

NBER WORKING PAPER SERIES

DO TRUST AND TRUSTWORTHINESS PAY OFF?

Joel Slemrod
Peter Katuscak

Working Paper 9200
<http://www.nber.org/papers/w9200>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2002

We would like to thank members of the University of Michigan Public Finance workshop, and particularly John Bound and Matthew Shapiro, for helpful comments on an earlier version of this paper. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2002 by Joel Slemrod and Peter Katuscak. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Do Trust and Trustworthiness Pay Off?
Joel Slemrod and Peter Katuscak
NBER Working Paper No. 9200
September 2002
JEL No. J30, H40

ABSTRACT

Are individuals who trust others better off than those who do not? Do trustworthy people prosper more than untrustworthy ones? We first pose these questions in a search model where individuals face repeated choices between trusting (initiating an investment transaction) and not trusting, and between being trustworthy (not stealing the investment) and cheating. We then derive predictions for the relationship between observed individual behavior, aggregate attitudes, and individual prosperity. Finally, we evaluate these predictions empirically using household-level data for eighteen (mostly developed) countries from the World Values Survey. We find that, on average, a trusting attitude has a positive impact on income, while trustworthiness has a negative impact on income. In addition, we find evidence of complementarity between these two attitudes and the aggregate levels of the complementary attitudes. Most strikingly, the payoff to being trustworthy depends positively on the aggregate amount of trust in a given country.

Joel Slemrod
Department of Economics
University of Michigan
and NBER
E-mail: jslemrod@umich.edu

Peter Katuscak
Department of Economics
University of Michigan
E-mail: pkatus@umich.edu

1 Introduction

The notions of trust and trustworthiness have received much recent attention in social science, stimulated in part by the work of Putnam (1993) and Fukuyama (1995), but with antecedents in, for example, Coleman (1990). Economists have for a long time recognized the critical role played by trust in economic performance. Arrow (1972), for example, remarks: “Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can plausibly be argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence.” In high-trust societies, individuals need to spend less resources to protect themselves from being exploited in economic transactions. Knack and Keefer (1997) argue that trusting societies tend to have stronger incentives to innovate and to accumulate both physical and human capital and, as a result, grow faster. Zak and Knack (2000) corroborate the positive effect of aggregate trust on growth. Alesina and La Ferrara (2000) investigate the individual-level determinants of trust within the United States and find that income and education are strongly positively correlated with trust.

The flip side of trust is trustworthiness. Glaeser et al (2000) distinguish between trusting behavior, which they define as “the commitment of resources to an activity where the outcome depends upon the cooperative behavior of others,” and trustworthy behavior, which “increases the returns to people who trust you.” The idea of reputation—the level of trust one is perceived to merit—has also been examined. As Axelrod (1986) puts it, an individual’s reputation derives from adherence to or vio-

lation of a norm that others view as a signal about the individual's future behavior in a wide variety of situations.

In this paper, we begin the task of linking the microeconomic theory to empirical evidence based on micro data. We start by developing an equilibrium search model in which individuals face repeated choices between trusting and not trusting, and between being trustworthy and cheating. Each person possesses individual-specific intrinsic predispositions to trust and trustworthiness. In addition to these intrinsic preferences, the person is strategic: he considers how his actions may affect his chance of developing and sustaining a current match and forming beneficial matches in the future. In equilibrium, his strategic actions are guided by the equilibrium distribution of his opponents' actions, i.e., by the equilibrium probability that a randomly chosen individual will trust, and will act in a trustworthy manner. Given the individual heterogeneity in the intrinsic predispositions to trust and trustworthiness, equilibrium entails both trusting and mistrusting individuals, and both trustworthy and cheating individuals.

There have been several recent theoretical contributions addressing the issue of trust. Tirole (1996) develops a dynamic model where there may exist a certain level of trust between individuals due to the considerations of individual and collective reputation.¹ Although there is some heterogeneity in the tendency towards trustworthy behavior in his model, he does not consider heterogeneity in the tendency towards trusting behavior. On the other hand, Chen (2000) develops a model in

¹Dixit (2001) studies the role of individual reputation and informational intermediaries in a similar framework.

which individuals differ in their intrinsic preferences for being honest, or trustworthy, as captured by the notion of a population distribution of trustworthiness. But his focus is on the role of trust in contracting, and he takes a reduced-form approach where dynamic effects are not explicitly considered.

The present model overcomes these limitations by integrating individual heterogeneity in the behavioral predisposition toward both trusting and trustworthy behavior with the dynamic considerations. In contrast to Tirole (1996), the introduction of heterogeneity in the behavioral predisposition for both trusting and trustworthy behavior leads to the joint determination of the behavioral predisposition cutoffs separating trusting and trustworthy behavior in equilibrium. One implication of this enrichment of the model is that a change in the distribution of either of the two predispositions affects the extent of both trusting and trustworthy behavior. In addition, the resulting two-sided behavioral heterogeneity allows us to compare the equilibrium monetary payoffs associated with acting in a trusting versus untrusting manner as well as acting in a trustworthy manner versus untrustworthy manner, which is the central question addressed in this paper.

Based on our theoretical framework, we estimate a model of the private return to trust and trustworthiness, using data from eighteen countries from the 1990 World Values Survey. We find evidence that the return to trustworthiness is negative on average and depends (in a statistically significant way) on the average amount of trust in the society. In particular, this return is negative in low-trust countries and positive in high-trust countries. We also find that the return to trust is positive on average and some of our results suggest that it is related in a positive way to the

average amount of trustworthiness in the society. However, this relationship appears to be statistically less robust than the previous one, although the sign pattern is consistent throughout various specifications. Strikingly, these results suggest the possibility that a country might be in an equilibrium trap where it is not in most people's interest to invest in either trust or trustworthiness.

The paper is structured in the following way: Section 2 develops the theoretical model. Section 3 discusses comparative statics results and empirical predictions of the model. Section 4 reviews previous empirical work. Section 5 describes the dataset we use. Section 6 contains our empirical results. Section 7 concludes. Technical proofs can be found in the Appendix.

2 Model

2.1 Setup

There is a continuum of individuals with a total measure normalized to 1. The output in the economy is created from business transactions. Each transaction has two parties to it: an initiator and a respondent. Each individual simultaneously participates in both roles. An initiator initiates a transaction by 'investing' 1 unit of a generic good, and a respondent, who if responding honestly, contributes to a successful completion of the transaction. In such case the total payoff from the transaction is $2a + 1$ and the net output of $2a$ is shared equally by the two parties, giving a net payoff a to each party. However, the respondent may also respond

dishonestly by 'stealing' the investment. In such case the net payoff to the initiator is -1 and the net payoff to the respondent is $1 - d$, where d measures inherent disutility from being dishonest, or intrinsic honesty. The value of d is individual-specific with support $[\underline{d}, \bar{d}]$ and with a continuous distribution function F that is strictly increasing on $[\underline{d}, \bar{d}]$. To make the dishonest response potentially attractive to at least some respondents, we assume that $a < 1 - \underline{d}$. In light of this possibility, the initiator may decide not to initiate a transaction in the first place. In such case the net payoff to the initiator is $-m$ and the net payoff to the respondent is 0 , where m is an individual-specific inherent propensity to trust, captured by the disutility m of mistrust. The value of m has support $[\underline{m}, \bar{m}]$ and a continuous distribution function G that is strictly increasing on $[\underline{m}, \bar{m}]$.²³ If a transaction is not initiated or if an initiated transaction is met with a dishonest response, there is no net output produced (the theft is just a transfer). The extensive form of the transaction game is pictured in Figure 1.

This setup tries to capture the distinction between trust and trustworthiness and the impact both of these have on individual prosperity. Successful completion of a transaction requires both a trusting approach of the initiator and a trustworthy approach of the respondent.⁴ In case either of them is missing, the transaction fails and no net output is produced (although some existing wealth might be redistributed).

Each period a subgroup of initiators interacts with a subgroup of respondents by

²The assumptions of F and G being strictly increasing on the two supports are only made for presentational convenience. None of the results is affected by dropping this assumption.

³Note that we do not assume that d and m are distributed independently across individuals. Indeed, they may be correlated. Whether they are correlated or not, however, is immaterial to the subsequent analysis since each individual acts independently in his initiator and respondent roles.

⁴Other interpretations of the transaction are possible, such as customer-firm.

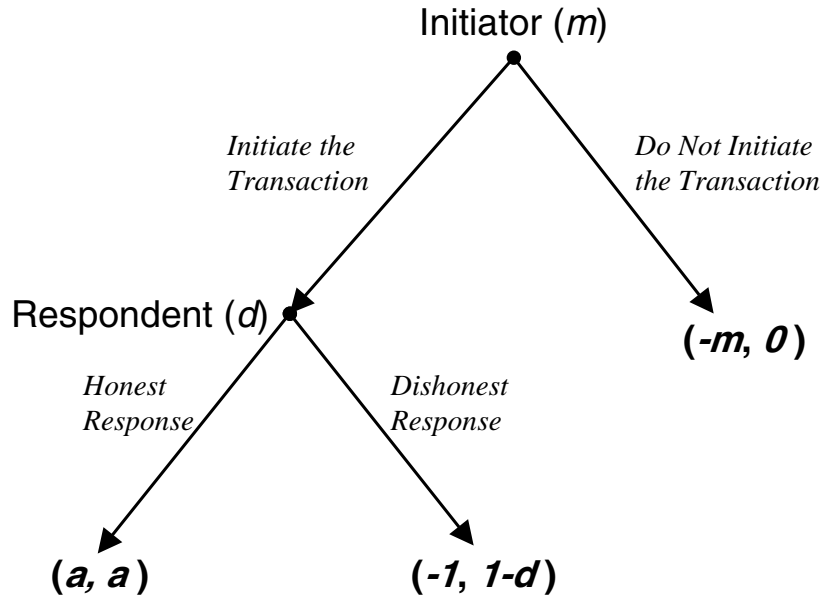


FIGURE 1: Extensive form of the transaction game

participating in an initiator-respondent match. Even though each individual has a dual role in each period, acting both as an initiator in one match and a respondent in another, it is helpful to separate these two roles and to think of the initiators and the respondents as two separate groups of the same size.⁵ At the beginning of each period there are equally sized groups of matched initiators and matched respondents and equally sized groups of unmatched initiators and unmatched respondents. Those matched participate in their "surviving" matches from the previous period. Each unmatched initiator gets matched with probability $\beta \in (0, 1)$ to some unmatched respondent and vice versa. Then, by the law of large numbers, β is also the fraction of both the searching initiators and the searching respondents who get matched in a new match within a period. If an initiator or a respondent is unmatched,

⁵This simplification is valid since the two matches of an individual (one in the initiator role and one in the respondent role) are generically different (their coincidence is a zero probability event).

his or her payoff for the current period is 0. If an initiator and a respondent are matched (in a new or a surviving match), they play the transaction game outlined above and collect their payoffs. If the transaction is completed successfully (i.e. it is initiated and responded to honestly), the match survives to the next period with probability $\alpha \in (0, 1)$. Otherwise it is dissolved and both participants will enter the next period as unmatched. The latter is also the case if the transaction is completed successfully but, conditional on that, the match does not survive until the next period for exogenous reasons, which happens with probability $1 - \alpha$. In turn, α is then also the fraction of matches with successfully completed transactions that actually survive to the next period. Intuitively, even if the match is "working", exogenous events such as population mobility or business turnover may cause the match to break up. All individuals are risk neutral and have a discount factor $\delta \in (0, 1)$. We assume that $\alpha > \beta$, i.e., that a working match is more likely to survive than a search is to result in a new match.

To make the analysis tractable, we restrict our attention to steady states and make the following additional assumptions.

Assumption 1: All the disclosed or inferred information about a partner in a transaction is specific to and lasts only during a given repeated match (i.e., there is no social learning and no memory).

Assumption 2: Both the initiators and the respondents condition their strategies only on what happened within the current period, on whether the present match is new or surviving and on the aggregate steady state probability of an initiator ini-

tiating (p) and a respondent responding honestly (q) *in a new match*.⁶

Assumption 3: If an initiator or a respondent is indifferent between two actions, she or he chooses to initiate or respond honestly, respectively.

In the next subsection, p and q are taken as given and the implied behavior of individual initiators and respondents is derived. The following subsection then aggregates these individual decisions to determine p and q endogenously.

2.2 Partial Equilibrium

At the beginning of a period, after both the initiators and the respondents have realized whether they are matched or not and whether the match is new or surviving, an initiator (of type) m may find herself in the following states with their associated discounted payoff values (or value functions):

I1: Not matched: $I_1(m, q)$

I2: Matched in a new match: $I_2(m, q)$

I3: Matched in a surviving match: $I_3(m, q)$

The value of q enters into the decisionmaking and value functions of initiators

⁶This Markovian strategy assumption is made in order to simplify the analysis. Given that a necessary condition for a match survival is a successful completion of the transaction in each period while the match lasts, a general strategy space would allow strategies to condition on the age of the match, since that is the only variable that may differ from one surviving match to another. Indeed, one could envision an equilibrium in which, conditional on match survival, initiators of type m initiate until period $x(m)$ and respondents of type d respond honestly until period $y(d)$, where $x(\cdot)$ and $y(\cdot)$ are (weakly) increasing and potentially infinitely valued. In such an equilibrium, given the age of a particular match, optimal initiator and respondent decisions would be determined by the intrinsic behavioral propensities and the updated distributions of match partner types (where the support of the latter only includes opponent types whose strategies prescribe cooperation until at least the realized age of the match). Intuitively, we focus on equilibria where $x(\cdot)$ and $y(\cdot)$ only assume values of zero or infinity. Although the general class of strategies and equilibria may be interesting from a purely theoretical standpoint, we believe that the subclass we focus on sufficiently captures the essentials of trust and trustworthiness in an equilibrium setting.

because it captures how likely one is to encounter an honest response in a new match.⁷

Based on the realization of whether he is matched or not, whether the match is new or surviving and whether the initiator has initiated a transaction or not, a respondent (of type) d may find himself in the following states with their associated discounted payoff values (or value functions):

R1: Not matched: $R_1(d, p)$

R2: Matched in a new match, but without an initiated transaction: $R_2(d, p)$

R3: Matched in a new match with an initiated transaction: $R_3(d, p)$

R4: Matched in a surviving match, but without an initiated transaction: $R_4(d, p)$

R5: Matched in a surviving match with an initiated transaction: $R_5(d, p)$

The value of p enters into the decisionmaking and value functions of initiators because it captures how likely one is to encounter a trusting initiator in a new match.⁸

First, consider the decisionmaking of a respondent d given p (for simplicity of notation, we omit p from the list of arguments in the value functions). In states R1, R2 or R4, there is no current decision to be taken and the respondent simply collects a payoff 0 in the current period. Then he goes searching at the beginning of the next period, since any existing match is dissolved at the end of the current

⁷The value of p does not enter into the initiators' decisionmaking because it only matters to them to the extent it affects q . Hence q is a sufficient statistic from the point of view of the initiators.

⁸The value of q does not enter into the respondents' decisionmaking because it only matters to them to the extent it affects p . Hence p is a sufficient statistic from the point of view of the respondents.

period. This implies that

$$R_1(d) = R_2(d) = R_4(d) \quad (1)$$

In addition to getting the payoff 0 in the current period, in the next period the respondent gets matched in a new match with probability β and, conditional on the latter, he will face an initiated transaction with probability p . Therefore the Bellman equation for $R_1(d)$ is

$$R_1(d) = 0 + \delta [(1 - \beta)R_1(d) + \beta(1 - p)R_2(d) + \beta p R_3(d)] \quad (2)$$

Using (1), this gives

$$R_1(d) = R_2(d) = R_4(d) = \frac{\beta \delta p}{1 - \delta + \beta \delta p} R_3(d) \quad (3)$$

If in state R3 or R5, the respondent must decide whether to respond to the initiated transaction honestly or dishonestly. If responding dishonestly, he collects $1 - d$ in the current period and goes searching at the beginning of the next period. This is quantitatively equivalent to collecting $1 - d$ currently and being in state R1 currently. If responding honestly, with probability $1 - \alpha$ the respondent collects a in the current period and the match dissolves, in which case the respondent goes searching at the beginning of the next period. This is quantitatively equivalent to collecting a currently and being in state R1 currently. With probability α the respondent collects a in the current period and the match survives.

To be able to judge the value conditional on the survival of the match, the

respondent has to form a belief about the initiator initiating in the next period in such scenario. This poses no complication if the current state is R5 because of the conditioning states used by the initiator in formulating her strategies (Assumption 2). In particular, the initiator will initiate again in the next period because she has done so in the current period. In other words, the fact that the initiator initiated in the current period perfectly reveals what action her strategy prescribes for surviving matches. Hence, in state R5 and conditional on the survival of the match, the respondent will again find himself in state R5 in the next period. Therefore the Bellman equation for $R_5(d)$ is

$$R_5(d) = \max \{1 - d + R_1(d); (1 - \alpha) [a + R_1(d)] + \alpha [a + \delta R_5(d)]\} \quad (4)$$

The first term in the maximand corresponds to the value of responding dishonestly while the second term corresponds to the value of responding honestly.

Analyzing (4) gives the following result:

Lemma 1 *All the respondents (irrespective of d) respond honestly in state R5.*

Intuitively, if in state R5, the respondent must have chosen to respond honestly in the previous period (in state R3) even in the presence of uncertainty about whether the initiator would or would not initiate in the current period. It then follows that the respondent will also opt to respond honestly once it is certain that the initiator will initiate in the next period.

This lemma and (4) then imply that

$$R_5(d) = (1 - \alpha) [a + R_1(d)] + \alpha [a + \delta R_5(d)]$$

which gives, by using (3),

$$R_5(d) = \frac{a}{1 - \alpha\delta} + \frac{(1 - \alpha)\beta\delta p}{(1 - \alpha\delta)(1 - \delta + \beta\delta p)} R_3(d) \quad (5)$$

If in state R3, however, it is not obvious from the fact that the initiator has initiated in the current period that she will do so also in the next period (if the match survives until then) because the initiator's strategy may prescribe different actions for states I2 and I3. Hence let $k \in [0, 1]$ be the belief of the respondent that the initiator will initiate in the next period if the match survives till then. Because k is defined conditionally on the initiator initiating and the respondent responding honestly in the current period, its value is, on the same conditioning set, independent of d and m that characterize the two participants to the current match. This is because, in the next period, the initiator will only observe that the match has survived, if it does, not a particular value of d . Similarly, when forming his belief, the respondent only observes that the transaction has been initiated in the current period, not a particular value of m . If the initiator does not initiate, the respondent will get to state R4. In the opposite case the respondent will get to

state R5. Then, in analogy with (4),

$$R_3(d) = \max \{1 - d + R_1(d); (1 - \alpha) [a + R_1(d)] + \alpha [a + (1 - k)\delta R_4(d) + k\delta R_5(d)]\} \quad (6)$$

Again, the first term in the maximand corresponds to the value of responding dishonestly while the second term corresponds to the value of responding honestly.

Before continuing the analysis of respondents' actions in state R3, it is useful to consider the decisionmaking of the initiators. Hence consider the decisionmaking an initiator m given q (for simplicity of notation, omit q from the list of arguments in the value functions). If in state I1, there is no current decision to be taken and the initiator simply collects a payoff 0 in the current period. In the next period she gets matched in a new match, hence getting into state I2, with probability β and does not get matched, hence getting into state I1, with probability $1 - \beta$. Therefore the Bellman equation for $I_1(m)$ is

$$I_1(m) = 0 + \delta [(1 - \beta)I_1(m) + \beta I_2(m)] \quad (7)$$

giving

$$I_1(m) = \frac{\beta\delta}{1 - \delta + \beta\delta} I_2(m) \quad (8)$$

If in states I2 or I3, the initiator must decide whether to initiate a transaction or not. If not initiating, she collects $-m$ in the current period and goes searching at the beginning of the next period. This is quantitatively equivalent to collecting $-m$ currently and being in state 1 currently. If in state I3 and initiating, it follows by

Lemma 1 that the respondent will respond honestly. Then with probability $1 - \alpha$ the initiator collects a in the current period and the match dissolves, in which case the initiator goes searching at the beginning of the next period. This is quantitatively equivalent to collecting a currently and being in state I1 currently. With probability α the initiator collects a in the current period and the match survives, in which case the respondent will be in state I3 in the next period. Therefore the Bellman equation for $I_3(m)$ is

$$I_3(m) = \max \{-m + I_1(m); (1 - \alpha) [a + I_1(m)] + \alpha [a + \delta I_3(m)]\} \quad (9)$$

The first term in the maximand corresponds to the value of not initiating while the second term corresponds to the value of initiating.

If in state I2 and initiating, however, it is not certain that the respondent will respond honestly. In particular, since the matching process is random, the probability of the respondent responding honestly is q . If the respondent responds dishonestly, the initiator will collect -1 in the current period and will go searching at the beginning of the next period. This is quantitatively equivalent to collecting -1 currently and being in state I1 currently. If the respondent responds honestly, reasoning analogous to state I3 applies. Therefore

$$I_2(m) = \max \{-m + I_1(m); (1 - q) [-1 + I_1(m)] + q ((1 - \alpha) [a + I_1(m)] + \alpha [a + \delta I_3(m)])\} \quad (10)$$

Again, the first term in the maximand corresponds to the value of not initiating, while the second term corresponds to the value of initiating.

Lemma 2 *All the initiators (irrespective of m) initiate in state I3.*

Intuitively, if in state I3, the initiator must have chosen to initiate in the previous period (in state I2) even in the presence of uncertainty about whether the respondent would respond honestly or dishonestly. Consequently, the initiator will also initiate once it is certain that the respondent will respond honestly.

This lemma and (9) then imply that

$$I_3(m) = (1 - \alpha) [a + I_1(m)] + \alpha [a + \delta I_3(m)]$$

which gives, by using (8),

$$I_3(m) = \frac{a}{1 - \alpha\delta} + \frac{(1 - \alpha)\beta\delta}{(1 - \alpha\delta)(1 - \delta + \beta\delta)} I_2(m) \quad (11)$$

Note the importance of Lemma 1 and Lemma 2: they imply that a successful completion of the transaction in the initial period of a match is a sufficient signal for both sides for developing trust and trustworthiness between them. This manifests itself in successful completion of the transaction in every subsequent period for as long as the match lasts.⁹

Now we can finalize the partial equilibrium analysis. Using the result of Lemma 2, k , which is the belief of the respondent in state R3 that the initiator will initiate

⁹Once trust and trustworthiness have been established, the expected survival time of a match is $\frac{\alpha}{1-\alpha}$ periods past the current period.

in the next period if the match survives until then, becomes 1 in (6). The latter, combined with (3) and (5), then gives

$$R_3(d) = \max \left\{ \frac{1 - \delta + \beta\delta p}{1 - \delta} (1 - d); \frac{1 - \delta + \beta\delta p}{(1 - \delta)(1 - \alpha\delta)} a - \frac{\alpha\beta\delta p}{1 - \alpha\delta} R_3(d) \right\} \quad (12)$$

where, as before, the first term in the maximand corresponds to the value of responding dishonestly, while the second term corresponds to the value of responding honestly.

Lemma 3 *A respondent d responds honestly in state $R3$ if and only if $d \geq d(p) \equiv 1 - \frac{a}{1 - \alpha\delta + \alpha\beta\delta p}$.*

This result is a straightforward implication of (12) and Assumption 3 and its proof is left to the reader.

Lemma 3 says that, in a new match, respondents with a relatively high level of intrinsic honesty will behave in a trustworthy way and reply honestly, while respondents with a relatively low intrinsic honesty will not behave in a trustworthy way and they will reply dishonestly. This is so because the latter group will find theft attractive because of their low "moral barriers", even though it entails termination of the match. On the other hand, the former group will not find theft attractive either because of their high "moral barriers" or because of reputation reasons, since it is more profitable for them to continue the match, even though they would behave dishonestly in a non-repeated setting. The threshold $d(p)$ is an increasing function of p , the aggregate steady state probability of an initiator initiating in a new match. Note that this implies that trust breeds untrustworthiness. The more trusting the

population, the more attractive acting in an untrustworthy way is. This is because the more likely the initiators are to trust strangers, the easier it is to get into state R3 (new match with an initiated transaction) if unmatched at the beginning of a period, and hence the less costly it is to forego reputation (and hence break up the match) by stealing relative to the gain from stealing, which is unchanged. Consequently, more respondents will choose to respond dishonestly in a new match with an initiated transaction. Note, however, that respondents with $d \geq 1$ never choose to behave dishonestly.

In a similar way, using (8) and (11) for substitution into (10) yields

$$I_2(m) = \max \left\{ -\frac{1-\delta+\beta\delta}{1-\delta}m; \frac{(q+aq-1+\alpha\delta-\alpha\delta q)(1-\delta+\beta\delta)}{(1-\delta)(1-\alpha\delta)} - \frac{\alpha\beta\delta q}{1-\alpha\delta}I_2(m) \right\} \quad (13)$$

where, as before, the first term in the maximand corresponds to the value of not initiating, while the second term corresponds to the value of initiating.

Lemma 4 *An initiator m initiates the transaction in state I2 if and only if $m \geq$*

$$m(q) \equiv \frac{(1-q)(1-\alpha\delta)-aq}{1-\alpha\delta+\alpha\beta\delta q}.$$

Again, this result is a straightforward implication of (13) and Assumption 3 and its proof is omitted.

Lemma 4 says that, in a new match, initiators with a relatively high level of intrinsic trust will behave in a trusting way and initiate, while initiators with a relatively low intrinsic trust will not trust and thus will not initiate. The threshold $m(q)$ is a decreasing function of q , the aggregate steady state probability of a respondent responding honestly in a new match. This is intuitive. The more likely the

respondents are to behave honestly in a new match, the more initiators will choose to trust them. In particular, a higher value of q implies a decreased likelihood of theft and an increased likelihood of a mutually profitable long-term relationship.

Lemmata 1 through 4 completely characterize initiator and respondent behavior in all states where they need to make a decision. They relate this behavior to aggregate measures of trusting (p) and trustworthy (q) behavior in new matches. Of course, the latter are not exogenous and should themselves be treated as aggregates of individual behavior. The next section turns to this task.

2.3 General Equilibrium Analysis

This section builds on Lemmata 1 through 4 in characterizing the general equilibrium in the society. The central idea is simple: an equilibrium is a pair (p, q) that is mutually consistent under the initiator-respondent interactions. To elaborate in more detail, consider a particular value of p and how this value maps to a value of q consistent with it. By Lemma 3, the respondents with $d \geq d(p)$ respond honestly if called to respond in a new match, while the others respond dishonestly. To simplify the language, call the former ones "trustworthy" and call the latter ones "untrustworthy". It follows that the measure of trustworthy respondents is $Q \equiv 1 - F[d(p)]$ and the measure of untrustworthy respondents is $1 - Q = F[d(p)]$.¹⁰

¹⁰To clarify the notation, note that q is the *conditional* steady state probability measure of trustworthy respondents, where the conditioning is based on the set of searching respondents. When multiplied by the steady state measure of searching respondents, it gives the measure of searching respondents who *stand ready* to behave in a trustworthy way in the current period (or any other specific period). On the other hand, Q is the *unconditional* steady state probability measure of trustworthy respondents. That is, it is the measure of respondents who *would* behave in a trustworthy way if they happened to find themselves in a new match with an initiated transaction in the current period (or any other specific period).

Since q is the aggregate steady state probability of a respondent responding honestly in a new match, it must be equal to the fraction of trustworthy searching respondents among the searching respondents. Because all the untrustworthy respondents search at the beginning of each period (since they never participate in a surviving match), we only need to find the fraction of trustworthy respondents that search in order to deduce q . Denote the latter fraction h^R . Since this fraction has to stay constant over time in a steady state, in any period the measure of new matches involving trustworthy respondents that survive until the following period has to be equal to the measure of surviving matches involving trustworthy respondents that get dissolved in the current period. As for the former, the fraction h^R of trustworthy respondents that search results in the fraction βh^R of trustworthy respondents involved in new matches, the fraction $p\beta h^R$ of trustworthy respondents involved in new matches experiencing an initiated transaction, the fraction $p\beta h^R$ of trustworthy respondents involved in new matches experiencing a successfully completed transaction and, finally, the fraction $\alpha p\beta h^R$ of trustworthy respondents involved in new matches that survive until the following period. As for the latter, the fraction $1 - h^R$ of trustworthy respondents participating in surviving matches results in the fraction $(1 - \alpha)(1 - h^R)$ of trustworthy respondents participating in surviving matches that get dissolved in the current period. In equilibrium, then, $\alpha p\beta h^R = (1 - \alpha)(1 - h^R)$, which gives

$$h^R = \frac{1 - \alpha}{1 - \alpha + \alpha\beta p} \quad (14)$$

Consequently, the measure of searching respondents is $F[d(p)] + \frac{1-\alpha}{1-\alpha+\alpha\beta p} \{1 - F[d(p)]\}$.

The previous analysis then implies that the value of q that is consistent with p is given by $q = T(p)$, where $T : [0, 1] \rightarrow [0, 1]$ is defined by

$$T(p) \equiv \frac{\frac{1-\alpha}{1-\alpha+\alpha\beta q} \{1 - F[d(p)]\}}{F[d(p)] + \frac{1-\alpha}{1-\alpha+\alpha\beta q} \{1 - F[d(p)]\}} = \frac{(1-\alpha) \{1 - F[d(p)]\}}{1 - \alpha + \alpha\beta p F[d(p)]} \quad (15)$$

$\forall p \in [0, 1]$.

Now consider a particular value of q and how this value maps to a value of p consistent with it. By Lemma 4, the initiators with $m \geq m(q)$ initiate in a new match, while the others do not. Call the former ones "trusting" and the latter ones "mistrusting". It follows that the measure of trusting initiators is $P \equiv 1 - G[m(q)]$ and the measure of mistrusting initiators is $1 - P = G[m(q)]$.¹¹ Since p is the aggregate steady state probability of an initiator initiating in a new match, it must be equal to the fraction of trusting searching initiators among the searching initiators. Because all the mistrusting initiators search at the beginning of each period (since they never participate in a surviving match), we only need to find the fraction of trusting initiators that search in order to deduce p . Denote the latter h^I . Then, following a similar logic as above, we get that

$$h^I = \frac{1 - \alpha}{1 - \alpha + \alpha\beta q} \quad (16)$$

¹¹Again, to clarify the notation, note that p is the *conditional* steady state probability measure of trusting initiators, where the conditioning is based on the set of searching initiators. When multiplied by the steady state measure of searching initiators, it gives the measure of searching initiators who *actually* behave in a trusting way in the current period (or any other specific period). On the other hand, P is the *unconditional* steady state probability measure of trusting initiators. That is, it is the measure of initiators who *would* behave in a trusting way if they happened to find themselves in a new match in the current period (or any other specific period).

Consequently, the measure of searching initiators is $G[m(q)] + \frac{1-\alpha}{1-\alpha+\alpha\beta q} \{1 - G[m(q)]\}$.

The previous analysis then implies that a value of p that is consistent with q is given by $p = V(q)$, where $V : [0, 1] \rightarrow [0, 1]$ is defined by

$$V(q) \equiv \frac{\frac{1-\alpha}{1-\alpha+\alpha\beta q} \{1 - G[m(q)]\}}{G[m(q)] + \frac{1-\alpha}{1-\alpha+\alpha\beta q} \{1 - G[m(q)]\}} = \frac{(1-\alpha) \{1 - G[m(q)]\}}{1-\alpha + \alpha\beta q G[m(q)]} \quad (17)$$

$\forall q \in [0, 1]$.

We now define a general equilibrium formally:

Definition 1 A general trust/trustworthiness equilibrium is a pair $(p^*, q^*) \in [0, 1]^2$ that satisfies $p^* = V(q^*)$ and $q^* = T(p^*)$.¹²¹³

¹²A reader may wonder why we do not impose an additional equilibrium condition requiring that the measure of searching initiators be equal to the measure of searching respondents, i.e., that

$$G[m(q^*)] + \frac{1-\alpha}{1-\alpha+\alpha\beta q^*} \{1 - G[m(q^*)]\} = F[d(p^*)] + \frac{1-\alpha}{1-\alpha+\alpha\beta p^*} \{1 - F[d(p^*)]\} \quad (18)$$

However, this additional requirement is redundant since this equality is guaranteed for any existing equilibrium. To see that, the two equilibrium conditions imply that

$$G[m(q^*)] = \frac{(1-p^*)(1-\alpha)}{1-\alpha+\alpha\beta p^* q^*} \quad (19)$$

and

$$F[d(p^*)] = \frac{(1-q^*)(1-\alpha)}{1-\alpha+\alpha\beta p^* q^*} \quad (20)$$

Using these results, both sides of the (18) are equal to

$$h = \frac{1-\alpha}{1-\alpha+\alpha\beta p^* q^*} \quad (21)$$

where h denotes a common measure of searching initiators and respondents in an equilibrium.

¹³Given a particular general equilibrium (p^*, q^*) of aggregate steady state probabilities of initiators initiating and respondents responding honestly in a new match, one can deduce the equilibrium measures of trusting initiators and trustworthy respondents, that is P^* and Q^* , by

$$P^* = 1 - G[m(q^*)] \quad (22)$$

and

$$Q^* = 1 - F[d(p^*)] \quad (23)$$

The next theorem establishes existence of a general equilibrium.

Theorem 1 *There exists a general trust/trustworthiness equilibrium.*

Although existence can be established by a routine argument, uniqueness is not guaranteed by the previous assumptions. To see this, note that because $d(\cdot)$ is increasing, T is decreasing. However, although $m(\cdot)$ is decreasing, an increasing property of V cannot be deduced. Intuitively, an increase in p has two effects. First, the trustworthiness threshold $d(p)$ increases and hence there will be less trustworthy and more untrustworthy respondents in the overall population of respondents. Second, as can be seen from (14), a lower fraction of trustworthy respondents will be searching. *Ceteris paribus*, each effect tends to reduce the share of trustworthy searching respondents among the searching respondents. That is, both effects work in the same direction and $T(p)$ decreases as a result. Similarly, an increase in q has two effects. First, the trust threshold $m(q)$ decreases and hence there will be more trusting and less mistrusting initiators in the overall population of initiators. Second, as can be seen from (16), a lower fraction of trusting initiators will be searching. *Ceteris paribus*, the first effect tends to increase the share of trusting searching initiators among the searching initiators, while the second effect has just the opposite impact. Hence the two effects work in opposite directions, and it is not in general possible to say which one will prevail.¹⁴ Figure 2 illustrates a case when both T and V are monotone, which results in a unique general equilibrium.

¹⁴Facing a possibility of multiple equilibria, one may wonder whether there would be a convenient sufficient condition that would rule such case out. (19) and (20) imply that

$$(1 - p^*)F[d(p^*)] = (1 - q^*)G[m(q^*)] \tag{24}$$

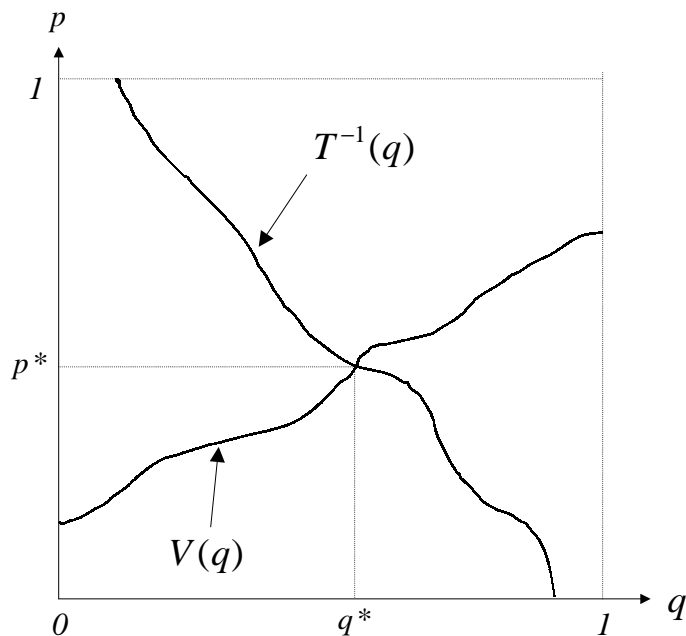


FIGURE 2: Illustration of general equilibrium

3 Empirical Predictions for Individual Prosperity

We are interested in how individual prosperity depends on whether one trusts or not and on whether one is or is not trustworthy. To be able to do that, we need to pick a measure of prosperity. Because the inherent utility of honesty and the inherent utility of trust are unobservable, our primary measure of individual prosperity is an expected, or average, payoff in a single period, net of either type of inherent utility.

Conceptually, there is a particular steady state equilibrium in the background.

in any general equilibrium. Suppose there are at least two equilibria. Then (24) has to be satisfied for at least two different equilibrium pairs (p_1^*, q_1^*) and (p_2^*, q_2^*) , where, without loss of generality, $p_1^* < p_2^*$. Since T is strictly decreasing, it must also be the case that $q_1^* > q_2^*$. Now consider a shift from the first to the second equilibrium. Because $m(\cdot)$ is decreasing, the right-hand side of (24) increases and hence so must the left-hand side to preserve the equality. Therefore a convenient sufficient condition to rule out multiplicity of equilibria is to require that $(1-p)F[d(p)]$ is nonincreasing in p . However, we do not wish to impose this assumption since it is not essential for our objectives.

In that equilibrium, an individual, be it an initiator or a respondent, may find herself or himself in five different outcomes at the end of a typical period.¹⁵ First, (s)he may be matched in a surviving match. Second, (s)he may be unmatched. Third, (s)he may be matched in a new match without an initiated transaction. Fourth, (s)he may be matched in a new match with an initiated transaction and an honest response. Fifth, (s)he may be matched in a new match with an initiated transaction, but with a dishonest response. Let the steady state probabilities of these five outcomes be, in the same order, $\pi^I(m, i)$ and $\pi^R(d, i)$, $i \in \{1, 2, 3, 4, 5\}$, for initiators of type m and respondents of type d , respectively. The per period payoffs for both the initiators and the respondents are a in outcome 1 and 4, 0 in outcome 2 and 3, and they are -1 for the initiators and 1 for the respondents in outcome 5. Hence the expected payoff $\Pi^I(m)$ for an initiator of type m is

$$\Pi^I(m) = [\pi^I(m, 1) + \pi^I(m, 4)] a - \pi^I(m, 5) \quad (25)$$

and the expected payoff $\Pi^R(d)$ for a respondent of type d is

$$\Pi^R(d) = [\pi^R(d, 1) + \pi^R(d, 4)] a + \pi^R(d, 5) \quad (26)$$

In order to be able to compute these expected payoffs, we need to find the steady state equilibrium probability distribution over the five outcomes for each initiator and for each respondent. This is the task to which we turn now.

¹⁵Note that the concept of outcome is different from the concept of state. While states are various ex ante decisionmaking situations, outcomes are various ex post payoff situations.

The first step in this process involves finding the fraction (or conditional measure) of initiators and the fraction of respondents of each type who search at the beginning of a generic period. Given the latter, the probability distribution over the individual outcomes for each type can be computed using the exogenous matching probability β together with an equilibrium degree of trusting and trustworthy behavior, both on the aggregate and the individual level. As in the previous section, one only needs to distinguish between trusting and mistrusting initiators and trustworthy and untrustworthy respondents, since behavior is uniform within each of these groups. Let $p : [0, 1] \rightarrow [0, 1]$ be a function that maps initiator types to equilibrium probabilities of (initiators of that type) initiating if having an opportunity. By Lemma 6,

$$p(m) \equiv \begin{cases} 1 & \text{if } m \geq m(q^*) \text{ (trusting initiators)} \\ 0 & \text{otherwise (mistrusting initiators)} \end{cases} \quad (27)$$

Also let $q : [0, 1] \rightarrow [0, 1]$ be a function that maps respondent types to equilibrium probabilities of (respondents of that type) responding honestly to an initiated transaction. By Lemma 5,

$$q(d) \equiv \begin{cases} 1 & \text{if } d \geq d(p^*) \text{ (trustworthy respondents)} \\ 0 & \text{otherwise (untrustworthy respondents)} \end{cases} \quad (28)$$

Now let $h^I : [0, 1] \rightarrow [0, 1]$ be a function that maps initiator types to equilibrium fractions (of initiators of that type) that search at the beginning of each period and

let $h^R : [0, 1] \longrightarrow [0, 1]$ be a function that maps respondent types to equilibrium fractions (of respondents of that type) that search at the beginning of each period.

Then, using the results of the previous section, we get that

$$h^I(m) = \begin{cases} \frac{1-\alpha}{1-\alpha+\alpha\beta q^*} & \text{if } m \geq m(q^*) \text{ (trusting initiators)} \\ 1 & \text{otherwise (mistrusting initiators)} \end{cases} \quad (29)$$

and

$$h^R(d) = \begin{cases} \frac{1-\alpha}{1-\alpha+\alpha\beta p^*} & \text{if } d \geq d(p^*) \text{ (trustworthy respondents)} \\ 1 & \text{otherwise (untrustworthy respondents)} \end{cases} \quad (30)$$

Given the functional forms of p^I , p^R , h^I and h^R , Figures 3 and 4 depict the probability distribution over the five outcomes for initiators of type m and respondents of type d , respectively.

Combining (25), (26) and the information from Figure 3 and Figure 4 then gives

$$\Pi^I(m) = [1 - h^I(m) + q^*p(m)\beta h^I(m)] a - (1 - q^*)p(m)\beta h^I(m) \quad (31)$$

and

$$\Pi^R(d) = [1 - h^R(d) + q(d)p^*\beta h^R(d)] a + [1 - q(d)]p^*\beta h^R(d) \quad (32)$$

Finally, after using (27), (28), (29) and (30) to substitute into (31) and (32), the expected per period payoff for a trusting initiator is

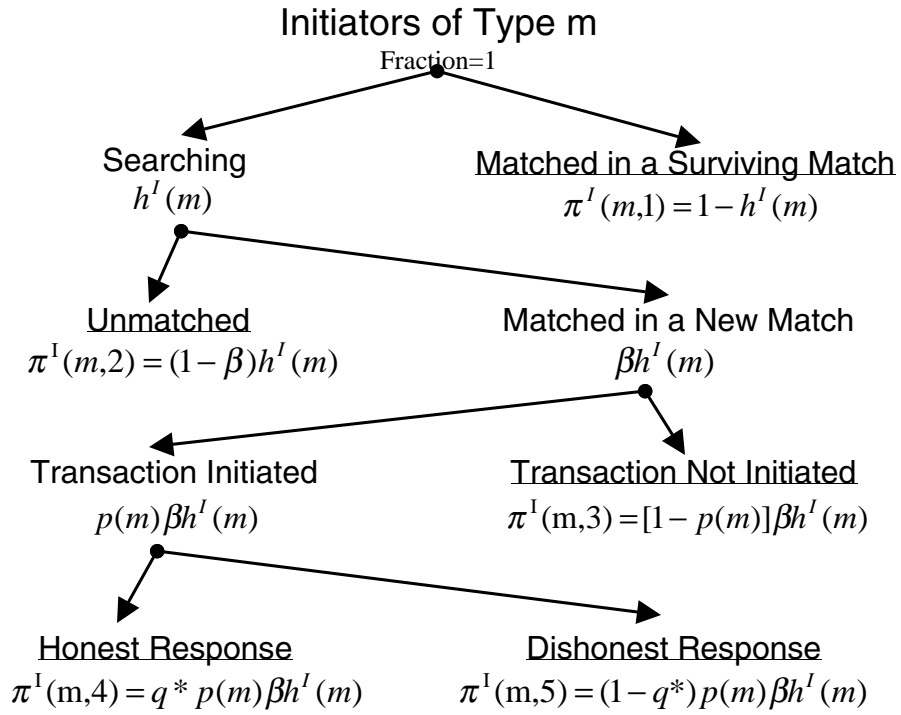


FIGURE 3: Probability distribution over outcomes for initiators of type m

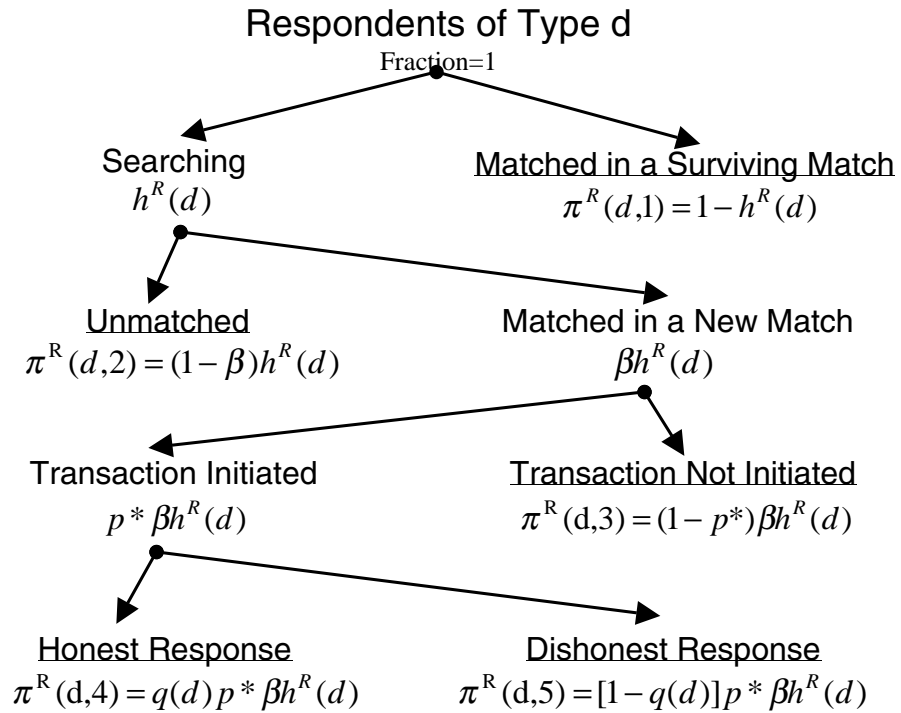


FIGURE 4: Probability distribution over outcomes for respondents of type d

$$\Pi_{trusting}^I = \frac{\beta [q^* a - (1 - q^*)(1 - \alpha)]}{1 - \alpha + \alpha \beta q^*}, \quad (33)$$

the expected per period payoff for a mistrusting initiator is

$$\Pi_{mistrusting}^I = 0, \quad (34)$$

the expected per period payoff for a trustworthy respondent is

$$\Pi_{trustworthy}^R = \frac{\beta p^*}{1 - \alpha + \alpha \beta p^*} a, \quad (35)$$

and the expected per period payoff for an untrustworthy respondent is

$$\Pi_{untrustworthy}^R = \beta p^*. \quad (36)$$

In general, $\Pi_{trusting}^I$ may be more or less than $\Pi_{mistrusting}^I$, depending on the value of the parameters and the shapes of F and G . So it is not possible to conclude in general whether trust does or does not increase expected income (i.e., "pay off"). However, for $q^* = 0$ we have $\Pi_{trusting}^I = -\beta < \Pi_{mistrusting}^I$, for $q^* = 1$ we have $\Pi_{trusting}^I = \beta a / (1 - \alpha + \alpha \beta) > \Pi_{mistrusting}^I$, and $\Pi_{trusting}^I$ is continuous and strictly increasing in q^* . Therefore we can conclude that trust does not pay off in low trustworthiness societies, but it does pay off in high trustworthiness societies. Furthermore, it pays off more the more trustworthy a society is. The threshold value of q^* that makes the two payoffs equal is $\hat{q} \equiv \frac{1-\alpha}{1-\alpha+a}$. Hence trust pays off if $q^* > \hat{q}$

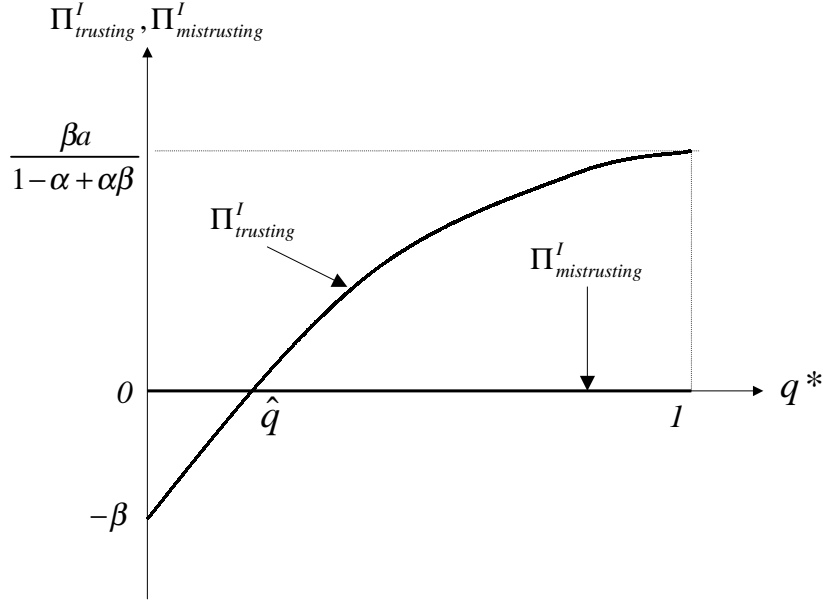


FIGURE 5: Expected equilibrium payoffs of trusting and mistrusting initiators as functions of q^*

and it does not pay off if $q^* < \hat{q}$. This is illustrated in Figure 5.

Similarly, $\Pi_{trustworthy}^R$ may be more or less than $\Pi_{untrustworthy}^R$ and hence it is not possible to conclude in general whether trustworthiness does or does not pay off. For $p^* = 0$ we have $\Pi_{trustworthy}^R = \Pi_{untrustworthy}^R$. Since $\Pi_{trustworthy}^R$ is strictly increasing and strictly concave in p^* , three scenarios, depicted in Figure 6, are possible. First, if $a \geq 1 - \alpha + \alpha\beta$, the "high" case, trustworthiness pays off no matter what the social level of trust p^* is. Second, if $1 - \alpha + \alpha\beta > a > 1 - \alpha$, the "medium" case, trustworthiness pays off for low levels of social trust, in particular $p^* \leq \hat{p} \equiv \frac{a - (1 - \alpha)}{\alpha\beta}$, but it does not pay off for high levels of social trust, i.e. when $p^* > \hat{p}$. Third, if $1 - \alpha > a$, the "low" case, trustworthiness does not pay off for any level of social trust.

Interpreting Figures 5 and 6 is complicated by the fact that the fundamental

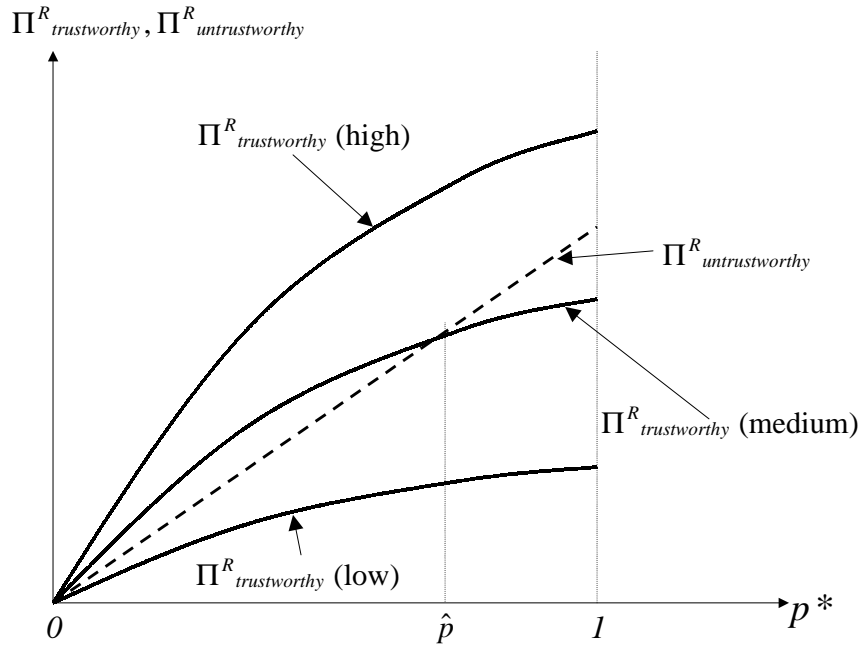


FIGURE 6: Payoffs of trustworthy and untrustworthy respondents as functions of p^*

aspects of countries that determine the equilibrium values of p^* and q^* - the distributions $G(\cdot)$ and $F(\cdot)$ of inherent attitudes, and the values of α , β , δ and a —also affect the difference in expected income between trusting and not trusting for any given level of q^* , and between being and not being trustworthy for any given level of p^* . For this reason, one cannot interpret Figures 5 and 6 as directly revealing how the expected difference in income depends on the values of p^* and q^* . Many of the variations in fundamentals that affect p^* and q^* will also affect the expected income of a person conditional on their attitude.

This stylized model of how individual attitudes combine to generate an equilibrium provides the framework we use in the empirical analysis of whether trust and trustworthiness pay off that we develop in Sections 5 and 6. In particular, we will

examine the model's prediction that the payoff to exhibiting each of these behaviors depends on the aggregate prevalence of the complementary attitude. Before we begin the empirical analysis, though, we briefly review the empirical literature that is related to our investigation.

4 Empirical Literature Review¹⁶

There is some empirical evidence that trust and civic duty among a country's citizens contribute to growth. Knack and Keefer (1997) tested the impact of these attitudes on both growth and investment rates in a cross-section of 29 countries, using measures of trust and civic norms from the World Values Surveys of 1981 and 1990. They find that social capital variables exhibit a strong and significant positive relationship to economic growth. As they note, the causality of this relationship could go in either direction: trust could be a product of optimism generated by high or growing incomes, or it could be that trust facilitates prosperity. However, they find that trust is more correlated with per capita income in later years than with income in earlier years, suggesting that the causation runs from trust to growth more so than vice versa.

Zak and Knack (2001) extend the Knack and Keefer framework by separately testing for the effect on growth of proxies for the presence of formal institutions, social distance, and discrimination and for whether their effect remains significantly

¹⁶**In this review, we focus on the impact of trust and trustworthiness on economic outcomes. There is also literature studying the determinants of trust. See Alesina and Ferrara (2000) for a recent contribution.**

correlated with growth controlling for measures of trust. They find that trust is positively and significantly related to growth even in the presence of measures of formal institutions or of social distance, but that most of the influence of the latter on growth occurs through their impact on trust. The one exception is a measure of property rights, which retains its independent positive association with growth even in the presence of a trust variable. They justify this finding by noting that this index includes government actions against private agents. In contrast, the trust measure is “likely to be little affected by perceptions of the trustworthiness of government. . .” (p. 316)

La Porta, Lopez-de-Silvanes, Shleifer, and Vishny (1999) find that, across countries, a one-standard deviation increase in the measure of trust increases judicial efficiency by 0.7 of a standard deviation and reduces government corruption by 0.3 of a standard deviation. Putnam (1993) examines cross-regional Italian data and concludes that local governments are more efficient where there is greater civic engagement.

In what follows, we use a somewhat different empirical strategy by examining household-level rather than country-level data from 18 countries. In particular, we estimate regression equations explaining household income with a specification that is based on the standard earnings equations from labor economics, but that is augmented to test for the impact of trust, trustworthiness and their interaction with aggregate levels of the complementary attitude. The theoretical model of Sections 2 and 3 frames our approach.

5 Data

Although the theory provides a consistent framework in which to evaluate data, it leaves open the precise relationship between income and personal and country characteristics. To shed empirical light on the issues discussed in the previous section, one needs measures of individual well-being, personal trust, trustworthiness and, preferably, some additional sociodemographic variables. To our knowledge, only two datasets provide this information: the National Opinion Research Center's General Social Survey (GSS) and the World Values Survey (WVS). In order to identify the impact of aggregate trust and trustworthiness within the society, we must use WVS, as it, unlike the GSS, provides individual-level data for multiple countries.

The purpose of the WVS is to facilitate cross-national comparisons of values, norms, and attitudes. The survey was conducted in multiple waves, with limited national modifications, in several dozen countries. It asked about attitudes concerning work, family, religion, politics, and contemporary social issues and gathered a limited amount of demographic data as well. Although the data are subject to the usual reservations about attitude surveys, and in particular cross-country attitude surveys, the data has been widely and fruitfully used by political scientists and sociologists, not to mention Knack and Keefer (1997) and Zak and Knack (2001). For an extensive, albeit incomplete, list of its use in research, see Inglehart, Basanez, and Moreno (1998). We use the data from the 1990-93 wave for 18 developed and

developing countries.¹⁷ We excluded the former communist countries because their economic and incentive structure as of the time of the survey was not conducive to trust and trustworthiness having much effect on individual prosperity.¹⁸ We supplement the WVS data with Summers and Heston (1991) Penn World Tables (PWT), Mark 5.6 to be able to make real income comparisons across countries.

Our measure of trust is based on the following WVS question: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" This question offered two responses: "can't be too careful" and "most people can be trusted". We associate the former answer with "mistrusting" individuals and the latter answer with "trusting" individuals. Based on these survey responses, we create a binary variable TRUST indicating the trusting individuals. Our measure of trustworthiness is based on the following WVS question: "Please tell me whether you think lying in your own interest can always be justified, never be justified, or something in between." This question offered 10 responses ordered from 1 (never justified) to 10 (always justified). In order to measure trustworthiness, we reversed the scale and call the resulting variable TRUSTW.

Glaeser et al. (2000) measure trust and trustworthiness by conducting experiments with monetary rewards. They find that the standard question used to measure

¹⁷We use the following countries: Austria, Belgium, Brazil, Britain, Canada, Chile, Finland, India, Italy, Japan, Mexico, The Netherlands, Portugal, South Africa, Spain, Turkey, USA, and West Germany.

¹⁸As for the remaining countries in the 1990-93 wave, we could not use Argentina, Denmark, Ireland, Nigeria, Norway, Sweden, and Switzerland because the income category thresholds that we use for measuring real household income (see below) were not available. We could not use France because the household income data records did not precisely match with the available income category thresholds. We could not use Iceland because of the missing household income data. Finally, we could not use South Korea because of the missing education data.

trusting behavior - used in the WVS as well as the GSS - does not have a significant correlation with trusting choices in either of two experiments. Two other questions, specifically about trusting strangers, do, though, predict trust (of strangers, in their experiments). Furthermore, the answers to questions about trustworthiness are not significantly related to trustworthy behavior. Surprisingly, a self-reported trusting attitude does appear to predict trustworthy behavior. Danielson and Holm (2002) conduct a similar experiment in Tanzania. They confirm that the standard survey question used to measure trust does not predict actual trusting behavior in their experimental setting. Unlike Glaeser et al. (2000), though, they find that the specific trust questions do not predict actual trusting behavior and that the general trust question does not predict trustworthy behavior. They also find that self-reported trustworthiness does in fact predict trustworthy behavior, but this effect disappears when donation motives are controlled for.

Glaeser et al. (2000) and Danielson and Holm (2002) conclude that empirical work based on the WVS/GSS survey questions about trust needs to be reinterpreted. While we take seriously the possibility that self-reported attitudes and behavior may not be highly correlated, we do find below that these self-reports help explain individual incomes with a systematic pattern, and so we conclude that they do reflect individual behavior in an important sense. Finally, although experimental evidence could certainly extend our knowledge of these issues, we expect that such evidence will not be available across countries in the near future, rendering the current study infeasible from this angle.

We measure individual prosperity by real household income based on the follow-

ing WVS question: "Here is the scale of incomes and we would like to know in what group your household is, counting all wages, salaries, pensions and other incomes that come in. Just give the letter of the group your household falls into, before taxes and other deductions." This question offered 10 country-specific ranges for income. We convert the thresholds into 1990 purchasing power parity U.S. dollars using the PWT measure of PPP-based exchange rates. Our measure of real household income is a midpoint of each range and 150% of the highest threshold for the top range. Summary statistics for household income, trust and trustworthiness by country are reported in Table 1.

Because individual trust and trustworthiness are certainly not to be the only determinants of individual income, we examine additional sociodemographic information provided by WVS. Our measure of respondent education is based on the following WVS question: "At what age did you or will you complete your full time education, either at school or at an institution of higher education? Please exclude apprenticeships." This question offered a 10-point scale ranging from 1 (12 years of age or earlier) to 10 (21 years of age or older). In addition, we use the data on respondent age and gender. It is important to note that the measure of income we investigate relates to the household, but both the attitude indicators and sociodemographic variables refer to the respondent. We will have more to say later about how that affects the interpretation of our results.

TABLE 1: Summary statistics of real household income, trust and trustworthiness by country

COUNTRY	HOUSEHOLD INCOME			TRUST			TRUSTWORTHINESS		
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Austria	1414	18.28	11.65	1301	0.318	0.466	1440	8.27	2.04
Belgium	1705	21.14	13.29	2576	0.332	0.471	2730	7.04	2.62
Brazil	1679	6.36	9.69	1766	0.067	0.249	1774	8.20	2.73
Britain	1101	24.80	17.47	1440	0.436	0.496	1473	8.17	2.08
Canada	1461	35.88	23.50	1673	0.524	0.500	1712	8.18	2.24
Chile	1470	6.74	7.22	1458	0.227	0.419	1495	8.65	2.23
Finland	586	29.52	15.13	558	0.627	0.484	580	8.13	1.97
India	2429	4.77	3.87	2365	0.343	0.475	2480	8.76	2.03
Italy	1424	21.52	23.12	1932	0.371	0.483	1987	8.33	2.19
Japan	896	32.39	18.98	911	0.417	0.493	968	8.75	1.83
Mexico	1451	19.29	29.68	1384	0.335	0.472	1513	6.76	2.95
Netherlands	790	23.60	15.26	965	0.558	0.497	1010	7.49	2.11
Portugal	1124	13.24	10.27	1149	0.214	0.410	1169	7.47	2.68
South Africa	2456	24.68	18.69	2594	0.283	0.450	2681	8.05	2.89
Spain	3431	11.19	8.04	3887	0.338	0.473	4042	8.07	2.32
Turkey	1007	12.50	18.27	1012	0.100	0.300	1017	8.72	2.23
USA	1696	34.83	22.44	1782	0.500	0.500	1821	8.59	2.02
West Germany	1932	23.68	14.27	1725	0.378	0.485	2013	7.46	2.34
TOTAL	28052	20.24	19.28	30478	0.354	0.478	31905	8.06	2.40

Notes: All summary statistics are weighted averages based on survey weights provided with the data, with the sums of weights equalized across countries.

6 Empirical Results

6.1 Baseline Results

Table 2 reports our baseline results.¹⁹ It presents the results of regressing the logarithm of real household income against variables that are standard in micro earnings equations plus indicators of the individual's level of trust and trustworthiness, sometimes interacted with the mean level of these variables in the respondent's country. All of the regressions include country dummy variables (coefficients of which are not reported here), and so all the estimated coefficients are identified from within-country variation only. In all reported specifications, these country dummies are jointly significant at 1 percent level. The specification in Column (1) contains only the standard variables in an earnings equation. The results are in line with the empirical literature discussed above, lending credence to the survey-based measures of income, education, age, and gender. The marginal return to the respondent's education level is always positive within the observed range (between 6 and 15 years), although decreasing. Based on the estimated coefficients, going from zero to ten years of education adds 87 percent to income. Furthermore, the marginal return is 11.1 percent per year at 0 years, and falls to 6.29 percent at 10 years. These results are within the range reported in the literature, as discussed earlier.²⁰ The respon-

¹⁹The regressions are calculated using observations unweighted within countries and with sums of weights equalized across countries. We have also estimated analogous regressions without any (cross-country) weight adjustment and with weighting within and across countries combined. None of the principal results reported in this section are affected by this change.

²⁰In the human capital earnings approach standard in labor economics, more recent estimates of the return to education fall anywhere between 0.023 (Isacsson (1999)) and 0.153 (Harmon and Walker (1995)) per additional year of schooling, depending on the dataset used, the set of control variables and the econometric technique. Card (1999) provides a good summary of this literature.

TABLE 2: Regressions results: The impact of trust and trustworthiness on real household income

	Dependent variable: Log of Real Household Income						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Education	0.111 (0.0175)***	0.111 (0.0180)***	0.109 (0.0176)***	0.110 (0.0181)***	0.112 (0.0180)***	0.106 (0.0176)***	0.108 (0.0181)***
Education Squared $\times 10^{-2}$	0.241 (0.0778)***	-0.247 (0.0799)***	-0.230 (0.0781)***	-0.244 (0.0802)***	-0.250 (0.0799)***	-0.217 (0.0781)***	-0.235 (0.0801)***
Age $\times 10^{-1}$	0.368 (0.0192)***	0.362 (0.0198)***	0.374 (0.0193)***	0.370 (0.0199)***	0.361 (0.0198)***	0.375 (0.0193)***	0.370 (0.0199)***
Age Squared $\times 10^{-3}$	-0.474 (0.0204)***	-0.469 (0.0211)***	-0.477 (0.0206)***	-0.474 (0.0212)***	-0.469 (0.0211)***	-0.479 (0.0205)***	-0.475 (0.0212)***
Male	0.0837 (0.0108)***	0.0866 (0.0111)***	0.0822 (0.0109)***	0.0855 (0.0111)***	0.0870 (0.0111)***	0.0836 (0.0109)***	0.0874 (0.0111)***
Trust		0.0733 (0.0118)***		0.076 (0.0118)***	-0.448 (0.197)**		-0.394 (0.198)**
Trustworthiness $\times 10^{-1}$			-0.117 (0.0246)***	-0.124 (0.0251)***		-0.362 (0.0718)***	-0.347 (0.0721)***
Trust*Av.Trustworthiness					0.0647 (0.0241)***		0.0581 (0.0243)**
Trustworthiness*Av.Trust						0.0750 (0.0193)***	0.0685 (0.0192)***
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26046	24544	25675	24235	24544	25675	24235
R-squared	0.42	0.42	0.42	0.43	0.42	0.42	0.43

Note 1: The regressions are calculated using observations unweighted within countries and with the sums of weights equalized across countries.

Note 2: Heteroscedasticity consistent robust standard errors are in parentheses. Significance level notation: * at 10%, ** at 5%, *** at 1%.

dent's age, which is undoubtedly partly a proxy for work experience, initially has a positive impact on income, but its impact peaks at age 38.8, and it has a marginal negative effect thereafter. The marginal return falls from 1.78 percent per year at age 20 to -2.96 percent per year at age 70.²¹ Households with male respondents have an 8.37 percent higher income. Although this is lower in absolute value than the findings in the literature²², the difference is unsurprising as our results relate to household income rather than the respondent's income. In the next subsection, we restrict the sample to include only those households in which the major earners are the respondents and observe a much larger estimated male-female income differential.

Columns 2 through 4 show the estimated return to individual trust and trustworthiness, ignoring any country-level interaction effect. The results suggest that trust, but not trustworthiness, is associated with higher income. Complete trust increases income by 7.33 percent compared to no trust at all. In contrast, complete trustworthiness decreases income by 10.53 percent compared to no trustworthiness at all.²³ A one standard deviation increase in trust increases income by 3.50 per-

Our marginal effect estimates lie within this range.

²¹Angrist and Krueger (1999), using 1990 Census and March 1990 CPS samples, report coefficient estimates on potential experience around 0.041 per year for the Census data and 0 or 0.013 for the CPS data, depending on whether they do or do not use the allocated CPS values. The coefficient estimates on potential experience squared are from -0.00057 to -0.00055 for the Census data and they are statistically insignificant for the CPS data. Our results are similar to these estimates.

²²Altonji and Blank (1999), using the Current Population Survey (CPS) data, estimate the coefficient on a female indicator variable to be -0.421 in 1979 when the additional controls are education, experience and region, and -0.348 when occupation, industry and job characteristics are controlled for as well. In 1995, these estimates are -0.272 and -0.221 , respectively. When using the National Longitudinal Survey of Youth data from 1994, the coefficient estimates on the female dummy are approximately -0.24 to -0.20 .

²³We obtain this numeric result by multiplying the coefficient of 1.17 percent by 9, which is the numeric difference between maximum (10) and minimum (1) measured trustworthiness.

cent, and a one standard deviation increase in trustworthiness decreases income by 2.81 percent. Column 4 shows that the point estimates are not notably changed by including both variables at once.

Columns 5 and 6 present our central results that allow for interaction between personal characteristics and country-level means of trust and trustworthiness. Column 5 reveals that the personal return to trust is larger, the greater is the prevalence of trustworthy people in the country. Moreover, the personal return to trusting behavior is negative unless these behavioral patterns are rewarded. The return to trust is negative in countries that have average trustworthiness below 6.92, and is positive in countries above that. All of the countries in our sample, except for Mexico, have means greater than 6.90, and hence the return to trust is almost always positive in our sample.

Column 6 shows that the same pattern applies to trustworthiness: its effect on individual income is negative unless one lives in a country where the level of trust is above 0.48. In contrast to column 5, though, in all but four countries the mean level of trust falls short of this figure, with the four exceptions being Canada, Finland, the Netherlands and the U.S. Thus, our results suggest that trustworthiness is in most countries not rewarded with higher income—dishonesty pays. How much it pays varies widely. In a very low trust country like Brazil, a one standard deviation increase in trustworthiness is associated with an 8.6 percent decrease in income.

Column 7 shows that the principal results from columns 5 and 6 are unchanged when both attitude variables and both interaction terms are included in the same equation, although statistical significance decreases due to multicollinearity. The

results in this column imply that the return to trust is positive if average trustworthiness exceeds the threshold of 6.78 and the return to trustworthiness is negative unless average trust exceeds 0.507.

As suggested earlier by the gender differential estimates, a potential problem with these results is that we use *household* real income as a dependent variable and *respondent* attitudes and demographic characteristics as independent variables. However, it has been shown (see, for example, Mare (1991) and the references contained therein) that most married or cohabiting couples are characterized by assortative matching by education, age and many other characteristics, thus lending more credibility to our results. Another potential problem is that we restrict the coefficient estimates to be the same across all the countries. It is also possible that trust and trustworthiness are endogenous to income. We address all of these issues in the next subsection.

6.2 Robustness Analysis

In this subsection, we submit our baseline results to four robustness checks. The first two of them pursue the possibility that the key interaction terms are estimated with bias because they are picking up country-specific variations in the effect on income of education, age, and gender. The third check returns to the issue of using household rather than individual income as our dependent variable. Finally, we discuss the implications for our results and remedies for dealing with a potential endogeneity problem due to the possibility of reverse causal impacts running from income to trust and trustworthiness.

In the first robustness test, we retain the structure of our baseline regressions, but allow the coefficients on education, education squared, age, age squared and gender to vary across countries. The coefficient estimates on the variables of interest after enriching the regression specification in this way are shown in Table 3.²⁴ The sign pattern is completely unaffected. The absolute value of all the estimates are scaled back toward zero, by between one-fifth and three-fifths. Because the standard errors of these estimates fall only slightly, the t-statistics all decline, so that the confidence with which we can say these coefficients are not zero also falls. Notably, though, the relative magnitude of the individual and interaction terms in equations (5) and (6) are only slightly changed, so that the estimated cutoff levels of aggregate trust and trustworthiness are not much different.

The second robustness test is even more rigorous. We conduct it in a two-step procedure. In the first step, we estimate regression equations (2), (3) and (4) separately for each country in the sample. This produces, for each equation, 18 separate estimates of the effect of trust or trustworthiness on real household income. These first-step estimates are presented in Panel A and Panel B of Table 4.²⁵ The third panel in this table contains the average values of trust and trustworthiness in each country for comparison purposes. In the second step, we regress the estimated trust coefficients against the country average level of trustworthiness, and vice versa.

²⁴The regressions are calculated using observations that are unweighted within countries, but with the sums of weights equalized across countries. We have also estimated analogous regressions without any (cross-country) weight adjustment, and with weighting within and across countries combined. In this case the coefficient estimates are quantitatively more sensitive to the particular weighting scheme employed. However, the results are affected qualitatively only to the extent of marginal changes in statistical significance.

²⁵These coefficient estimates are based on unweighted observations.

TABLE 3: Regressions results: impact of trust and trustworthiness on real household income when the coefficients on education, education squared, age, age squared and gender are allowed to vary across countries

	Dependent variable: Log of Real Household Income						
	(2)	(3)	(4)	(5)	(6)	(7)	
Trust	0.0665 (0.0112)***		0.0680 (0.0113)***	-0.3140 (0.1956)		-0.2738 (0.1970)	
Trustworthiness $\times 10^{-1}$		-0.0581 (0.0237)**	-0.0644 (0.0242)***		-0.183 (0.0701)***	-0.171 (0.0706)**	
Trust*Av.Trustworthiness				0.0472 (0.0240)**		0.0423 (0.0242)*	
Trustworthiness*Av.Trust					0.0380 (0.0188)**	0.0329 (0.0188)*	
Observations	24544	25675	24235	24544	25675	24235	
R-squared	0.50	0.49	0.50	0.50	0.49	0.50	

Note 1: The regressions are calculated using observations unweighted within countries and with the sums of weights equalized across countries.

Note 2: Heteroscedasticity consistent robust standard errors are in parentheses. Significance level notation: * at 10%, ** at 5%, *** at 1%.

Note 3: Regression equations are numbered according to Table 2.

This procedure imposes much less structure on the form of the interaction between individual characteristics and the average level of the complementary attitude.

The results of the second-stage regressions are shown in Table 5. Panel A reveals that the country-specific estimates of the effect of trust on real income are not significantly related to the average country level of trustworthiness, although the coefficient estimates are positive. This is likely a consequence of the fact that there are only 18 countries in our sample, and hence it is difficult to reach standard levels of significance. However, Panel B shows that, even with only 18 observations, the estimated country-specific effect of an individual's trustworthiness on real income is positively related to the average trust in the country, and the coefficient is significantly different from zero at a 5% confidence level. Thus, this two-step procedure does not corroborate the results of specification (5) in Tables 2 and 3, but it does corroborate the findings in specification (6).

The third robustness check returns to the issue of having *household* real income as a dependent variable and *respondent* attitudes and demographic characteristics as independent variables. As we mentioned already, this issue is likely to be less important for the education and age variables (due to assortative matching of household members) than it is for the gender variable. In this check, we run our baseline specifications (as in Table 2), but we include only those households where the respondent coincides with a major or equal wage earner within the household. The results are shown in Table 6. Compared to Table 2, the coefficient estimates are similar and the country dummies are also jointly significant at 1 percent level. A major difference is that the simple and interaction coefficients on trust now lose

TABLE 4: First stage: Trust and trustworthiness coefficients by country

COUNTRY	A: Trust coefficient			B: Trustworthiness coefficient			C: Country averages			
	Trustworthiness not included		Trustworthiness included	Trust not included		Trust included	Trust	Trustworthiness		
	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.		
Austria	0.0789	2.0629	0.0786	2.0216	-0.0088	-1.1291	-0.0094	-1.1663	0.318	8.266
Belgium	0.0543	1.7694	0.0635	2.0513	-0.0108	-1.8851	-0.0136	-2.2960	0.332	7.038
Brazil	0.0445	0.3773	0.0422	0.3568	-0.0116	-1.1192	-0.0115	-1.1035	0.067	8.198
Britain	0.1018	2.0307	0.1051	2.0947	0.0021	0.1635	0.0004	0.0319	0.436	8.165
Canada	0.1297	3.5106	0.1337	3.6082	-0.0126	-1.5591	-0.0114	-1.3961	0.524	8.178
Chile	0.1663	3.0502	0.1647	3.0055	0.0150	1.5362	0.0135	1.3558	0.227	8.652
Finland	0.0430	0.9212	0.0392	0.8372	0.0179	1.3506	0.0149	1.1570	0.627	8.128
India	-0.0059	-0.1853	-0.0023	-0.0722	-0.0150	-2.4410	-0.0163	-2.5946	0.343	8.762
Italy	0.1466	4.2170	0.1397	3.9651	0.0099	1.2870	0.0037	0.4778	0.371	8.327
Japan	0.1583	3.7541	0.1420	3.3091	0.0035	0.2974	0.0016	0.1322	0.417	8.747
Mexico	-0.0575	-0.7492	-0.0469	-0.6089	-0.0009	-0.0795	0.0025	0.1969	0.335	6.761
Netherlands	0.1056	2.1435	0.1032	2.0792	0.0147	1.3085	0.0111	0.9597	0.558	7.492
Portugal	0.0122	0.2456	0.0105	0.2107	-0.0040	-0.5215	-0.0039	-0.4998	0.214	7.469
South Africa	-0.0341	-0.8289	-0.0400	-0.9690	-0.0437	-6.3683	-0.0429	-5.9532	0.283	8.047
Spain	0.0748	3.3765	0.0800	3.5728	-0.0177	-3.6194	-0.0191	-3.7708	0.338	8.071
Turkey	-0.0720	-0.6373	-0.0756	-0.6660	-0.0146	-1.0304	-0.0151	-1.0582	0.100	8.724
USA	0.1106	2.9761	0.1144	3.0617	0.0003	0.0303	-0.0031	-0.3378	0.500	8.593
West Germany	0.0698	2.4208	0.0662	2.2289	-0.0030	-0.5128	-0.0046	-0.6996	0.378	7.459

Note 1: The regressions are calculated using unweighted observations.

Note 2: The first pair of columns in Panel A corresponds to specification (2) in Table 2. The first pair of columns in Panel B corresponds to specification (3) in Table 2. The second pairs of columns in both Panel A and Panel B correspond to specification (4) in Table 2.

Note 3: Heteroscedasticity consistent robust standard errors are in parentheses. Significance level notation: * at 10%, ** at 5%, *** at 1%.

TABLE 5: Second stage: The impact of average trustworthiness and trust on personal return to trust and trustworthiness

Dependent variable:	A: Trust coefficient		B: Trustworthiness coefficient	
	Trustworth. not included	Trustworth. included	Trust not included	Trust included
Average Trustworthiness	0.0362 (0.0321)	0.0312 (0.0310)		
Average Trust			0.0437 (0.0181)**	0.0383 (0.0169)**
Constant	Yes	Yes	Yes	Yes
Observations	18	18	18	18
R-squared	0.09	0.07	0.19	0.17

Note 1: Ordering of columns corresponds to Table 4.

Note 2: Heteroscedasticity consistent robust standard errors in parentheses. Significance level notation: * at 10%, ** at 5%, *** at 1%.

statistical significance in specifications (5) and (7), although they retain the now-familiar sign pattern. Note that the conclusions from this robustness check mirror the ones from the previous check: the personal return to trust is *not* statistically significantly related to the average level of trustworthiness in society, but the personal return to trustworthiness *is* positively and statistically significantly related to the average level of trust in society. Also note that, as expected, the estimated impact on income of being a male is now much higher, by a factor of four, compared to Table 2.

Finally, we return to the issue of what the WVS trust (and, to a lesser degree, trustworthiness) responses really measure. Glaeser, in particular, has argued that higher-income people are more likely to say they trust others, in part because rich people can effectively punish those who act in an untrustworthy way towards them. That is, they can “afford” to trust. To the extent this is true, there is causation

TABLE 6: Regressions results: The impact of trust and trustworthiness on real household income when only households with major earner respondents are included

	Dependent variable: Log of Real Household Income						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Education	0.149 (0.0278)***	0.145 (0.0291)***	0.148 (0.0279)***	0.144 (0.0292)***	0.146 (0.0291)***	0.145 (0.0279)***	0.142 (0.0291)***
Education Squared $\times 10^{-2}$	-0.376 (0.122)***	-0.364 (0.128)***	-0.373 (0.122)***	-0.361 (0.128)***	-0.369 (0.127)***	-0.358 (0.122)***	-0.351 (0.128)***
Age $\times 10^{-1}$	0.623 (0.0346)***	0.620 (0.0362)***	0.626 (0.0348)***	0.623 (0.0365)***	0.619 (0.0363)***	0.627 (0.0348)***	0.623 (0.0365)***
Age Squared $\times 10^{-3}$	-0.691 (0.0341)***	-0.688 (0.0356)***	-0.692 (0.0343)***	-0.689 (0.0358)***	-0.688 (0.0356)***	-0.694 (0.0343)***	-0.690 (0.0359)***
Male	0.346 (0.0206)***	0.345 (0.0214)***	0.350 (0.0207)***	0.349 (0.0215)***	0.345 (0.0214)***	0.351 (0.0207)***	0.350 (0.0215)***
Trust	0.0652 (0.0185)***	0.0652 (0.0185)***	0.0689 (0.0186)***	0.0689 (0.0186)***	-0.395 (0.377)	-0.395 (0.377)	-0.338 (0.378)
Trustworthiness $\times 10^{-1}$			-0.103 (0.0396)***	-0.114 (0.0413)***		-0.429 (0.104)***	-0.417 (0.107)***
Trust*Av.Trustworthiness					0.0571 (0.0459)		0.0503 (0.0459)
Trustworthiness*Av.Trust						0.0983 (0.0264)***	0.0916 (0.0270)***
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12579	11869	12399	11715	11869	12399	11715
R-squared	0.45	0.44	0.45	0.45	0.44	0.45	0.45

Note 1: The regressions are calculated using observations unweighted within countries and with the sums of weights equalized across countries.

Note 2: Heteroscedasticity consistent robust standard errors are in parentheses. Significance level notation: * at 10%, ** at 5%, *** at 1%.

running from income to the trust (and possibly trustworthiness) response, and the coefficients estimated here do not measure the structural effect of attitudes on income.

A natural approach to this concern is to identify a set of instrumental variables that are correlated with trust (or trustworthiness), but which are not influenced by income. With this objective, we investigated a wide variety of survey responses.²⁶ In the case of trust, the instruments that we considered either lacked statistically significant explanatory power for trust and trust interacted with aggregate trustworthiness in the first stage or, alternatively, led to very imprecise (as measured by the standard errors) and unstable (across various specifications) estimates of the coefficients on predicted trust and trust interacted with aggregate trustworthiness in the second stage. In contrast, we could identify apparently appropriate instruments for trustworthiness, and two-stage least squares estimates yielded qualitatively similar results to those reported in the paper.²⁷ Thus, with respect to the trust measure we regard the issue of potential reverse causation as unsettled, and an important topic for future research. We are less concerned that this is a problem for our estimates of the impact of individual trustworthiness and its interaction with aggregate trust

²⁶To be specific, we investigated answers to whether the respondent considers himself/herself to be a religious person, whether religion is important in his/her life, whether he/she was raised religiously, whether he/she thinks it is important to teach children responsibility, whether he/she would mind having (each category individually) people with criminal record, emotionally unstable people, heavy drinkers, and drug addicts as neighbors, whether he/she considers having friends and acquaintances to be an important aspect of life, whether he/she considers meeting people to be an important job attribute, and whether he/she has confidence in the civil service.

²⁷In particular, using religious variables as instruments for trustworthiness and their interactions with aggregate trust as instruments for trustworthiness interacted with aggregate trust, the coefficient on trustworthiness in specifications (3) and (4) of Table 2 was reduced to about -0.1 , and the sign pattern and the relative magnitude of the coefficients on trustworthiness and trustworthiness interacted with aggregate trust was preserved in specifications (6) and (7), with all these estimates being statistically significant.

on income.

7 Conclusions

Earlier research has established that countries with a high proportion of trusting citizens tend to have higher per capita income and to grow faster. What had not been established is the incentive people have to act in a trusting and trustworthy manner. In this paper we address this issue by developing an equilibrium model of trust and trustworthiness and empirically investigating whether the return to behaving this way depends on the mean amount of the complementary behavior in the society. We find evidence that the personal return to trustworthiness is nonpositive in countries with low levels of average trust, but it is increasing in the latter and eventually positive. On the other hand, the personal return to trust is positive on average, but we are unable to conclude at standard levels of statistical significance that this return is increasing in the average level of trustworthiness in the society.

What our empirical results suggest is that personal attitudes matter for personal prosperity and that the payoff to trustworthiness is related to the aggregate trust level in society. These empirical regularities should inform future theories of the role and interaction of trust and trustworthiness. Although these results are interesting and suggestive in many dimensions, more research is necessary to further refine these findings. First, given the difficulty in linking the survey measure of trust to observed behavior in trust game experiments, we need to verify the validity of our

current measure in other settings and/or develop and implement a more reliable survey instrument. Second, more work is necessary to sort out the causal links between trust, trustworthiness and prosperity.

References

- [1] Alesina, A. and E. La Ferrara. 2000. "The Determinants of Trust." NBER Working Paper No. 7621, March 2000.
- [2] Altonji, J. and R. Blank. 1999. "Race and Gender in the Labor Market." In Handbook of Labor Economics, eds. O. Ashenfelter and D. Card.
- [3] Angrist, J. and A. Krueger. 1999. "Empirical Strategies in Labor Economics." In Handbook of Labor Economics, eds. O. Ashenfelter and D. Card.
- [4] Arrow, K. 1972. "Gifts and Exchanges." Philosophy and Public Affairs 1, no. 4: 343-67.
- [5] Axelrod, R. 1986. "An Evolutionary Approach to Norms." American Political Science Review 80, no. 4: 1095-1111.
- [6] Card, D. 1999. "The Causal Effect of Education on Earnings". In Handbook of Labor Economics, eds. O. Ashenfelter and D. Card.
- [7] Chen, Y. 2000. "Promises, Trust, and Contracts." Journal of Law, Economics, & Organization: 209-232.

- [8] Coleman, J. 1990. *Foundations of Social Theory*. Cambridge, Massachusetts: Harvard University Press.
- [9] Danielson, A. and H. Holm. 2002. "Trust in the Tropics? Experimental Evidence from Tanzania." Lund University Department of Economics Working Paper No. 2002:12, April 2002.
- [10] Dixit, A. 2001. "On Modes of Economic Governance." CESifo Working Paper No. 589.
- [11] Fukuyama, F. 1995. *Trust*. New York: Basic Books.
- [12] Glaeser, E. L., D. Laibson, J. A. Scheinkman, and C. L. Soutter. 2000. "Measuring Trust." *Quarterly Journal of Economics* 115, no. 3 (August): 811-46.
- [13] Harmon, C. and I. Walker. 1995. "Estimates of the Economic Return to Schooling for the United Kingdom." *American Economic Review* 85(5), December 1995: 1278-86.
- [14] Inglehart, R., M. Basanez and A. Moreno. "Human Values and Beliefs: A Cross-Cultural Sourcebook." Ann Arbor: University of Michigan Press, 1998.
- [15] Isacson, G. 1999. "Estimates of the Return to Schooling in Sweden from a Large Sample of Twins." *Labour Economics* 6, no. 4 (November): 471-89.
- [16] Knack, S., and P. Keefer. 1997. "Does Social Capital Have an Economic Pay-off? A Cross-Country Investigation." *Quarterly Journal of Economics* 112, no. 4 (November): 1251-88.

- [17] La Porta, R, F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. 1999. "The Quality of Government." *Journal of Law, Economics, and Organization* 15, no. 1: 222-79.
- [18] Mare, R.. 1991. "Five Decades of Educational Assortative Mating." *American Sociological Review* 56: 15-32.
- [19] Putnam, R. (with R. Leonardi and R. Y. Nanetti). 1993. *Making Democracy Work*. Princeton: Princeton University Press.
- [20] Summers, R. and A. Heston. 1991. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988." *Quarterly Journal of Economics* 106, no. 2 (May): 327-68.
- [21] Tirole, J. 1996. "A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality)." *Review of Economic Studies* 63: 1-22.
- [22] Zak, P. and S. Knack. 2001. "Trust and Growth." *Economic Journal* 111 (April): 295-321.

A Appendix

Proof of Lemma 1. Consider a respondent d . Suppose that the current match lasted for two or more periods before the current period. In that case the respondent must have been solving the problem identical to (4) in the previous period and he must have chosen to respond honestly. Consequently, he will choose to respond

honestly in the current period too. (This is, due to Assumption 3, true even if the respondent is repeatedly indifferent between the two actions.) Now suppose that the current match lasted for only one period before the current period. It is easy to see from (4) (also using Assumption 3) that it is optimal for the respondent to respond dishonestly if and only if

$$\alpha(1 - \delta)R_1(d) > -(1 - \alpha\delta)(1 - d) + a .$$

If $d \geq 1$, this entails $R_1(d) > a / [\alpha(1 - \delta)]$. However, the best possible sequence of payoffs the respondent can get is a forever (because replying dishonestly pays off $1 - d \leq 0 < a$), which has discounted value $a/(1 - \delta) < a / [\alpha(1 - \delta)]$. But this contradicts $R_1(d) > a / [\alpha(1 - \delta)]$. Therefore if $d \geq 1$, the respondent will respond honestly in state 5. Intuitively, if $d \geq 1$, the respondent will respond honestly whenever the transaction has been initiated, since he has too high a disutility from behaving dishonestly. If $d < 1$, then (1) and (4) imply that $R_5(d) > R_1(d) = R_4(d)$, which implies that the value associated with honest response in (4) is no less than the value associated with honest response in (6), i.e. that

$$(1 - \alpha) [a + R_1(d)] + \alpha [a + \delta R_5(d)] \geq$$

$$(1 - \alpha) [a + R_1(d)] + \alpha [a + (1 - k)\delta R_4(d) + k\delta R_5(d)]$$

Because in the previous period the match was new and the respondent chose to respond honestly, it must be the case that

$$(1 - \alpha) [a + R_1(d)] + \alpha [a + (1 - k)\delta R_4(d) + k\delta R_5(d)] \geq 1 - d + R_1(d)$$

and hence also

$$(1 - \alpha) [a + R_1(d)] + \alpha [a + \delta R_5(d)] \geq 1 - d + R_1(d) .$$

Consequently, the respondent will choose to respond honestly in the current period.

■

Proof of Lemma 2. Consider an initiator m . Suppose that the current match lasted for two or more periods before the current period. In that case the respondent must have been solving the problem identical to (9) in the previous period and she must have chosen to initiate. Consequently, she will choose to initiate in the current period too. (This is, due to Assumption 3, true even if the initiator is repeatedly indifferent between the two actions.) Now suppose that the current match lasted for only one period before the current period. It is easy to see from (9) (also using Assumption 3) that it is optimal for the initiator not to initiate if and only if

$$\alpha(1 - \delta)I_1(m) > a + m .$$

If $m \geq 0$, this entails $I_1(m) > a/[\alpha(1 - \delta)]$. However, the best possible sequence of payoffs the initiator can get is a forever (because not initiating pays off $-m \leq$

$0 < a$), which has discounted value $a/(1 - \delta) < a/[\alpha(1 - \delta)]$. But this contradicts $R_1(d) > a/[\alpha(1 - \delta)]$. Therefore if $m \geq 0$, the initiator will initiate in state 3. Intuitively, if $m \geq 1$, the initiator will initiate whenever matched if the respondent is known to respond honestly, since she has a disutility from not initiating. If $m < 0$, (9) and (10) imply that

$$\begin{aligned}
I_2(m) &\leq \max \{-m + I_1(m); (1 - q) [-m + I_1(m)] \\
&\quad + q(1 - \alpha) [a + I_1(m)] + q\alpha [a + \delta I_3(m)]\} \\
&= qI_3(m) + (1 - q) [-m + I_1(m)] \\
&\leq qI_3(m) + (1 - q)I_3(m) \\
&= I_3(m)
\end{aligned}$$

Because in the previous period the match was new and the initiator chose to initiate, it must be the case that

$$(1 - q) [-1 + I_1(m)] + q(1 - \alpha) [a + I_1(m)] + q\alpha [a + \delta I_3(m)] \geq -m + I_1(m)$$

and hence also

$$I_2(m) = (1 - q) [-1 + I_1(m)] + q(1 - \alpha) [a + I_1(m)] + q\alpha [a + \delta I_3(m)] .$$

Combining the previous three results gives

$$I_3(m) \geq -m + I_1(m) .$$

Consequently, the initiator will choose to initiate in the current period. ■

Proof of Theorem 1. Let $W : [0, 1]^2 \rightarrow [0, 1]^2$ be defined by

$$W(p, q) \equiv (V(q), T(p)) \quad \forall (p, q) \in [0, 1]^2$$

Since $F : R \rightarrow [0, 1]$ and $G : R \rightarrow [0, 1]$ are both continuous, it follows from this definition and (23) and (22) that the mapping W is continuous. Since $[0, 1]^2$ is a closed, bounded and convex subset of R^2 , it follows by the Brouwer Fixed Point Theorem that W has a fixed point, i.e., a general trust/trustworthiness equilibrium exists. ■