The Trade-off of Selfishness and Empathy in the Human Brain

Dan Levine[1]

Department of Psychology, University of Texas,
Arlington, TX 76019-0528, USA

Abstract

Self-interest and concern for others are both woven into the human evolutionary makeup, and both involve all levels of the brain. A neural network model has previously simulated the tradeoff between those two considerations and the balance between them achieved through the operation of the executive system centered in the prefrontal cortex. More recent behavioral and neural data are leading to refinement of this network model, which is in progress. These recent data are also suggesting both policies and customs that might tip the balance society-wide more toward cooperation and concern for one another.

Introduction

My interest in biological neural networks began when I was a student in the late 1960s and early 1970s. Like many of my fellow American students, I saw the big problems of the world and wanted to make a difference in solving them. War seemed to me a terrible waste, and so did the poverty of some while others were wealthy. And we were beginning to be concerned that industrial growth was making our air unfit to breathe and our water unfit to drink.

My father was a scientist who had done ground-breaking research on the biochemical action of insulin. He told me once that for him the excitement of doing science was looking at phenomena in the real world and asking "Why?" So I started asking "Why?" about war and income inequality and environmental damage. It did not seem to me that the answers could be found solely in pragmatic and economic considerations. There also needed to be a psychological component to human behavior that generates these undesirable outcomes.

Yet the other side of the "Why?" question is why do most humans (with the possible exception of some sociopaths) *find* these outcomes undesirable? Why do we frequently empathize with each other, have concern for the welfare of others – even people genetically unrelated to us – cooperate with each other?

---

[1] Dan Levine is Research Professor of Psychology at the University of Texas at Arlington, a Fellow of the International Neural Network Society (INNS), having served as President of INNS in 1998 and on the INNS Board of Governors from 1995 to 2008. He is the author of 'Healing the Reason-Emotion Split' (Routledge, 2021) and 'Introduction to Neural and Cognitive Modeling' (3rd edition, Routledge, 2019). Dan's research centres on cognitive-emotional interactions in human decision making. See: http://blog.uta.edu/levine, levine@uta.edu.

The 1960s/1970s student culture set out to make each of the professions we were entering (in my case mathematical and scientific research) "relevant" toward solving social and political problems. So I set out to ask what sorts of social arrangements promote people living together peacefully and cooperatively. To begin to answer that question, it was necessary to understand better the nature of human brains and minds. That led me halfway through graduate school to study neuroscience and psychology, and to tie those fields to my background in mathematics via neural network modeling.

The Triune Brain and the Conflict Systems Model

Evolution has favored the development of self-interest because humans, like other animals, need to be able to feed themselves, reproduce, and defend their lives and territory. Yet evolution has also favored the development of concern for each other, because humans, like other animals, need to be able to cooperate in efforts needed for survival and reproduction, which are facilitated by families and friendships. The self-interest system and the cooperation system sometimes work together but sometimes are in conflict. The development of the mammalian (especially human) cerebral cortex has enabled executive control of the relative activities at any given time of the self-interest and cooperative systems. This evolutionary complexity is the basis for the *conflict systems neurobehavioral model* (Cory, 1999, 2004).

The search for a neural basis of conflicting evolutionary systems goes back to the pioneering nineteenth and early twentieth century work of William James (James, 1890/1981). When James wrote it was popular to believe that humans, being more "advanced" than other mammals, had no need for instincts. James disagreed with that belief. Like Charles Darwin, he emphasized our similarities with other animals as much as our differences, making instincts one of his three prime categories of mental function along with thoughts and emotions. He noted that instincts, emotions, and thoughts are closely interrelated: for example, insights on intellectual problems are among the strongest generators of emotions, and instincts carry with them appropriate emotional reactions.

James did not have enough known neuroscience at his disposal to tie thoughts, emotions, and instincts into an overall theory of the mind and brain. The task of building such a theory was taken up later by the behavioral neuroscientist Paul MacLean, who published between the 1960s and 1990s (see in particular, MacLean, 1985, 1990). MacLean emphasized that in evolution the human brain retained a great many processes from other animals but built on top of those processes. Humans and other mammals inherited from reptiles the brain structures that enable a variety of common behavioral patterns. These patterns include, for example, establishing and defending territory; ritual postures of domination and surrender; foraging, hunting, and hoarding food; forming social groups and social hierarchies; greeting and grooming one another; and courtship, mating, and breeding. Yet adding more complex brain structures, in the limbic system and the cortex, enables mammals to engage in behaviors not typical of reptiles, such as parental care.

From decades of observing many species of animals, particularly monkeys and lizards, MacLean developed the notion of the *triune brain* consisting roughly of three parts that arrived at different stages of evolution. At the deepest levels are the brain stem, midbrain, and basal ganglia forming the "reptilian brain" (or R-complex) involved in automatic, instinctive, often routine behavior. Just above the R-complex is the limbic system, the center of the "old mammalian brain," involved in emotions such as fear, love, and anger. Finally, at the very top is the cerebral cortex, also called the "new mammalian brain," which is necessary for our verbal and intellectual abilities.

Cory (1999, 2004) attributed our self-preservation instincts to the R-complex and our empathic capacities to the growth of the limbic system. In Cory's framework, our ability to empathize with and be concerned for others starts with parental concern for offspring. We should add to our discussion the caveat that the word "empathy" has a variety of meanings and connotations; specifically, emotional empathy (the tendency to feel good about another person's happiness or bad about their unhappiness) and cognitive empathy (the ability to understand another person's perspective) do not always correlate or activate the same brain areas (see Levine, 2021, for review). It is plausible that many sociopaths possess cognitive but not emotional empathy, and the reverse may be true for many individuals on the autistic spectrum.

In the evolution from reptiles to mammals, the expansion of parental behavior is connected by MacLean (1990, Chapter 21) with expansion of a part of the limbic system that includes the thalamus, cingulate cortex, and their connections. The cingulate is a versatile part of the brain whose functions include integration of affect and cognition (Levine, 2021) and resolution of conflicting information. Stevens et al. (2011) found a particular type of neuron plentiful in the cingulate that could play a role in its varied functions:

> … Von Economo neurons … found only in cingulate (pACC and MCC) and insular cortices. Von Economo neurons are present in great apes and humans, but in no other primates. They are more numerous in humans … They are much larger than pyramidal neurons, suggesting faster transmission of information between brain regions, and possibly more connections. … It has been suggested that Von Economo neurons perform an adaptive function by helping humans and great apes act quickly on an instinctual/intuitive level in social situations.

Neural Network Modeling of the Triune Brain and Conflict Systems

Levine and Jani (2002) and Levine (2006) developed a neural network model of the trade-off between two criteria in a competitive system, labeling the two criteria selfishness and empathy. The network is defined by a dynamical system of nonlinear differential equations and simulated in MATLAB. That model is built on the previous work of Grossberg and Levine (1975) on modeling shunting excitation and inhibition with biases. Specifically, if x_1 is the activity of a selfishness node and x_2 the activity of an empathy node, those two variables satisfy the differential equations

$$dx_1/dt = -0.5x_1 + (B_1 - x_1)\,\mathrm{sig}(x_1) - 0.2x_2 \qquad (1a)$$
$$dx_2/dt = -0.5x_2 + (B_2 - x_2)\,\mathrm{sig}(x_2) - 0.2x_1 \qquad (1b)$$

where B_1 and B_2 denote bias weights (for selfishness and empathy, respectively), and sig is a sigmoid function. Specifically, sig(x) is first set to equal $1/(1 + \exp(-20x + 10))$ then truncated to 1 if that function value is greater than 1, and to 0 if that function value is less than 0. In Equations

(1a) and (1b), the first term on the right hand side denotes decay; the second term denotes shunting excitation and causes x_i to be bounded above by B_i for $i = 1, 2$; and the third term denotes competition from the other node.

Now how are the bias weights B_1 and B_2 chosen? Cory (1999, 2004) notes that for each of us, the trade-off between self-interest and concern for others is a balancing act, so that if either one is neglected for too long, we change the bias in the direction of favoring the neglected criterion. The top graph in Figure 1 shows the behavior of the selfishness and empathy node activities x_1 and x_2 if biases start out at $B_1 = 10$ and $B_2 = 1$ and then switch to the opposite when x_2 gets below a fixed threshold, then switch back when x_1 gets below that threshold, and so forth. As the graph shows, the biases oscillate periodically and never reach an equilibrium. The bottom graph of Figure 1 shows that equilibrium between selfishness and empathy is achieved when a third node is added to the network representing an idealized frontal lobe executive. The frontal node causes biases in favor of empathy or selfishness to shift when either variable is too low, but in a gradual instead of a sudden manner.

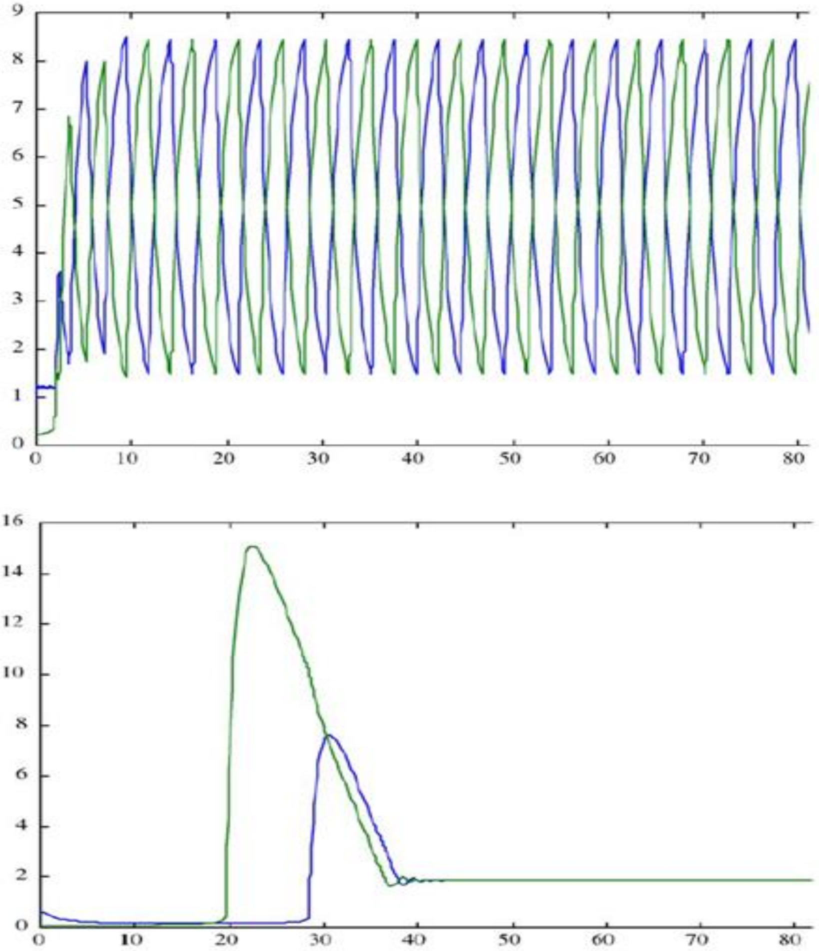

Figure 1. Activities of selfishness and empathy nodes in the network of Levine and Jani (2002). Top: without frontal lobe node. Bottom: with frontal lobe node. (Adapted from Journal of Socio-Economics, 35, Daniel S. Levine, Neural Modeling of the Dual Motive Theory of Economics, pp. 613-625, Copyright 2006, with permission from Elsevier.)

What Distinguishes Bullies

The labeling of the nodes by Levine and Jani (2002) as selfishness and empathy was arbitrary. They were simply competing biases that could have had any interpretation, and we attached those labels to the biases. So currently we are working on extending the network simulations to incorporate behavioral and neural data on the bases for empathy and selfishness.

To get a better idea of the possible neural bases for selfishness and empathy, we need to examine the neuroscience and behavior of people who are the most extreme on the side of a bias toward selfishness. Those would be bullies, people who actually seem to take delight in making life worse for others. Michael Perino is a behavioral neuroscientist who has studied bullies extensively. On the neural level, Perino et al. (2019) showed that self-reported adolescent bullies were particularly sensitive to social exclusion, with higher activity in several reward-related brain regions, including the ventral striatum, amygdala, medial prefrontal cortex, and insula when viewing scenes of social exclusion, than when viewing scenes of social inclusion. On the behavioral level, Perino (personal communication, October 2023) reported that bullies as compared to non-bullies, show greater sensitivity to relative, rather than absolute, reward. For example, given Scenario A, where you and a friend or co-worker both earn $100,000 a year, and Scenario B, where you earn $80,000 and the friend or co-worker earns $40,000, the majority of people choose A, but bullies choose B. In other words, they are willing to sacrifice some of their own rewards if it means someone else is less rewarded than they are.

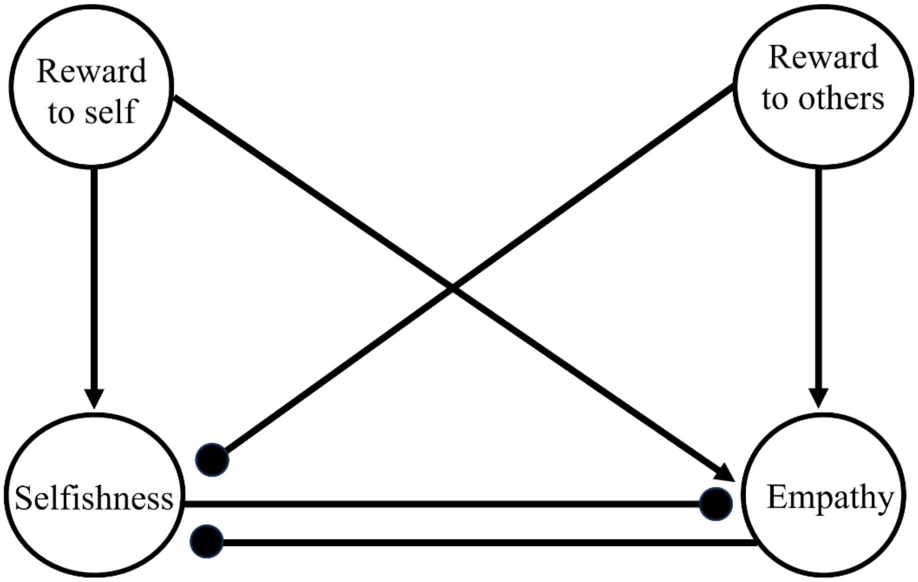

Figure 2. Network that incorporates idealized definitions of selfishness (wanting to maximize relative reward) and empathy (wanting to maximize reward for both oneself and others). Arrows denote excitation or positive connection; filled circles denote inhibition or negative connection.

Our current computational studies in progress are based on Perino's insights about bullies' relative reward sensitivity. As shown in Figure 2, our expanded neural network incorporates a definition of the selfishness node as sensitive to relative reward, that is, responding positively to whatever rewards one obtains from the external world and negatively to rewards that others obtain. The empathy node, by contrast, responds positively both to rewards one obtains, and rewards others obtain. The caveat should be added that selfishness often means indifference to rewards or punishments going to other people, rather than actively desiring they not be rewarded. However, the predilection for relative reward seems to be part of the competitive side of human nature which is common in the human race and not restricted to those who would classify as bullies, so we take that definition of selfishness for simplicity.

Neural network simulations in progress by Chaithanya Kota and myself are directed toward extending the analysis of Levine and Jani (2002) to include the network of Figure 2, with or without the addition of the simulated prefrontal executive node. The aim is to study the network behavior under varying conditions of reward for the person represented by the network and for other people that person may interact with. This network will still not incorporate what is known about specific brain regions and neural transmitters involved in selfishness and empathy, which will hopefully be the subject of network studies further in the future.

Possible Neural Bases for Selfishness and Empathy Biases

Grossberg and Levine (1975) analyzed mathematically the type of biased competitive neural network of which Equations (1a) and (1b) are an example. These authors identified three generic sources for the biases B_i: (1) Development (what stimuli are experienced); (2) Attention (what stimuli are salient, emotionally or otherwise); (3) Statistical errors in network design. In this exposition we focus on source (2), selective attention.

At the time Grossberg and Levine appeared, there was no developed theory for the operation of these biases in the brain. But we can make a few general statements about brain regions based on knowledge that has accrued since then. We can conjecture that the B_i exert modulatory influences on connections to the x_i. The x_i are representations of high-level concepts such as empathy and selfishness that have emotional significance, so we tentatively locate those nodes in the orbitofrontal cortex, the prime cortical processor of emotional and social information.

The attentional modulation that produces the biases B_i comes from a variety of sources in other parts of the brain. Clearly these sources include the emotional centers in the anterior cingulate and other parts of the limbic system (Cory, 1999, 2004; MacLean, 1985, 1990). These sources also include two hormones whose effects can be considered complementary: oxytocin and cortisol.

Broadly speaking, oxytocin – which is found only in mammals and was originally found in connection with maternal milk secretion – tends to be associated with bonding and cooperation between humans or other animals. For example, Insel and Winslow (1998) studied two related species of voles, one of which has pair bonds between females and males with both partners involved in parenting young, the other of which is promiscuous with uninvolved fathers. The species that has pair bonding has more oxytocin receptors in brain reward areas than the promiscuous species. Uvnäs-Moberg (1998) found that oxytocin increases in humans with lactation and good sexual intercourse – and suggested it also increases with positive social bonding. Kosfeld et al. (2005) showed that intranasal oxytocin increases trust in humans – although the later work of De Dreu et al. (2011) showed that such trust can be selective toward one's own ethnic group.

Oxytocin also depresses the amount of circulating cortisol, a hormone typically released in stressful situations. Conversely, excessive stress decreases the amount of circulating oxytocin, and if this stress is chronic (as with childhood abuse or combat exposure) it can decrease the capacity for affiliative behavior (Henry & Wang, 1998; Perry et al., 1995). Cortisol is involved with the neural transmitter norepinephrine and the sympathetic part of the autonomic nervous system in the classic fight-or-flight response. Oxytocin, by contrast, lowers sympathetic system activity and raises activity of the parasympathetic part of the autonomic system, facilitating responses to stress that involve tending and befriending rather than fighting or fleeing (Taylor et al., 2000).

One could almost say humans have both an "oxytocin system" promoting bonding and empathy, and a "cortisol system" promoting aggressive self-preservation. Clearly, the two systems complement each other, and we need both of them. Yet the type of society that promotes cooperation and bonding needs to be one that is biased toward the "oxytocin" side, with "cortisol" entering the picture when stresses and challenges occur.

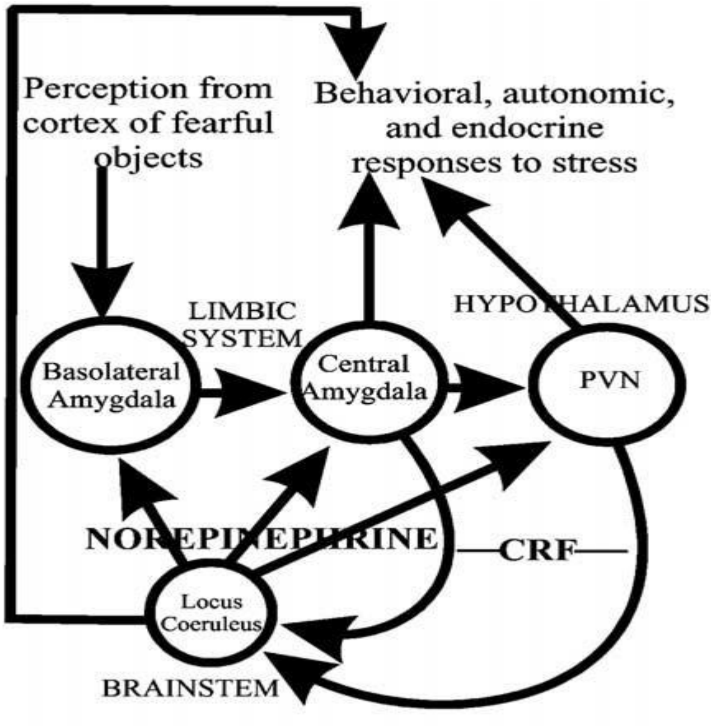

Figure 3. Part of the interactive feedback system between CRF (the precursor to cortisol) and norepinephrine stress-related systems in subcortical emotion-related regions of the brain. (Brain and Mind, Riane Eisler and Daniel S. Levine, 3, 9-52, 2002, reproduced with permission from SNCSC.)

Both cortisol and oxytocin have known effects on subcortical emotional regions of the brain, in addition to their effects on the prefrontal cortex. For cortisol (actually, a substance called CRF which is a biochemical precursor to cortisol), many of the subcortical pathways were mapped out by Koob (1999) and are described in Figure 3. The perception from the cortex of a fearful object activates a pathway that goes from the basolateral nucleus of the amygdala to the amygdala's central nucleus to the paraventricular nucleus (PVN) of the hypothalamus. The central nucleus and PVN in turn activate behavioral, autonomic, and endocrine responses to stress. The central nucleus and PVN also release CRF and send it as a neurotransmitter to the locus coeruleus, the norepinephrine-releasing nucleus of the midbrain. Noradrenergic signals from the locus coeruleus in turn activate all three of the amygdala-hypothalamic regions involved and potentiate the behavioral, autonomic, and endocrine responses. These pathways create the positive feedback loop involved in the fight-or-flight response, sometimes called the *hypothalamic-pituitary-adrenocortical (HPA) axis*.

The subcortical pathways for the bonding-related effects of oxytocin are best understood in combination with another substance called vasopressin, which is a neuropeptide like oxytocin. The network of Figure 4 was devised by Eisler and Levine (2002) based on data about brain regions to which oxytocin and vasopressin bind in voles. Both peptides have different binding patterns in the brain for the pair-bonding prairie vole than for the promiscuous montane vole, and Figure 4 is based on the prairie vole patterns. Oxytocin, as befits its connection to "mother's milk," has more connection to maternal function, whereas vasopressin has more connection to paternal functions like protection. Yet Cho et al. (1999) found that both hormones are required in both sexes for pair bonding in prairie voles, a datum that argues in favor of fluid gender roles in humans.

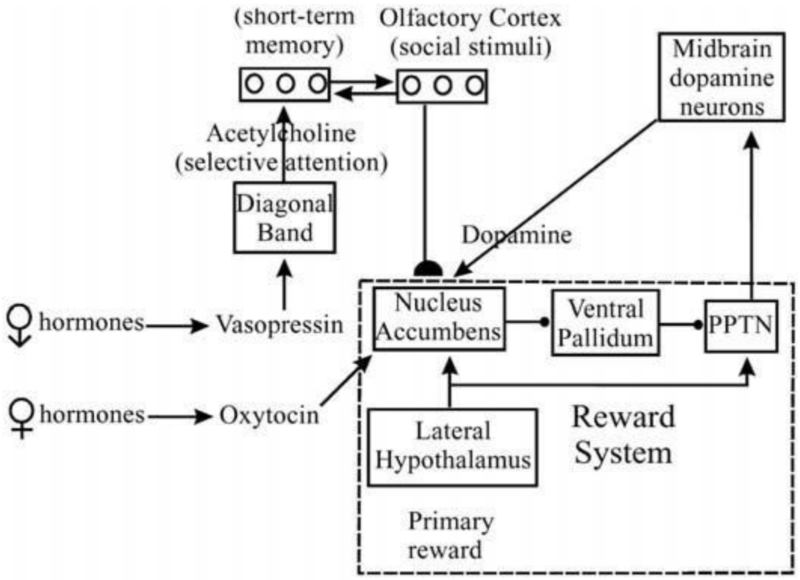

Figure 4. Proposed network for subcortical bonding effects of oxytocin and vasopressin. PPTN is the pedunculopontine tegmental nucleus, a part of the midbrain. Ventral pallidum and nucleus accumbens are parts of the basal ganglia. Both of these areas, along with the lateral hypothalamus and the nucleus accumbens, parts of the neural circuit for processing rewards. Arrows between boxes represent excitatory (glutamatergic?) connections; filled circles represent inhibitory (GABAergic?) connections; semicircles represent modifiable connections. (Brain and Mind, Riane Eisler and Daniel S. Levine, 3, 9-52, 2002, reproduced with permission from SNCSC.)

Conclusion: Societal Implications

The neural pathways that are sketched in this article partly answer the "Why?" questions posed at the beginning about selfishness, empathy, and interactions between them. Now we return to the other question of how neuroscience can illuminate which social arrangements – including both official government policies and informal customs – promote empathy and cooperation, keeping selfishness within bounds.

A huge body of behavioral results on both humans and non-human animals indicate that people respond, often in lasting ways, to the amount of friendliness or hostility in their social environments. Chronic childhood stress inclines people as adults toward greater fight-or-flight or dissociative responses and away from caring responses (e.g., Perry et al., 1995). Conversely, children raised in caring societies and homes typically become caring adults (e.g., Eisenberg, 1992).

Much has been written about the importance of social, political, and economic policies that create a strong safety net, reward caring more than killing, and invest in the welfare of children (e.g., Eisler, 2002, 2007; Taylor, 2002). All these policies increase the empathy bias and decrease the selfishness bias in the network of Figure 2 for each of us. But for policies to be effective, they must be accompanied by subtle manipulation of our customs and prevailing habits of thought. One of our more unfortunate pervasive human habits is to conflate bullying with strong leadership – which accounts for the attractiveness of authoritarian dictators to large numbers of people.

Perino et al. (2019) found that the adolescent bullies he studied were not deficient in empathic accuracy, that is, in picking up social cues about whether the people they interacted with were happy or unhappy. By contrast, these authors said that

> Aggression stemming from socially dissonant responses — e.g., taking pleasure in others' pain or expressing displeasure when seeing others' happiness—is more explanatory than traditional empathic explanations (p. 1473).

They suggested that for many of these teenagers, bullying was a deliberate social strategy with the aim of attaining leadership or exalted social positions for themselves. Hence, they added, the most effective type of intervention in the lives of bullies is to show them how they can attain leadership without bullying. I would add, based in part on my own experience with academic departments and houses of religious worship, that this should be combined with encouraging non-bullies to discern signs of bullying behavior and reject opportunities to entrust leadership positions (e.g., department chair or clergyperson) to candidates who show those signs. All too frequently I have seen colleagues or fellow worshippers drawn to a bully because they see the bully as dynamic, charismatic, a person who gets things done or "shakes things up." Finding the right combination of charismatic leadership with the caring emotional side of empathy is difficult but must be pursued for organizations to flourish.

The structure of neural pathways for both selfishness and empathy indicates that there are no intrinsically "good" or "evil" parts of our brains. Self-interest and concern for others involve overlapping brain regions at all three levels of MacLean's (1990) triune brain. We need to keep constantly training our prefrontal executive systems, and when possible each other's, to maintain an optimal balance between selfishness and empathy.

References

Cho, M. M., DeVries, C., Williams, J.R., & Carter, C.S. (1999). The effects of oxytocin and vasopressin on partner preferences in male and female prairie voles (Microtus ochrogaster). *Behavioral Neuroscience*, *113*, 1071-1079.

Cory, G. A., Jr. (1999). *The Reciprocal Modular Brain in Economics and Politics: Shaping the Rational and Moral Basis of Organization, Exchange, and Choice*. New York: Plenum Press.

Cory, G. A., Jr. (2004). *The Consilient Brain: The Bioneurological Basis of Economics, Society, and Politics*. New York: Kluwer Academic.

De Dreu, C. K. W., Lindred, L., Gerben, G., Van Kleef, A., Shalvi, S., & Handgraaf, M. J. J. (2011). **Oxytocin** promotes human ethnocentrism. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 1262-1266.

Eisenberg, N. (1992). *The Caring Child.* Cambridge, MA, USA: Harvard University Press.

Eisler, R. (2002). *The Power of Partnership: Seven Relationships that will Change your Life.* Novato, CA, USA: New World Library.

Eisler, R. (2007). *The Real Wealth of Nations: Creating a Caring Economy.* San Francisco: Berrett-Koehler.

Eisler, R., & Levine, D. S. (2002). Nurture, nature, and caring: We are not prisoners of our genes. *Brain and Mind, 3*, 9-52.

Grossberg, S., & Levine, D. S. (1975). Some developmental and attentional biases in the contrast enhancement and short-term memory of recurrent neural networks. *Journal of Theoretical Biology, 53,* 341-380.

Henry, J.P., & Wang, S. (1998). Effects of early stress on adult affiliative behavior. *Psychoneuroendocrinology*, *23*, 863-875.

Insel, T. R., & Winslow, J. T. (1998). Serotonin and neuropeptides in affiliative behaviors. *Biological Psychiatry, 44*, 207-219.

James, W. (1890/1981). *The Principles of Psychology (Vols. 1 and 2),* published originally in 1890 by Henry Holt, New York; reprinted in 1981 by Harvard University Press, Cambridge, MA, USA.

Koob, G. F. (1999). Corticotropin-releasing factor, norepinephrine, and stress. *Biological Psychiatry, 46*, 1167-1180.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673-676.

Levine, D. S. (2006). Neural modeling of the dual motive theory of economics. *Journal of Socio-Economics, 35*, 613-625.

Levine, D. S. (2021). *Healing the Reason-Emotion Split: Scarecrows, Tin Woodmen, and the Wizard*. New York and London: Routledge.

Levine, D. S., & Jani, N. G. (2002). *Toward a neural network theory of the triune brain.* In G. A. Cory, Jr., & R. Gardner (Eds.), *The neuroethology of Paul MacLean: Convergences and frontiers* (pp 383-394). Westport, CT: Praeger.

MacLean, P. D. (1985). Brain evolution relating to family, play, and the separation call, *Archives of General Psychiatry*, *42*, 405-417.

MacLean, P. D. (1990). *The Triune Brain in Evolution*. New York: Plenum Press.

Perino, M. T., Guassi Moreira, J. F., & Telzer, E. H. (2019). Links between adolescent bullying and neural activation to viewing social exclusion. *Cognitive, Affective & Behavioral Neuroscience, 19*(6), 1467-1478, 2019.

Perry, B. D., Pollard, R. A., Blakley, T. L., Baker, W. L., & Vigilante, D. (1995). Childhood trauma, the neurobiology of adaptation, and "use-dependent" development of the brain: How "states" become "traits." *Infant Mental Health Journal, 16*, 271-291.

Stevens, F. L., Hurley, R. A., & Taber, K. H. (2011). Anterior cingulate cortex: Unique role in cognition and emotion. *Journal of Neuropsychiatry and Clinical Neuroscience, 23*, 119-124.

Taylor, S. E. (2002). *The Tending Instinct: How Nurturing is Essential for Who We Are and How We Live*. New York: Henry Holt and Company.

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review, 107*, 411-429.

Uvnäs-Moberg, K. (1998). Oxytocin may mediate the benefits of positive social interaction and emotion. *Psychoneuroendocrinology, 23*, 819-835.