Journal of Economic Methodology,
Vol. 17, No. 2, June 2010, 133–146

Routledge
Taylor & Francis Group

Inductive modeling using causal studies in neuroeconomics: brains on drugs

Moana Vercoe[a] and Paul J. Zak[b]*

[a]Center for Neuroeconomics Studies, Claremont Graduate University, Claremont, CA 91711-7176, USA; [b]Department of Economics, Claremont Graduate University, Claremont, CA 91711-7176, USA; Department of Neurology, Loma Linda University Medical Center, Loma Linda, CA 92354, USA

This paper introduces a new approach to economic analysis. We show how to move from deductive to inductive modeling and thereby reunite economics with approaches used in the natural sciences. This paper presents the empathy-generosity-punishment model as an example of research based on observation, experimentation, and the elimination of alternatives. Inductive modeling in neuroeconomics allows the identification of the physiologic mechanisms that produce behavior. Unlike most neuroeconomics studies, we show how to establish causation by using drugs to manipulate brain activity. This approach is demonstrated using three experiments that circumscribe the brain processes behind prosocial behavior.

Keywords: oxytocin; testosterone; social preferences; modeling, induction

JEL Codes: D87; B84; C9

1 Introduction

The traditional approach in economics has been to propose mathematical models of behavior that can be used to derive theorems that identify causal mechanisms. Empirical analyses can then be used to test causal claims to build a case that all or most of the model's implications are consistent with the data. When building models, a researcher hopes for novel, surprising or non-intuitive results that, when tested, can confirm the model's veracity in a non-obvious way. This approach has led to a plethora of competing models that are either not tested, or if tested often explain the data equally well. Moreover, traditional economic models are deductive, rather than inductive: we are taught to imagine situations and use these imaginings to derive models.

By comparison, the standard in the natural sciences, especially in biology, is to derive models inductively using observation as the starting point. The inductive method begins with the running of controlled experiments in which one factor at a time is varied in order to determine its significance. An early proponent of induction as a method of inquiry, Francis Bacon (1561–1626), characterized this approach as 'the selective process of elimination among a number of alternative possibilities' (1605, III, p. 340). Only after this controlled experimentation is a general model identifying the mechanisms of action proposed.

*Corresponding author. Email: paul.zak@cgu.edu

ISSN 1350-178X print/ISSN 1469-9427 online

DOI: 10.1080/13501781003756675
http://www.informaworld.com

Observation produced from nearly 40 years of laboratory experiments in economics has shown consistent deviations from deduced equilibria, especially in game theoretic settings (Camerer 2003). For example, many of the models that seek to explain deviations from the predictions of traditional models that assume that economic actors are narrowly self-interested, imagine that people have preferences for fair outcomes (e.g. Rabin 1993; also see Camerer 2003). The preference for fairness is imposed in these models rather than being found from controlled experiments that systematically vary fairness conditions or intentions. It is our view that economics can do better.

Experimental economists have proposed inductive models (e.g. Andreoni 1990; Camerer 2003), though this approach is rare. This paper proposes that inductive models are 'good science' and that such models can be made more precise by using causal studies in neuroeconomics. In seeking to improve economic models, we turn to Alfred Marshall who in 1890 stated that economics 'is simply a branch of biology broadly interpreted' ([1890] 1961, p. 772). In this paper we show how to go even further: by administering drugs to participants in experiments we are able to 'turn up' or 'turn down' brain activity in order to show that these manipulations directly cause changes in behavior. Using this approach, causal relations are identified, providing a direct route to building inductive economic models.

We demonstrate this approach by building an inductive model, the empathy-generosity-punishment (EGP) model. This model generalizes our studies of prosocial behaviors and explains the mechanisms behind a particular prosocial behavior, generosity towards a stranger in a one-shot setting. The model's implications are then tested in three neuroeconomics experiments that administer drugs to participants. The key mechanism identified in the EGP model that generates generosity is confirmed in these manipulations that cause a parameter in the model to vary. We emphasize that identifying causal relationships using drug infusions is rare in neuroeconomics, but it has great value in identifying the mechanisms driving observed behavior. It also moves away from using participants' post-hoc reflections and rationalizations that are known to be inconsistent and imprecise (Smith 1998).

In addition to using induction, there is a second more subtle modeling choice in the EGP model that incorporates recent findings in brain science. Rather than assume that people are (hyper-) rational, something that is metabolically costly and slow, we use a modern version of satisficing or bounded rationality (Simon 1957) in which limited information and cognitive constraints result in decisions based on heuristics (memory) and feelings (emotions). We have termed the phenomenon where individuals devote scarce cognitive resources in decision-making only when the expected payoff offsets the cognitive cost of the decision 'rational rationality' (Zak 2008). In rational rationality, individuals are conditional satisficers who will settle for 'good enough' outcomes in some circumstances, but expend energy and cognitive resources to achieve better outcomes in others. In the EGP model, rational rationality manifests when decision-makers receive affective signals that impact their decisions. Indeed, the drug manipulations we describe here subtly impact on emotional states while preserving cognitive abilities. These manipulations are unrecognizable to participants on post-experiment debriefing. As a case in point, and out of ethical concerns, one of the authors (PZ) took each drug before giving it to participants in experiments and felt no cognitive or emotional effects nor were any reported by those observing him. Nevertheless, our tests of the EGP model show that the infused drugs can affect behavior.

2 Modeling prosocial behavior

The model presented here is based on the following axiom: *A person's physiologic state affects his or her decisions*. This seems to us to be noncontroversial and broadly well-supported in biology.

The question we seek to understand using this axiom is why from a behavioral perspective in a blinded, one-shot setting someone would share resources with a stranger. The behavioral reasons behind helping behaviors when reputation can be built or interactions are with family or friends are well understood (Morhenn, Park, Piper, and Zak 2008). The EGP model predicts the starker case of why one would offer resources to a stranger that one will not again encounter and when such behavior cannot be used to build one's reputation as an altruist.

2.1 Foundations of the EGP model

In 2002 our lab began running neuroeconomics experiments that have consistently shown a high degree of prosocial behavior. Indeed, 98% of several thousand participants in our experiments are conditional reciprocators (Zak 2005). Our initial studies using the trust game (Berg, Dickhaut, and McCabe 1995) showed that when one could give up money and transfer it to a stranger because this would triple the amount of the transfer the other person received, the brain of the receiver had a surge in the neuroactive hormone oxytocin (OT). This transfer is considered to measure trust since one has to give up resources to make the stake grow for the other person. We found that larger transfers denoting trust produced a larger increase in OT. In turn, higher OT was associated with a greater return transfer (trustworthiness) to the person who initially showed trust (Zak, Kurzban, and Matzner 2004; Zak, Kurzban, and Matzner 2005). We have replicated the relationship between OT release and trustworthy behavior by varying the conditions under which choice occurs (Morhenn et al. 2008).

By measuring brain activity during these experiments, we have been able to characterize a brain circuit that we call HOME (Human Oxytocin Mediated Empathy) that appears to produce and sustain prosocial behaviors. A full description of this circuit is beyond the scope of this article (for this, see Zak in press), but we briefly describe the biochemical cascade in HOME here. HOME activates with the release of the neuroactive hormone oxytocin (OT). OT is an evolutionarily ancient molecule that is a critical component of the mammalian attachment system motivating care for offspring. More generally, OT is active in brain regions associated with emotions that regulate approach/withdrawal behaviors. In the HOME circuit, OT causes the release of two neurotransmitters, dopamine (DA) and serotonin (SERT). Dopamine is associated with goal-directed behaviors and reinforcement learning, making prosocial behaviors internally rewarding. At the same time, SERT has positive effects on mood and reduces anxiety that may occur around other people. Figure 1 shows the three primary components of HOME: OT, DA, and SERT.

Next, we had to understand what causes OT to be released and the HOME circuit to activate. Studies of both animal behavior and human psychology have shown that observing another's distress is a motivating factor in prompting costly assistance (e.g. Batson 1991; Sober and Wilson 1998; Preston and de Waal 2002; De Waal 2006). There is a consensus that a typical psychological state produced when observing another in distress is empathy (de Waal 2008). The term 'empathy' traces back to the work of two German philosophers, Rudolph Lotze (1817–1881) and Theodor Lipps (1851–1914). While Lotze's conceptualization is an emotional contagion, Lipps's (1903) concept of

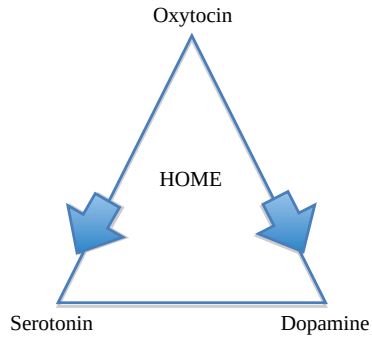

Figure 1. HOME activates when OT is released following a positive social stimulus, motivating and reinforcing helping behavior by facilitating mid-brain dopamine release, and down-regulating distress by modulating serotonin release.

Einfühlung involves the perception of an emotional gesture in another that directly activates that same emotion in the perceiver.

An emotional reaction to another's distress is the mechanism that Adam Smith argued supports moral behaviors in *The Theory of Moral Sentiments*:

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it. Of this kind is pity or compassion, the emotion we feel for the misery of others. (1759, p. 47)

Smith furthers argued that the 'healing consolation of mutual sympathy' (1759, p. 19) is analogous to a utility flow from engagement that alleviates the distress of another. Modern psychology has developed this in Dovidio's arousal: cost-reward model of helping behavior (Dovidio 1984; Dovidio, Piliavin, Gaertner, Schroeder, and Clark 1991) that posits that the motivation to help another requires an awareness of the need for assistance. The recognition of distress in another calls upon Adam Smith's 'fellow feeling' and Lipps's 'in feeling'. In economics, the pleasure of helping another has been called warm glow utility (Andreoni 1990).

Brain imaging studies locate the empathic response in reaction to the perceived pain in another in the observer's pain matrix, with especially strong activation in the anterior cingulate cortex and anterior insula (Decety et al. 1997; Singer et al. 2004, 2006; Hein and Singer 2008). That is, the perception of pain in another produces discomfort in the viewer that manifests as pain. Because we seek to avoid pain, activation in the perceiver's pain matrix provides a motivation to engage to relieve the other's distress. Helping others can be viewed as self-serving as it derives from a desire to reduce the observer's own pain.

Batson has found that low to moderate amounts of distress increase empathy and the likelihood of a prosocial response, while high degrees of distress are so aversive that they extinguish the desire to help and motivate instead a desire to escape (Batson, Fultz, and Schoenrade 1987; Batson 1991; Batson and Oleson 1991). Studies from our lab have found a similar nonlinearity in the physiology linking distress and empathy. In a study where participants watched a 100-second highly emotional video about a father and his four-year-old son who has terminal brain cancer, subjects reported both distress and increased empathy (Barraza and Zak 2009). Those who simply watched the video had a 157% increase in the neurohormone oxytocin as well as a spike in the stress hormone

cortisol. The change in oxytocin was associated with the subjective experience of empathy ($r = 0.20 > 0$, $p = 0.01$, $N = 145$) after controlling for the distress one felt. At the same time, distress reduced the change in oxytocin ($p = 0.05$), even while distress and empathy were highly correlated ($r = 0.81 > 0$, $p < 0.001$). Those who were empathically engaged were more likely to share resources with a stranger ($r = 0.24 > 0$, $p = 0.05$) and to donate some of their earnings to charity ($r = 0.356$, $p = 0.004$). This indicates that empathy is associated with OT release and is a motivator for prosociality.

An additional motivator for prosocial behavior is the threat of punishment. In *The Theory of Moral Sentiments*, Adam Smith wrote, 'That whatever appears to be the proper object of gratitude appears to deserve reward; and that, in the same manner, whatever appears to be the proper object of resentment, appears to deserve punishment' (1759, p. 79). This suggests an innate drive to punish in response to certain behaviors. An fMRI study tracking changes in brain activity in response to ungenerous offers in the one-shot ultimatum game reported activation of the anterior insula, an area associated with the feelings of disgust (Sanfey, Rilling, Aronson, Nystrom, and Cohen 2003). Activation increased with increasingly unfair offers and predicted the decision to reject the offer, thereby punishing the one making the inadequate offer.

Recent work in neuroeconomics has shown that individuals (especially men) derive pleasure from punishing those who are stingy even if punishment requires that one pays a cost (De Quervain et al. 2004; Singer et al. 2004; Delgado, Frank, and Phelps 2005; Zak, Borja, Matzner, and Kurzban 2005). This can be called moralistic punishment (Kurzban, DeScioli, and O'Brien 2007). These findings suggest that the threat of punishment is a second reason for prosociality.

2.2 The EGP model

The EGP model is based on a single interaction between two individuals denoted Decision-Maker 1 (DM1) and Decision-Maker 2 (DM2). DM1 must decide how much of a fixed amount of resources, M, to offer to DM2, while DM2 then decides to accept or reject DM1's offer. A rejection will occur if the offer is seen as unacceptably low. By allowing DM2 to accept or reject the offer, the EPG model includes the possibility of costly punishment. Without loss of generality, we assume that if DM2 rejects an offer, then all available resources are lost to both individuals. This set of choices is captured in experiments using the ultimatum game (UG) introduced by Güth, Schmittberger, and Schwarze (1982).

Based on findings from our lab and others reviewed above, punishment is included in the model as a preference. Let $P(b_2)$ be the utility DM2 derives from punishing DM1 for making an inadequate offer of benefits b_2 to him. Assume that $P(b_2)$ is decreasing, continuous and strictly concave. DM2 also derives utility from consuming benefits $U(b_2)$, where $U(b)$ is increasing, continuous and strictly concave. The more DM2 is offered, the higher his utility from consuming b_2 and the less pleasure he receives from punishing DM1. In this setting, DM2 chooses the supremum of the set $\{U(b_2), P(b_2)\}$. The concavity restrictions generate a well-defined, unique solution to DM2's decision problem. Assume that if $U(b_2) = P(b_2)$, then DM2 will choose not to punish. Let us define the indifference value b_2^* such that $U(b_2) = P(b_2)$ and assume that b_2^* is common knowledge.

DM1's decision is to determine a split of a finite amount of M benefits for himself, b_1, and benefits for DM2, b_2, taking into account that the split must be acceptable to DM2. The findings reviewed above suggest that DM1 may get utility from the

transfer to DM2. Formally, this is:

$$\text{Max}_{b1b2} U(b_1) + \alpha(\tau) U(b_2)$$

$$\text{s.t.} \quad b_1 + b_2 = M$$

$$b_2 > b_2^*$$

The two functions $\alpha(\tau)$ and $P(b_2)$ are the key innovations in the EGP model. We will call $\alpha(\tau)$ empathy. The empathy function depends on the parameter τ that measures the distress of DM2 that is perceived by DM1. The results surveyed in section 2.1 indicate that moderate levels of perceived distress increase empathy and assistance to others. Yet, high degrees of observed distress cause avoidance rather than engagement and result in little or no assistance. The function $\alpha(\tau)$ is included to capture the effect of OT release; it is known that high stress inhibits OT release and the engagement of HOME. Thus, we proposed that $\alpha(\tau)$:[0,1] $\rightarrow$ [0,1] is a continuous hyperbolic function with the following properties, $\alpha(0) > 0$, $\alpha(1) = 0$, and $\tau^* = \text{argmax}\ \alpha(\tau)$, with $\alpha(\tau^*) > \alpha(0)$. The EGP model predicts that as $\alpha < \alpha(\tau^*)$ rises, the benefits offered to DM2, b_2, increase.

In addition, DM1 obtains utility from offering benefits to DM2 if $\alpha(\tau) > 0$ and $b_2 > b_2^*$; that is, if empathy is engaged and the offer to DM2 is not so stingy that it is rejected. Thus, the EGP model produces a continuum of equilibria indexed by α. Generosity is defined as an offer of b_2 by DM1 that exceeds b_2^*. This fits the common definition of generosity as 'liberality in giving'.

In reviewing the literature, we have found that the EGP model has cousins. Economists have asked why people are prosocial nearly since the founding writings in our field, and we have discovered the EGP model is similar to a model introduced in a footnote by the Irish statistician and economist Francis Ysidro Edgeworth (1845–1926) in *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. In Edgeworth's model, utility is derived from a combination of one's own consumption and the weighted utility from consumption by another (1881, p. 53). He included a parameter α on the other's utility that he called 'effective sympathy' and was assumed constant. A number of contemporary theorists have incorporated other-regarding preferences into related models as motivation for prosocial behavior (Andreoni 1990; Rabin 1993; Fehr and Gachter 2000; Sally 2000, 2001; Levitt and List 2007). But we emphasize that these models are still deductive: they use the 'guess and verify' approach rather than being based on systematic experimentation.

3 Testing the EGP model

The EGP model predicts that generosity will occur when:

(1) person 1 perceives distress in person 2;
(2) the distress elicits an empathic response in person 1 sufficient for action, but not so great as to provoke avoidance [$0 < \alpha(\tau) < \alpha(\tau^*)$];
(3) the offer of help is large enough to avoid punishment from person 2 ($b_2 > b_2^*$).

We test the predictions of the EGP model by physiologically manipulating empathy, α. Note that all the manipulations we use to test the EGP model produce different effects over a women's menstrual cycle. As a result, the three experiments described to test the model only use men as participants. Our discussion will therefore use the masculine pronoun

throughout and our findings cannot be extended to women without additional experiments that as yet have not been done.

Determining generosity requires a measure of DM2's rejection or punishment threshold, b_2^*. A high b_2^* punishes DM1 for an ungenerous offer, but at a cost to DM2. To measure b_2^* we modified the UG by having participants make decisions as both DM1 and DM2. After decisions, participants were randomly assigned the role of DM1 or DM2 to determine their payouts.

This adaptation of the UG allows an assessment of individual generosity by comparing each participant's proposal as DM1 to his punishment threshold as DM2. The standard version of the UG requires that DM1s infer the mental state of the DM2s in their dyads to generate proposals that will not be rejected (Firth and Firth 1999). Introducing a self-reported punishment threshold into the UG primed DM1s to be more deliberate in considering the reactions of DM2s because they themselves had considered their responses in the role of DM2. Our adaption motivates perspective-taking.

A second decision task, the dictator game (DG) was included in our experiments as a control. This one-shot interaction is similar to the UG with participants randomly assigned to dyads where one is designated DM1 and the other DM2. In the DG, DM1 starts with an endowment while DM2 starts with nothing. DM1 is asked to make a transfer of some part of his endowment to DM2. Unlike the UG, in the DG this proposal is unilateral because DM2 cannot reject it – the ability to punish is removed. Transfers in the DG are understood to measure altruism (Camerer 2003).

The DG is used as a control for two reasons. First, the EGP model does not generate predictions for altruism. As a result, the model is focused on generosity and punishment, rather than more generally on altruism. Second, including the DG permits us to show that our drug infusions do not affect all behaviors, but only affect some behaviors. This provides evidence that even those given moderate doses of drugs are cognitively intact. Third, generosity is a subset of altruism, and measuring altruism lets us extract the impact of empathy on generosity separate from altruism.

4 Manipulating physiology

The neuroscience literature suggests three substances that might affect the empathy parameter α and the punishment threshold b_2^* in the EGP model. Table 1 reports basic information about each experiment and we describe each of these in turn below.

4.1 Manipulation 1: oxytocin

An obvious candidate substance for increasing α empathy in the UG is OT. Since 2004, research from our lab has established the role played by OT in interactions involving trusting strangers by measuring OT released in blood (Zak et al. 2004, 2005; Zak 2008;

Table 1. Number of subjects receiving physiologic manipulations and the total number of each decision made in each experiment, and the average age of subjects.

Substance	N receiving hormone	N receiving placebo	Behavioral sample size	Mean age (years)
Oxytocin	34	34	68	21.80
Arginine vasopressin	27	20	188	21.77
Testosterone	25	25	200	20.80

Morhenn et al. 2008). These studies established for the first time the influence of OT on non-reproductive behaviors in humans. To establish the causal link between trust and OT, in another study we infused 24IU of OT intranasally into men. We found that this more than doubled the number of subjects who trusted a stranger with all their money (Kosfeld, Heinrichs, Zak, Fischbacher, and Fehr 2005). OT infusion had no effect on mood, cognition, or in nonsocial lottery-choice tasks. OT appeared to subtly alter the balance between trust and distrust of strangers, pushing people towards greater levels of trust. In the present experiment, we hypothesized that OT would increase generosity but have no effect on punishment or altruism.

4.2 Manipulation 2: arginine vasopressin

Also known as antidiuretic hormone, arginine vasopressin (AVP) is pivotal in maintaining water balance within the body. When released in the brain, AVP is associated with peri-reproductive behaviors, primarily in males. In socially monogamous mammals, AVP facilitates attachment of males to females and their offspring. This manifests as reactive aggression when other males approach mates or offspring (Young, Nilsen, Waymire, MacGregor, and Insel 1999; Young and Wang 2004; Bester-Meredith, Martin, and Marler 2005; Carter 2007; Donaldson and Young 2009).

In humans, elevated AVP in cerebrospinal fluid correlates with histories of aggressive behavior (Coccaro, Kavoussi, Hauger, Cooper, and Ferris 1998). Intranasal AVP administration in men causes aggressive facial expressions in response to an unfamiliar man, but affiliative facial expressions when interacting with an unfamiliar woman (Thompson, George, Walton, Orr, and Benson 2006). Like OT, AVP can be measured in blood and cerebrospinal fluid and can be infused intranasally. We hypothesized that AVP would have no impact on empathy and therefore would not alter generosity, but would affect the threshold for punishing others for being ungenerous.

4.3 Manipulation 3: testosterone

The EGP model predicts generosity will decrease when empathy is reduced. A direct way to do this is to inhibit OT binding to its receptor. Testosterone administration (T) has been shown to inhibit OT binding (Insel, Young, Witt, and Crews 1993) and this is the approach we take here.

Studies of salivary T in humans have found that high T men, compared to men with lower T, are more likely to engage in a range of antisocial behaviors, including physical altercations, divorce, spending less time with their children, and having more sexual partners (Dabbs and Dabbs 2000; Newman and Josephs 2009). When their status is threatened, high T men become dominant and aggressive (Popma et al. 2007; Mehta and Josephs 2006). High T men generally appear to be more aggressive and less prosocial than men with lower T levels (Harris, Rushton, Hampson, and Jackson 1996).

In examining T and behavior in our lab we have shown that distrust, as evidenced by small monetary transfers in the trust game, provokes a spike in blood plasma levels of the bioactive metabolite of T, dihydrotestosterone (DHT; Zak, Borja et al. 2005). This indicates a physiologically aggressive response to unfair behavior. DHT levels correlated with reduced return transfers in men, but not women. A recent study using salivary assays found that high T males were more likely to reject low offers in the UG than lower T males (Burnham 2007).

We hypothesized that pharmacologically elevating T in men would cause them to be less generous. We also hypothesized that T would raise the punishment threshold,

making men more willing to bear the cost of punishing others for making ungenerous offers.

Owing to space constraints, the material and methods of each experiment are not reported here. The interested reader is referred to the publications reporting these experiments for details (Zak, Stanton, and Ahmadi 2007; Stanton 2007; Zak et al. 2009).

5 Results

Table 2 summarizes the effects of the three physiologic manipulations on offers in the UG, generosity and punishment. The discussion below offers more details on these findings. All these studies were run double-blind with placebo controls. For the OT and AVP studies, normal saline (salt water) was used as the placebo, for the T study, an inert gel that looked and smelled like testosterone gel (trade name: Androgel®) was used.

5.1 Oxytocin

Generosity was 80% higher among participants who received OT compared to those who received the placebo (OT \$1.89, SD \$1.06; placebo \$1.06, SD \$1.29; two-tailed Mann-Whitney $p = 0.005$). Proposed splits of \$10 by DM1s receiving OT in the UG were on average 21% greater than proposals made by participants receiving the placebo (OT \$4.86, SD \$1.06; placebo \$4.03, SD \$1.29; two-tailed Mann-Whitney $p = 0.005$; $N = 68$). OT had no effect on the punishment threshold (OT \$3.03, SD \$1.69; placebo \$2.91, SD \$1.74; two-tailed Mann-Whitney $p = 0.78$; group mean \$2.97). Only two participants in this study proposed the UG subgame perfect Nash equilibrium of \$1, and both had received the placebo. These results confirm that OT increased generosity in the EGP model as we hypothesized.

5.2 Arginine vasopressin

The average proposal in the UG did not differ between participants receiving AVP and those receiving the placebo (AVP \$5.14, SD \$1.68; placebo \$5.44, SD \$1.61; two-tailed Mann-Whitney $p = 0.76$; $N = 188$). Counter to our hypothesis, AVP had no effect on the punishment threshold (AVP \$3.73, SD \$1.61; placebo \$3.54, SD \$1.75; two-tailed Mann-Whitney $p = 0.54$). Mean generosity was 36% lower among participants who received AVP compared to those on the placebo, but because of large standard errors this was not statistically significant (AVP \$1.40, SD \$2.80; placebo \$1.90, SD \$2.92; two-tailed Mann-Whitney U-test $p = 0.53$). AVP does not appear to affect generosity or punishment in the EGP model.

5.3 Testosterone

Average proposals in the UG were 9% lower for participants on T compared to themselves on the placebo (T \$4.63; placebo \$5.08; one-tailed paired t-test, $N = 200$, $p = 0.001$).

Table 2. Impact of each physiologic manipulation on UG decisions relative to placebo.

Substance	UG offer	Generosity	Punishment threshold	Altruism (DG offer)
Oxytocin	+21%	+80%	0	0
Arginine vasopressin	0	0	0	0
Testosterone	−9%	−27%	0	0

Notes: Generosity is the UG offer minus the punishment threshold. All reported effects are statistically significant at $p < 0.05$.

Compared to themselves, average generosity was 27% lower when participants were on T (T \$1.57; placebo \$2.15; $p = 0.001$), but the punishment threshold was unchanged (T \$3.05; placebo \$2.92; $p = 0.61$). Since a within subject experimental design was used in this study, we tested for order (drug or placebo first) and learning effects (changes in behavior between sessions). Although participants tended to be more generous their second time in the lab, all results maintained significance after controlling for learning and order effects (Zak in press).

Because we obtained blood samples, we were able to test for a parametric relationship between T and behavior. Generosity negatively correlated with all measures of T (total T: $r = -0.25$; $p = 0.001$; free T: $r = -0.19$; $p = 0.007$; DHT: $r = -0.31$; $p = 0.001$). A parametric relationship was also found between all measures of T and the punishment threshold (total T: $r = 0.194$; $p = 0.006$; free T: $r = 0.153$; $p = 0.031$: DHT: $r = 0.228$; $p = 0.001$). These tests reveal that generosity and punishment scale with a man's level of T. For example, men in the lowest decile of DHT had average generosity of \$3.65 compared to generosity of \$0.55 for men in the highest decile of DHT (85% lower). Similarly, men in the lowest decile of DHT had average punishment threshold of \$2.15 compared to a punishment threshold of \$4.00 for men in the top decile of DHT (86% higher).

5.4 Control analyses

Our three physiologic manipulations had no impact on altruism as measured by offers made in the DG. Table 3 presents average DG offers for participants receiving drugs and placebo. This shows that our drug manipulations narrowly affect generosity and/or punishment, rather than having global effects on behavior or cognition.

Since generosity is a subset of altruism, for each physiologic manipulation we used least-squares regressions to test if the drug continued to affect generosity after accounting for altruism. We found this to be the case for OT and generosity (OT coeff. 0.648, two-tailed t-test $p = 0.014$; DG coeff. $= 0.376$, two-tailed t-test $p = 0.0001$; $N = 68$, $R^2 = 0.43$). We also found that the lack of a relationship between AVP and generosity persisted when DG offers were included (AVP coeff. 0.47, two-tailed t-test $p = 0.18$; DG coeff. $= -0.46$, two-tailed t-test $p = 0.0001$; $N = 232$, $R^2 = 0.16$). Finally, the parametric relationship between all measures of T and reduced generosity maintained significance when DG offers were included (total T: $\beta = -0.400$, $p = 0.001$; free T: $\beta = -0.057$, $p = 0.013$; DHT: $\beta = -0.001$, $p = 0.001$).

The survey data on personality traits collected in the experiments showed no significant differences in attachment, empathy or anger between participants receiving the drug and those receiving the placebo in any of the three experiments. Statistical tests also revealed that personality traits had no impact on generosity or punishment.

Table 3. Altruism as demonstrated by average offers in the DG.

Substance	Drug	Placebo
Oxytocin	\$3.77	\$3.58
Arginine vasopressin	\$3.54	\$3.48
Testosterone	\$3.34	\$3.56

Note: All differences are statistically insignificant (p values range from 0.51 to 0.90).

6 Conclusions

We have argued that for economics to be an evolutionary science – as Thorstein Veblen famously proposed in 1898 – we must begin building inductive models that are problem-driven, rather than imagination-driven. Imagining unusual scenarios for models can be useful in small doses, but as economists we have taken this much too far. By running controlled experiments that change conditions one at a time, the impact of each on behavior can be catalogued. But building behavioral models based on presumed mechanisms is not enough. Neuroeconomics studies can be used to identify the reasons why different conditions produce different behaviors. Experiments in neuroeconomics using drugs to cause changes in brain activity are, we believe, the most direct way populate models with the correct behavioral mechanisms.

We demonstrated this approach by presenting the EGP model that was based on years of our experiments studying prosocial behaviors. Our tests of the model used OT, AVP and T as candidates to increase or decrease empathy following known physiologic mechanisms to test if empathy causes prosociality. We found evidence for this hypothesis. We found no overall difference in the threshold for moralistic punishment for men on T or AVP, though this threshold did increase with a man's level of T. Even though AVP had no effect on generosity or punishment, it served as a positive control – not every drug manipulation has impact on the behavior under study. Another set of controls used the DG to measure the impact of our manipulations on altruism. None were found, and altruism did extinguish the effect of OT or T on generosity.

Physiologic manipulations need to be tailored to the behavior of interest; we had insufficient space to fully justify our use of OT, AVP and T, but these substances were chosen based on both animal and human studies – many from our own lab – showing that social behaviors are associated with endogenous release of these hormones. The experiments reported here revealed that OT and T directly caused changes in social behaviors, consistent with the EGP model.

The analysis presented in this paper is only a starting point for an improved methodology in economics. Other neural substrates can be tested to deepen our understanding of the physiology causing many types of economic decisions. Alfred Marshall ([1890] 1961) wrote that 'The modern economic organism is vertebrate; and the science which deals with it should not be invertebrate' (p. 769). We heartily agree.

References

Andreoni, J. (1990), 'Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving', *The Economic Journal*, 100, 464–477.

Bacon, F. (1605), 'The Advancement of Learning', *http://darkwing.uoregon.edu/~ rbear/adv1.htm*.

Barraza, J.A., and Zak, P.J. (2009), 'Empathy Toward Strangers Triggers Oxytocin Release and Subsequent Generosity', *Annals of the New York Academy of Sciences*, 1167, 182–189.

Batson, C.D. (1991), *The Altruism Question: Toward a Social-Psychological Answer*, Hillsdale, NJ: Lawrence Erlbaum.

Batson, C.D., and Oleson, K.C. (1991), 'Current Status of the Empathy-Altruism Hypothesis', in *Review of Personality and Social Psychology, Vol. 12: Prosocial behavior*, ed. M.S. Clark, Newbury Park, CA: Sage, pp. 62–85.

Batson, C.D., Fultz, J., and Schoenrade, P.A. (1987), 'Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences', *Journal of Personality*, 55, 19–39.

Berg, J., Dickhaut, J., and McCabe, K. (1995), 'Trust, Reciprocity, and Social History', *Games and Economic Behavior*, 122–142.

Bester-Meredith, J.K., Martin, P.A., and Marler, C.A. (2005), 'Manipulations of Vasopressin Alter Aggression Differently Across Testing Conditions in Monogamous and Non-Monogamous Peromyscus Mice', *Aggressive Behavior*, 31, 189–199.

Burnham, T.C. (2007), 'High-Testosterone Men Reject Low Ultimatum Game Offers', *Proceedings of the Royal Academy (B)*, 274, 2327–2330.

Camerer, C.F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, NJ: Princeton University Press.

Carter, C. Sue (2007), 'Sex Differences in Oxytocin and Vasopressin: Implications for Autism Spectrum Disorders?' *Behavioural Brain Research*, 176, 170–186.

Coccaro, E.F., Kavoussi, R.J., Hauger, R.L., Cooper, T.B., and Ferris, C.F. (1998), 'Cerebrospinal Fluid Vasopressin Levels Correlates with Aggression and Serotonin Function in Personality-Disordered Subjects', *Archives of General Psychiatry*, 55, 708–714.

Dabbs, J.M., and Dabbs, M.G. (2000), *Heroes, Rogues, & Lovers: Testosterone and Behavior*, New York: McGraw-Hill.

De Quervain, D.J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004), 'The Neural Basis of Altruistic Punishment', *Science*, 305, 1254–1258.

De Waal, F.B.M. (2006), *Primates and Philosophers: How Morality Evolved*, Princeton, NJ: Princeton University Press.

——— (2008), 'Putting the Altruism Back Into Altruism: The Evolution of Empathy', *Annual Review of Psychology*, 59, 279–300.

Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., and Fazio, F. (1997), 'Brain Activity During Observation of Actions: Influence of Action Content and Subject's Strategy', *Brain*, 120, 1763–1777.

Delgado, M.R., Frank, R.H., and Phelps, E.A. (2005), 'Perceptions of Moral Character Modulate the Neural Systems of Reward During the Trust Game', *Nature Neuroscience*, 8, 1611–1618.

Donaldson, Zoe R., and Young, L.J. (2009), 'Sociality Oxytocin, Vasopressin, and the Neurogenetics of Sociality', *Science*, 322, 900–904.

Dovidio, J.F. (1984), 'Helping Behavior and Altruism: An Empirical and Conceptual Overview', in *Advances in Experimental Social Psychology* (Vol. 12), ed. L. Berkowitz, New York: Academic Press, pp. 361–427.

Dovidio, J.F., Piliavin, J.A., Gaertner, S.L., Schroeder, D.A., and Clark, R.D. III (1991), 'The Arousal: Cost-Reward Model and the Process of Intervention', in *Review of Personality and Social Psychology. Volume 12: Prosocial Behavior*, Newbury Park, CA: Sage, pp. 86–118.

Edgeworth, F.Y. ([1881] 1967), *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*, New York: Augustus M. Kelley.

Fehr, E., and Gachter, S. (2000), 'Cooperation and Punishment in Public Goods Experiments', *American Economic Review*, 90, 980–994.

Frith, C., and Frith, U. (1999), 'Interacting Minds - A Biological Basis', *Science* 286(5445), 1692–1695.

Güth, W., Schmittberger, R., and Schwarze, B. (1982), 'An Experimental Analysis of Ultimatum Bargaining', *Journal of Economic Behavior and Organization*, 3(4), 367–388.

Harris, J.H., Rushton, H.J., Hampson, E., and Jackson, D.N. (1996), 'Salivary Testosterone and Self-Report Aggressive and Pro-Social Personal Characteristics in Men and Women', *Aggressive Behavior*, 22, 321–331.

Hein, G., and Singer, T. (2008), 'I Feel How You Feel But Not Always: The Empathic Brain and its Modulation', *Current Opinion in Neurobiology*, 18, 153–158.

Insel, T.R., Young, L., Witt, D.M., and Crews, D. (1993), 'Gonadal-Steroids Have Paradoxical Effects on Brain Oxytocin Receptors', *Journal of Neuroendocrinology*, 5, 619–628.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., and Fehr, E. (2005), 'Oxytocin Increases Trust in Humans', *Nature*, 435, 673–676.

Kurzban, R., DeScioli, P., and O'Brien, E. (2007), 'Audience Effects on Moralistic Punishment', *Evolution and Human Behavior*, 28, 75–84.

Levitt, S.D., and List, J.A. (2007), 'What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?', *Journal of Economic Perspectives*, 21, 153–174.

Lipps, T. ([1903] 1906), *Ästhetik. Psychologie des Schönen und der Kunst* (2 vols.), Hamburg/Leipzig: Leopold Voss.

Marshall, A. ([1890] 1961), *Principles of Economics* (9th ed.), London: Macmillan.
Mehta, P.H., and Josephs, R.H. (2006), 'Testosterone Change After Losing Predicts the Decision to Compete Again', *Hormones and Behavior*, 50, 684–692.
Morhenn, V.B., Park, J.W., Piper, E., and Zak, P.J. (2008), 'Monetary Sacrifice Among Strangers Is Mediated by Endogenous Oxytocin Release After Physical Contact', *Evolution and Human Behavior*, 29, 375–383.
Newman, M., and Josephs, R.A. (2009), 'Testosterone as a Personality Variable', *Journal of Research in Personality*, 43, 258–259.
Popma, A., Vermeiren, R., Geluk, C.A.M.L., Rinne, T., van den Brink, W., and Knol, D.L. (2007), 'Cortisol Moderates the Relationship Between Testosterone and Aggression in Delinquent Male Adolescents', *Biological Psychiatry*, 61, 405–411.
Preston, Stephanie D., and de Waal, Frans B.M. (2002), 'Empathy: Its Ultimate and Proximate Bases', *Behavioral and Brain Sciences*, 25, 1–72.
Rabin, Matthew (1993), 'Incorporating Fairness Into Game Theory and Economics', *The American Economic Review*, 83, 1281–1302.
Sally, D. (2000), 'A General Theory of Sympathy, Mind-Reading, Social Interaction, With an Application to the Prisoner's Dilemma', *Social Science Information*, 39, 567–634.
——— (2001), 'On Sympathy and Games', *Journal of Economic Behavior and Organization*, 44, 1–30.
Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., and Cohen, J.D. (2003), 'The neural basis of economic decision-making in the ultimatum game', *Science*, 300, 1755–1758.
Simon, Herbert A. (1957), *Models of Man: Social and Rational, Mathematical Essays on Rational Human Behavior in a Social Setting*, New York: John Wiley & Sons.
Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., and Frith, C.D. (2004), 'Empathy for Pain Involves the Affective but not Sensory Components of Pain', *Science*, 303, 1157–1162.
Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, J.R., and Frith, C.D. (2006), 'Empathic Neural Responses are Modulated by the Perceived Fairness of Others', *Nature*, 439, 466–469.
Smith, A. (1759] 1982), *The Theory of Moral Sentiments*, eds. D.D. Raphael and A.L. Macfie, Vol. I of the Glasgow Edition of the Works and Correspondence of Adam Smith. Indianapolis: Liberty Fund.
Smith, V.L. (1998), 'The Two Faces of Adam Smith', *Southern Economic Journal*, 65, 1–19.
Sober, E., and Wilson, D.S. (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: Harvard University Press.
Stanton, A.A. (2007), 'Neural Substrates of Decision-Making in Economic Games', *Journal of Dissertations*, 1(1), e1–59.
Thompson, R.R., George, K., Walton, J.C., Orr, S.P., and Benson, J., (2006), 'Sex-specific Influences of Vasopressin on Human Social Communication', *Proceedings of the National Academy of Sciences USA*, 103(20), 7889–7894.
Veblen, Thorstein B. (1898), 'Why Is Economics Not an Evolutionary Science?', *Quarterly Journal of Economics*, 12(3), 373–397.
Young, L.J., Nilsen, R., Waymire, K.G., MacGregor, G.R., and Insel, T.R. (1999), 'Increased Affiliative Response to Vasopressin in Mice Expressing the V1a Receptor From a Monogamous Vole', *Nature*, 400, 766–768.
Young, L.J., and Wang, Z. (2004), 'The Neurobiology of Pair Bonding', *Nature Neuroscience*, 7(10), 1048–1054.
Zak, P.J. (2005), 'Trust: A Temporary Human Attachment Facilitated by Oxytocin', *Behavioral and Brain Sciences*, 28(3), 368–369.
Zak, P.J. (2008), 'The Neurobiology of Trust', *Scientific American*, (June), 88–95.
Zak, P.J. (in press), 'The Physiology of Moral Sentiments', *Journal of Economic Behavior and Organization*.
Zak, P.J., Borja, K., Matzner, W.T., and Kurzban, R. (2005), 'The Neuroeconomics of Distrust: Sex Differences in Behavior and Physiology', *American Economic Review*, 95, 360–364.
Zak, P.J., Kurzban, R., and Matzner, W.T. (2004), 'The Neurobiology of Trust', *Annals of the New York Academy of Sciences*, 1032, 224–227.
Zak, P.J., Kurzban, R., and Matzner, W.T. (2005), 'Oxytocin is Associated with Human Trustworthiness', *Hormones and Behavior*, (48), 522–527.

Zak, P.J., Kurzban, R., Ahmadi, S., Swerdloff, R.S., Park, J.W., Efremidze, L., Redwine, K., Morgan, K., and Matzner, W. (2009), 'Testosterone Administration Decreases Generosity in the Ultimatum Game', *Public Library of Science ONE*, 4(12), e8330. doi:10.1371/journal.pone.0008330.

Zak, P.J., Stanton, A.A., and Ahmadi, S. (2007), 'Oxytocin Increases Generosity in Humans', *Public Library of Science ONE*, 2, e1128.