Trust and the Distribution of Caution

Breuer, Janice Boucher and McDermott, John
University of South Carolina

August 2009

Trust and the Distribution of Caution

Janice Boucher Breuer and John McDermott*

August 11, 2009

Department of Economics, University of South Carolina, Columbia, SC 29208

Abstract

Trust is often considered a determinant of economic performance. The exogeneity of trust, however, is questionable. We develop a model with heterogeneous agents to determine aggregate trustworthiness, trust, and output. People differ according to their risk aversion (caution). The distribution of risk aversion across individuals – along with the threat of punishment – is critical in the process by which trust is formed. More cautious societies supposedly have less trust. We call this the direct effect. There is an indirect effect that works through trustworthiness and it leads to *more* trust. Paradoxically, the net effect may be positive – more caution leads to more trust – especially in

*We wish to thank participants at the University of South Carolina Department of Economics Seminar series, especially Nancy Buchan, Melayne McInnes, and Chun-Hui Miao, for helpful comments.

homogenous societies. In heterogeneous societies, the reverse is true. We also show that trust and output are simultaneously determined, and not monotonically related across countries.

JEL Codes: Z1, C7

1 Introduction

In recent decades, several empirical studies have linked trust to various performance measures across countries. But what does this work mean if trust is endogenous? Trust derives meaning in an environment of uncertainty over the trustworthiness of a counterparty, and requires a commitment to make resources vulnerable to the behavior of that counterparty. This suggests that trust is determined by both individual risk aversion and the perception of others' trustworthiness. Our purpose in this paper is to clarify the nature of these relationships.

Linking trust to trustworthiness is not a new idea. This idea was present in the work of Gambetta (1988), Coleman (1988), Putnam (1993), and Fukuyama (1995). More recently, Glaeser et al. (2000) find that trust and trustworthiness rise when experimental subjects are closer to each other socially. Breuer and McDermott (2008) show that a measure of intrinsically trustworthy agents is a good predictor of high-quality institutions, trust, and output. In his inquiry into the determinants of trust, Fehr (2008) reads the evidence from experiments as consistent with the idea that beliefs about others' trustworthiness are crucial for the formation of trust. Tabellini (2006) constructs an evolutionary model of value transmission across generations to show how cooperation arises from innate preferences and social distance.

Caution has more recently been linked to trust. The idea is that people who do not trust may be simply fearful – have an excess of caution or risk aversion. This literature may have been prompted, in part, by dissatisfaction with the standard trust question.[1] In any case, in their experiments, Schechter (2007) and Karlan (2005) find support for a link between trust and caution, whereas Eckel and Wilson (2004) do not.[2]

Our contribution is to introduce a continuum of heterogenous agents by risk aversion into a basic trust game to derive aggregate measures of trustworthiness, trust, and output. We think of the distribution of caution across individuals as an artifact of culture and we categorize societies by the mean and dispersion of this distribution. Our results are derived in this cross cultural context.

In the game, there are two types of agents, Type A and Type B. Type A agents extend trust to B agents who may respond trustworthily or not. To develop a role for caution, we endow each agent with her own risk aversion,

[1]This is question a165 in the World Values Survey (2006). It has been used in the General Social Survey as well.

[2]The literature on trust in economics is vast. See Fehr (2008) and Buchan (2009) for literature reviews. One branch relates measures of trust at the national or regional level to economic performance. Seminal papers include Knack and Keefer (1997), La Porta et al. (1997), and Zak and Knack (2001). Another branch uses experiments and surveys to measure trust in individuals. Examples include Ashraf et al. (2006), and Berg et al. (1995), as well as those noted in the text.

identically distributed across both A and B types. The endowment of risk aversion is a key factor in the decisions of both A and B because it influences the expected utility of choices for each player. We introduce one other wrinkle regarding B-agents. A small fraction of B agents are *intrinsically trustworthy*. All of the others are *opportunists*. The intrinsically trustworthy behave honestly regardless of their aversion to risk. Opportunists, however, are conditional operators in the sense that they maximize expected utility given the institutional environment. We introduce institutions for capture and punishment to clarify their role in undermining or supporting not only trustworthy behavior but, by extension, trust.[3]

Out of this structure, we derive an equilibrium amount of trustworthy behavior and an equilibrium amount of trust. Then, we investigate how trustworthy behavior and trust are influenced by differences in the mean and standard deviation of the risk distribution. The usual story in the literature is that a more cautious society trusts less, since trust involves making one's self vulnerable to cheating. However, our model brings out a countervailing force that can dominate in certain cases. Because more cautious players have a greater fear of capture and punishment, a larger number of type B players will behave honestly. By backward induction, more A players will trust. It is quite possible that more cautious societies will, paradoxically, trust more. This is more likely to be true in groups of homogeneous societies – those with

[3]See Drago et al. (2009) for empirical support for the hypothesis that crime is sensitive to expected punishment.

a low standard deviation of the distribution of caution. More heterogeneous societies behave conventionally – those with greater average levels of caution trust less.

Our results are important for empirical work relating trust to macroeconomic performance. First, we contend that trust is endogenous and so empirical results demonstrating the importance of trust to output, investment, or other outcomes may be biased. Second, our results suggest that trust and caution are related in a complex manner. We cannot simply say that countries that demonstrate little trust are just very cautious.[4] There is no way to infer trust from an observation of *average* caution. Both results suggest that the empirical work on trust and its importance to economic outcomes needs further study.

The paper is organized as follows. In Section 2 we describe the agents, set out the trust game, and constrain the parameters. In Section 3, we focus on the opportunists' choice of whether or not to cheat on a contract. Section 4 defines the distribution of risk aversion (caution) and shows how the mean and variance affect trustworthiness. In this section, we link trustworthiness to different types of cultures based on their innate caution and degree of cultural diversity. Section 5 is concerned with the determination of trust, given that trustworthiness has already been determined and is known to A

[4] A different point concerns the validity of the standard trust question (a165 in the World Values Survey (2006)). Miller and Mitamura (2003), using survey methods across Japanese and American students, find evidence that the question captures caution *instead* of trust.

agents. In Section 6 we construct a measure of output per capita and show how it is related to the distribution of caution. This allows us to show how trust and output are related as we vary the moments of the distribution across countries. Section 7 concludes.

2 The Game

Agents of type A consider initiating transactions with agents of type B that they encounter randomly. To do so requires trust. No information about B is available to A prior to her decision to trust; she only has aggregate information on B's trustworthiness, as described below.[5] Failure to trust means that no transaction is initiated or realized by the two parties in that period. If an agent fails to make a transaction with another, we assume that each produces 1 unit of a consumption good, which they consume that period. We think of this outcome as the result of working alone: it is not efficient, but it is sufficient for survival.

If a transaction is realized, then output in excess of 2 units is created and split between A and B in a manner to be described below. Regardless of type, Agent j receives utility of:

$$u\left(c_j, \rho_j\right) = \frac{c_j^{1-\rho_j} - 1}{1 - \rho_j} \tag{1}$$

[5]This makes the solution of our game straightfoward. In Tirole (1996), by contrast, A receives a signal about the probability that B has cheated in the past. This gives rise to multiple equilibria based on parameter values. That work does not, however, deal with heterogeneity of risk aversion.

where $c_j \geq 1$ is her consumption and $-\infty < \rho_j < \infty$ is her rate of relative risk aversion. Our agents are heterogeneous. Both A and B types are distributed continuously according to their rate of risk aversion ρ. We will often refer to ρ as an individual's level of *caution.* Notice that when a transaction is not realized, each individual receives $c_j = 1$ so utility is equal to zero, no matter what the value of ρ:

$$u(1, \rho_j) = 0 \tag{2}$$

We also assume that a fraction of the B-agents are *intrinsically trustworthy* and will never cheat in a transaction.[6] There is ample support for this assumption in the literature.[7] The rest of the B-agents are *opportunists* who may cheat or not, depending on the expected payoff.

The game is shown in Figure 1. Player A moves first from Point **A**. She must decide whether to offer a contract to B for the joint production of output. A decision to trust B means that the outcome for player A is dependent on the behavior of player B and that B is contractually obligated to fulfill the contract.[8] If player A trusts, the game continues to node **B**; if she does not, the game ends at Point **a**. Both A and B receive a payoff of 1 in

[6] Alternatively, we could assume that a certain fraction have a value for ρ_j so large that they would never risk cheating.

[7] See Frank (1987), Sen (1977), and Tirole (1996) for examples. More recently, in Tabellini (2008) some types are naturally more trustworthy than others.

[8] Without a contractual obligation, player B's behavior cannot be characterized as either trustworthy or not, and ipso facto, capture and punishment are moot. Institutions would be irrelevant. In our view, trust is based on an agreed contract between A and B, whether explicit or implicit. Otherwise, the game is a game of 'hope.' See Breuer and McDermott (2008). Fehr (2008), Ben-Ner and Putterman (2008), and Karlan (2005) design experiments or conduct surveys and field work on trust and trustworthiness where an obligation is involved.

that case. The payoff to this outcome is shown in Figure 1 as the bracketed pair $[A_{payoff}, B_{payoff}] = [1, 1]$ at Point **a**.

Does A trust B and proceed? Her problem is that she does not know B's type: he may be intrinsically trustworthy or he might be an opportunist. If she knew B were trustworthy, she would trust, initiate, conclude a transaction, and earn income of $(1 + y_m)$ following path **ABb**. Person B would also earn $(1 + y_m)$. The payoff again is noted in brackets at the path's end, and the game ends. The open circle at node **B**, indicates that the choice is made by nature: with probability R_{TW} player A encounters an intrinsically trustworthy B player.

If B is an opportunist – with probability $(1 - R_{TW})$ – but acts honestly, and fulfills the contract, then both A and B also receive the standard amount $(1 + y_m)$, just as if B were intrinsically trustworthy. In Figure 1, this corresponds to path **ABCc**. But an opportunist may not act honestly. If he cheats, he will be caught with exogenous probability Q. This parameter is our measure of "institutions": Q is high in societies that succeed in finding and convicting criminals.

When B cheats, α units of output are transferred from A to B. A deadweight loss of β is also experienced by A. If B is caught – with probability Q – he faces two penalties: ϕ is paid to the state as punishment (this also represents a deadweight loss) and λ is paid to make restitution to A. In this case, B's total payoff is $(1 + y_m + \alpha - \phi - \lambda)$. Person A receives the amount $(1 + y_m - \alpha - \beta + \lambda)$. This is represented by path **ABCDd** in Figure 1.

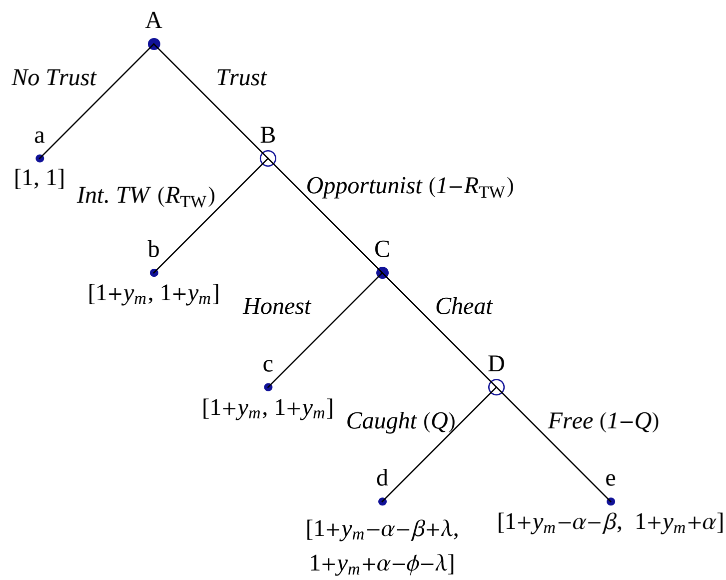

Figure 1: Game of Trust and Trustworthiness

If B cheats and is *not* caught – an event that occurs with probability $(1 - Q)$ – then B receives $(1 + y_m + \alpha)$ while A gets $(1 + y_m - \alpha - \beta)$. This possibility is captured with path **ABCDe** in Figure 1. Again, we represent a choice made by nature with an open circle at node **D**.

The payoffs must conform to several conditions to ensure a solution. First, it is necessary that

$$\lambda + y_m < \alpha + \beta < 1 + y_m \tag{3}$$

This condition ensures that a trustor who is cheated, whether the cheater is caught or not, receives income in the interval $(0, 1)$. Consumption cannot be negative and the initiator must prefer the work-alone option to be being cheated, even if the thief is punished. Note that this requires $\lambda < 1$.

We also require:

$$\alpha < \phi + \lambda < 1 + y_m + \alpha \tag{4}$$

When this is satisfied, the income received by an opportunist that cheats and is caught is in the interval $(0, 1 + y_m)$. This ensures that consumption is positive but that cheating and getting caught is worse than cooperation (acting trustworthy).

Based on her own level of risk aversion, Player A must use the information available to her to calculate the expected utility from trusting. To see what she must know to make a rational decision, we must first turn to the problem facing the opportunist.

3 The Opportunist's Choice

The decision to cheat is equivalent to the purchase of a lottery ticket: with probability Q the opportunist gets caught and ends up with $(1 + y_m + \alpha - \phi - \lambda)$; with probability $(1 - Q)$ he goes free and gains $(1 + y_m + \alpha)$. Thus, the expected utility from cheating for the opportunist is:

$$Q\, u\,(1 + y_m + \alpha - \phi - \lambda, \rho_j) + (1 - Q)\, u\,(1 + y_m + \alpha, \rho_j) \tag{5}$$

On the other hand, playing trustworthy yields utility of:

$$u\,(1 + y_m, \rho_j) \tag{6}$$

The opportunist will be trustworthy only if (6) is greater than (5). That is, for person j honesty is rational if:

$$\begin{aligned} u\,(1 + y_m, \rho_j) - \Big\{ Q\, u\,(1 + y_m + \alpha - \phi - \lambda, \rho_j) + \\ + (1 - Q)\, u\,(1 + y_m + \alpha, \rho_j) \Big\} > 0 \end{aligned} \tag{7}$$

The interesting feature of this condition is that its satisfaction depends on the value of ρ_j so that it can hold for some opportunists, but not for others. Agents with a high value of ρ will, other things equal, act honestly compared to agents with a low value of ρ. If the chance of apprehension Q is sufficiently high, *all risk averse* opportunists will honor the trust extended to them. We derive this value by first setting $\rho = 0$ – representing the agent who is risk neutral. Then, we set (7) equal to zero and solve for Q, which yields:

$$Q_M = \frac{\alpha}{\phi + \lambda} \tag{8}$$

This is the condition that makes the risk neutral opportunist indifferent between cheating and playing honest. For $Q > Q_M$, any risk averse opportunist would play trustworthy.[9] From condition (4) we know that Q_M is less than one.

More generally, for any vector $(Q, y_m, \alpha, \phi, \lambda)$, there is a critical value of ρ – call it ρ_c – such that an agent with $\rho_j > \rho_c$ will *play trustworthy*. We find this critical value by setting (7) to zero and letting the resulting expression implicity define ρ_c:

$$\rho_c = C(Q) \tag{9}$$

where $C' < 0$. The critical value ρ_c also depends on y_m, α, ϕ, and λ but, since we concentrate on Q, we ignore them in (9). In countries where the chance of apprehension and conviction Q is high, the marginal opportunist has a low level of risk aversion.

We illustrate many of our results below with numerical examples. Already, we have made one important calibrating assumption: the work-alone option yields output (and consumption) of 1. This normalization is the

[9] Our focus on Q is somewhat arbitrary. We could just as easily express (8) as conditions on ϕ , λ, or α such that for any $\phi > \phi_M \equiv \frac{\alpha}{Q} - \lambda$, $\lambda > \lambda_M \equiv \frac{\alpha}{Q} - \phi$ or $\alpha < \alpha_M \equiv Q(\phi + \lambda)$, risk averse opportunists would play honest.

benchmark by which the other variables and parameters must be compared. We are, moreover, constrained in our choice of parameters by conditions (3) and (4). We begin by setting two values:

- $y_m = 3.0$
- $\lambda = 0.3$

The first means that a completed transaction is three times more productive than the work-alone process. The second says that restitution from B to A (if B is caught) is 30% of the output from the work-alone option.

To set α and β, we adopt the following rules. We assume that the sum $\alpha + \beta$ is exactly halfway through the interval defined by (3), which is (3.3, 4.0), given our choices for y_m and λ. Then, we assume that $\alpha = 2\beta$: the amount stolen by B in a dishonest transaction (α) is twice the deadweight loss that is also inflicted on A (β). Similarly, we assume that the sum $\phi + \lambda$ is three-quarters of the range determined by (4). Since we have already set λ, this is sufficient to determine the penalty suffered by B if he is caught (ϕ).[10]

The values that come out of this exercise are:

- $\alpha = 2.433$
- $\beta = 1.217$
- $\phi = 5.133$

[10] The reason we do this is so we can change y_m and/or λ and the other three parameters will automatically change to maintain the constraints in a reasonable way. Later, we show that a doubling of y_m and λ do not alter our main results.

The baseline parameter values imply that $Q_M = \frac{2.43}{5.13} = .448$, so if society captures (and convicts) criminals 45% or more of the time, *no* risk averse opportunists would cheat ($\rho_c = 0$). In our baseline case, we consider a country that has weaker institutions such that the probability of capture and conviction is $Q_1 = \frac{Q_M}{2} = .224$. In this case, the cut-off value is calculated from (7) to be $\rho_{c1} = 1.18$. Everyone with risk aversion greater than 1.18 in the baseline society will act honestly.

4 The Distribution of Caution and Trustworthy Behavior

We now turn to the distribution of risk aversion – or *caution* – in society, which allows us to link individual behavior to aggregate outcomes. We assume the distribution of caution ρ within a country is normal[11] with mean μ and standard deviation σ:

$$\rho \sim N(\mu, \sigma) \tag{10}$$

We think of both moments as reflecting something fundamental about a nation's culture, or blend of cultures. In risk-loving, *entrepreneurial* societies, for example, μ is low. In traditional, *cautious* societies, μ is high. In *homogeneous* societies, σ is low, regardless of the value of μ. *Heterogeneous* societies, on the other hand, are made up of many different cultures and have a high

[11] We have also used a uniform distribution. The results are very similar.

value of σ. Table 1 shows the four-way categorization based on the moments. We can think of the United States as an example of the first kind of society (the upper left corner of Table 1) and Japan as an example of the second (the lower right corner).[12]

Table 1: Societal Categories and the Distribution of Caution

		Mean Level of Caution	
		μ_{low}	μ_{high}
Standard Deviation	σ_{high}	Entrepreneurial Heterogeneous	Cautious Heterogeneous
of Caution	σ_{low}	Entrepreneurial Homogeneous	Cautious Homogeneous

For our baseline case, we set the mean to $\mu_1 \equiv .60$ and the standard deviation to $\sigma_1 = .50$ because these values are similar to those found by Harrison et al. (2007) in their Danish field experiments.[13] This distribution $f(\rho)$ is shown in Figure 2. These values ensure that when the chance of getting caught is $Q_1 = \frac{Q_M}{2}$ (so $\rho_c = 1.18$, as previously discussed) then less than half will play fair since $\mu < \rho_c$. The dark area shows the fraction of *opportunists* who act honestly when $Q = Q_1$. In this case, 21.1% of opportunists act trustworthily. The lighter area is the proportion of opportunists who choose

[12] There is little hard evidence that Japan is, on average, more cautious as a society than the US. However, there is some survey evidence that does suggest that this is the case. One piece of evidence comes from the World Values Survey (2006). Question c009 asks people to select important features of a job: 15% in the US chose "Job security"; 34% did so in Japan. More evidence comes from Hofstede's Index of Uncertainty Avoidance: the US score was 46%; Japan's was 92%. These are measures of μ, if anything. We have no information about σ except the general perception of cultural homogeneity in Japan.

[13] Holt and Laury (2002, 2005) and Harrison et al. (2005) all find similar values with different experimental populations.

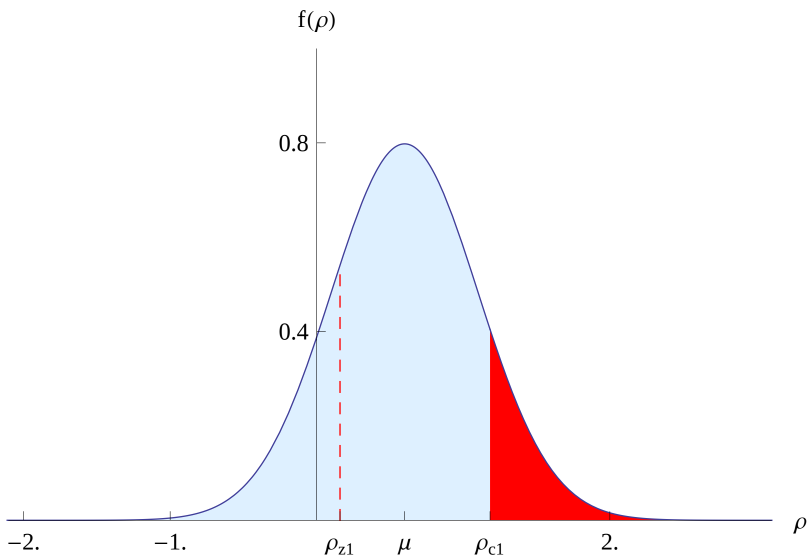

Figure 2: Distribution of Opportunists by Caution

to cheat.

We call the fraction of opportunists acting in a trustworthy manner V_{TW}. When we refer to a nation's *trustworthiness* we mean $V_{TW} + R_{TW}$, the sum of the endogenous and exogenous components of trustworthy behavior. V_{TW} depends on three key parameters:

$$V_{TW} = V\left(\rho_c, \mu, \sigma\right) = V\left(C\left(Q\right), \mu, \sigma\right) \tag{11}$$

where $V_1 < 0$, $V_2 > 0$, and $V_3 \lesseqgtr 0$. The first two derivatives need little explanation. If either the cut-off value falls or the mean rises, trustworthiness in the country will increase. The effect of σ, however, is more complex and we deal with it below.

Our main interest lies in in the effects of the parameters of the distribution of risk aversion on trustworthiness and, eventually, trust. In keeping with the conceptual matrix in Table 1, we consider values of the mean and standard deviation in the following ranges:

$$0 < \mu < 2.36$$

$$.25 < \sigma < 1.5$$

Figure 3 shows two extreme cases, one with the (μ, σ) pair of (2.36, 0.25), and the other with (0, 1.5). As noted above, the first of might represent a country like Japan, a *cautious* and *homogeneous* society; the second might describe the United States, which is considered to be more *entrepreneurial* and *diverse*. Now, imagine that both nations have the same institutions, represented by $Q = .135$. This would result in a cut-off value of $\rho_{c2} = 1.8$ in both countries. The more cautious and homogeneous nation – the one on the right – will have a larger share of its population of opportunists playing trustworthy than the risk-loving and heterogeneous nation.

We can gain insight into the effects of μ and σ on trustworthy behavior from Figure 4, which shows various *Iso*-V_{TW} lines. Every point represents a different country: to continue our example, Japan would be located in the lower right corner; the US in the upper left corner. Every line shows the combination of μ and σ that keeps V_{TW} equal to the indicated value, given the same value for Q.

First, consider the effect of μ on V_{TW}. As μ increases, given the value of Q

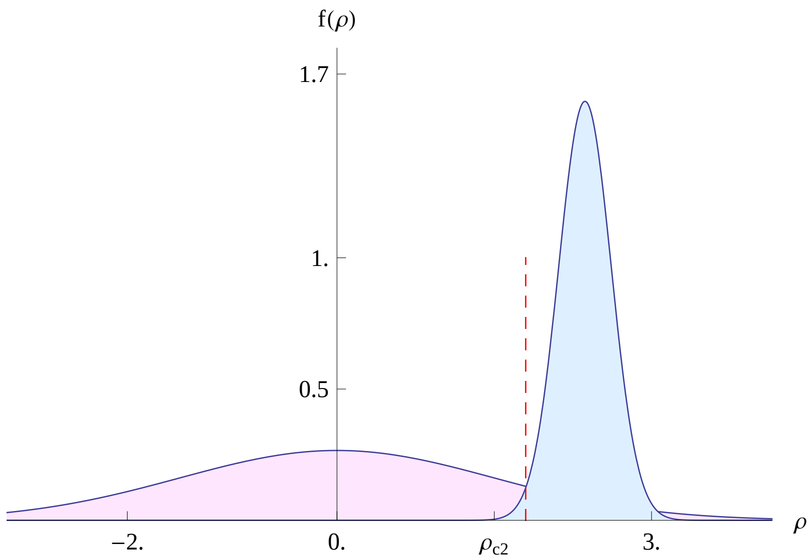

Figure 3: Extreme Densities of Caution

– and ρ_c – we see from Figure 4 that the proportion of opportunists playing trustworthy rises ($V_2 > 0$). That is, moving horizontally – from countries with low caution to those with high caution – we cross increasingly high values of *Iso*-V_{TW} curves. Greater caution corresponds to more trustworthy behavior. The standard deviation influences the *sensitivity* of this effect (lower σ makes the rate of conversion from untrustworthy to trustworthy faster), but the direction is always positive. This effect is completely general and does not depend on the values of the parameters chosen to illustrate it.[14]

The effect of σ is not as straightforward. As noted previously, an increase in σ may increase or decrease the proportion of opportunists who act trust-

[14]If we double both y_m and λ, our two independent parameters, Figure 4 is essentially unchanged: there is a slight increase in ρ_c but nothing else. The other parameters (α, β, and ϕ) are constrained by the rules outlined in Section 2 to guarantee that they remain in the necessary ranges.

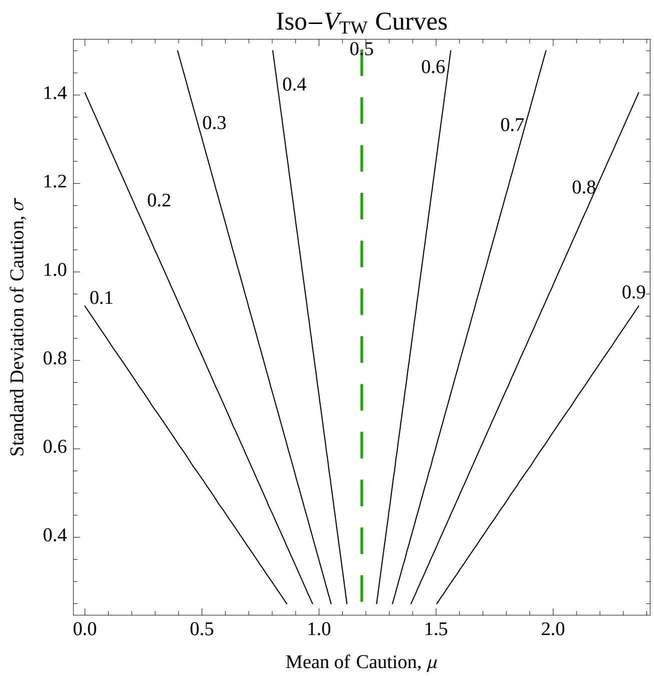

Figure 4: The Effect of μ and σ on Trustworthiness

worthy ($V_3 \lesseqgtr 0$). It depends only on the location of μ relative to ρ_c. To see this, start with the central case of $\mu = \rho_c$. Since the normal distribution is symmetric, exactly 50% of opportunists play trustworthy, no matter what the value of σ. An increase in the standard deviation will not alter the proportions of opportunists who play trustworthy and those who do not. This is represented by the vertical line above $\rho_{c1} = 1.18$ in Figure 4.

Now assume that $\mu < \rho_c$. In this case, V_{TW} rises as the standard deviation σ rises. Using Figure 4, as σ rises, we are considering increasingly heterogeneous societies. Heterogeneity leads to greater trustworthiness when risk aversion is low.

We may think of this phenomenon in the following way. When $\mu < \rho_c$, the mass of trustworthy opportunists is concentrated in the upper tail of the distribution. The central area and lower tail of the distribution are filled with opportunists who are less cautious and take a risk by cheating. Now, a higher σ increases the probability mass in the tails, while removing it from the center. This shift, in essence, pulls equal numbers of people away from the mean – who were cheating – and pushes them toward the upper and lower tails. Those cheaters who cross ρ_c change status and now play fair. People are pushed from the center toward the lower tail, too, but these opportunists do not change status: they continue to cheat. As a result, V_{TW} rises.

If $\mu > \rho_c$, then V_{TW} falls as σ increases. We see this on the righthand side of Figure 4: as we move vertically through the *Iso*-V_{TW} curves, their values diminish. Increasingly heterogeneous societies are *less* trustworthy now. We can make an argument based on the probability density as we did above: as σ increases, more people move away from the mean and cross over ρ_c to the lower end of the density. As they do so, more opportunists who were trustworthy become cheaters. Those remaining above ρ_c do not change status – so here the proportion who act trustworthy diminishes.[15]

Earlier we noted that, based on average caution and perceived homogeneity, Japan would be located in the lower right corner of Figure 4 and the US would be in the upper left corner. The figure suggests that Japanese citizens should demonstrate trustworthy behavior in larger numbers than their American counterparts. Based on four questions in the World Values Survey (2006) that ask about attitudes toward cheating, this does seem to be the case. In three of the four questions, the Japanese answered in a manner that indicated they were less willing to cheat. On the fourth, the two cultures tied.[16]

These results suggest that a full explanation of trustworthy behavior requires knowledge of both moments of these national distributions. In the next section, we build on these results to show how trust is determined.

[15] We also tried the uniform distribution and the log normal distribution (which is appropriate if one wishes to rule out the case of $\rho < 0$). Both worked very much like the normal.

[16] The questions are F114 – F117. They ask if it is "ever justified" to cheat in certain situations: claiming government benefits, riding a bus, paying taxes, or accepting a bribe. The tie occured on the bribe question.

5 Trust

5.1 The Initiator's Choice and Equilibrium Trust

Now return to the decision faced by person A, who is deciding whether or not to trust B. We assume that A knows R_{TW}, the proportion of possible partners who are intrinsically trustworthy. We also assume that she knows enough about the distribution of opportunists $f(\rho)$, the probability of getting caught Q, and the penalties to cheaters, ϕ and λ, to calculate the value of V_{TW} at any moment.

In order to concentrate on the influence of the distribution of caution, we adopt a simple view of the way that initiators are able to obtain information. In particular, every new match is random and no information is conveyed by the respondent, unlike in work of Tirole (1996) on collective reputations. Each initiatior is informed about the means R_{TW} and V_{TW} but of nothing else.

Initiators have utility function (1) just like opportunists. We assume throughout that *the initiator's distribution of caution is the same as that for the opportunists* $f(\rho)$. This is an important assumption and we make it primarily to establish this case as the starting point: there seems to be no good reason to differentiate the population of opportunists from the initiators by risk aversion distribution. Moreover, it seems reasonable that in any economy, people play both the role of agent A and agent B, depending on the circumstances. There is no reason why their risk distribution should change with their role.

If A does not trust, she gets a payoff of 1, which means a utility of $u(1,\rho) = 0$. If she does trust and initiate, she chooses a lottery ticket with the expected payoff:

$$\begin{aligned} E_i(U_T) = \quad & R_{TW}\, u(1+y_m,\rho_i) + \\ & (1-R_{TW}) \Big\{ V_{TW} u(1+y_m,\rho_i) + \\ & (1-V_{TW}) \Big[Qu(1+y_m-\alpha-\beta+\lambda) + \\ & (1-Q)\, u(1+y_m-\alpha-\beta) \Big] \Big\} \end{aligned} \tag{12}$$

As long as this expression exceeds 0, she will trust. The value of (12) is inversely related to the individual's level of caution ρ. For any set of parameters, there is a value of ρ that is so high that an individual will *not* trust. We call this critical value ρ_z. Individuals with a value of $\rho > \rho_z$ are too afraid ("cautious") to risk getting cheated.

We derive the threshold value ρ_z by setting (12) equal to zero, and regarding the result as a relation that defines ρ_z as the implicit function:

$$\rho_z = Z(R_{TW}, V_{TW}) \tag{13}$$

where $Z_1 > 0$, $Z_2 > 0$.[17] The threshold – and therefore the overall level of trust – rises if there are more intrinsically trustworthy people R_{TW} or if

[17] We ignore y_m, α, β, and λ – even though they affect this decision – to better focus on the risk Q and the parameters μ and σ.

opportunists are more likely to play fair V_{TW}. Although R_{TW} is exogenous, V_{TW} is given by Equation (11).

To find the threshold numerically, we need to set the value for R_{TW}:

- $R_{TW} = .10$

Given this value and the other baseline parameters we specified earlier, we find that the threshold for trust is $\rho_{z1} = .16$ (see Figure 2). All initiators with a ρ value *below* ρ_z will trust; they are sufficiently brave to accept the gamble of being cheated. This is the the area under the curve to the left of the dashed line above ρ_{z1} in Figure 2. We calculate this to be 18.8%.

Call the fraction who trust P_T. From the discussion above, we may infer that:

$$P_T = \Pi\left(\rho_z, \mu, \sigma\right) \tag{14}$$

where ρ_z is given by (13). In this expression, $\Pi_1 > 0$ and $\Pi_2 < 0$, and $\Pi_3 \lesseqgtr 0$. We discuss the separate effects below.

5.2 Comparative Differences

There are four exogenous influences on a society's level of trust: Q, R_{TW}, μ, and σ. The first of these is institutional and the other three are cultural.

5.2.1 Institutional Differences

The only institutional factor we have considered is the probabilty that a cheater will be caught. Consider several societies with the same distribution of caution, but different institutions. Those with higher levels of Q will have lower values of ρ_c by (9), which increases the fraction of opportunists who play honestly V_{TW} by (11). This, in turn, leads to higher values of ρ_z by (13) and then to P_T by (14). A greater chance of punishment makes more opportunists honest, and this perception of trustworthy behavior leads initiators to trust more.

5.2.2 Innately Trustworthy

Countries with a larger share of innately trustworthy types R_{TW} have higher levels of trust. The mechanism is straightforward and works by increasing the expected success rate for any initiated transaction. Societies with high R_{TW} have a high cut-off ρ_z by (13), which, other things the same, generates more trust.

5.2.3 The Moments of the Distribution of Caution

We now address the more complex relationship between trust and the moments of the distribution of risk behavior. In general, μ and σ have two effects, a *direct* effect and an *indirect* effect on trust. By "direct effect", we mean the effect on P_T holding the cut-off ρ_z fixed; that is, the last two partial derivatives in (14). Trust is complicated by the "indirect effect": the effect of μ and σ on *player B's trustworthiness and therefore on the initiator's perceptions about the trustworthiness of others.* These work through V_{TW} in (11), which then affect ρ_z – see (13) – which, in turn, affects trust in (14).

The indirect effect thus follows the following chain of causation:

$$(\mu, \sigma) \rightarrow V_{TW} \rightarrow \rho_z \rightarrow P_T \tag{15}$$

We begin our discussion with the effect of the mean on trust.

The Effect of the Mean

The direct effect of the mean is the easiest to understand since it refers to the partial derivative of μ in (14), which is always negative. In countries with higher values of μ, given ρ_z, *initiators* are more cautious on average, so they trust *less*. A higher mean of risk aversion directly reduces trust. This effect is often pointed out in the literature to suggest that cross-country empirical measures of trust may really be measuring caution in those cultures.

A high value of μ, however, also means that more *opportunists* are cautious so a random match is more likely to be trustworthy. With greater trustworthy behavior, the cut-off value ρ_z is higher (see Figure 2). This raises initiatiors' beliefs about the trustworthiness of others and therefore their willingnesss to trust. This is the *indirect effect* of a rise in μ on trust. The indirect effect may well dominate the direct effect so that the net impact of μ on P_T may be positive. High-caution societies may trust *more* than those that are less cautious.

For different values of μ the dominant effect depends on the standard deviation of the distribution of risk, σ. At low values of σ, the indirect effect

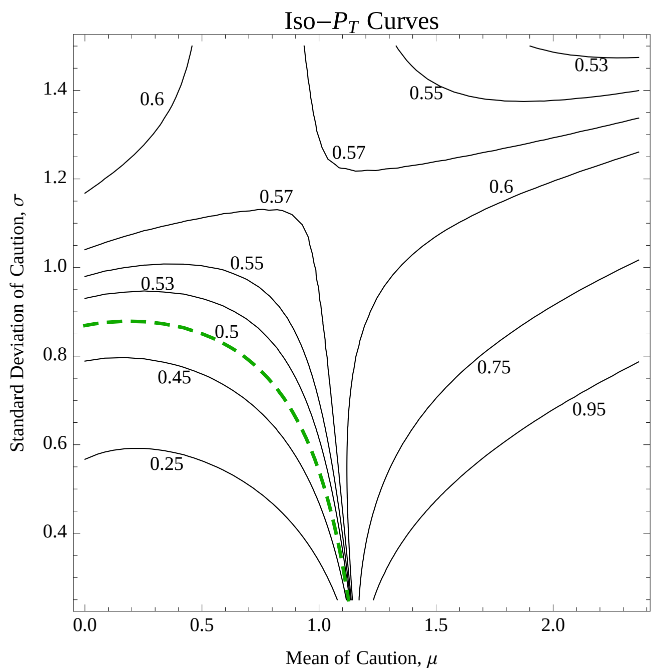

Figure 5: The Effect of μ and σ on Trust

on P_T of a rise in μ is stronger, so trust *rises* as μ increases.[18] At high values of σ, the direct effect dominates and so trust declines as μ increases. We illustrate these effects using the set of $Iso - P_T$ curves in Figure 5. We highlight the contour where $P_T = .50$. Above this contour, trust exceeds 0.50; beneath the contour, trust is less than 0.50.

Consider what happens to trust when the standard deviation is relatively low, say $\sigma = .40$, and μ increases. We see that in societies with higher values of μ – moving horizontally across the graph – there is more trust. The indirect, positive effect dominates, and is powerful: for any increase in μ, the effect on P_T is quite strong. If, on the other hand, $\sigma = 1.4$, the direct,

[18]At very low values of μ, the net effect is essentially zero.

negative effect of μ on P_T dominates, but it is weak: it takes a large increase in μ to provide a modest decrease in P_T.

Why does the standard deviation matter so much? If σ is large, then as μ increases there are relatively few new opportunists playing fair, and the trusters' enhanced caution dominates the perception of a more honest environment. In economies with a low value of σ, however, the opposite occurs: ρ_z is quite sensitive to an increase in μ – so an initiator's augmented caution is overcome by more trustworthy behavior – the perception of which raises trust.

The Effect of the Standard Deviation

Examination of Figure 5 shows that in societies with low μ, greater variability of caution – a vertical movement through the *Iso-*P_T curves – results in greater trust. On the other hand, in those societies where μ is high, greater variability of caution reduces trust. Although we may still think in terms of the direct and indirect effects, there is no simple way to explain how they interract. Even the direct effect of σ is ambiguous ($\Pi_3 \lesseqgtr 0$).

The pattern in Figure 5 is robust, however, to two kinds of changes. When we use either a uniform distribution or a log normal distribution (both with the same baseline μ and σ), the results are very similar. This is also true when we double the values of y_m and λ (keeping the other parameters within their permitted ranges).

5.3 Trust and Caution

We argued earlier that Japan would be located in the lower right corner of Figure 5, based on its high average caution and relative cultural homogeneity. The US would be located in the upper left corner, assuming that it is less cautious and more diverse. Given the values of the $Iso - P_T$ contours in Figure 5, we expect $P_T^{US} < P_T^J$: there would be less trust in the US than in Japan. Using the standard trust question from the World Values Survey (2006), this does appear to be the case. In the US, 36% answered the trust question affirmatively as compared to 43% in Japan.

6 Aggregate Analysis

We began this paper noting that there have been many studies linking trust to economic outcomes like per capita output. In this section, we show what our model implies about aggregate economic activity.

Assume that there are N encounters every period between a player A and a player B. Total output, on average, is governed by the payouts discussed in Section 2 and will depend on the proportions of the individuals who trust, who are innately trustworthy, and who behave trustworthily, as well as the losses and gains accruing to A and B. Total output is:

$$Y = (1 - P_T)\,2N + P_T N \Big\{ R_{TW} 2\,(1 + y_m) +$$
$$+ (1 - R_{TW}) \Big[V_{TW} 2(1 + y_m)$$
$$+ (1 - V_{TW}) \big\{ Q\,[2(1 + y_m) - \beta - \phi] +$$
$$+ (1 - Q)\,[2(1 + y_m) - \beta] \big\} \Big] \Big\} \quad (16)$$

Divide both sides by N and simplify to find an expression for *income per encounter* y:

$$y = 2 + P_T\,[2y_m - (1 - R_{TW})\,(1 - V_{TW})\,(\beta + Q\phi)] \quad (17)$$

where P_T is given by (14) and V_{TW} by (11).[19]

When the probability of punishment is extremely high, honest behavior and trust will prevail: $V_{TW} = P_T = 1$. From (17), this means output per capita $y = 2 + 2y_m = 8.0$, given our baseline value of $y_m = 3.0$. If, on the other hand, there were no trust ($P_T = 0$) then output would only be 2.0 units. In this sense, the model assigns a fourfold increase in output to trust. This difference is purely arbitrary, in the sense that it does depend on our baseline calibration. If we doubled the values of y_m and λ – and adjusted the other parameters so that the constraints (3) and (4) were still satisfied – the

[19]A version of this equation appeared in Breuer and McDermott (2008), who used it to estimate the relative importance of trust and trustworthiness across countries. Notice that because α and λ are transfers between A and B, they do not affect aggregate output.

maximum effect of trust would be 7 to 1. However, we are most interested in the *relative* effect on y of changing the exogenous parameters Q, R_{TW}, μ, and σ, not on the overall effect of trust.

We first consider the link between institutions and economic outcomes. Although Q enters negatively in (17), this is a partial effect that only accounts for the reduction in y following the event of capture – the deadweight loss ϕ. The indirect, positive effects of Q working through V_{TW} and P_T are actually stronger so that the total effect is positive.[20] Better institutions, in the form of a greater rate of conviction for crime, raise output per person by increasing the number of completed transactions.

A greater fraction of natively trustworthy agents R_{TW} necessarily increases aggregate output. This follows from (17), using (13) and (14). Both the direct effect and the two indirect effects – in raising V_{TW} and P_T – work in the same direction to increase the number of transactions.

Finally, we consider the effect of the distribution of caution on y. In Figure 6 we show a series of *Iso-y* curves in (μ, σ) space. Holding σ constant, countries with high levels of caution μ always have greater levels of y. This means that the direct, negative effect of making initiators trust less is never the dominant effect. Higher average caution gives countries higher levels of V_{TW} and P_T, which means more transactions and greater numbers of those concluded honestly. Output is greater in these countries. However,

[20]This result was robust to other assumptions about the distribution of risk aversion and the values for y_m and λ.

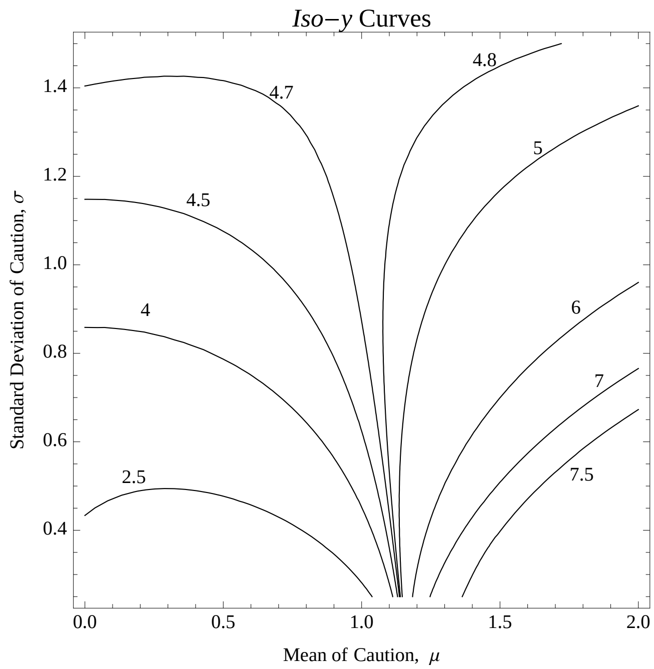

Figure 6: The Effect of μ and σ on y

in societies with a large standard deviation σ, the effect of μ in raising y is extremely weak. When $\sigma = 1.4$, for example, it takes a near tripling of μ to raise y from 4.7 to 5.0.

The effect of the standard deviation of caution on output per capita depends on the level of μ. For nations with a low value of μ – the ones we are calling entrepreneurial – those with the highest σ will also have the highest y. Societies that have a high value of μ – cautious societies – demonstrate the reverse: among these, the ones who have a high σ will have lower living standards.

What can we say, in the end, about trust and economic performance? In our model, both are endogenous, so the relationship is not monotonic.

Consider two countries with a high σ (say, 1.5 for both) but different values for μ: Country 1 is imprudent or entrepreneurial – has a low μ – while Country 2 is cautious – has a high μ. As we have just seen, y is higher in Country 2 than in Country 1 (see Figure 6). Trust, however, is larger in Country 1 than in Country 2 (Figure 5). We would observe an inverse relationship between trust and output in a sample of internally heterogeneous countries. Had we taken groups of nations with low values for σ, we would have found a direct relationship between output and trust across countries.

7 Conclusion

In this paper we constructed a theory of trust based on heterogeneity in the risk aversion preferences of the population, which we consider to be a deep cultural attribute. We showed that attitudes toward risk affect the desire to be trustworthy and the willingness to trust. Our results call into question the heuristic that more cautious societies trust less. Although this partial effect is present, it is often dominated by an indirect effect. Greater average caution will cause people to be more trustworthy, which prompts cautious initiators to overcome their hesitancy to trust. We showed that the net effect of caution on trust depended systematically on the standard deviation of caution. If there is greater homogeneity in attitudes toward risk in a society – a lower standard deviation – the indirect effect is more likely to dominate and more caution will lead to more trust.

Trustworthiness is instrumental in building trust in society. Society's behavior toward risk, as well as its institutions, contributes to trustworthiness, since acting honestly is, to some degree, a matter of balancing return and risk. We showed that output per capita is always positively related to trustworthiness. Trust and output, however, are not monotonically related. It is possible that in culturally diverse societies where risk attitudes vary widely, higher trust is associated with lower standards of living.

References

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

Ben-Ner, A. and Putterman, L. (2008). Trust, communication and contracts: An experiment. Working Papers 0206, Industrial Relations Center, University of Minnesota (Twin Cities Campus).

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142. available at http://ideas.repec.org/a/eee/gamebe/v10y1995i1p122-142.html.

Breuer, J. B. and McDermott, J. H. (2008). Trustworthiness and Economic Performance. *SSRN eLibrary*.

Buchan, N. (2009). The complexity of trust: cultural environments, trust, and trust development. forthcoming in the Handbook of Cultures, Organizations, and Work.

Coleman, J. S. (1988). Social capital and the creation of human capital. *The American Journal of Sociology*, 94(1):95–120.

Drago, F., Galbiati, R., and Vertova, P. (2009). The Deterrent Effects of Prison: Evidence from a Natural Experiment. *Journal of Political Economy*, 117(2):257–280.

Eckel, C. C. and Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization*, 55(4):447–465.

Fehr, E. (2008). On the economics and biology of trust. IZA Discussion Papers 3895, Institute for the Study of Labor (IZA).

Frank, R. H. (1987). If homo economicus could choose his own utility function, would he want one with a conscience? *American Economic Review*, 77(4):593–604. available at http://ideas.repec.org/a/aea/aecrev/v77y1987i4p593-604.html.

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. Free Press.

Gambetta, D. (1988). *Can We Trust Trust?* Basil Blackwell. Reprinted in electronic edition from Department of Sociology, University of Oxford, chapter 13, pp. 213-237".

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846. available at http://ideas.repec.org/a/tpr/qjecon/v115y2000i3p811-846.html.

Harrison, G. W., Johnson, E., McInnes, M. M., and Rutstrom, E. E. (2005). Risk aversion and incentive effects: Comment. *American Economic Review*, 95(3):897–901.

Harrison, G. W., Lau, M. I., and Rutström, E. E. (2007). Estimating risk attitudes in denmark: A field experiment. *Scandinavian Journal of Economics*, 109(2):341–368.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Holt, C. A. and Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, 95(3):902–912.

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*, 95(5):1688–1699.

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly Journal of Economics*, 112(4):1251–88. available at http://ideas.repec.org/a/tpr/qjecon/v112y1997i4p1251-88.html.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1997). Trust in large organizations. *American Economic Review*, 87(2):333–38. available at http://ideas.repec.org/a/aea/aecrev/v87y1997i2p333-38.html.

Miller, A. S. and Mitamura, T. (2003). Are surveys on trust trustworthy? *Social Psychology Quarterly*, 66(1):62–70.

Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural paraguay. *Journal of Economic Behavior & Organization*, 62(2):272–292.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, 6(4):317–344.

Tabellini, G. (2006). Culture and institutions: economic development in the regions of europe. Levine's Working Paper Archive 321307000000000241, UCLA Department of Economics. available at http://ideas.repec.org/p/cla/levarc/321307000000000241.html.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3):905–950.

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):1–22.

World Values Survey (2006). European and world values survey four-wave integrated data file.

Zak, P. J. and Knack, S. (2001). Trust and growth. *Economic Journal*, 111(470):295–321.