The Neuroeconomics of Trust*

By

Paul J. Zak
Center for Neuroeconomics Studies
Department of Economics
Claremont Graduate University
and
Department of Neurology
Loma Linda University Medical Center

Forthcoming in: Two Minds. Intuition and Analysis in the History of Economic Thought, Roger Frantz, editor. London: Springer.

* This work would not be possible without the many outstanding collaborators and graduate students with whom I have had the privilege of working. I also thank Jang Woo Park for comments on an earlier draft, and Roger Franz for the invitation to contribute this chapter.

1. Introduction

You must trust and believe in people or life becomes impossible.

Anton Chekhov

The traditional view in economics is that individuals respond to incentives, but absent strong incentives to the contrary selfishness prevails. Moreover, this "greed is good" approach is deemed "rational" behavior; without extreme self-interest, the standard models predict that money will be left on the table during a transaction and therefore an equilibrium cannot have been reached. For example, standard principal-agent models predict that absent monitoring, employees will shirk to the extent possible since working is presumed to produce a negative utility flow. Nevertheless, in countless firms on every day of the week, employees labor away without overt monitoring; for example, those who telecommute. This is not to say that some shirking does not occur, but that human beings behave a bit differently than in models of "rational economic agents" for reasons that are not well-understood (though many possible explanations have been advanced, see Camerer, 2003).

Similarly, a substantial body of research has examined variations in efficiency by individuals within a firm, called X-inefficiency (XE) by Leibenstein (1966) (reviewed in Franz, 1997; and Franz, this book). The consensus view is that XE arises from bounded rationality and psychological motives that militate against efficiency. One way to reduce XE is to provide incentives for individuals to behave "more rationally." Unfortunately, how this is done (and what it even means) is difficult to identify. Nevertheless, many estimates of the degree of XE are moderate (Franz, this book), suggesting that employees, most of the time, are reasonably efficient.

A third example of a failure of the fully rational agent model is the degree of cheating during intertemporal transactions with asymmetric information (Zak & Knack, 2001). Choosing a money manager or investment advisor to invest on one's behalf typically results in an informational asymmetry regarding subsequent returns. The investor can estimate returns imperfectly because the type and timing of each transaction may difficult to establish, while the advisor knows the actual return but may not to report it's true value to the investor. While this problem exists, casual observation suggests that, for a given institutional setting, "most" investors do not appear to be cheated, at least not grossly, although spectacular exceptions have been widely reported. Indeed, Zak & Knack (2001)

demonstrate that low rates of investment occur primarily because of weak formal and informal institutions that inadequately enforce contracts. Nevertheless, even in institutional environments that do not enforce contracts well, a substantial number of investments still occur, presumably without undo duress, suggesting that some or even many money managers are reasonably trustworthy (or, alternatively, that investors are poor monitors of advisors, but since investors have a substantial incentive to monitor, this explanation is unlikely).

A possible explanation for the substantial amount of "irrational" behavior observed in markets (and elsewhere) is that humans are a highly social species and to an extent value what other humans think of them. This behavior can be termed trustworthiness—cooperating when someone places trust in us. Indeed, we inculcate children nearly from birth to share and care about others. In economic nomenclature, reciprocating what others expect us to do may provide a utility flow itself (Frey, Matthias & Stutzer, 2004). Loosely, it is possible that it "feels good" to fulfill others' expectations in us. If such a cooperative instinct exists, it must be conditioned on the particular environment of exchange, including the history of interactions (if any) with a potential exchange partner. If conditional cooperation were not the case, individuals would be gullible, and the genes that code for gullibility would not have survived over evolutionary time (Boyd et al. 2003).

Instead, conditional on the parties involved in trade, budget and time constraints, and the social, economic, and legal institutions in place, individuals may exhibit high degrees of cooperation or nearly complete selfishness. This leads one to ask which institutional arrangements promote or inhibit trustworthiness. A second question is, for a fixed institutional environment, what are the mechanisms that allow us to decide who to trust, and when to be trustworthy? Relatedly, for a given institutional setting, why is there variation among individuals if the incentives to trust or be trustworthy are identical?

This chapter sketches a neuroeconomic model of trust and provides several forms of evidence in support of this model. Neuroeconomics (Zak, 2004) is an emerging transdisciplinary field that utilizes the measurement techniques of neuroscience to understand how people make economic decisions. This approach is of particular interest in studying trust because subjects in a laboratory who can choose to trust others and be trustworthy are unable to articulate why they make their decisions. Taking

neurophysiological measurements during trust experiments permits researchers to directly identify how subjects make decisions even when the subjects themselves are unaware of how they do this. Readers are referred to Zak (2004) for a full description of neuroscientific techniques used to measure brain activity. These tools open the black box inside the skull and provide radical new insights in economics. Trust is among the most interesting of the topics being studied.

2. Institutions, Public Policy, and Generalized Trust

Generalized trust is defined as the probability that two randomly chosen people will trust each other in a one-time interaction. Evidence for generalized trust across different institutional settings can be obtained from the World Values Survey (and its imitators). Figure 1 plots the proportion of those who answered yes to the question "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" The data show an order-of-magnitude variation, with 3% of Brazilians and 5% of Peruvians responding affirmatively, while 66% and 60% of Norwegians and Swedes, respectively, asserting that others can be trusted.

There is a simple explanation for trustworthiness during repeated bilateral interactions. The Folk Theorem states that cooperative behaviors can be sustained when there are mutual gains from cooperation as long as repeat interactions occur for an indefinite future. This explanation is problematic when analyzing generalized trust because many transactions occur only once, or repeat for only a finite number of times. Why in these settings do people still trust one another?

When I began to investigate how individuals decide when to trust and be trustworthy in 1998, I was surprised to find there was very little written about trust by economists. Psychologists have studied trust, but this literature focused more on individual attributes rather than on the setting of particular interactions. The magnitude of the variation in the data in Figure 1 strongly suggested to me that trust varied not because Brazilians were different from Norwegians, but because the setting in which interactions took place was different. Because I have a background in biology, I searched this literature and discovered a rich set of findings I could draw upon to build a biologically-consistent, or bioeconomic (Zak & Denzau, 2001) model of trust.

The resulting paper, Zak & Knack (2001), built a dynamic general equilibrium model based on Hamilton's Rule from evolutionary biology that identifies how much one is expected to care about another's welfare as a function of the proportion of shared genes. It extends Hamilton's Rule to account for how variations in exchange environments affect the likelihood that one's transaction partner will be trustworthy when information is asymmetrically distributed and contracts are costly to enforce. The model shows that the degree of generalized trust in a country is inversely related to the transactions costs associated with enforcing an investment contract. In particular, trust depends on the *social environment* (how similar or dissimilar are those in a transaction; for example, think of the high degree of ethnic homogeneity in Norway, and how strongly social norms are enforced); the *legal environment* (how effectively contracts are enforced by formal institutions; for example, how readily redress can be obtained if one party of to the transaction believes that he or she obtained an unfair outcome); and the *economic environment* (as incomes rise, people will behave as if they trust others more because their time cost to investigate their trading partner rises; as income inequality rises, it is more likely that one's trading partner will be untrustworthy because differences between parties to exchange, and therefore incentives to cheat, are greater).

The extensive empirical tests done by Zak & Knack (2001) show that the exchange-environment variables identified in the theoretical model explain 76% of the variation in the cross-country trust data plotted in Figure 1. It also shows that societies that are less heterogeneous (in income, language, ethnicity, etc.) have higher trust because social ties between parties who are similar informally enforce contracts. For similar reasons, societies that are fair (have less economic discrimination) have higher trust. Alternatively, sufficiently strong formal institutions that enforce contracts can promote high levels of trust even in highly heterogeneous societies like the U.S. Lastly, the economic positions of trading partners affect the degree to which they will trust others and be trustworthy.

The Zak & Knack model shows that trust is directly related to economic growth by reducing transaction costs and facilitating investment. Empirically, trust is among the powerful factors economists have discovered that promote growth. The analysis in Zak & Knack (2001) shows that a 15 percentage point increase in the proportion of people in a country who think others are trustworthy raises income per person by 1% per year *for*

every year thereafter. For example, if trust in the U.S. increased from its current level of 36% to 51%, average income would rise by about $400 per year thereafter due to the additional business investment and job creation. The impact of trust on living standards is quantitatively large; $400 per year corresponds to an additional $30,000 in average lifetime income.

Zak & Knack (2001) also show that if trust is sufficiently low (below 30% for the average country in Figure 1), then the investment rate will be so low that living standards will stagnate or even decline. This "poverty trap" is primarily due to ineffective formal institutions that result in low levels of generalized trust. The model shows that the threshold level of trust necessary for positive economic growth is increasing in per capita income. As a result, it appears to be difficult to escape from a low-trust poverty trap without outside intervention.

In a sequel paper, Knack & Zak (2003) asked if there were cost-effective policies that governments could implement to raise trust levels. Cost-effective policies were defined as those that produce a greater increase in income by raising trust (which raises investment) than they cost to implement. Knack & Zak found that many policies are able to raise trust, and some do so by affecting multiple aspects of the environment of exchange identified in Zak & Knack (2001). For example,

- *education* has three effects: increasing the quality of formal institutions that enforce contracts, decreasing income inequality, and directly raising trust by raising incomes;
- *Press freedoms* and *civil liberties* increase the quality of civil institutions and thereby trust;
- *Telephones* and *roads* directly raise trust by increasing social ties between interacting parties; and
- *Income transfers* reduce inequality and thereby raise trust.

The analysis in Knack & Zak (2003) shows that levels of generalized trust can be affected by public policy. Unfortunately, few of the policies examined were cost-effective. Note that the determination of cost-effectiveness included only the effect of the policy on trust itself and in this way on incomes. These calculations therefore underestimated the true

benefits of each policy. For example, a new road may raise trust by increasing social interactions, but it also has a direct effect on growth by reducing the cost of getting goods to market; the latter is ignored in the foregoing analysis to focus solely on trust-based growth policies. This narrow view of cost-effectiveness was chosen to see if trust-based development policies existed

Two policies unambiguously increase incomes by raising trust more than they cost to implement: education and income transfers. The former occurs because of the three ways that education raises trust, producing a nearly 500% average return on the cost of paying for an additional year of education for the countries depicted in Figure 1. Surprisingly, income transfers produce an approximately 50% return by raising trust, taking into account administrative costs (it costs roughly one dollar to transfer one dollar). This does not account for possible disincentive effects from transfers and likely is driven by the very low trust among countries with very unequal income distributions. A third factor, freedom, was found to have a powerful effect on trust by increasing the number of social interactions and making institutions and individuals more accountable. Unfortunately, there is no agreed upon way to determine the cost of freedom. As a result, the cost-effectiveness of policies that, for example, increase press freedoms, are difficult to determine.

Because of my interest in the biological factors that drive trustworthy behavior, I investigated whether biological factors directly impact generalized trust (Zak & Fakhar, 2005). Using a large set of data on biological environments across countries, using a theory (described in Section 3 below) that neuroactive hormones "guide" humans as to when they should be trustworthy, Fakhar and I found omnibus variables that were related to generalized trust. We built these variables using factor analysis using the high degree of correlation between related environmental measures. Two factors, *ecological* and *phyto*, were statistically related to trust. *Ecological* measures pollution in the physical environment of exchange. It is dominated by measures of "xenoestrogens" or synthetic estrogen-mimics (such as the pesticide DDT), and is strongly negatively related to trust. *Phyto* is an index of phytoestrogen consumption. Phytoestrogens are plant-based estrogens found in soybeans, legumes, wine, tea and many other foods, and we find they are strongly positively related to trust. This is consistent with findings from biology showing that estrogens affect social behaviors. The correlations Fakhar and I found maintain statistical significance when

income is controlled. Interestingly, these biological factors are orthogonal to the institutional factors that Zak & Knack (2001) show affect trust. Thus, the biological environment represents a distinct pathway that affects the likelihood that others will be trustworthy.

The results of Zak & Knack (2001), Knack & Zak (2003), and Zak & Fakhar (2005) demonstrate that the likelihood of two individuals who do not know each other exhibiting trust depends crucially on the social, legal, biological, and economic environments. This extends the narrowly rational models in economics by showing that although people respond to incentives, they do so without making consciously deliberated decisions. These analyses do not provide evidence that individuals respond *similarly* to changes in institutions. Nor do they address the mechanisms through which one person decides to trust another because of the level of aggregation. We turn to these issues next.

3. Experimental Findings

This section surveys a variety of experimental studies that support the thesis that human beings are "wired" to be conditionally cooperative. President Abraham Lincoln said "... people, when rightly and fully trusted, will return the trust," A substantial number of behavioral experiments by economists and psychologists have characterized the high degree of trust and trustworthiness in the laboratory consistent with Lincoln's view that humans tend to reciprocate trust. A typical experimental task to investigate trust and trustworthiness is the "trust game" (Berg, Dickhaut & McCabe, 1995). All the experimental evidence presented here uses variants of this game so its structure is presented in detail. Subjects (typically students) are recruited for an experiment and all those who show up at the laboratory receive $10 for agreeing to participate for an hour to an hour and a half. It is important that subjects' identities are masked so that neither other participants nor the researchers can associate a particular person with his or her choices. Otherwise, subjects may change their choices to "please" experimenters or avoid confrontations with other participants. For example, Smith (1998) discusses the substantial increase in cooperative behaviors in games run single vs. double blind show.

The game is fully described to participants prior to play, usually through a series of examples. There is strong ethic in experimental economics to avoid deception, and

experiments using this game nearly always follow this norm. Subjects are then randomly assigned to dyads. Within each dyad, subjects are randomly given the role of decision-maker 1 (DM1) or decision-maker 2 (DM2). DM1 then is prompted (often via software, but sometimes using written instructions) to send an integer amount (including zero) of his or her $10 show-up earnings to the DM2 in his/her dyad. Both subjects are instructed that whatever DM1 sends to DM2 is deducted from DM1's account and *tripled* in DM2's account. For example, if DM1 sends $8, he or she keeps $2, and DM2 then has $34 (=$8×3 + $10 show up amount). DM2 is then told how much DM1 sent him/her and the total in his/her account, and then is prompted to send some integer amount (including zero) back to DM1.

Subjects are informed before the experiment begins that they will (typically) make a single decision after which the interaction ends (a variant is having DM1 and DM2 make N decisions with N different individuals). The single decision structure controls for the possible effects of reputation that can sustain trust. Providing an endowment to both DM1 and DM2 reduces the incentive for subjects to make transfers to equalize earnings within a dyad. Finally, the show up amount is typically emphasized as being paid to compensate participants for spending an hour in the lab so they don't view this as gambling with "house money."

The consensus in the literature is that the transfer from DM1 to DM2 is a (costly) signal of trust. The mostly likely reason that DM1 would sacrifice some or all of his/her show up earnings is to indicate to DM2 that the "pie just got larger based on my sacrifice." There is an expectation that DM2 understands this and will act accordingly by sharing the larger pie. Nevertheless, there is no guarantee that DM2 will return any money, and no external enforcement mechanism. Thus, DM1's choice reflects his or her view of the human predilection for reciprocity. The return transfer from DM2 to DM1 is commonly viewed as a measure of trustworthiness (or reciprocity). To be trustworthy in this game entails a 1:1 dollar cost to DM2. Subjects know that the transfer from DM2 to DM1 is not tripled, and each dollar sent comes out of DM2s account. It is the costliness of the choices that make this an interesting way to quantify trust and trustworthiness. It also captures the notion that individuals trust each other because there is potentially mutual benefit.

The subgame perfect (SGP) Nash equilibrium for this game is found by iterating backwards. If DM2 prefers more money than less, then he/she will keep everything DM1 sends. DM1, anticipating this, is predicted to send nothing to DM2. Although the SGP Nash equilibrium predicts no trust and no trustworthiness, this is at odds with the data from the large number of times this experiment has been run, including for stakes up to $1,000 in the US and for three months average salary in developing countries (Smith, 1998; Camerer, 2004). Typically, three-quarters of DM1s will send some money to DM2s, and an even higher proportion of DM2s return some money to DM1s. Indeed, in the experiments run by my lab, which are typical of findings from other labs, DM1s who exhibit trust leave with approximately $14, or 40% more than their $10 show up earnings. DM2 average earnings are even more, about $17, because they are typically trusted by the DM1 in their dyad but do not equally share the largess.

There is clearly a problem with the SGP Nash equilibrium in this game since those who play out of equilibrium earn more money. Though John Nash did not directly analyze the trust game (which is a sequential-play prisoner's dilemma), his well-publicized illness reveals why the SGP Nash equilibrium concept does not apply here. As most people know, John Nash suffers from the neuropsychiatric disorder schizophrenia. Schizophrenics are typically socially withdrawn, and analogously the SGP Nash equilibrium for the trust game does not recognize that the game is embedded in a social interaction. DM2s nearly always return some money to DM1s because of the social obligation incurred by the sacrifice made by DM1 to signal trust. Put differently, DM1s appear to make transfers using their understanding of the typical human behavior that follows when someone does something "nice" for you; that is, you are obligated to return the favor. This does not always happen, but it nearly always does: in my experiments, roughly 90% of DM2s return at least some money to the DM1s they are paired with.

In contrast, "economic man" has no social conscience and therefore plays the SGP Nash strategy. One such subject appeared in a recent experiment in my lab. Some of my experiments (described in detail below) involve blood draws. Subjects in these experiments know before they participate that the $10 show up earnings compensate them for spending up to 1.5 hours in the lab and for the needle stick(s) and four tubes of blood we will take from them. Once subjects have made their choices regarding the degree of trust or

trustworthiness and we have obtained blood from all of them, the subjects leave the lab and we centrifuge the tubes and extract out plasma and serum for analysis. Each experiment session has 16 to 20 subjects, so when blood is obtained from the final subject, there is a rush to begin processing up to 80 tubes of blood. There are some subjects from whom it is difficult to obtain blood, for example, those with very small veins, and those with a layer of fat covering the veins. The subject in question was a chubby male and the phlebotomist had to stick him four times before a vein was found. Meanwhile, I and my graduate students were hovering, ready to get to work. After we collected his blood, I apologized to him for the multiple needle sticks and thanked him for participating. He said he was elated to be in the experiment and asked if he could return for another session (no). "Elation" is not what I had ever observed for subjects who suffer through four needle sticks. Out of curiosity, I checked his behavioral data since I knew he was the last subject in that session. This participant was a DM2 who had had maximal trust placed in him by the DM1 in his dyad (the DM1 had sent his/her entire $10 show up earnings to him and kept nothing). Nevertheless, this DM2 was completely untrustworthy, being unwilling to share any money with the person who had trusted him. He left the lab with the maximal earnings of $40 (=$10*3 + $10). This was economic man, and he unabashedly played the SGP Nash strategy. Of course he enjoyed this experiment!

Although this anecdote illustrates that untrustworthy economic men (and women, see below) do exist, a mystery in the literature before I began running my experiments was why there were so *few* untrustworthy subjects in the trust game (Smith, 1998). If this mystery could be solved, it would likely identify the mechanism through which people decide to be trusting and trustworthy. Then such a mechanism could be manipulated, for example, by designing exchange environments that utilize it. This led me to think that there might be a *physiologic* mechanism that motivates subjects to be trustworthy. Now I'll explain the blood draws.

As discussed above, experimental subjects are unable to describe why they make trusting decisions, so if a physiologic process was driving trustworthiness, it would have to work below the level of conscious thought—very much counter to the model of the thoughtful economic agent. I was unable to find evidence of such a mechanism in the experimental economics and psychology literatures, but there were some hints in the

neuroscience literature studying rodents. Some rodent species are highly sociable, living in groups, and often forming long-term pair bonds where both males and females care for offspring. For example, the prairie vole, a rodent living in the Midwestern US, exhibits these behaviors. Interestingly, a genetically and geographically closely related species, the montane vole, shows none of these traits—males are solitary, promiscuous, and avoid their offspring. These behaviors were first studied in the 1980s by several labs, and by the 1990s the consensus in this literature was that these pro-social behaviors were the result of a hormone called oxytocin (OT). OT has target receptors in both the peripheral organs and in the brains of mammals. I wondered if the trusting behaviors in the lab were being caused by OT. In other words, I hypothesized that strangers in the trust game may have been forming temporary "attachments" to each other, much as OT causes attachment in prairie voles. Unfortunately, the distribution of OT receptors is not well conserved across species, so extrapolation from voles to human behavior was only speculative. An experiment was needed.

Prior to my experiments, the behavioral effects of OT had been little studied in humans. This is primarily because OT is medically uninteresting unless a women is giving birth or breastfeeding. Oxytocin means "fast birth" in Greek and this hormone contracts the uterus during parturition; women often get synthetic OT (drug trade name: pitocin) to speed up birth. It also promotes the release of breast milk. I set up an experiment to test if OT rose when subjects received a signal of trust and motivated subjects reciprocate and be trustworthy.

My collaborators and I used an anonymous one-shot trust game to see if even in the starkest case of (mostly) depersonalized one-time exchange OT mediated trustworthiness. Trust games played face-to-face show nearly 100% trustworthiness, removing the behavioral variation, so we did not use this approach even though the rodent literature emphasized the importance of visual and olfactory cues to promote oxytocin release. We randomized subjects to play the standard one-shot trust game or a control game in which DM1 publicly pulled a ping-pong ball numbered 0, 1, ..., 10 from an urn and this amount was taken from his/her account and tripled in DM2's account. This control game replicates the standard game but removes the intentionality of DM1s choice to sacrifice money to send a signal of trust. It accounts for the possibility that simply receiving money may raise DM2 OT.

When discussing the findings below, I call the standard intentional choice experiments the Intentional condition, and the random choice experiments the Random Draw condition. Note that sample sizes are moderate as the direct cost of obtaining the data (blood draw supplies, subject payments and hormone assays) are relatively expensive, around $300 per subject (this does not include the cost of necessary specialized equipment such as a refrigerated centrifuge, an ultracold freezer, etc.). The reader is referred to the published work cited below for details on blood acquisition, handling, and assays.

As reported in Zak, Kurzban & Matzner (in press; and 2005) and as shown in Figure 2, OT levels in DM2s who receive an intentional trust signal are almost double that in DM2s in the Random Draw condition. This difference is highly statistically significant (F-test, one-tailed, N=38, p=0.00001), and occurred even though the average amount of money transferred from DM1s to DM2s is the same between conditions (F-test, two-tailed, p>0.87). Relatedly, there is a high degree of reciprocity (trustworthiness) when DM2s receive intentional transfers from DM1s. The correlation between the amount received by DM2s and amount they return to DM1s in the Intention condition is .80 (different than zero, two-tailed t-test, p=0.00001, N=19). This contrasts with the Random Draw condition, in which this correlation is 0.20 and is not statistically different from zero (two-tailed t-test, p>0.40, N=19).

In addition, DM2 OT levels were strongly related to their behavior in the Intention condition. Estimating a multiple regression model of relative trustworthiness (the amount returned by DM2 to DM1/three times the transfer DM1 sent to DM2), both OT and OT^2 are highly statistically significant (t-test, p<.03, R^2=0.39). Using relative trustworthiness as the dependent variable controls for the amount the DM2 received from DM1. The inclusion of OT^2 accounts for physiologic saturation. The significance of OT and OT^2 hold whether or not control variables such as age are included (estimated parameters of all controls variables are statistically indistinguishable from zero). There was no overall difference in the trustworthiness of males and females, as some behavioral experiments have found (Croson & Buchan, 1999).

Because OT is known to interact with many other hormones as the body seeks homeostasis, we also measured nine other hormones to determine if the behavioral effects we found were directly caused by OT or by some other hormone affecting OT, or by OT affecting another hormone. For example, OT suppresses the release of one of the primary

human stress hormones, cortisol, so DM2s might have been more trustworthy because they were less stressed physiologically. None of the other hormones were related to OT levels or DM2 behavior, with one exception. Randomly, some women in our experiment were ovulating (progesterone > 3ng/ml) but none were pregnant (which is another time progesterone is high) by testing their levels of human chorionic gonadotropin (β-hCG). Progesterone has been shown to inhibit the uptake of OT by its receptor. This natural experiment where some female participants were ovulating allowed progesterone to disrupt the effect of OT on DM2 behavior: these women got the same OT surge when receiving a signal of trust but were less trustworthy (one-tailed t-test, $p<.04$). This is solid evidence for the direct and causal effect of OT on trustworthy behavior.

OT is a highly reactive hormone; without a stimulus it is present only in minute amounts. It is released in pulses when needed and has a very short half-life (3-5 minutes). Not surprisingly, we did not find any relationship between basal OT levels of DM1s and the signal of trust they sent. These are "basal" levels because DM1s did not receive a social stimulus as did DM2s. Contrarily, DM2 OT is "activated" by the social signal. The lack of a relationship between OT and DM1 behavior is also consistent with an evolutionary account of OT. Suppose high OT individuals were more likely to give away resources to strangers. Over evolutionary time these individuals would be targets for predation and the genes responsible for this behavior would mostly disappear. Contrast this with DM2s. They are *conditionally* trustworthy—OT rises *after* they receive a signal of trust (and rises roughly in proportional to the signal). OT appears to motivate DM2s to behave in a pro-social manner rather than play the SGP Nash strategy. It is also worth noting that OT receptors in the brain are in evolutionarily old regions, well below the cerebral cortex. This provides a reason why subjects are unable to tell us why they are trustworthy—they simply have a sense that this is the thing to do.

We also gave our experimental subjects an extensive survey inquiring about demographics, social behaviors, sexual behaviors (since OT is a reproductive hormone), and psychological profiles. Of 200 questions, almost none were related to OT levels or behavior in the trust game. Trust was related to three questions on whether DM1s thought others were mostly trustworthy or honest, but none of the survey questions were related to DM2 behavior.

Scottish novelist and poet George MacDonald (1824-1905) appeared to understand the physiologic value of being trusted when he wrote "Few delights can equal the presence of one whom we trust utterly." The evidence presented above supports my hypothesis that signals of trust cause OT to be released. OT appears to induce a temporary attachment by DM2 to the DM1 who has trusted him or her, much as OT induces mothers to attach to infants and vice-versa. This temporary attachment might be called empathy. It literally feels good when someone trusts you, and that good feeling causes most of us to be trustworthy.

3.1 Where is economic man?

Friedrich Nietzsche (1844-1900) wrote that "People who have given us their complete confidence believe that they have a right to ours. The inference is false, a gift confers no rights." Nietzsche has provided the perfect rationale to be an economic man or woman. My lab has run approximately 200 subjects through the trust and blood draw experiment, and this large sample has allowed us to provide some insights into which subjects behave like economic men or women, i.e. those DM2s who take all or nearly all of what they are sent. Figure 3 shows the data on DM2 OT and trustworthiness, with 5 outlier economic men/women circled. The identified DM2s received trust signals, had correspondingly high levels of OT, but somehow suppressed the urge to be trustworthy. Why did they do this?

I recently reported (Zak, in press) that these subjects (3 male and 2 female) appear to have personality traits that are quite different than the average subject in the experiment. I examined if on any of the survey questions, these subjects were more than one standard deviation from the mean of the entire sample (i.e. including them). I found that they were exceptionally emotionally labile, experiencing large mood swings, and were usually sexually active. They said that they believed others were trustworthy and evaluated themselves as very trustworthy—perhaps a bit of self-deception as the survey was completed before the choice in the experiment. They also were more likely to agree that accumulating wealth while others lived in poverty was acceptable.

These results come from a small sample and should be taken with some skepticism, but they are suggestive that personality traits may influence who plays the SGP

Nash strategy. Contrarily, in the typical DM2s who are trustworthy, most of the variation in the behavioral data are explained by OT levels. Much of my current work seeks to characterize how the influences of nature and nurture interact to produce trustworthy or untrustworthy individuals.

3.2 Trust in the brain

There may be more than one system in the brain that permits us to trust others and be trustworthy. Oliver Williamson (1993) coined the term "calculative trust" to denote the ability to use one's experience to estimate the likelihood someone will be trustworthy. If there is a calculative trust substrate in the brain, it is likely distinct from the OT system (though perhaps informed by it) as OT receptors are densest in regions of the brain associated with emotional responses and autonomic regulation.

In an early and important contribution to neuroeconomics, McCabe et al (2001) had subjects play a binary-choice version of the trust game inside a magnetic resonance imaging (MRI) scanner. McCabe and colleagues measured blood flow changes in the brain, an indirect measure of neural activity, when subjects interacted with another human or with a computer that moved with stated probabilities (see Zak, 2004 for a fuller description of this measurement technique, known as functional MRI). These researchers focused on an area in the medial prefrontal cortex (BA10) shown in previous studies to be associated with "theory of mind." Theory of mind is the ability that most humans older than four years old have that allows them to anticipate what others will do by putting themselves in someone else's situation. Small children as well as most autistics are unable to do this and have associated deficits in social interactions. In the trust game, using theory of mind, a DM1 could probabilistically forecast what a DM2 would do.

Comparing regional neural activity for DM1s and DM2s who choose to trust/be trustworthy to analogous choices when subjects were told they were playing against a computer, McCabe and colleagues found greater neural activity in BA10. They also found greater neural activity in BA10 when a subject played against another human and cooperated vs. did not cooperate. The interpretation of these findings is that greater prefrontal activity is needed to forecast what another person will do and trust them compared to taking the sure payoff when playing the SGP Nash strategy or when

interacting with a computer that plays using known probabilities. This theory of mind activation is a neural substrate associated with calculative trust.

A similar study was published in 2002 by Rilling et al. Rilling's group studied neural activity using functional MRI in 36 women playing a binary-choice sequential prisoner's dilemma game (i.e. the trust game). Contrasting play against a human to play with a computer, they found greater neural activation in dopamine-innervated midbrain regions, frontal regions associated with attention and error monitoring, as well as frontal regions that process emotions. Midbrain regions rich in dopamine receptors are the primary areas active during rewarding behaviors. These authors conclude that among the women studied, cooperation itself is rewarding, but requires the mediation of the conflicting concerns of making more money but behaving in socially less acceptable ways.

The findings of Rilling and colleagues are consistent with a central nervous system role for OT during decisions to be trustworthy. OT facilitates the release of dopamine during maternal to infant bonding—such attachment must be rewarding if mothers are to care for infants, and for infants to seek maternal care. Similarly, a spike in OT and subsequent dopamine release occur during sexual intimacy in order to motivate reproduction and pair-bonding. It is literally (internally) rewarding to be trustworthy.

The effect of exogenous OT infusion on human trusting behaviors was recently studied by myself and a team at the University of Zurich (Kosfeld et al., 2005). We ran a trust game in which DM1s could transfer 0, 4, 8 or 12 monetary units; each monetary unit was worth .40 Swiss Francs. 128 men received either 24IU of intranasal oxytocin, or placebo in a double-blind design. After waiting 50 minutes for the drug to load, subjects played four rounds of the trust game, being rematched with a different player in each round. We found that DM1s who received exogenous OT were significantly more trusting than those on placebo. For example, in the placebo group, 21% chose to trust maximally (transferring 12 MUs), while 45% in the OT group exhibited maximal trust. On average, DM1 trust was 17% higher in the OT group than the placebo group, a statistically significant difference (one-sided Mann-Whitney test $p<.03$).

This exogenous manipulation demonstrates causally that OT can induce DM1s to be more trusting. This appears to occur by reducing the anxiety associated with placing trust in a stranger. Consistent with the findings of Zak et al (in press; and 2005), exogenous

OT infusion had no effect on DM2 trustworthiness. Why? The majority of DM2s received a signal of trust and had endogenous OT release. OT receptors were therefore mostly bound up with OT, and additional exogenous OT would therefore have no physiologic effect.

4. Implications and Conclusions

In my cross-country work, the most highly correlated variable associated with generalized trust is self-reported happiness (see Figure 4; correlation different than zero at $p < .01$, two-tailed t-test). Why are happy people trusting (or vice-versa)? The evidence presented in this chapter strongly suggests that nature has designed us to be conditional cooperators because it literally feels good. This positive feedback is how OT facilitates bonding of mother to child, spouses to each other, and my experiments have shown, causes strangers who are shown tangible evidence of trust placed in them to temporarily attach to each other. The SGP Nash equilibrium in the trust game does not obtain because the equilibrium's assumptions are inconsistent with human nature.

The understanding of the mechanisms producing cooperative behaviors among humans that my lab and other labs are developing has manifold applications in economics and indeed to many human endeavors. Most importantly, in reducing poverty (Zak & Knack, 2001). Trust and trustworthiness are also a solution to the low level of shirking in principal-agent relationships, and the mostly fair dealing observable in transactions with asymmetric information. More generally, trust arises in the quotidian human interactions of all types that standard models of self-interest in economics and biology cannot explain, such as tipping the waitress in a city you will not visit again.

So why do we trust? Modern life is nearly impossible without it, and certainly in modern economies with largely impersonal exchange conditional trust is necessary for transactions to occur. Understanding the neuroeconomics of trust can aid in the design of institutions to promote interpersonal trust. This includes using face-to-face negotiations during transactions whenever possible, organizational designs that promote activities that permit employees to form bonds such as outdoor adventures, and a recognition that children and family are important. An effective way to raise trust, which is used by many organizations including agencies of the U.S. government, is on-site massage therapy. It is

not only the psychological effect of the employer "caring" about employees, but that this caring manifests in human touch that raises oxytocin and productivity. Further applications can be found in Zak (2003).

At the national level, trust can be raised by emphasizing the importance of education, reducing inequalities, and promoting freedom and democracy. National institutions that allow and encourage individuals to achieve their goals directly promote trust and therefore the creation of wealth. This is reflected in the higher rates of return on national stock markets for countries that have higher levels of generalized trust (Zak, 2003).

English philosopher Bertrand Russell (1872-1970) wrote that "The most valuable things in life are not measured in monetary terms. The really important things are not houses and lands, stocks and bonds, automobiles and real state, but friendships, trust, confidence, empathy, mercy, love and faith." The research reviewed here extends Russell's statement. Friendships, confidence, empathy, mercy, love and faith all follow from trust and are likely mediated by oxytocin. As social scientists apply these findings to institutional design, not only will productivity be raised, but so will happiness.

References

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10, 122-142.

Boyd, R., Gintis, H., Bowles, S, Richerson, P.J. 2003. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the USA.* 100(6): 3531-3535.

Camerer, C., 2003, *Behavioral Game Theory*, Princeton University Press.

Croson, R., Buchan, N., 1999. Gender and culture: international experimental evidence from trust games. *American Economic Review* 89, 386-91.

Frantz, R. 1997. *X-Efficiency. Theory, Evidence and Applications.* Boston: Kluwer Academic.

Frey, B.S., Benz, M. Stutzer, A. 2004. Introducing Procedural Utility: Not Only What, but Also How Matters. *Journal of Institutional and Theoretical Economics,* 160(3): 377-401.

Leibenstein, H. 1966. Allocative Efficiency vs. 'X-Efficiency'. *American Economic Review*, 56: 392-415.

McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the USA* 98(20):11832-5.

Kosfeld, M., Heinrichs, M., Zak, P.J. U. Fischbacher, Fehr, E. 2005. Oxytocin—A Biological Basis for Trust. *Nature*, 435(2):673-676, 2005. *doi10.1038/nature03701.*

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D., 2002. A neural basis for social cooperation. *Neuron,* 35: 395-405.

Smith, V., 1998. The two faces of Adam Smith. *Southern Economic Journal*, 65: 1-19.

Williamson, O. 1993. Calculativeness, trust, and economic organization. *Journal of Law and Economics*, 36:453-486.

Zak, P.J. 2003. Trust. *Capco Institute Journal of Financial Transformation*, 7:13-21.

Zak, P.J. 2004. Neuroeconomics. *Philosophical Transactions of the Royal Society B*, 359:1737-1748.

Zak, P.J., In press. Trust: A Temporary Human Attachment Facilitated by Oxytocin. *Behavioral and Brain Sciences*.

Zak, P.J., Kurzban, R., Matzner, W.T. In press. The Neurobiology of Trust. *Annals of the New York Academy of Sciences*, 1032.

Zak, P.J., Denzau, A.T. 2001. Economics is an Evolutionary Science. In *Evolutionary Approaches in the Behavioral Sciences: Toward a Better Understanding of Human Nature*, Albert Somit and Stephen Peterson, Editors, JAI Press, pp. 31-65.

Zak, P.J., Kurzban, R., Matzner, W.T. 2005. Oxytocin is Associated with Human Trustworthiness. *Hormones and Behavior*, i*n press*.

Figures

Figure 1: Survey data on trust in 1994 from 42 countries with varying institutional environments.

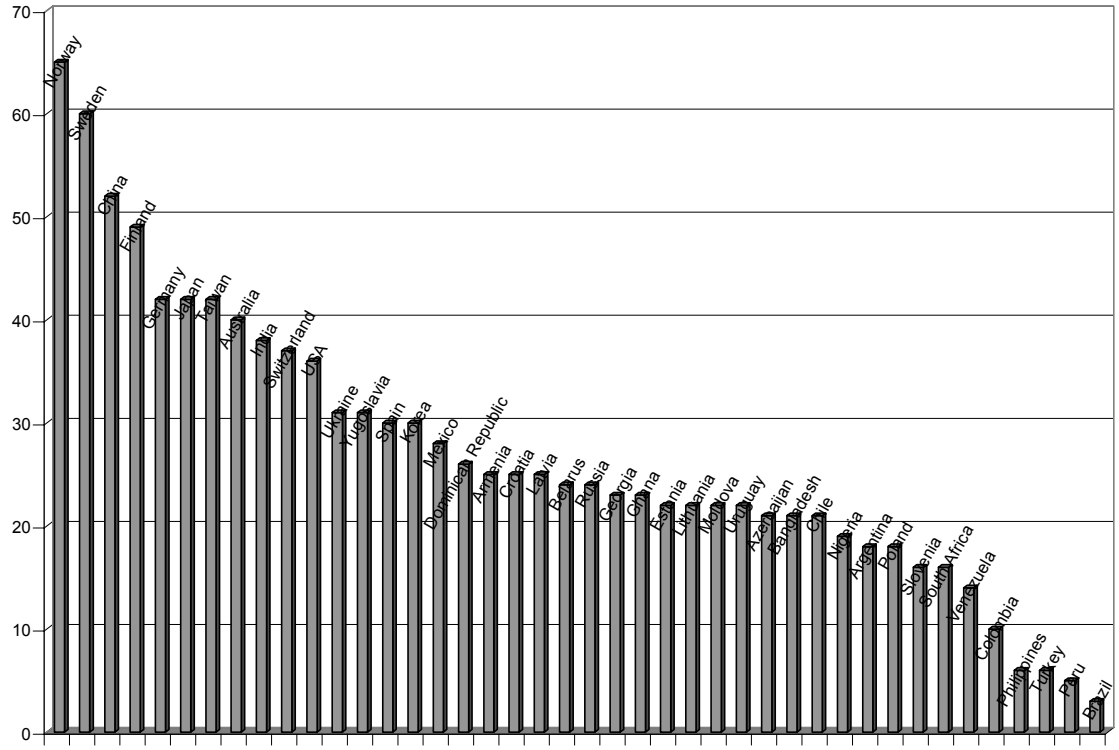

Figure 2: OT levels and standard errors for DM2s in the standard one-shot anonymous intentional trust condition and the random draw (unintentional transfer) condition. In the Intention condition DM1s voluntarily transfer money to DM2s. In the Random Draw condition the transfer from DM1 to DM2 was determined by a public draw of a numbered ball.

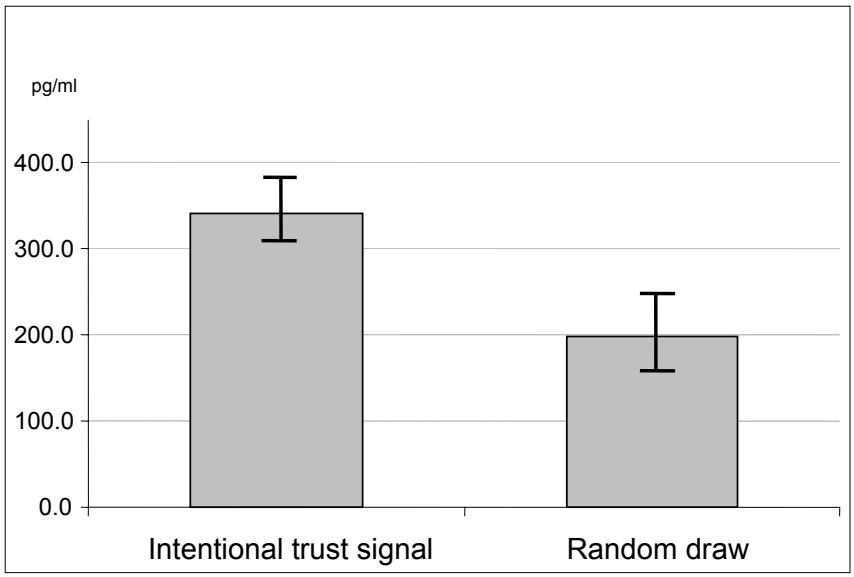

Figure 3: DM2 trustworthiness and OT levels. The five subjects in the circle received signals of trust, had a surge in OT, but behaviorally were untrustworthy. They are the classic economic men and women.

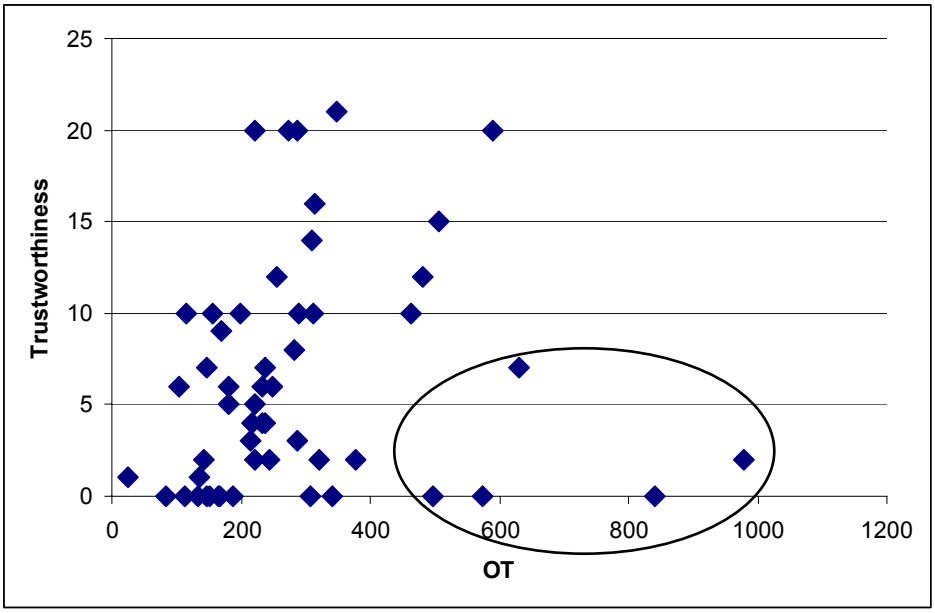

Figure 4: Self-reported happiness is strongly related to generalized trust across countries. This is consistent with the experimental evidence showing OT is released when someone trust us. OT facilitates the release of the neurotransmitter dopamine that is associated with rewarding behaviors.