The Neuroeconomics of Distrust: Sex Differences in Behavior and Physiology

By PAUL J. ZAK, KARLA BORJA, WILLIAM T. MATZNER, AND ROBERT KURZBAN*

Trust is an essential component of transactions that occur over time. The amount of trust accorded to others is typically conditioned on multiple factors, including knowledge of the other party, the history of interactions, and the context of exchange. Determining who to trust and who to distrust is especially important in modern societies with largely impersonal exchange (Vernon L. Smith, 2003). In fact, trust is among the strongest predictors of whether a country will successfully develop: poor countries are by-and-large low-trust countries. This occurs because low trust inhibits investment and thereby the creation of wealth (Zak and Stephen Knack, 2001).

Unfortunately, subjects in laboratory settings are unable to articulate clearly why they decide to trust or distrust a trading partner. In order to discover why human beings trust or distrust others, economists have begun obtaining physiologic measurements during trust experiments (Zak, 2005). This new transdisciplinary field is called neuroeconomics (Kevin McCabe et al., 2001; Zak, 2004; Colin F. Camerer et al., 2005). Recently, Zak et al. (2004, 2005) reported that people who received a signal of trust in an experimental game had higher levels of the neuroactive hormone oxytocin (OT) than those who received similar amounts of money absent a trust signal. In addition, higher OT levels were associated with an increased reciprocation of trust (i.e., greater trustworthiness). Animal models have shown that OT promotes prosocial behaviors by producing a pleasurable sensation.

Because humans are highly social creatures, there may be both positive and negative physiologic controls over social behaviors, as there are for other important behaviors. For example, the hormones ghrelin and leptin are primary regulators of nourishment, promoting food intake and signaling satiety to terminate food consumption, respectively. This paper provides evidence for a hormone that is associated with a negative social interaction, distrust.

We test two hypotheses:

H1) *Receipt of signals of distrust will be associated with an increase in dihydrotestosterone* (*DHT*);

H2) *The relationship between distrust signals and DHT will be stronger in men than in women.*

DHT is a biologically active metabolite of testosterone (T), with T and DHT circulating in both men and women. In men, T is principally produced in the testes; in women T is synthesized in the adrenals and ovaries. Ninety-eight percent of T is protein-bound and is therefore biologically inactive as only free T binds to T receptors. The enzyme 5-alpha reductase transforms T into DHT. DHT can be considered "high octane" testosterone. Indeed, DHT, not T, causes the development of secondary sexual characteristics in males during puberty (e.g., muscle growth, facial hair, vocal-chord thickening, etc.). In adults, basal T and DHT are 5–50 times higher in men than in women.

T and DHT are highly reactive hormones in both sexes, rising prior to an athletic match or when winning at chess, and falling in defeat (James M. Dabbs and Mary G. Dabbs, 2000). Terrence Burnham (2003) found that males with higher basal T were more likely to reject unfair offers in the ultimatum game. In this study we examined activated rather than basal DHT; that is, we measured hormone levels after

* Zak: Center for Neuroeconomics Studies and Department of Economics, Claremont Graduate University, Claremont, CA 91771-6165, and Department of Neurology, Loma Linda University Medical Center, Loma Linda, CA 92354 (e-mail: paul@pauljzak.com); Borja: Center for Neuroeconomics Studies and Department of Economics, Claremont Graduate University; Matzner: Department of Economics, Graziadio School of Business and Management, Pepperdine University, Los Angeles, CA 90045 (e-mail: wmatzner@aol.com); Kurzban: Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104 (e-mail: kurzban@psych.upenn.edu).

a stimulus was presented to subjects. This allowed us to compare DHT levels between men and between women as amount of distrust varied. We do not compare DHT levels across sexes as the latter is nonsensical given basal differences.

I. Methods

Distrust was examined using the "trust game" (Joyce Berg et al., 1995), which we describe in detail. Subjects were recruited for the experiment, provided written informed consent, and assigned an identity-masking code by a lab administrator. All those who showed up earned \$10 for agreeing to spend up to 1.5 hours in the lab and for allowing us to take four tubes of blood from them. Participants were instructed in the decisions that they would be asked to make through a series of examples and assured there was no deception in the experiment. Subjects were then randomly matched into dyads. Within each dyad, subjects were randomly assigned the role of either decision-maker 1 (DM1) or decision-maker 2 (DM2). DMs interacted only by computer and sat in partitioned computer stations in a large lab.[1]

When the experiment began, DM1s were prompted via software to send an integer amount of his or her \$10, including zero, to the DM2 in his/her dyad. Both DMs were told that whatever DM1 sent would be taken out of his/her account and *tripled* in DM2's account. Subsequently, DM2s were told how much the DM1 transferred to him/her and were prompted to return an integer amount, including zero, to the DM1 in the dyad. All subjects were informed that they would make a single decision that affected how much money they would earn during the experiment.

Trust is indexed as the transfer from DM1 to DM2. Trustworthiness is the return transfer from DM2 to DM1. We define *distrust* as the money DM1 did not transfer to DM2 relative to what DM1 could have transferred; that is, distrust $= 30 - 3$(DM1 transfer to DM2). This measure is a simple linear metric of the signal of distrust received by DM2 from DM1, with a minimum distrust of zero (complete trust), and maximal distrust of 30. The subgame perfect Nash equilibrium of this game predicts maximal distrust, but across many experiments and experimenters this is uncommon (Camerer, 2003).

Two conditions were run. The Intention condition was the standard game described above where subjects make intentional choices. The Random Draw condition, which is used as a control, has each DM1 publicly pull a ball numbered 0, 1, ... , 10, from an urn, and this amount was deducted from DM1's account and tripled in DM2's account. DM2s were aware in the Random Draw condition that the money they received from DM1s was random rather than an intentional choice. This condition isolates the consequences of intentions to distrust from the effect of receiving more or less money.

Subjects made decisions sequentially, and following each decision, they were led to an anteroom where 28 ml of blood was drawn from an arm vein. The blood was immediately placed on ice and then centrifuged at 1,500 rpm at 4°C for 12 minutes. Plasma and serum were extracted and stored at −70°C prior to analysis (methods are fully described in Zak et al. [2005]).

II. Results

A total of 212 individuals participated in the experiment: 144 in the Intention condition and 68 in the Random Draw condition. Fifty percent of participants were female, and average subject age was 22. All major ethnic groups were represented, and each session had roughly a 50–50 mix of male and female subjects. Note that subjects did not know the sex of the other person in their dyad.

The generation of distrust signals by DM1s showed a distinct gender difference in the Intention condition. Female DM1s distrusted DM2s more than men: female distrust averaged 18.24, while male distrust averaged 14.03 (different at $p = 0.03$, one-tailed t test, $N = 74$). In other words, female DM1s exhibited less willingness to trust others using a monetary transfer.

Next we turn to DM2s. One of the novelties of the research reported here is that we examine the physiologic effect of the *receipt of signals of distrust* by DM2s, rather than simply reporting

[1] The experiment was run at the California Social Science Experimental Laboratory (CASSEL) at UCLA. We thank David Levine and Bill Zame for use of the lab.

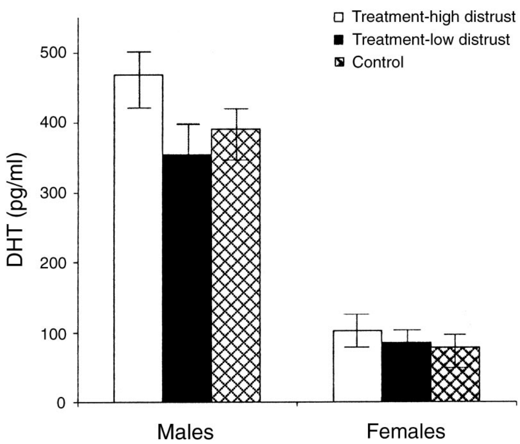

FIGURE 1. DHT LEVELS IN MALE AND FEMALE DM2S

Notes: DHT levels in male DH2s are proportional to the distrust signal they received; DHT levels in female DM2s responded only weakly to distrust signals and do not differ statistically from DHT levels in female DM2 controls. Error bars indicate ±SE.

DM2 behavior as in previous analyses of the trust game. Turning to our main result, we find a sex difference in the physiologic response to the receipt of distrust signals. Figure 1 depicts DHT levels in the Intention condition for men and women receiving a high distrust signal (distrust ≥ 15) versus a low signal (distrust < 15), as well as DHT in the Random Draw condition. Average male DHT in the high-distrust group was 468.2 pg/ml, while males in the low-distrust group had an average DHT of 353.1 pg/ml (different at $p = 0.02$, one-tailed t test). DHT levels were also marginally greater in the male high-distrust group than DHT for men in the Random Draw condition ($p = 0.06$, one-tailed t test). There was no statistical difference between male DM2 DHT in the low-distrust group compared to the Random Draw condition ($p = 0.19$, one-tailed t test). Note that the average transfers DM2s received in the Intention and Random Draw conditions were statistically identical ($p = 0.34$, F test), so these results are not being driven by differences in the monetary transfer between the two conditions.

For females, DHT levels in DM2s did not differ between those who received a high distrust signal (mean 103.3 pg/ml) and those who received a low distrust signal (mean 86.0 pg/ml) (different at $p = 0.17$, one-tailed t test). These levels were not different than DHT for DM2 women in the Random Draw condition (mean 79.7 pg/ml) ($p = 0.15$, one-tailed t test).

The correlation between the distrust signal received and DHT for DM2 men in the Intention condition was positive and significant ($r = 0.32$, $p = 0.05$). There is no relationship for DM2 males in the Random Draw condition between the distrust index and DHT ($r = -0.01$, $p = 0.52$). Likewise, DHT for DM2 women was unrelated to the received distrust signal ($r = 0.10$, $p = 0.91$).

Lastly, we asked if DHT was related to DM2 behaviors—their reactions to signals of distrust. Examining the percentage of money DM2s returned to DM1s (i.e., relative to what they received from DM1s) in the Intention condition, men sent back 25 percent while women gave back 42 percent (different at $p = 0.008$, one tailed t test). This suggests, but does not prove, that male behavior was influenced by the physiologic reaction they had to the signal of distrust. Surprisingly, even in the Random Draw condition that removed the intentional signal of distrust, female DM2s returned significantly more money to the DM1s in their dyads, $4.67 (28 percent), while male DM2s returned only $1.80 (10 percent; amounts different at $p = 0.005$).

The relationship between signals of distrust and T levels is substantially weaker in both men and women. As expected, for males the correlation between T and DHT was high ($r = 0.42$), while the relationship in women was weaker ($r = 0.24$). For neither men nor women was there a significant relationship between T and distrust among DM2s (males: $r = -0.01$, $p = 0.63$; females: $r = 0.00006$, $p = 0.91$). This indicates that it is the bioactive metabolite of T, DHT, that responds to the receipt of signals of distrust in men.

III. Discussion

We have presented evidence that men and women produce and respond to signals of distrust differently. The sex differences in behavior we report are in line with those from other experiments that admit cooperative behaviors (Rachel Croson and Nancy Buchan, 1999; Catherine Eckel and Phillip Grossman, 2001). What we add to this literature is preliminary evidence for a physiologic driver for the observed sexually differentiated responses to signals of distrust.

Our physiologic findings showed that men, but not women, responded to distrust with increased levels of DHT. The relationship between testosterone and aggression is well established in animals, although more equivocal in humans (Dabbs and Dabbs, 2000). This suggests that men in our experiment had an aggressive reaction when they received a signal of distrust, while women did not. We measured nine other hormones, and none of them was related to the distrust signals received by women or men. During debriefing, women reported that they disliked being distrusted, but we did not find a physiologic signature for this. While our results allow us to draw a causal inference about the relationship between signals of distrust and DHT in males, additional work is required to establish causality between the physiology and behavior associated with the receipt of signal of distrust.

The results reported here, in conjunction with those in Zak et al. (2004, 2005), have begun to identify the physiologic changes that occur in response to being trusted or distrusted, and they build on a growing understanding of the biological basis for social decisions.

REFERENCES

Berg, Joyce; Dickhaut, John and McCabe, Kevin. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 1995, *10*(1), pp. 122–42.

Burnham, Terrence. "Caveman Economics: Proximate and Ultimate Causes of Non-materially Maximizing Behavior." Working paper, Harvard University, 2003.

Camerer, Colin F. *Behavioral game theory: Experiments in strategic interactions*. Princeton, NJ: Princeton University Press, 2003.

Camerer, Colin F.; Loewenstein, George and Prelec, Drazen. "Neuroeconomics: How Neuroscience Can Inform Economics." *Journal of Economic Literature*, 2005, *43*(1), pp. 9–64.

Croson, Rachel and Buchan, Nancy. "Gender and Culture: International Experimental Evidence from Trust Games." *American Economic Review*, 1999 (*Papers and Proceedings*), *89*(2), pp. 386–91.

Dabbs, James M. and Dabbs, Mary G. *Heroes, rogues, and lovers: Testosterone and behavior*. New York: McGraw-Hill, 2000.

Eckel, Catherine and Grossman, Phillip. "Chivalry and Solidarity in Ultimatum Games." *Economic Inquiry*, 2001, *39*(2), pp. 171–88.

McCabe, Kevin; Houser, Daniel; Ryan, Lee; Smith, Vernon and Trouard, Theodore. "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." *Proceedings of the National Academy of Sciences*, 2001, *98*(20), pp. 11832–35.

Smith, Vernon L. "Constructivist and Ecological Rationality in Economics." *American Economic Review*, 2003, *93*(3), pp. 465–508.

Zak, Paul J. "Neuroeconomics." *Philosophical Transactions of the Royal Society B*, 2004, *359*(1451), pp. 1737–48.

Zak, Paul J. "The Neuroeconomics of Trust," in R. Frantz, ed., *Two minds. Intuition and analysis in the history of economic thought*. Berlin: Springer, 2005 (forthcoming).

Zak, Paul J. and Knack, Stephen. "Trust and Growth." *Economic Journal*, 2001, *111*(470), pp. 295–321.

Zak, Paul J.; Kurzban, Robert and Matzner, William T. "The Neurobiology of Trust." *Annals of the New York Academy of Sciences*, 2004, (1032), pp. 224–27.

Zak, Paul J.; Kurzban, Robert and Matzner, William T. "Oxytocin is Associated with Human Trustworthiness." *Hormones and Behavior,* 2005 (forthcoming).