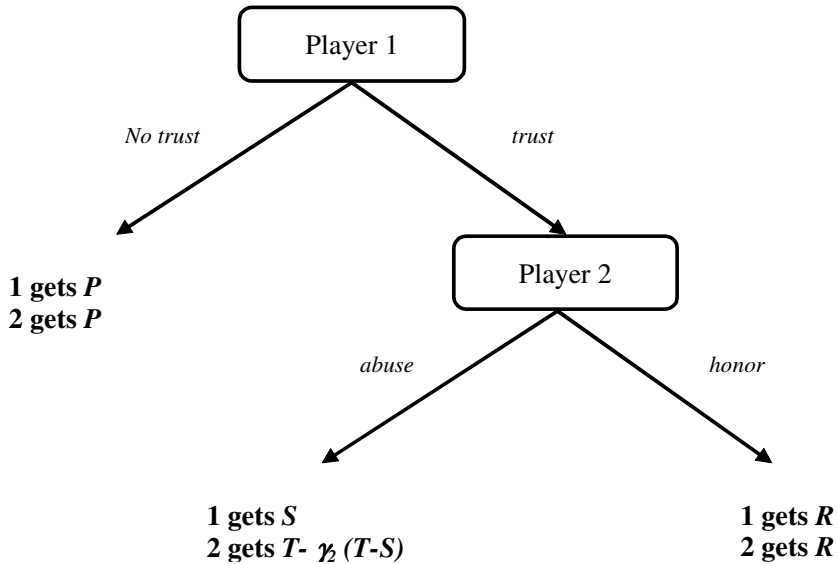
Yamagishi, T. and Yamagishi, M. (1994) Trust and commitment in the United States and Japan. Motivation and emotion., 18, 2: 129-166.

**FIGURE 1: TRUST GAME (EXAMPLE AND GENERAL FORMAT)**

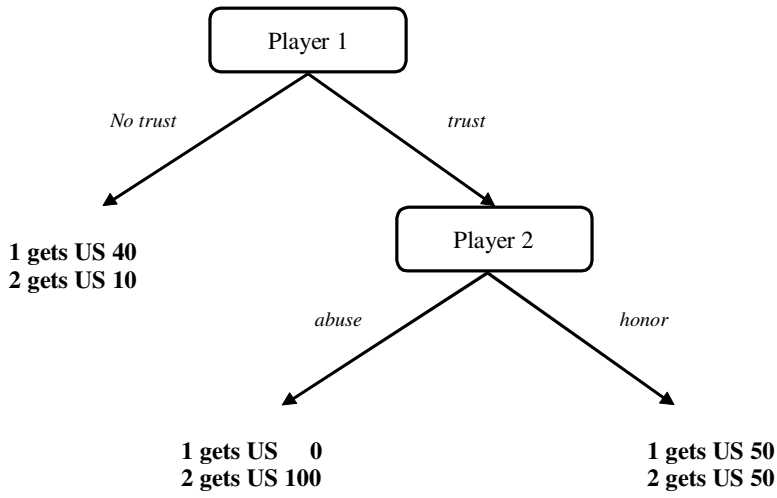
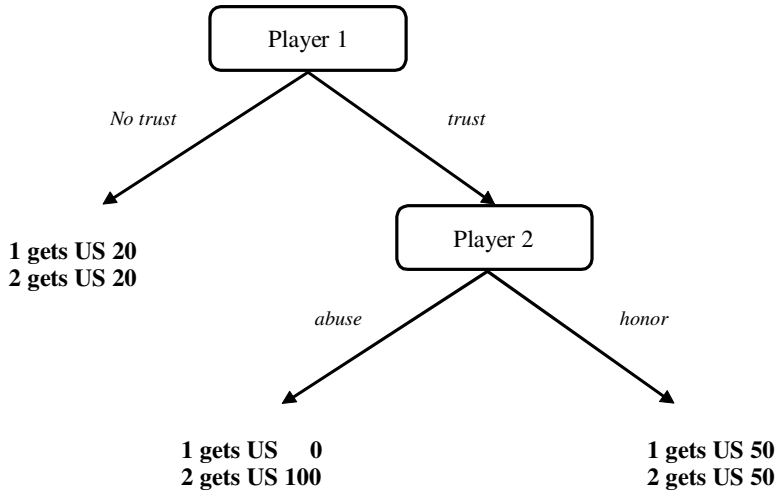


**FIGURE 2: AN IMPLEMENTATION OF THE “EFFECTIVE FORM” OF THE TRUST GAME WHEN INCLUDING GUILT**

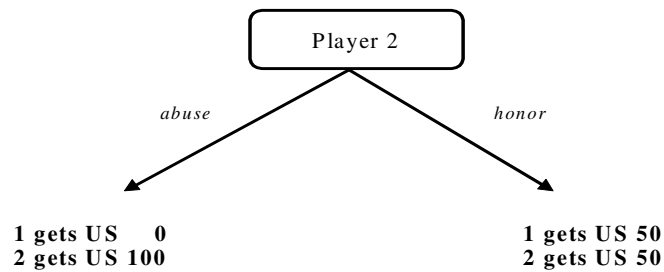
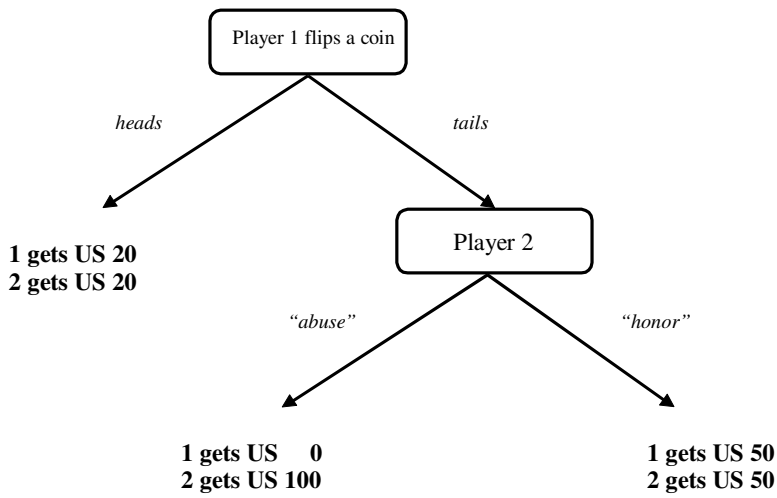
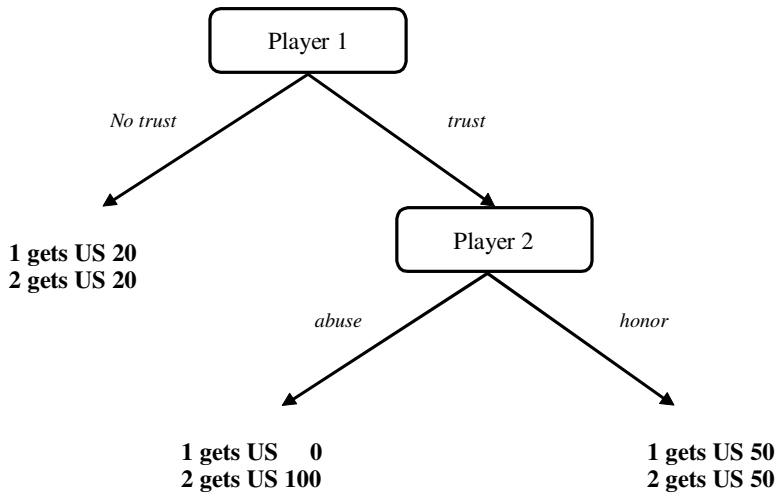
Player 2 is assumed to experience a loss of utility when abusing trust. This loss is larger depending on the difference between player 2 and player 1’s monetary payoff ( $T-S$ ). It also depends on the person ( $\gamma_2$ ). As before, the assumption is that  $S < P < R < T$ . It is straightforward to show that player 2 honors trust when  $\gamma_2 > (T-R)/(T-S)$ . Player 1 trusts player 2 if player 1 assesses that the probability that player 2 will honor trust is larger than  $(P-S)/(R-S)$ .



**FIGURE 3: TRUST GAMES WITH DIFFERENT  $(P,P)$  OUTCOMES: TESTING THE BOOKKEEPING HYPOTHESIS**



**FIGURE 4: TRUST GAME, COIN-FLIP GAME, AND CONTROL CONDITION. TESTING THE INTENTIONALITY HYPOTHESIS**



**Table 1: Net effects of disposition on (the probability of) trustfulness and trustworthiness.**

	Hypothesis	Trustfulness	Trustworthiness
<i>General</i>			
Male	?	+0.07	-0.09
Age	-	0	0
<i>Indicators of pro-social behavior</i>			
Donor Codicil	+	+0.10	0
Blood donor	+	0	0
<i>Might versus morality</i>			
Being a student of economics	-	0	0
Knowledge about game theory	-	0	0
Beta-graduation (math etc)	-	0	0
<i>Ring measure of social orientation</i>			
Pro-social (reference: others)	+	+0.20	+0.19

**Table 2: Net effects of socio-economic characteristics on (the probability of) trustfulness.**

	Trustfulness
<i>General characteristics</i>	
Male	0
Education	0
Income	0
Age	0
<i>View on life</i>	
Job is related to economics	0
Education related to economics	0
Religion	
- Roman Catholics	-0.06
Political pref. (higher= more rightwing)	-0.02
Donates to charity	+0.13
Living in city or in village	0
<i>Control over short-term gratification</i>	
Smokes	0
Drinks	0
<i>Previous breach of trust</i>	
Parents divorced	0
Mother worked when respondent 14	+0.11
Number of previous relationships	0
Number of previous marriages	0

**Table 3: Net effects of perceived trustworthiness on (the probability to) trust, and of perceived trustworthiness on (the probability of) actual trustworthiness.**

	Hypothesis	Trustfulness	Player 2's Trustworthiness
<i>Anticipation</i>			
Player 2 appears trustworthy	+	+0.12	.
Player 2 appears trustworthy	?	.	0

**Table 4: Net effects of monetary payoffs and of payoff indices (risk and temptation) on (the probability to) trust.**

	Hypothesis	Trustfulness	Hypothesis	Trustworthiness
<i>Model 1: Monetary payoffs</i>				
S-payoff	+	+0.01	-	0
P-payoff	-	-0.02	+	0
R-payoff	+	+0.01	+	+0.03
T-payoff	-	0	-	-0.02
<i>Model 2: Indices of monetary payoff</i>				
Risk $((P-S)/(R-S))$		-	0	0
Temptation $((T-R)/(T-S))$		-	-	-0.94

**Table 5: Net effects of testing the bookkeeping and intentionality hypotheses.**

	Hypothesis	Trustworthiness
<i>Bookkeeping hypothesis</i>		
$P_1-S$ payoff difference	+	0
$R_2-P_2$ payoff difference	+	0
$P_1-P_2$ payoff difference	+	0
<i>Intentionality hypothesis</i>		
Game versus control condition	+	+0.25
Active player 1 versus coin-flip	+	+0.18