Fool Me Once, Shame on You; Fool Me Twice, Shame on Oxytocin

Mauricio R. Delgado[1,*]
[1]Department of Psychology, Rutgers University, 101 Warren Street, Newark, NJ 07102, USA
*Correspondence: delgado@psychology.rutgers.edu
DOI 10.1016/j.neuron.2008.05.005

In this issue of *Neuron*, a study by Baumgartner et al. investigates the influence of oxytocin on trust behavior and its neural mechanisms. The authors report that, following breaches of trust, oxytocin facilitates prosocial behavior while modulating neural signals in the amygdala and caudate nucleus. The findings have implications for an array of mental disorders where social behavior is compromised.

To trust or not to trust is a social dilemma that impacts our way of life. While expressing trust is essential to building social relationships that are important for personal fulfillment and success, decisions to trust can also backfire and result in a lack of reciprocity and eventual resentment toward those that violated trust. Such breaches of confidence can lead to the development of betrayal aversion, potentially influencing how future social interactions are evaluated and trust decisions are executed. In this issue of *Neuron*, Baumgartner et al. (2008) attempt to understand the neurobiology underlying trust behavior following a breach in trust by combining pharmacological manipulations with neuroimaging techniques and an economic paradigm called the "trust game"(Camerer and Weigelt, 1988; Berg et al., 1995).

A typical trust game involves a one-shot social interaction between two players, an investor and a trustee. The investor is faced with a decision to keep a sum of money (e.g., $10) or share it with a trustee. If shared, the investment is tripled ($30) and the trustee now faces the decision to repay the trust by sending back a larger amount of money (e.g., $15 for each participant) or to defect and violate trust by keeping the money, leaving the investor with nothing to show for his display of trust. The social dilemma for the investor is a clear one, as it is potentially more profitable to trust, but doing so leaves the investor susceptible to the risk of a breach in trust. Evidence suggests that humans are traditionally averse to these types of risks (Bohnet and Zeckhauser, 2004), and that this behavior may be modulated by the neuropeptide oxytocin (Kosfeld et al., 2005). A hormone released during social touch and childbirth, oxytocin has long been known for its role in social attachment and facilitation of social interactions (Insel and Young, 2001). More recently, intranasal applications of oxytocin have been demonstrated to increase one's tendency to engage in social risks in a trust game, while having no effect on a similar but nonsocial risk game (Kosfeld et al., 2005).

Oxytocin receptors are abundant in the amygdala (Huber et al., 2005), a structure involved in emotion and fear learning (Phelps and LeDoux, 2005), and oxytocin administration leads to decreased amygdala blood oxygenated level dependent (BOLD) responses to fearful stimuli (Kirsch et al., 2005). Thus, one interpretation of the Kosfeld et al. (2005) findings is that oxytocin may aid an individual in overcoming the betrayal aversion that is inherent in such social economic exchanges, increasing the prosocial behavior of sharing. In conjunction with the amygdala, the striatum, particularly the caudate nucleus, a structure involved in reward-related learning and decision-making (Balleine et al., 2007), is also hypothesized to be involved in trust behavior. More specifically, the caudate nucleus has been linked with acquiring reputations or learning to associate a positive outcome (e.g., payoff in trust game) with a particular action (e.g., sharing with trustee X) during repeated iterations of a trust game (King-Casas et al., 2005). Failure to take into account the current feedback during such social interactions (which may reflect betrayal) leads to diminished responses in the caudate nucleus and a lack of behavioral adaptation in the trust game (Delgado et al., 2005).

In their study, Baumgartner and colleagues propose that oxytocin reduces betrayal aversion that results from breaches of trust by modulating subcortical targets involved in fear learning and reward-related processing, namely the amygdala and the caudate nucleus (Baumgartner et al., 2008). The authors administered oxytocin or placebo to 49 male participants acting as investors in multiple rounds of a trust game (with different trustees) while simultaneously undergoing an fMRI scan. Participants played either a trust game or a risk game in which similar financial risks were taken without a social component (i.e. with a computer). In order to investigate the role of oxytocin following breaches of trust, the experiment was divided into a prefeedback and postfeedback phase. In between the two phases, participants received feedback information indicating that roughly 50% of their decisions (in both trust and risk games) had resulted in poor investments—that is, their trust had been breached (trust game) or their gamble did not pay off (risk game).

As expected, participants in the placebo group decreased their expression of trust (measured as amount of money invested) after discovering that their prior displays of trust had been violated; that is, placebo participants shared less in the trust game during the postfeedback phase compared with the prefeedback phase. In contrast, participants that received oxytocin maintained their prosocial behavior of sharing in the trust game, irrespective of breaches of trust. This behavioral difference between placebo and oxytocin group in the postfeedback trust

game was marked by neural differences in the hypothesized regions. Compared with the placebo group, the oxytocin group showed less activation in the amygdala and caudate nucleus, in support of the idea that the mechanism by which oxytocin affects social behavior is through a decrease in fear mechanisms associated with betrayal aversion (Kirsch et al., 2005) concurrent with a decrease in immediate feedback processing necessary for guiding future decisions (Delgado et al., 2005; King-Casas et al., 2005). Importantly, these behavioral and neural differences were apparent during the trust game, but not the risk game, further suggesting that the effect of oxytocin is exclusive to social risks.

The report by Baumgartner and colleagues represents an ambitious and significant development in the literature, integrating different methodologies (pharmacological and neuroimaging) and disciplines (neuroscience and economics). The study highlights the strength of oxytocin in facilitating social interactions after trust has been violated, by potentially lowering defense mechanisms associated with social risks and by overcoming negative feedback that is important for adapting behavior. While a degree of wariness may protect one from harm, being able to "forgive and forget" is an imperative step in maintaining long-term relationships. This study, therefore, has significant implications for understanding mental disorders where deficits in social behavior are observed. Betrayal aversion, for example, could serve as a precursor to social phobia, a disorder characterized by aversion to social interactions; the reported oxytocin finding could provide a bridge for potential clinical applications.

Other questions for future exploration may focus on what constitutes betrayal and how different types of feedback may modulate the stability of the oxytocin finding. Simple manipulations such as different magnitudes (e.g., 80% defection rate), valence (e.g., positive and negative feedback), or even rate of the feedback (e.g., every trial) may or may not influence the documented oxytocin effect. Betrayal, however, may be stronger when preceded by social expectations. A good test of the role of oxytocin in overcoming betrayal aversion, therefore, might involve social interactions where expectations exist, such as a breach in confidence by a loved one. Future investigations may also pursue potential sex differences in interpreting breaches of trust during administration of oxytocin.

Trust is essential to building social relationships and breaches of trust have a profound impact on social behavior and mental health. It is worth noting, however, that while oxytocin may facilitate prosocial behavior by potentially reducing betrayal aversion, often times this is not advantageous. As the old proverb states, "Fool me once, shame on you; fool me twice, shame on me."

REFERENCES

Balleine, B.W., Delgado, M.R., and Hikosaka, O. (2007). J. Neurosci. *27*, 8161–8165.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Neuron *58*, this issue, 639–650.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Games Econ. Behav. *10*, 122–142.

Bohnet, I., and Zeckhauser, R. (2004). J. Econ. Behav. Organ. *55*, 467–484.

Camerer, C., and Weigelt, K. (1988). Econometrica *56*, 1–36.

Delgado, M.R., Frank, R.H., and Phelps, E.A. (2005). Nat. Neurosci. *8*, 1611–1618.

Huber, D., Veinante, P., and Stoop, R. (2005). Science *308*, 245–248.

Insel, T.R., and Young, L.J. (2001). Nat. Rev. Neurosci. *2*, 129–136.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., and Montague, P.R. (2005). Science *308*, 78–83.

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., Gruppe, H., Mattay, V.S., Gallhofer, B., and Meyer-Lindenberg, A. (2005). J. Neurosci. *25*, 11489–11493.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., and Fehr, E. (2005). Nature *435*, 673–676.

Phelps, E.A., and LeDoux, J.E. (2005). Neuron *48*, 175–187.