{5}

The Prosocial Brain

PERCEIVING OTHERS IN NEED AND ACTING ON IT

Alek Chakroff and Liane Young

In early 2011, a 9.03 magnitude earthquake struck east of Tōhoku, Japan, creating a tsunami that caused widespread destruction and the deaths of over 15,000 people. Like many tragedies, this one elicited an outpouring of aid from people around the world. American citizens alone donated over $650 million to the relief effort (Japan Center for International Exchange, 2012). Locally, hundreds of elderly Japanese citizens volunteered to clean up the Fukushima Daiichi nuclear power plant, arguing that they were less susceptible to radiation poisoning than younger volunteers and in any case had fewer remaining years for radiation effects to emerge (Lah, 2011). Why did American citizens offer aid to people who would likely never return the favor or even meet their benefactors? Why did elderly Japanese citizens form what was known as the "suicide corps," cleaning up toxic waste at their own ultimate personal risk?

Prosocial behavior, or behavior carried out with the intention to help others (Eisenberg, Fabes, & Spinrad, 2006), may be rooted in different reasons, some more noble than others. We may help others anticipating direct or indirect reciprocal benefits or, more generally, social credit; we may help others to avoid negative social consequences; we may help others for purely instrumental reasons (e.g., tax write-offs); and we may help others out of an altruistic motivation, a selfless desire to increase their well-being, without anticipated selfish benefit (Batson et al., 1989).

Debate rages about how a noncolonial social species could have evolved altruistic tendencies such that individuals incur costs to benefit others. Fitness comparisons at the level of individuals—the individual altruist and the individual miser—reveal the miser as the clear winner. Yet, groups of altruists are thought to outperform groups of misers (Boyd & Richerson, 2002; Sober & Wilson, 1998), though altruism may also emerge particularly robustly in groups that share genetic variance (West, Mouden, & Gardner, 2011). Altruism could therefore be selected for at the cultural and genetic level. Regardless of the evolutionary origins of altruistic behavior, the present review concerns the psychologically proximate motivations for prosocial behavior. Indeed, ultimate "ends" will not necessarily correspond with proximate "means" (Mayr, 1993; Von Hippel & Trivers, 2011). Thus, even in the absence of true biological altruism, we need not predict that individuals never act out of "genuine" altruistic motivations.

What psychological capacities does an individual need in order to engage in prosocial behavior as defined above (Eisenberg et al., 2006)? Acting to intentionally benefit others first requires that one perceive others—other living beings with needs and desires—in other words, acting to benefit others requires social cognition. We offer aid to living creatures, not rocks; friends, not foes; family more than strangers. However, perceiving a need in others does not always lead to sufficient motivation to help. Millions around the world are starving to death, and members of developed countries have the power to save lives by donating some or most of their income to charitable organizations (Singer, 1997). Why do they not donate more? It may be difficult to empathize with the plight of unseen others across the globe—we simply do not feel bad enough. Alternatively, we may be less likely to help those who have no power to return the favor. Prosocial behavior requires not only the perception of an opportunity to help, but the motivation to act. This motivation may be rooted in empathetic processes through feeling the pain of another and desiring to reduce the pain. The motivation to help others may also require the ability to anticipate some future reward, whether that reward is anticipated reciprocity (e.g., you scratch my back) or the "warm glow" of being a good person.

This review is organized around the distinct psychological processes underlying prosocial behavior. First, we outline the processes that are likely required for prosocial action to occur: perception of life, minds, and negative mental states in others as well as anticipation of positive outcomes. We discuss the neural mechanisms underlying each of these psychological processes and use this discussion as a lens through which to interpret research directly relating prosocial behavior to the brain. Finally, we discuss ambiguities in the interpretation of the current research and suggest future directions.

Psychological Underpinnings of Prosocial Behavior

Would you be more likely to intervene to stop a child from pulling wings off a fly, or to stop a child from pulling leaves off a fallen tree branch? To make this distinction, one must distinguish animate from inanimate forms. What if the child were trying to pull the wings off a bird? We distinguish between different kinds of minds, and some may be perceived as more capable of experiencing pain. We care about the needs of others and about our relative influence on their affective state. We are adept at distinguishing between kinds of life forms, kinds of minds, and kinds of mental states—and we adjust our prosocial behavior accordingly. In the following sections, we provide an overview of the neural processes underlying the perception of animate objects, minds, mental states, and reward anticipation and we discuss the contribution of these processes to prosocial behavior.

PERCEIVING LIFE

Humans evolved in a harsh environment, in which the ability to distinguish living from nonliving enabled people to find food and to avoid becoming food. Animacy perception is automatic. For a demonstration, watch a low-budget cartoon where a small triangle shakes and quickly moves away from a larger triangle, which then proceeds in the same direction, in what appears, unmistakably, to be a hostile pursuit (Heider & Simmel, 1944). Although the shapes bear little surface resemblance to animate agents (e.g., no eyes, no limbs), we infer agency and experience from motion that is contingent (e.g., reactive to others) or self-generated (Gelman & Gottfried, 1996; Luo & Baillargeon, 2005, Scholl & Tremoulet, 2000). We also process surface cues to animacy, such as faces, automatically (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Hadjikhani, Kveraga, Naik, & Ahlfors, 2009) and from infancy on (Johnson, Dziurawiec, Ellis, & Morton, 1991). In fact, we detect agents even where they are not technically present—when we see faces in the clouds or the Virgin Mary in our toast (i.e., pareidolia).

The neural underpinnings of animacy perception have been assessed using techniques such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), aimed at measuring, primarily, the speed and spatial location of neural processing, respectively. Research using EEG has shown that faces can be neurally distinguished from nonfaces approximately 170 milliseconds after presentation (Bentin et al., 1996), even when the "face" is a sparse schematic of three dots arranged like eyes and a mouth (Hadjikhani et al., 2009). However, more subtle cues to animacy (e.g., used to distinguish the face of a live human being from the face on a mannequin) are integrated later, around 400 milliseconds after presentation (Wheatley, Weinberg, Looser, Moran, & Hajcak, 2011).

Meanwhile, fMRI research has shown that face processing takes place near the posterior fusiform gyrus (FG; Kanwisher, McDermott, & Chun, 1997), while biological motion is processed in the posterior superior temporal sulcus (pSTS; Grossman et al., 2000; Wheatley, Milleville, & Martin, 2007). Both the FG and the pSTS are sensitive to subtle cues to animacy. For example, Looser, Guntupalli, and Wheatley (2012) presented participants with pictures of humans and dogs, both animate and inanimate (e.g., realistic dolls), and found that the FG and pSTS exhibit similar responses to animate humans and dogs, but dissimilar responses to animate versus inanimate targets, suggesting that animacy is a primary driver of these regions, regardless of form (e.g., doglike versus humanlike). By contrast, regions in the inferior and lateral occipital cortex were more sensitive to the overall form of objects (including animate entities) and responded similarly to targets with a similar form, such as an animate human and an inanimate doll (Looser et al., 2012). These occipital responses can also be used to distinguish different biological kinds (e.g., birds versus bugs; Connolly et al., 2012). Together, these results reveal complementary roles for these regions: The FG and pSTS distinguish animate agents from inanimate objects, whereas the inferior and lateral occipital cortex distinguish among kinds of animate and inanimate entities.

PERCEIVING MINDS

Bugs and birds alike are living entities, but birds may seem to possess greater mental capacity. We judge some creatures to be more capable of conscious experience (e.g., pain and suffering) than others and some creatures to be more capable of action or agency than others (Gray, Gray, & Wegner, 2007). These judgments carry consequences for empathy and moral cognition. Entities seen as highly capable of experience are more readily recognized as victims. It is worse to harm a bird than a bug. Entities seen as highly capable of agency are also deemed more morally responsible for their behavior. Theft is worse if carried out by an adult man versus a small child. Some evidence suggests that these dimensions may be inversely related; entities high in agency are also low in experience, and vice versa. In an unnerving demonstration of this trade-off, participants were more willing to inflict pain on Mother Teresa, relative to a neutral target, presumably because "super agents" are not easily seen as "patients," capable of experiencing pain (Gray & Wegner, 2009).

When judging others, we assess not only their capacity for agency and experience but also their specific intent to help or hinder us (warmth) and their ability to carry out those intentions (competence; S. T. Fiske, Cuddy, & Glick, 2007). We spontaneously sort people along the dimensions of warmth and competence: friends versus foes, strong versus weak. Compare your feelings toward the homeless (low warmth, low competence), the elderly (high warmth, low competence), businessmen (low warmth, high competence), and middle-class Americans (high warmth, high competence). Research indicates that the elderly and the middle class are perceived as similarly friendly (warmth) but different in their capacity to act on generally good intentions (competence). By contrast, neither the businessmen nor the homeless are perceived as particularly well intentioned but, because of perceived differences in competence, are regarded with envy and disgust, respectively.

Dimensions of warmth and competence roughly correspond with those of trustworthiness and dominance, as assessed through trait inferences of human faces (Todorov, Said, Engell, & Oosterhof, 2008). People judge warmth and competence based on people's facial expressions and configurations (Todorov, Gobbini, Evans, & Haxby, 2007; Todorov et al., 2008), semantic stereotypical knowledge of the person (Contreras, Banaji, & Mitchell, 2012; Mitchell, Ames, Jenkins, & Banaji, 2009), and assessments of similarity to oneself (Mitchell, Macrae, & Banaji, 2006). We judge those who are happy, similar to us, and/

or part of our group to be highest in warmth and trustworthiness. Inferring social traits (and social cognition more generally) consistently recruits the medial prefrontal cortex (mPFC; Amodio & Frith, 2006; Contreras et al., 2012; Mitchell et al., 2009; Mitchell et al., 2006). For example, similar or in-group others are more likely to elicit activation in the ventral mPFC (Krienen, Tu, & Buckner, 2010; Mitchell et al., 2006), a region also implicated in self-referential processing (Jenkins & Mitchell, 2011; Kelley et al., 2002; Morrison, Decety, & Molenberghs, 2012). Perceiving close others also increases activity in "reward areas" such as the ventral striatum (VS; Cloutier, Heatherton, Whalen, & Kelley, 2008; Mende-Siedlecki, Said, & Todorov, 2012), which are often activated when a participant receives or anticipates some kind of reward, such as money, food, or social contact (Fehr & Camerer, 2007; Knutson & Cooper, 2005). Meanwhile, dissimilar or out-group others are likely to elicit activation in dorsal mPFC (Mitchell et al., 2006), as well as regions associated with negative affect, such as the amygdala and anterior insula (Harris & Fiske, 2006; Lieberman, Hariri, Jarcho, Eisenberger, & Brookheimer, 2005; Mende-Siedlecki, Said, & Todorov, 2012). Other work suggests that individuals judged to be both cold and incompetent elicit less activation in the mPFC in general, indicating a potential neural signature of dehumanization (Harris & Fiske, 2006).

PERCEIVING MENTAL STATES

We perceive not only different kinds of minds and stable mental dispositions, but also transient mental states. We make inferences about people's feelings, desires, beliefs, and intentions (Amodio & Frith, 2006; Saxe & Powell, 2006). Mental state inferences rely on a network of brain regions including the mPFC, precuneus (PC), and bilateral temporoparietal junction (TPJ), though these regions are theorized to support distinct functions. For example, while the mPFC supports the processing of dispositional traits and preferences (Contreras et al., 2012; Mitchell et al., 2009), the right TPJ supports the inference of transient beliefs and intentions (Saxe & Powell, 2006; Young & Saxe, 2009).

Of particular interest here is perception of unpleasant mental states in others (e.g., pain or suffering), an ability required for emotional perspective taking (e.g., Batson et al., 1989). Suppose you watch a video of someone being beaten with a hammer. Even with full knowledge that the video is staged, your heart rate will increase, your palms will sweat, and you may grimace or even recoil at the sight (Cushman, Gray, Gaffey, & Mendes, 2012; see also Rozin, Millman, & Nemeroff, 1986). Emotional perspective taking may result in: (1) empathy, where one feels the same emotion as another, (2) sympathy, where one feels concern *for* another, or (3) personal distress, a self-focused response where one feels negatively because of another (Batson, 2009; Decety & Lamm, 2009; Eisenberg & Eggum, 2009). Perceiving others in pain often activates the same brain regions involved in the personal experience of pain, including the anterior insula (AI) and anterior cingulate cortex (ACC; De Vignemont & Singer, 2006; Jackson, Meltzoff, & Decety, 2005). As automatic as pain processing is, it may also be modulated by contextual knowledge. For example, as discussed previously, we are more sensitive to the pain of moral patients or victims, versus moral agents or perpetrators (Gray & Wegner, 2009). This differential sensitivity is reflected in increased activity in AI and ACC when perceiving pain in victims versus perpetrators (Decety, Echols, & Correll, 2010). Empathic reactions are also modulated by group membership (i.e., reduced activity in AI and ACC in response to the painful experiences of out-group members; Xu, Zuo, Wang, & Han, 2009). These effects may be particularly strong in the context of intergroup conflict or competition, resulting even in the recruitment of reward regions in response to the physical or emotional suffering of out-group members (Cikara, Botvinick, & Fiske, 2011; Cikara, Bruneau, & Saxe, 2011; Singer et al., 2006).

ANTICIPATING REWARD

Prosocial behavior requires the perception of another being who could benefit from one's own behavior. However, as we have seen, simply identifying an opportunity for prosociality is no guarantee that one will act on it. We may be more motivated to behave prosocially when the cost of such behavior is low and the anticipated benefits are high. These benefits may come in the form of reciprocation or public credit, or as a private positive feeling (i.e., "warm glow"). The reward system can be roughly divided into two processes: the anticipation of reward and the processing of a rewarding outcome. These two processes, prediction and the processing of prediction error, are associated with activity in distinct brain regions (Knutson & Cooper, 2005). The anticipation of reward is associated with activity in the VS and the nucleus accumbens, a subregion of the VS; the processing of a rewarding outcome is associated with activity in the ventral mPFC. Both regions seem to be involved in the processing of reward across multiple domains, for example, reflecting the rewarding aspects of food, money, or social contact (Fehr & Camerer, 2007; Knutson & Cooper, 2005). While this generality hints that prosocial behavior may indeed be associated with activity in reward brain regions, it also increases the difficulty of making strong inferences about the *kind* of reward that is anticipated by the prosocial individual. For example, one person may help another because he anticipates reputational benefits; another may do so because he finds prosocial behavior intrinsically rewarding, that is, it feels good to "do what's right." Both of these expected outcomes are rewarding in a sense and may be associated with activity in brain regions such as the VS and ventral mPFC. Notably, this ambiguity in the interpretation of brain data applies not only to the reward system, but also to regions involved in social cognition and the perception of negative mental states. We revisit this discussion after reviewing research directly associating brain activity and prosocial behavior.

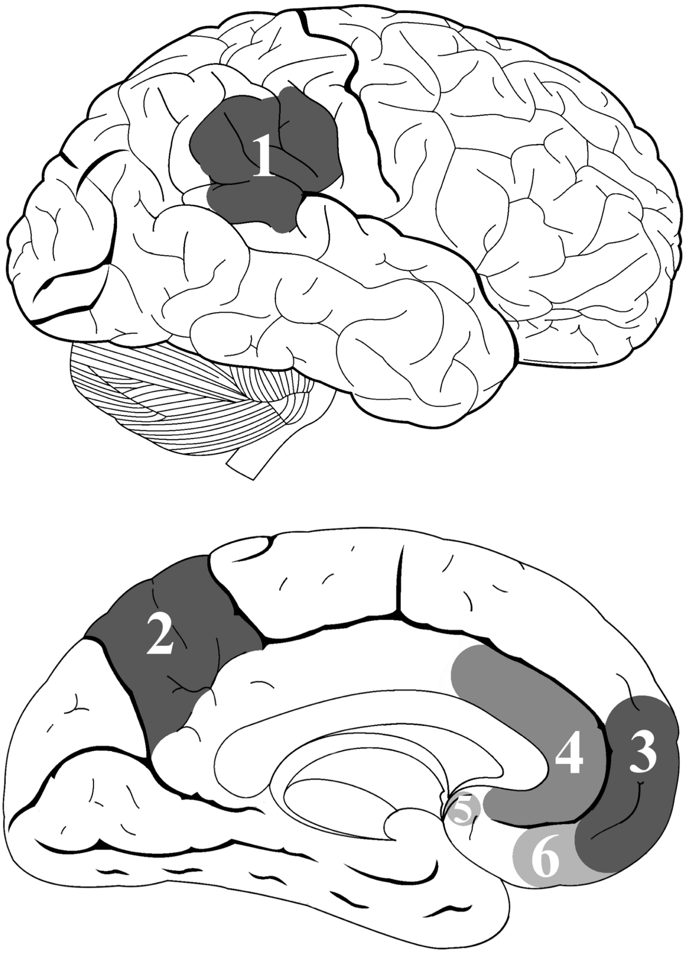

FIGURE 5.1 **The Prosocial Brain** *Prosocial behavior is associated with processing in brain regions implicated in social cognition, such as the bilateral temporoparietal junction (TPJ, 1), precune us (PC, 2) and medial prefrontal cortex (mPFC, 3); brain regions involved in processing conflict and discomfort, such as the anterior cingulate (ACC, 4) and anterior insula (AI, not pictured); and brain regions involved in reward processing, such as the ventral striatum (VS, 5) and ventral mPFC (6).*

Modified from Gray (1918, Figures 727 and 728) by Oona Räisänen.

The Prosocial Brain

Earlier, we provided a brief review of the psychological and neural processes that enable people to perceive life, minds, and mental states in others. Here we discuss research revealing a link between social cognition and prosocial behavior and in particular evidence showing that greater involvement of regions for mind perception (e.g., mPFC, TPJ), the perception of unpleasant states in others (e.g., AI and

ACC), and the anticipation of reward (VS) is associated with prosocial sentiments and behavior (see Figure 5.1). Notably, there has been little neuroscientific research associating the perception of life or animacy and prosocial behavior. This may be because, on the face of it, people are clearly motivated to aid living creatures rather than inanimate objects such as rocks. Although people sometimes admit to “animistic” tendencies, judging inanimate objects to have animate traits (e.g., my printer hates me), these judgments often go beyond mere animacy attribution, leading, ultimately, to inferences about the presence of minds as well (see Waytz, Gray, Epley, & Wegner, 2010).

Notably, the anticipated role of mentalizing and reward brain networks in prosocial behavior dovetails with work linking prosociality to neurochemicals such as oxytocin and dopamine. Oxytocin interacts with social cognitive brain regions such as the mPFC and reward regions such as the VS (Bethlehem, Honk, Auyeung, & Baron-Cohen, 2013) and plays a critical role in modulating social behavior. Polymorphisms on the oxytocin receptor (OXTR) gene are associated with individual differences in prosocial behavior (Israel et al., 2009), and oxytocin administration increases trust and generosity (De Dreu et al., 2010; Israel, Weisel, Ebstein, & Bornstein, 2012; Zak, Stanton, Ahmadi, 2007). Genetic variance in the receptors of dopamine, a neurotransmitter implicated in reward processing, is also associated with prosocial tendencies (Knafo, Israel, & Ebstein, 2011).

Brain regions involved in perception of other minds and unpleasant mental states are consistently predictive of both empathic responses and prosocial behavior. Rameson, Morelli, and Lieberman (2012) assessed prosocial tendencies by instructing participants to complete a diary over two weeks, documenting their own helpful behavior (e.g., holding a door open, lending money) directed toward friends and strangers. Participants later judged the negativity of sad images while undergoing fMRI. Correlations revealed that both empathy and self-reported helping behavior were associated with greater activity in the mPFC and ACC. Reported helping of *friends* was associated with activity in a much more robust network, including the mPFC, the ACC, and the PC, whereas reported helping of *strangers* was associated with only a relatively small section of the mPFC.

Other research directly targets the relationship between prosocial behavior and brain activity in examining participants’ actual prosocial behavior toward a victim based on participants’ neural response to that victim’s plight. Masten, Morelli, and Eisenberger (2011) scanned participants while they viewed a game where three players pass a virtual “ball” back and forth. Partway into the game, players 1 and 2 began passing only to each other, effectively excluding player 3. Participants later sent e-mails to all players, and participants’ e-mails to the excluded player were rated for prosocial (e.g., consoling) content. Viewing socially salient exclusion interactions versus inclusion interactions resulted in greater activity in the mPFC. Activity in the dorsal mPFC, along with the right AI (associated with negative affect), predicted the prosocial content of e-mails. This research presents a novel perspective by (1) measuring actual prosocial behavior, rather than self-reported tendencies and (2) predicting prosocial behavior directed toward a victim based on an empathic response to that same victim's plight. Yet, this research is limited in its focus on relatively low-cost prosocial behavior (e.g., sending an e-mail).

Waytz, Zaki, and Mitchell (2012) scanned participants while they made judgments about other individuals (e.g., "Does this person like to gossip?") and later completed a monetary distribution task that allowed participants to pay a monetary cost to benefit those individuals while also maximizing utility (e.g., you receive $1 or they receive $1.50). Following their scan session, participants were given an opportunity to spend time helping others complete a tedious quiz. Both kinds of costly prosocial behavior were predicted by activity in the dorsal mPFC.

Prosocial behavior is associated not only with functional differences (i.e., in brain activity) but also with structural differences in brain architecture. Using a monetary allocation task, Morishima, Schunk, Bruhin, Ruff, and Fehr (2012) related prosocial behavior to cortical gray matter volume using voxel-based morphometry (VBM) analysis. The task design allowed for distinct kinds of prosocial behavior to be examined. Gray matter volume in the right TPJ (involved in social cognition, as discussed earlier) was predictive of prosocial behavior across participants, but only when participants were in an advantageous position (e.g., giving to others if they are poorer but not wealthier than oneself). Interestingly, this region was insensitive to prosocial behavior that could be modeled by reciprocity (e.g., tit-for-tat strategy).

Neuropsychiatric studies of lesion patients have consistently implicated the mPFC in the regulation of social behavior. Most famously, Phineas Gage became antisocial and belligerent following an injury that disrupted function of mPFC and surrounding regions. Recent research on patients with ventral mPFC damage shows increased impulsivity and reduced sensitivity to social norms, suggesting a link between ventral mPFC lesions and "acquired psychopathy" (Bechara, Tranel, Damasio, & Damasio, 1996; Damasio, Tranel, & Damasio, 1991; Krajbich, Adolphs, Tranel, Denburg, & Camerer, 2009). Krajbich et al. (2009) found lower rates of generosity among six patients with varying degrees of ventral mPFC damage, relative to patients with damage to other regions. Ventral mPFC patients also behaved in ways suggesting that they were less trusting of others and felt less guilt after behaving selfishly. Thus, compromised ventral mPFC function may lead to deficits in prosocial behavior.

Finally, neuroscientific evidence suggests that engaging in prosocial behavior is rewarding (Harbaugh, Mayr, & Burghart, 2007; Moll et al., 2006; Rilling et al., 2002; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004; Tabibnia, Satpute, & Lieberman, 2008; see Lee, 2008, for review). Moll et al. (2006) found similar patterns of activity in the VS (a "reward" area) both when participants received money and when participants donated money toward a charity. Moreover, activity in this region predicted the frequency (and consequent amount) of donations made by participants. Harbaugh et al. (2007) found VS activity when participants made costly donations toward a charity, even when those donations were mandatory.

Activity in this reward region was also greater for voluntary versus mandatory donations and, reflecting Moll et al. (2006), the amount of activity in VS scaled up with the magnitude of donation. These results are consistent with behavioral work documenting a positive relationship between spending money on others and happiness—a relationship that was particularly strong when the recipients of one's spending were one's family or close friends (Aknin, Sandstrom, Dunn, & Norton, 2011). However, as stated previously, donation-dependent reward activity may reflect either anticipated social benefits or the intrinsic value of prosocial behavior.

Multiple Prosocial Motivations

In the neuroscientific research reviewed earlier, prosocial behavior was associated with activity in brain regions involved in social cognition, perception of mental states in others, and anticipation of reward. However, this work alone cannot distinguish between different kinds of proximal motivations for prosocial behavior. For example, increased activity in social cognitive brain regions such as the mPFC could reflect either an empathetic focus on a victim or a kind of strategizing whereby one assesses the social benefits attainable by helping the victim. The finding that the mPFC activates more robustly for close versus distant others (Mitchell et al., 2006; Krienen et al., 2010; Rameson et al., 2012) is consistent with both accounts: close others may elicit more empathy from us, and close others are also more likely to reciprocate prosocial behavior. To inform the discussion of possible motivations of prosocial behavior, we look beyond the neuroscientific literature and highlight relevant behavioral work. We first review research that supports a strategic account of prosocial behavior and then we turn to research supporting an altruistic account.

QUID PRO QUO

It pays to be nice. Being on the receiving end of a prosocial act is rewarding (Rilling et al., 2002; Rilling et al., 2004; Tabibnia et al., 2008) and may make the recipient want to return the favor (Axelrod & Hamilton, 1981; Cialdini, 2006). More indirect benefits may come in the form of reputational credit or prestige, making conspicuous prosociality an adaptive social strategy (Barclay & Willer, 2007; Hardy & Vugt, 2006). Accordingly, people are sensitive to the reputational effects of prosocial behavior. They are more likely to donate to strangers face-to-face than to unseen strangers (Bohnet & Frey, 1999; Rege & Telle, 2004) and to named than unnamed strangers (Charness & Gneezy, 2008). People are also more likely to behave well when their actions are known to an identified third party (Andreoni & Petrie, 2004; Franzen & Pointer, 2012; Piazza & Bering, 2008; Satow, 1975). People are even sensitive to the mere idea of accountability, acting more prosocially in the presence of a picture of watchful eyes (Bateson, Nettle, & Roberts, 2006; Ernest-Jones, Nettle, & Bateson, 2011; Haley & Fessler, 2005) and when primed with the concept of God or a watchful spirit (Gervais & Norenzayan, 2012; Mazar, Amir, & Ariely, 2008; Piazza, Bering, & Ingram, 2011; Shariff & Norenzayan, 2007). Conversely, people may withhold help or act selfishly when their identity is obscured by darkness (Hirsh, Galinsky, & Zhong, 2011; Zhong, Bohns, & Gino, 2010) or when they are unidentified in a group (Darley & Latane, 1968; Postmes & Spears, 1998; J. E. Singer, Brush, & Lublin, 1965).

This sensitivity emerges early in development. Children age six and under are more willing to share toys with a friend than a stranger (Knight & Chao, 1991; Moore, 2009). Leimgruber, Shaw, Santos, and Olson (2012) gave 5-year-olds a choice between receiving a toy on the one hand and receiving a toy and also allowing a peer to receive a toy on the other. The children were more willing to "give" the toy to the peer (at no direct cost to themselves) when the generous behavior was visible to the recipient, but not otherwise. As with adults, prosocial behavior directed at children results in positive outcomes or social benefits. Children are sensitive to others' previous actions, preferring to play with helpful versus unhelpful others (Hamlin & Wynn, 2011) and helping a former collaborator more than a known noncollaborative peer (Hamann, Warnecken, & Tomasello, 2012).

The reputational credit bestowed on a prosocial agent scales up with the costliness of their behavior (Barclay & Willer, 2007; Hardy & Vugt, 2006). The two most admired people of the 20th century, Mother Teresa and Martin Luther King Jr., are known for their selfless commitment to prosocial causes (Gallup, 1999). We revere those who sacrifice the most for others, and we may in turn be motivated to engage in costly acts in pursuit of admiration. But can this help explain why anyone would ever sacrifice his or her life to benefit others? The soldier who jumps on a grenade is not likely thinking about cashing in on the social credit his act will earn him. Reputation may help explain how altruistic motives could have developed in the first place (Nowak & Sigmund, 2005; Panchanathan & Boyd, 2004). However the question remains—what was going through the minds of these extreme altruists?

GOOD FOR GOODNESS' SAKE

Prosocial behavior may be proximally motivated through a need to maintain a positive moral identity or self-concept (Aquino & Reed, 2002; Gino, Schweitzer, Mead, & Ariely, 2011; Mazar et al., 2008; Monin & Miller, 2001; Sachdeva, Iliev, & Medin, 2009; Young, Chakroff, & Tom, 2012). Most people want to see themselves as moral individuals and may try to act in accordance with this ideal self-concept. Deviations from this ideal may create an unpleasant state of dissonance, which can serve as a proximate motivator to act morally (Stone & Cooper, 2001). We may behave unethically (or withhold prosocial behavior) only to the extent that we can still justify our moral identities to ourselves (Mazar et al., 2008). Thus, people may behave unethically when they can do so indirectly. For example, people are more likely to cheat to earn tokens worth cash, rather than cash itself (Mazar et al., 2008); some will steal soda rather than money, which could be used to buy soda (Ariely, 2008). If one expresses egalitarian sentiments (e.g., hiring a person of a minority group), one may be less likely to express similar egalitarian sentiments in a secondary task, presumably because one has already "proven" one's prosocial sentiments (Monin & Miller, 2001; Sachdeva et al., 2009). This kind of moral "licensing" seems to depend on the perception of moral behavior as a "credit," rather than an expression of one's moral nature. Other research highlights people's need to feel consistent with their own moral self-concepts. Young at al. (2012) demonstrate that people are more willing to donate to charity after they have recounted instances of their own prosocial behavior, relative to people who recalled instances of their own past antisocial behavior.

Providing monetary incentives for prosocial behavior often backfires, resulting in less prosociality overall, despite what seems to be a better incentive structure. For example, Gneezy and Rustichini (2000a) found that providing monetary incentives for students collecting money for charity actually decreased performance. In another domain, incentivizing parents to be on time to pick their children up from school (by fining late parents in this case) resulted in more late parents (Gneezy & Rustichini, 2000b). Taking the perspective of the prosocial agents, however, renders these results less surprising. Following an example from A. P. Fiske and Tetlock (1997), how would you feel if, after preparing and serving a feast for friends and family, a guest offered to show his appreciation by writing you a check? How motivated would you be to cook for him in the future? Presumably, the original motivation for the feast was a desire to please others or, more generally, to do good—in other words the feast was not a service for a fee but an expression of love or good will. Monetary compensation for good deeds reframes moral or social events as business transactions and may therefore undermine intrinsic motivations to do good. This reframing may affect not only one's motivation to cook in the future but also others' perceptions of the act ("he did it for the money"). Thus, any social benefits of the prosocial behavior, such as reputational credit, may be undermined by secondary incentives. In an economist's terms, the intrinsic incentives (e.g., morality) were crowded out by the extrinsic ones (e.g., money; Frey & Jergen, 2001). The consequences of this effect extend far beyond daycare and dinner. Researchers provided monetary incentives to facilitate conflict resolution among Israeli and Palestinian participants. As above, the incentives backfired, resulting in emotional outrage and support of violent opposition to the resolution (Ginges, Atran, Medin, & Shikaki, 2007).

In sum, people may try to behave in accordance with their own moral self-concept (Stone & Cooper, 2001; Young et al., 2012). Discrepancies between one's ideal and one's behavior may be obscured through indirect action (Ariely, 2008; Mazar et al., 2008). External incentives can also undermine intrinsic motivations to do good or to be good, resulting in reduced prosociality. In other instances, people who act for reputational credit may feel licensed to behave selfishly once they feel they have earned sufficient social credit (Monin & Miller, 2001; Sachdeva et al., 2009). Self-concept maintenance may therefore be in the service of reputation-building at the level of adaptive function (Mazar et al., 2008), but, again, these ultimate ends need not correspond with proximate motivations (Mayr, 1993; Von Hippel & Trivers, 2011).

SAINTS AND STRATEGISTS

Some people are nicer than others. This is obvious and also reflected in intersubject variation in prosociality documented in the studies reviewed above. But perhaps more interesting is that different people are nice for different reasons. Some people may act prosocially for reputational reasons (e.g., social credit) and others out of intrinsic motivations (e.g., to be good, to do good). There is variation in our dispositional "moral identity" (Aquino & Reed, 2002) and "social value orientation" (Van Lange, 1999), with consequences for base levels of prosociality and also for susceptibility to the reputational effects outlined above.

Gino et al. (2011) measured unethical behavior in participants who scored high or low on "moral identity" (Aquino & Reed, 2002), measured through agreement with statements such as "Being someone who is [generous, kind, fair, friendly] is an important part of who I am." Participants had an opportunity to cheat on an exam by falsely reporting their score. Additionally, some participants completed the exam after having completed a cognitively depleting task that reduces self-control capacity (cf. Baumeister, Bratslavsky, Muraven, & Tice, 1998). Participants with a low moral identity behaved more unethically and were also more affected by the self-control depletion manipulation. By contrast, there was no significant effect of self-control depletion on participants with high moral identity, suggesting that for "honest" participants, being honest was the default (and easy) response.

In a neuroimaging study, Greene and Paxton (2009) were able to sort participants into "honest" and "dishonest" groups based on cheating behavior. They then examined the neural responses to task trials featuring an opportunity to cheat, relative to trials in which there was no opportunity to cheat. For "dishonest" participants, refraining from cheating when provided the opportunity recruited the ACC (associated with conflict and negative affect) and the dorsal lateral PFC (associated with cognitive control). By contrast, for "honest" participants, refraining from cheating when given the opportunity looked no different from refraining from cheating when not given the opportunity. In other words, the brains of "honest" participants behaved as though there was never an opportunity to cheat.

Of course, refraining from antisocial behavior is not the same as engaging in prosocial behavior. Indeed, increased prosocial tendencies can sometimes track with *increased* antisocial tendencies (Hirsh, Galinsky, & Zhong, 2011). In a more direct test of individual differences in prosocial motivations, Simpson and Willer (2008) classified participants into "egoists," who report a desire to maximize their own rewards only, and "altruists," who report a desire to maximize rewards for themselves and others (Van Lange, 1999). Participants were given an opportunity to donate to an unknown other, either privately, or when watched by a third party who would later donate money to the participant. Participants did donate more under public than private conditions, but this effect was much larger for the "egoists" than "altruists," suggesting that "egoists" are more sensitive to public reputational concerns. In a second experiment (Simpson & Willer, 2008), participants played the role of a third party, watching a donation between two parties, and then donated to the former donor. Participants were told either that the donor knew he was being watched, or that he thought he acted anonymously. Participants donated more to an anonymous donor than to a public donor. However, this difference was also stronger for "egoists" than "altruists." As before, "altruists" would donate similarly large amounts to the former donor regardless of condition. Finally, Young et al. (2012) found that priming a positive moral self-concept leads to greater prosociality (see above). However, this effect was especially powerful for participants who did not mention reputational considerations (e.g., whether others appreciated their acts) in recounting their past good deeds (reflecting positive moral self-concept), suggesting that some individuals may be motivated to do good "for goodness' sake."

Consistent with the individual differences documented above, between honest and dishonest individuals, altruists and egoists, Rand, Greene, and Nowak (2012) demonstrated differences between participants who reported being generally trusting of others in their social environment versus untrusting. Priming trusting (but not untrusting) participants to rely on their gut intuitions led to higher donations in a public goods game, whereas priming them to rely on deliberate reflection led to lower contributions. In another experiment, comparing the amount of time trusting participants used to make their choice generated a similar behavioral profile—when responding quickly, trusting participants contributed more.

Finally, regardless of one's prosocial disposition, one may be motivated to act prosocially toward different agents for different reasons. One buys the first round of drinks for friends, expecting someone else to get the next round (e.g., direct reciprocity). But a parent may help her child without stipulation and in the absence of any direct social benefits. This may be because some find it intrinsically rewarding to help family. In a recent study (Telzer, Masten, Berkman, Lieberman, & Fulgini, 2011), participants made costly donations to family members while undergoing fMRI. Consistent with other work, donation was associated with activity in social cognitive brain regions such as the mPFC. Furthermore, for participants with strong stated family obligations, there was enhanced functional coupling between social cognitive and reward brain regions during donation.

Consistent behavioral findings come from Maner and Gailliot (2007), who found that the best predictors of prosocial behavior were moderated by the relationship between the prosocial agent and the recipient. Participants were presented with hypothetical opportunities to help a family member versus stranger (e.g., helping a family member versus stranger who has been evicted from their apartment). Participants rated their empathic concern for the person in need, as well as their feelings of "oneness" with that individual. While empathic concern best predicted willingness to help a family member, "oneness" best predicted willingness to help a stranger. Put another way, we may help strangers only to the extent that we feel similar or close to them, whereas we help family members out of unconditional empathic concern, regardless of similarity.

Conclusions

Prosocial behavior depends on numerous mental processes working in concert. We must first perceive an opportunity for acting prosocially and then we must motivate ourselves to act. If prosocial behavior is intended to benefit another, opportunities for prosocial behavior require the perception of another (agent or mind) who might benefit from such an act. We help those in need and we may assess need based on our perception of negative mental states in others. Finally, we find helping others rewarding and we will only act when sufficiently motivated by some kind of anticipated benefit.

The research reviewed here presents a rough outline of psychological and neurological underpinnings of prosocial behavior and also points to avenues for future research. For example, why do we help some more than others? The present data cannot distinguish between "empathetic" and "strategic" accounts. On an empathetic account, we find it easier to take the perspective of a similar other, leading to increased empathy and motivation to help. Activity in the mPFC may index psychological closeness, while activity in AI indexes perception of others' pain and perhaps our own empathic concern. By contrast, on a strategic account, we are motivated to help others when our actions will directly or indirectly help us at some point in the future. Cooperative or similar others may be more likely to reciprocate or to contribute toward our group's interests, which include our own interests. Likewise, the suffering of a group member may hurt the group as a whole (See Tomasello & Vaish, 2012). For example, a soccer player who sees a teammate fail may experience anguish and feel motivated to help him in order to ensure the team's (and his own) eventual success. On this account, mPFC activity may reflect the processing of group identity (e.g., is this my teammate?) or perspective taking in order to assess the likelihood of reciprocity (see Gilin, Carpenter, & Galinsky, in press).

Although prosocial behavior is rewarding, the reward may take the form of anticipated public benefits (i.e., knowing that one looks good in the eyes of others) or the private knowledge that one is a good and moral person (i.e., moral self-concept). We seem to act in order to appear good to others (Barclay & Willer, 2007; Hardy & Vugt, 2006) and also to ourselves (Mazar et al., 2008; Young et al., 2012). These motivational rewards may function differentially across prosocial agents (e.g., altruists versus egoists; Simpson & Willer, 2008) and also across the beneficiaries of prosocial behavior (e.g., family versus strangers; Maner & Gailliot, 2007). Similarly, empathetic and strategic motivations may not be mutually exclusive but could differ by person or by situation.

Future research should continue to resolve this ambiguity, perhaps by comparing prosocial behavior directed at different agents (e.g., family, friends, and strangers), under public versus private conditions. Indeed, prosocial behavior is multidimensional, encompassing diverse helpful acts directed at different people for multiple reasons. Understanding prosocial behavior across these dimensions will surely contribute to a richer account of when we do good, when we fail, and why.

References

Aknin, L. B., Sandstrom, G. M., Dunn, E. W., & Norton, M. I. (2011). It's the recipient that counts: Spending money on strong social ties leads to greater happiness than spending on weak social ties. *PloS One*, *6*(2), e17018.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277.

Andreoni, J., & Petrie, R. (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics*, *88*(7), 1605–1623.

Aquino, K., & Reed, A. II, (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440.

Ariely, D. (2008). *Predictably irrational*. New York: HarperCollins.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396.

Barclay, P., Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1610), 749–753.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414.

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). Cambridge, MA: MIT Press.

Batson, C. D., Batson, J. G., Griffitt, C. A., Barrientos, S., Brandt, J. R., Sprengelmeyer, P., & Bayly, M. J. (1989). Negative-state relief and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *56*(6), 922.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252–1265.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral cortex*, *6*(2), 215–225.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.

Bethlehem, R. A., van Honk, J., Auyeung, B., & Baron-Cohen, S. (2013). Oxytocin, brain physiology, and functional connectivity: A review of intranasal oxytocin fMRI studies. *Psychoneuroendocrinology*, *38*(7), 962–974.

Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *The American Economic Review*, *89*(1), 335–339.

Boyd, R., & Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology*, *215*(3), 287–296.

Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior and Organization*, *68*(1), 29–35.

Cialdini, R. B. (2006). *Influence: The psychology of persuasion*. New York: Harper Business.

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us versus them social identity shapes neural responses to intergroup competition and harm. *Psychological Science*, *22*(3), 306–313.

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them intergroup failures of empathy. *Current Directions in Psychological Science*, *20*(3), 149–153.

Cloutier, J., Heatherton, T. F., Whalen, P. J., & Kelley, W. M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *Journal of Cognitive Neuroscience*, *20*(6), 941–951.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., . . . Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*(8), 2608–2618.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience, 7,* 764–770.

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*(1), 2.

Damasio, A. R., Tranel, D., & Damasio, H. (1991). Somatic markers and the guidance of behavior: Theory and preliminary testing. In H. S. Levin, H. M. Eisenberg, A. L. Benton (Eds.), *Frontal Lobe Function and Dysfunction* (pp. 217–229). New York, NY: Oxford University Press.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4 part 1), 377.

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., . . . Feith, S. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, *328*(5984), 1408–1411.

De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441.

Decety, J., Echols, S., & Correll, J. (2010). The blame game: The effect of responsibility and social stigma on empathy for pain. *Journal of Cognitive Neuroscience*, *22*(5), 985–997.

Decety, J., & Lamm, C. (2009). Empathy versus personal distress: Recent evidence from social neuroscience. In J. Decety & W. Ickes (Eds.), *The social neuroscience of* empathy (pp. 199–213). Cambridge, MA: MIT Press.

Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71–83). Cambridge: MIT Press.

Eisenberg, N. & Fabes, R. A. (1998). Prosocial development. In W. Damon & N. Eisengerb (Eds.), *Handbook of child psychology*. New York, NY: Wiley & Sons.

Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior*, *32*(3), 172–178.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: The neural circuitry of social preferences. *Trends in Cognitive Sciences*, *11*(10), 419–427.

Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology*, *18*(2), 255–297.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Franzen, A., & Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior and Organization*, 81, 74–81.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589–611.

Gallup, G (1999). *The Gallup Poll*. Wilmington, DE: Scholarly Resources. 248–249.

Gelman, S. A., & Gottfried, G. M. (1996). Children's causal explanations of animate and inanimate motion. *Child Development*, *67*(5), 1970–1987.

Gervais, W. M., & Norenzayan, A. (2012). Like a camera in the sky? Thinking about God increases public self-awareness and socially desirable responding. *Journal of Experimental Social Psychology*, *48*(1), 298–302.

Gilin, D., Maddux, W. W., Carpenter, J., & Galinsky, A. D. When to use your head and when to use your heart: The differential value of perspective-taking versus empathy in competitive interactions. *Personality and Social Psychological Bulletin* 39(1), 3–16.

Ginges, J., Atran, S., Medin, D., & Shikaki, K. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences*, *104*(18), 7357–7360.

Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes*, *115*(2), 191–203.

Gneezy, U., & Rustichini, A. (2000a). Pay enough or don't pay at all. *Quarterly Journal of Economics*, *115*(3), 791–810.

Gneezy, U., & Rustichini, A. (2000b). A fine is a price. *Journal of Legal Studies*, *29*(1), 1–18.

Gray, H. (1918). *Anatomy of the human body*. Philadelphia, PA: Lea & Febiger.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619.

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences*, *106*(30), 12506–12511.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, *12*(5), 711–720.

Hadjikhani, N., Kveraga, K., Naik, P., & Ahlfors, S. P. (2009). Early (N170) activation of face-specific cortex by face-like objects. *Neuroreport*, *20*(4), 403.

Haley, K. J., & Fessler, D. M. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*(3), 245–256.

Hamann, K., Warneken, F., & Tomasello, M. (2012). Children's developing commitments to joint goals. *Child Development*, *83*(1), 137–145.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30–39.

Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*(5831), 1622–1625.

Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, *32*(10), 1402–1413.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 243–259.

Hirsh, J. B., Galinsky, A. D., & Zhong, C. B. (2011). Drunk, powerful, and in the dark: How general processes of disinhibition produce both prosocial and antisocial behavior. *Perspectives on Psychological Science*, *6*(5), 415–427.

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Reibold, M., Laiba, E.,... Ebstein, R. P. (2009). The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task. *PLoS One*, *4*(5), e5535.

Israel, S., Weisel, O., Ebstein, R. P., & Bornstein, G. (2012). Oxytocin, but not vasopressin, increases both parochial and universal altruism. *Psychoneuroendocrinology*, *37*(8), 1341–1344.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, *24*(3), 771–779.

Japan Center for International Exchange. (2012). US giving for Japan's 3/11 disaster exceeds $665 million. Retrieved from http://www.jcie.org/311recovery/usgiving-2.html

Jenkins, A. C., & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience*, *6*(3), 211–218.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*, 1–19.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, *14*(5), 785–794.

Knafo, A., Israel, S., & Ebstein, R. P. (2011). Heritability of children's prosocial behavior and differential susceptibility to parenting by variation in the dopamine receptor D4 gene. *Development and Psychopathology*, *23*(1), 53.

Knight, G. P., & Chao, C. C. (1991). Cooperative, competitive, and individualistic social values among 8- to 12-year-old siblings, friends, and acquaintances. *Personality and Social Psychology Bulletin*, *17*(2), 201–211.

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*(4), 411–417.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience*, *29*(7), 2188–2192.

Krienen, F. M., Tu, P. C., & Buckner, R. L. (2010). Clan mentality: Evidence that the medial prefrontal cortex responds to close others. *Journal of Neuroscience*, *30*(41), 13906–13915.

Lah, K. (2011). Japanese seniors volunteer for Fukushima "suicide corps." *CNN World*. Retrieved from http://articles.cnn.com/2011-05-31/world/japan.nuclear.suicide_1_nuclear-plant-seniors-group-nuclear-crisis?_s=PM:WORLD

Lee, D. (2008). Game theory and neural basis of social decision-making. *Nature Neuroscience*, *11*(4), 404–409.

Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PloS One*, *7*(10), e48292.

Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, *8*(6), 720–722.

Looser, C. E., Guntupalli, J. S., & Wheatley, T. (2012). Multivoxel patterns in face- sensitive temporal regions reveal an encoding schema based on detecting life in a face. *Social Cognitive and Affective Neuroscience*. Epub ahead of print, July 13, 2012. doi: 10.1093/scan/nss07

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*(8), 601–608.

Maner, J. K., & Gailliot, M. T. (2007). Altruism and egoism: Prosocial motivations for helping depend on relationship context. *European Journal of Social Psychology, 37*(2), 347–358.

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for "social pain" and subsequent prosocial behavior. *Neuroimage*, *55*(1), 381–388.

Mayr, E. (1993). Proximate and ultimate causations. *Biology and Philosophy*, *8*(1), 93–94.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. J*ournal of Marketing Research*, *45*(6), 633–644.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *8*(3), 285–299.

Mitchell, J. P., Ames, D. L., Jenkins, A. C., & Banaji, M. R. (2009). Neural correlates of stereotype application. *Journal of Cognitive Neuroscience*, *21*(3), 594–604.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*(4), 655–663.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences*, *103*(42), 15623–15628.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*(1), 33.

Moore, C. (2009). Fairness in children's resource allocation depends on the recipient. *Psychological Science*, *20*(8), 944–948.

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*, *75*(1), 73–79.

Morrison, S., Decety, J., & Molenberghs, P. (2012). The neuroscience of group membership. *Neuropsychologia*, *50*(8), 2114–2120.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*,*437*(7063), 1291–1298.

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, *432*(7016), 499–502.

Piazza, J., & Bering, J. M. (2008). Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behavior*, *29*(3), 172–178.

Piazza, J., Bering, J. M., & Ingram, G. (2011). "Princess Alice is watching you": Children's belief in an invisible person inhibits cheating. *Journal of Experimental Child Psychology*, *109*(3), 311–320.

Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, *123*(3), 238.

Rameson, L. T., Morelli, S. A., & Lieberman, M. D. (2012). The neural correlates of empathy: Experience, automaticity, and prosocial behavior. *Journal of Cognitive Neuroscience*, *24*(1), 235–245.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430.

Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, *88*(7), 1625–1644.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*(2), 395–405.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport*, *15*(16), 2539–2243.

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, *50*(4), 703.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, *20*(4), 523–528.

Satow, K. L. (1975). Social approval and helping. *Journal of Experimental Social Psychology*, *11*(6), 501–509.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*(8), 692–699.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309.

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*(9), 803–809.

Simpson, B., & Willer, R. (2008). Altruism and indirect reciprocity: The interaction of person and situation in prosocial behavior. *Social Psychology Quarterly*, *71*(1), 37–52.

Singer, J. E., Brush, C. A., & Lublin, S. C. (1965). Some aspects of deindividuation: Identification and conformity. *Journal of Experimental Social Psychology*, *1*(4), 356–378.

Singer, P. (1997). The drowning child and the expanding circle. *New Internationalist*, *289*, 28–30.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*(7075), 466–469.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior* (No. 218). Cambridge, MA: Harvard University Press.

Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, *37*(3), 228–243.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, *19*(4), 339–347.

Telzer, E. H., Masten, C. L., Berkman, E. T., Lieberman, M. D., & Fuligni, A. J. (2011). Neural regions associated with self control and mentalizing are recruited during prosocial behaviors towards the family. *Neuroimage*, *58*(1), 242–249.

Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*(1), 163–173.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460.

Tomasello, M., & Vaish, A. (2012). Origins of human cooperation and morality. *Annual Review of Psychology,* 64, 231–255.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*(2), 337.

Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, *34*(1), 1.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388.

Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *The Journal of Neuroscience*, *32*(22), 7646–7650.

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, *32*(4), 231–262.

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychological Science, 18*, 469–474.

Wheatley, T., Weinberg, A., Looser, C., Moran, T., & Hajcak, G. (2011). Mind perception: Real but not artificial faces sustain neural activity beyond the N170/VPP. *PloS One*, *6*(3), e17960.

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, *29*(26), 8525–8529.

Young, L., Chakroff, A., & Tom, J. (2012). Doing good leads to more good: The reinforcing power of a moral self-concept. *Review of Philosophy and Psychology*, *3*(3), 325–334.

Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*(10), 2065–2072.

Zak, P. J., Stanton, A. A., & Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS One*, *2*(11), e1128.

Zhong, C. B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science*, *21*(3), 311–314.