Norms and the Flexibility of Moral Action

Oriel FeldmanHall[1], Jae-Young Son[1], and Joseph Heffner[1]
[1]Department of Cognitive, Linguistic, Psychological Sciences, Brown University, Providence RI 02906

Corresponding Author
Oriel FeldmanHall
Brown University
190 Thayer St.
Providence, RI 02906
oriel.feldmanhall@brown.edu

ABSTRACT
A complex web of social and moral norms governs many everyday human behaviors, acting as the glue for social harmony. The existence of moral norms helps elucidate the psychological motivations underlying a wide variety of seemingly puzzling behavior, including why humans help or trust total strangers. In this review, we examine four widespread moral norms: fairness, altruism, trust, and cooperation, and consider how a single social instrument—reciprocity—underpins compliance to these norms. Using a game theoretic framework, we examine how both context and emotions moderate moral standards, and by extension, moral behavior. We additionally discuss how a mechanism of reciprocity facilitates the adherence to, and enforcement of, these moral norms through a core network of brain regions involved in processing reward. In contrast, violating this set of moral norms elicits neural activation in regions involved in resolving decision conflict and exerting cognitive control. Finally, we review how a reinforcement mechanism likely governs learning about morally normative behavior. Together, this review aims to explain how moral norms are deployed in ways that facilitate flexible moral choices.

Key words: moral, social, norms, reciprocity, altruism, fairness, cooperation, trust, emotions, learning

INTRODUCTION

Consider the various social norms that dictate how you behave in your daily life: you refrain from having conversations in a theatre, you dress conservatively in a place of worship, you tip your waiter after good service, and you keep a secret when a friend tells you something confidential. Humans share a set of social and moral beliefs that govern how we behave, from mundane chitchat during a movie to the most consequential behaviors that dictate whether we harm or help others. Norms help create social cohesiveness and an understanding of shared expectations

that support and shape identities at both a societal and individual level. For this reason, norms are critically important for determining whether social communities function well and efficiently.

Humans rely on a set of complex, evolved, and learned norms to encourage community members to adopt certain perspectives that can guide and promote prosocial interactions (Cialdini, 2003; Cialdini, Reno, & Kallgren, 1990; Goldstein, Cialdini, & Griskevicius, 2008; Nichols, 2004; Sherif, 1936). These moral norms are so important for social functioning that there appears to be a sort of universal moral grammar, through which certain moral norms are sacredly held. This has been demonstrated in multiple research fields (Bicchieri, 2006), including affective neuroscience (Chang & Smith, 2015), cognitive development (Kohlberg & Hersh, 1977; Van de Vondervoort & Hamlin, 2018), cross-cultural studies (Hauser, 2006; Mikhail, 2007), and work on non-human primates (de Waal, 2009). The growing literature suggests that this seemingly elaborate system of natural jurisprudence is relatively stable over time and extends across social groups (Mikhail, 2008; Sripada, 2008).

Here we discuss the importance of social and moral norms, what types of values they convey, and how their existence can alter behavior. We begin by defining social and moral norms: what are they, how do they develop, and how are they sustained? We then go on to discuss the set of moral norms we consider to be foundational for harmonious and successful social living. We propose that a single mechanism—reciprocity—underpins the adherence to, and enforcement of, most moral behaviors. Using a game theoretic perspective, we illustrate how these norms act as a driving force behind flexible moral behavior (Melnikoff & Bailey, 2018), whereby different classes of behavioral patterns can arise depending on which norm is activated (Ajzen, 1991). We review neural evidence that people find it intrinsically rewarding to comply with moral norms, before examining how emotions can enhance reciprocal behaviors and the adherence to moral norms. Finally, we discuss how moral norms are likely learned and sustained through reward and punishment contingencies based on expectations of reciprocity.

WHAT ARE SOCIAL AND MORAL NORMS?

Social norms are ubiquitous and endemic to social life. They provide a standard for behavior based on mutual and widely shared psychological attitudes, expectations, and beliefs about how members of society ought to behave (House, 2018). At the broadest level, these norms help to promote harmonious living, in which the rights of others are taken into account (Ullmann-Margalit, 1978). They prescribe mores (e.g., wear black to a funeral) and sometimes even consequential rules (e.g., while in America, drive on the right side of the road) about what people should and should not do in various social situations (Turiel, 1983). Since deviations from social norms often elicit informal (or even formal) social sanctions, they are a useful explanatory tool for describing many of our everyday social behaviors.

Moral norms can be considered a subset of social norms in that they explicitly govern behaviors that have positive or negative outcomes for both the self and others. For example, social norms, such as 'do not chew gum at the table', typically appeal to a wide set of behaviors without necessitating harm be prevented (Turiel, 1983). In contrast, moral norms, such as 'behave fairly', dictate that individuals navigate through the world without harming others (Schein & Gray, 2017). In some cases, moral norms act in opposition to ingrained desires (e.g., biological urges), which are generated to promote survival (Darwin, 1859; Dawkins, 1989). Enhancing self-benefit—for example, increasing one's wealth, power and prestige—is one avenue by which biological urges can be expressed. If increasing one's wealth leads an individual to deviate from morally normative patterns, negative consequences for others may ensue (harm is applied, money stolen, and so forth). The existence of moral norms, which aim to promote the well-being of others and the community at large, can help attenuate these negative consequences by tempering (either through suppression or regulation of) these self-enhancing desires.

Accordingly, if a core component of morality is that humans share a set of codes and beliefs that dampen selfish inclinations, it is important to examine what those moral strictures might be. We make the case that there are four fundamental moral norms—fairness, altruism, trust, and cooperation—that play a prominent role in shaping many everyday social interactions. While there are other possible candidate norms that could be included (e.g., norms of respect, justice,

harm, and so forth), these four norms are sufficiently general enough to be applicable to a wide array of moral behavior (for example, trusting that an individual will not be harmed by others), while also having enough specificity to capture unique behavioral patterns across them. Here we argue that these norms of fairness, altruism, trust, and cooperation are all subserved by, and rooted in, a single mechanism—reciprocity—that enables people to make flexible moral decisions across a range of social contexts.

Reciprocity as a mechanism

Reciprocity has traditionally been operationalized either as individual beliefs about the structure of the world, or as a culturally-mandated standard of behavior. In regards to the first, reciprocity is often construed within a framework of a 'just world' (Lerner, 1980), whereby people believe in a system of social exchanges that reach a fair equilibrium over time (Gouldner, 1960). Such a belief in universal justice implies that destructive individuals who violate expectations of reciprocity will eventually face consequences for disturbing the equilibrium. On the other hand, from a culturally-mandated normative standpoint, reciprocity is widely perceived as a moral 'ought' (Eisenberger, Lynch, Aselage, & Rohdieck, 2004; Gouldner, 1960; Tsui & Wang, 2002). This framework argues that reciprocity operates either by responding to negative actions with negative treatment, or by responding to positive actions with positive treatment. One particularly potent example of negative reciprocity is when punishment is levied on those who do not comply with moral norms (Fehr & Fischbacher, 2004; Gintis, 2000).

However, instead of viewing reciprocity as a norm in and of itself, it may be more appropriate to refashion the concept of reciprocity as a mechanism that motivates adherence to a suite of moral norms (Cropanzano & Mitchell, 2005; Leimgruber, 2018). From this framework, reciprocity can powerfully and flexibly drive different behaviors, from rewarding those who help, to punishing those who harm (Dufwenberg & Kirchsteiger, 2004; Gintis, Henrich, Bowles, Boyd, & Fehr, 2008; Gouldner, 1960; Nowak, 2006; Rabin, 1993; Rand & Nowak, 2013). Below, we discuss the roles that contextual factors, emotional experiences, and learning play in influencing how reciprocity supports flexible moral action.

THE NORMS THAT GOVERN MORAL BEHAVIOR

In the following section, we use a game theoretic approach to examine fairness, altruism, trust, and cooperation. This is done for two reasons. First, behavioral economic games allow researchers to observe how individuals anticipate, infer, and act on what others do (Von Neumann & Morgenstern, 1945). Because each game has a series of discrete rule sets, researchers can control and manipulate the structure and context of any game (Camerer, 2003). To the extent that people's decisions are exquisitely sensitive to the intricacies and contextual minutia of the game environments, researchers can observe how different norms and expectations alter social and moral behavior by modifying the games' rule sets. Second, the strategic interactions and behaviors that fall out of economic games are mathematically expressed on a universal scale: we know with precision how much money an individual is willing to altruistically offer another, how much punishment is conferred upon a perpetrator in the wake of a fairness violation, and how much people care to trust or cooperate with an unfamiliar partner. Together, these two dimensions of economic games provide a powerful testbed for understanding moral norms and their influence on behavior.

Fairness

It is difficult to imagine how groups of individuals would manage to divide resources in a harmonious way without appealing to a shared standard of fairness (Charness & Rabin, 2002; Fehr & Schmidt, 1999). While resources can be divided meritoriously (e.g., according to an individual's effort or contribution), the overarching norm of fairness mandates that, all things considered, resources ought to be divided equitably amongst community members. Evidence for this fairness norm is abundant. Strangers routinely split resources equitably in the absence of social sanctions (Fehr & Fischbacher, 2003; Roth, Prasnikar, Okuno-Fujiwara, & Zamir, 1991; Zelmer, 2003), notions of fairness are universally appealed to across cultures (Henrich et al., 2005; Henrich et al., 2010), human infants are sensitive to (and expect) the equitable distribution of resources (Sommerville & Enright, 2018), and even non-human animals (e.g., primates, dogs,

and birds) are attentive to unequal outcomes between members of their own species (Brosnan & de Waal, 2014).

How humans resolve fairness transgressions has been a central question in behavioral economics for decades. Economists have traditionally used the Ultimatum Game (Güth, Schmittberger, & Schwarze, 1982) to demonstrate that after experiencing a fairness violation (Bicchieri & Chavez, 2010), people are willing to forgo even large sums of money to punish norm violators (Camerer, 2003). In the Ultimatum Game, two players partake in an economic exchange. One player acts as the Proposer and makes an offer to the other participant, the Responder. The Responder can then either accept or reject the offer. If accepted, the money is split as proposed. If rejected, then neither player receives any money, which effectively punishes the Proposer for offering an unfair split. The most rational decision for the Responder is to accept any offer from the Proposer no matter how small, since some money is better than no money. However, people routinely go against monetary self-interest to reject unfair offers (Fehr & Fischbacher, 2004; Fehr & Gächter, 2002; Henrich et al., 2006; Herrmann, Thöni, & Gächter, 2008; Jordan, Hoffman, Bloom, & Rand, 2016; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), indicating a strong preference for upholding fairness norms.

Despite this preference, individuals frequently split resources unequally to benefit themselves, revealing a dueling desire to act in one's own self-interest (Camerer & Thaler, 1995; Kahneman, Knetsch, & Thaler, 1986). This class of unfairness is termed advantageous inequality (when one receives more than another), which stands in contrast to disadvantageous inequality (when one receives less than another). Such unequal distributions are not ideal for lasting partnerships, since receiving less than another signals a disadvantageous relationship that should potentially be terminated, while receiving more reward might risk undermining future goals because one's partner could feel exploited (Nishi, Shirado, Rand, & Christakis, 2015). Accordingly, although self-interest may initially lead an individual to prefer advantageous inequality, minimizing both types of unfairness helps individuals and societies stabilize long-term ventures (Piketty, 2017; Tavoni,

Dannenberg, Kallis, & Löschel, 2011). The mutual expectation of fair treatment therefore leads individuals on both sides of the dyad to prefer equal payoffs.

As a consequence, concerns about maintaining fairness create incentives for individuals to punish those who violate fairness norms, even if the transgression does not affect one's own welfare. Indeed, the desire to punish is so strong that even third-party members—who have no clear vested interest in the equal distribution of resources between others—are still willing to incur a cost to ensure that those peddling unfair allocations are punished (Cronk, Chagnon, & Irons, 2000; Fehr & Fischbacher, 2004). This may be due in part to the notion that failing to punish can itself be punishable (Kandori, 1992).

Decisions to punish unfairness—whether as a victim or third-party member—ultimately demonstrate negative reciprocity, whereby the amount of punishment approximately equals the harm caused (Carlsmith, Darley, & Robinson, 2002; Fehr & Gächter, 2000a, 2000b). As a mechanism for enforcing equality, negative reciprocity encourages individuals to offer fair distributions (Azar, Lahav, & Voslinsky, 2015). Furthermore, people who engage in negative reciprocity can procure positive reputational benefits (Gintis, Smith, & Bowles, 2001). For example, individuals who punish are trusted more, and in return, behave in a more trustworthy manner (Jordan et al., 2016). In contrast, the threat of public exposure of unfair behavior (and thus the possibility of accruing a negative reputation), leads individuals to make more fair offers (Bolton & Zwick, 1995; Straub & Murnighan, 1995). These findings illustrate that negative reciprocity, through punishment, helps enforce and maintain norms of fairness, and by extension the overall well-being of social communities (Gürerk, Irlenbusch, & Rockenbach, 2006; Herrmann et al., 2008).

Altruism

Some accounts of natural selection argue that survival requires self-benefit be prioritized at all costs. At first blush, acts of altruism—choosing to help another at a cost to the self (de Waal, 2008)—seem to significantly reduce one's evolutionary fitness (Darwin, 1859). However, an

influential concept known as 'kin selection' posits that altruistic individuals' genes propagate when prosocial behaviors are performed, which aids in the survival of genetically-related individuals (Trivers, 1971; Wilson, 2000). Accumulating evidence now demonstrates that altruistic behavior is not confined to kin selection strategies, and many species expend valuable resources to help unrelated others (FeldmanHall, Mobbs, et al., 2012; Pitman et al., 2017; Preston & de Waal, 2002; Quervel-Chaumette, Dale, Marshall-Pescini, & Range, 2015; Warneken, Hare, Melis, Hanus, & Tomasello, 2007). Given these findings, it has been subsequently argued that altruism may have evolved for the good of the social community and not just individual genes (Nowak, Tarnita, & Wilson, 2010).

Applying a slight modification to the Ultimatum Game described above elegantly illustrates this point. If the option to reject the offer is removed, the social exchange becomes a Dictator Game (Camerer, 2003) where the Receiver must accept any offer no matter how small. Although the rational decision is for a Dictator (analogous to the Proposer in the UG) to offer the smallest amount of money (since the split is always realized as-is), Dictators routinely go against such monetary self-interest and offer around 28% of their initial endowment (Engel, 2011). Critically, these acts of generosity observed in the lab reflect real-world concerns for altruism (Benz & Meier, 2008; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005; Moll et al., 2006). In America, approximately 60% of households give to charity each year (at a rate of about 4% of a household's income), which totals to more than $250 billion a year (Meer, Miller, & Wulfsberg, 2016).

On the surface, charitable giving does not seem to be a self-beneficial act. Money is given to unknown others, oftentimes in distant countries where there is little chance of meeting those who received the donation. In these cases, it is unlikely that one's altruistic behavior will be directly reciprocated by that specific individual. However, when viewed through the lens of an indirect reciprocity mechanism, the existence of altruistic behaviors has important implications for how we expect humans to behave and be treated in a community (Simpson & Willer, 2008). One example is that individuals hold expectations that people will be behave in generous ways (Brañas-Garza, Rodríguez-Lara, & Sánchez, 2017), and violating these expectations may result in

punishment by third-parties (Fehr & Fischbacher, 2004). Thus, acting selflessly by donating to others provides an advantage to the altruist in that there will be some kind of indirect, downstream benefit (or avoidance of admonishments) from the community at large.

Compellingly, human social groups that act altruistically appear to fare better than those who do not (Ostrom, 2014). Take welfare states for example: even though many Western societies are large and complex, members are intimately dependent on one another, as there are social expectations that people who are more fortunate will help those who are less fortunate (Wilensky, 1974). This norm of altruism ranges from long-term governmental edicts to fleeting one-on-one relationships (Barr, 2012). If an individual is drowning in a lake or falls off a subway platform, people nearby will risk their lives to help the distressed individual (Marsh et al., 2014). These acts of altruism are typically performed without the belief that the beneficiary will directly return the favor. Rather, the expectation (even if implicitly held) is that someone else will display a similarly altruistic act if the altruist were later in a situation and needed help (Nowak & Sigmund, 1998, 2005). Indeed, removing the ability to directly reciprocate a generous act can motivate individuals to 'pay it forward' by helping another in need (Gray, Ward, & Norton, 2014; Hackel & Zaki, 2018).

Such costly indirect altruism is believed to be a key factor in the evolution of human cooperation (Nowak, 2006), and simulations of Dictator Game behavior reveal that indirect and direct generosity is driven by the anticipation of such uncertain future relationships (Delton, Krasnow, Cosmides, & Tooby, 2011; Zisis, Di Guida, Han, Kirchsteiger, & Lenaerts, 2015). This is in part believed to be causally influenced by feelings of moral obligation, social responsibility (Schwartz, 1977), and the knowledge that others are behaving in generous ways (Bartke, Friedl, Gelhaar, & Reh, 2017). For example, activating norms of altruism induces greater helping, and fluctuations in the environment (e.g., the level of a target's expressed distress, number of individuals who can readily help) can either amplify or attenuate altruistic decisions (Cameron & Payne, 2011; Darley & Batson, 1973; FeldmanHall, Dalgleish, Evans, & Mobbs, 2015; Gottlieb & Carver, 1980; Preston, 2013; Preston & de Waal, 2002). Other motivations, such as the desire for social prestige and

reputation (Olson, 1965), or avoiding social ostracism (Becker, 1974), are also known to influence altruism and can be considered positive or negative reciprocity, respectively.

Trust

Trust spans a multitude of situations, cultures, and disciplines, as it is an integral feature of relationships between spouses, friends, teachers and students, and governments and civilians (Cottrell, Neuberg, & Li, 2007). Trust facilitates positive social interactions and has been suggested to be one of the foundations of an efficient economy; there is a strong correlation between economic growth and the percentage of citizens who generally trust others (Knack & Keefer, 1997). This is unsurprising given that a significant aspect of any economic transaction is the ability to trust and cooperate with non-related others (Arrow, 1974). At the dyadic level, deciding to trust—lending money to a friend or sharing personal information with an acquaintance—allows for the formation of partnerships that can produce mutual advantages to maximize an individual's social fitness (Trivers, 1971) and overall societal well-being (Fehr & Camerer, 2007). However, decisions to trust are inherently risky because of the unpredictability and uncertainty of partners' responses during social exchanges (Kosfeld et al., 2005). For example, an untrustworthy individual may fail to repay a loan, or gossip about another's personal information. Without trust, however, neither markets nor social relations could thrive, as there would be an unwillingness to risk something of value in exchange for a later reward.

As with fairness and altruism, trust can be measured using a simple economic game (Berg, Dickhaut, & McCabe, 1995). A typical Trust Game involves a one-shot social interaction between two players, an Investor and a Trustee. The first player, the Investor, is initially faced with a decision to keep an endowment of money (e.g., $10) or share part of it with the Trustee. If shared, the investment is multiplied (often by a factor of four), and the Trustee faces the difficult decision to repay the trust by sending back up to half of the increased sum, or to violate that trust by keeping all the money, leaving the Investor with nothing. The social dilemma for the Investor is clear: though it is more profitable to trust if it will be reciprocated, doing so leaves the Investor susceptible to the risk of a breach in trust, and ultimately, the loss of money. This game can be

adapted for repeated play, such that social sanctions, communication between players, reputation, and relationships can all be manipulated.

In traditional formulations, the Investor normally trusts approximately 50% of their endowment to a Trustee, and the Trustee typically returns 50% of the expanded pie (Camerer, 2003). In one-shot games where there is no opportunity for social sanctions or reputation building through repeated play, it is rather surprising that Trustees return so much of the money, especially since many economists would argue that a rational, self-interested person should return nothing. That Trustees do not exhibit this behavioral pattern—even in situations where individuals are playing together only once and doing so anonymously—suggests the existence of (and adherence to) a moral norm of reciprocal trust (Cox, 2004; Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014; McCabe, Rigdon, & Smith, 2003). Moreover, if Trustees were only motivated by altruistic generosity, then their typical return should map onto the 28% given by Dictators in the Dictator Game (Engel, 2011). Thus, the expectation of direct reciprocity, the critical component of any Trust Game, appears to exist on both sides of the dyad: an individual invests her money because she believes that it will be reciprocated (Ma, Meng, & Shen, 2015), and partners reciprocate the increased monetary sum because there is a strong expectation of reciprocity (Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009; Chang, Smith, Dufwenberg, & Sanfey, 2011; Fareri, Chang, & Delgado, 2012, 2015).

Of course, the degree to which individuals value norms of trust can vary. Even when the parameters of a task are held constant, there are some individuals who resolutely adhere to reciprocal trust norms and others who deviate from this norm (Baumgartner et al., 2009; Cesarini et al., 2008). There are other cases in which individuals might doggedly reciprocate trust in one situation, but swiftly forgo reciprocal behavior when the situation changes (Melnikoff & Bailey, 2018). For example, if a Trustee knows an Investor made a highly risky decision to trust, the Trustee will reciprocate with more money, illustrating the exquisite sensitivity people have to normative signals (Van Den Bos, van Dijk, Westenberg, Rombouts, & Crone, 2009). Evidence that

individuals behave in accordance with, or deviate from, a moral norm depending on the context, suggests that adhering to a moral code of trust can be quite malleable.

Cooperation

Norms can also powerfully influence cooperative behavior (Ostrom, 2014). Behavior in one-shot cooperation problems such as the Prisoner's Dilemma or Public Goods Game reveals that people typically cooperate, despite understanding that it is in one's best self-interest to maximize reward by defecting (Andreoni & Miller, 1993; Blake, Rand, Tingley, & Warneken, 2015; Harrington, 1995; Sally, 1995). From a purely economic perspective, this is a puzzling behavior, as it suggests that the social norm of cooperation is more motivating than maximizing favorable outcomes for the self. Even group size (i.e., playing with four individuals as opposed to 100) does not substantially change the rate at which individuals cooperate (Isaac, Walker, & Williams, 1994), indicating that there is a desire to maintain cooperation even in large, anonymous, and complex settings.

Prominent theories suggest that cooperation exists because of a reciprocal tit-for-tat pattern of behavior (Fehr & Fischbacher, 2004; Fehr & Gächter, 2000a, 2000b; Hamilton & Axelrod, 1981). Once there is an initial signal to cooperate, others will cooperate in return. This notion of conditional cooperation is supported by strong empirical evidence: when communication of intentions is allowed between partners, high levels of cooperation follow suit (Bohnet & Frey, 1999; Messick & Brewer, 1983; Ostrom & Walker, 2003; Sally, 1995). In contrast, cooperation languishes when external rules and sanctions are directly and explicitly imposed, compared to systems that allow internal norms to spontaneously develop (Yamagishi, 1988). These cooperative patterns can be manipulated by expectations of either direct or indirect reciprocity. Repeated play, for instance, typically garners greater rates of cooperative behavior (Fudenberg, Rand, & Dreber, 2012; Nowak, Sasaki, Taylor, & Fudenberg, 2004; Rand & Nowak, 2013). In these cases, individuals form a belief that their fellow partners will cooperate if they cooperate, a form of direct reciprocity. Cooperation can also arise out of indirect reciprocal actions, such as when an individual cooperates knowing that other individuals will be privy to this information (Gächter & Fehr, 1999; Mao, Dworkin, Suri, & Watts, 2017). Such a system allows individuals to enhance

their reputation by cooperating more, thereby procuring the downstream benefits that are associated with positive social standing (Pfeiffer, Tran, Krumme, & Rand, 2012).

As with other norms, patterns of cooperation can vary depending on the setting (Hilbe, Chatterjee, & Nowak, 2018; Ostrom, 2014). Contextual factors—such as whether others around you are cooperating (Fowler & Christakis, 2010; Mao et al., 2017), whether the norm of cooperation has been primed (Capraro, Smyth, Mylona, & Niblo, 2014; Peysakhovich & Rand, 2015), whether resources are abundant (Van Vugt & Samuelson, 1999), whether the size of temptation to freeride is small (Van Lange, Liebrand, Messick, & Wilke, 1992), or whether an individual is a member of a collectivist culture where there are strong norms of reciprocity amongst in-group members (Hofstede, 1980; Leung, 1997)—all positively contribute to an individual ultimately cooperating. Importantly, many of these contextual factors can also shape perceptions of reciprocity. For example, when social cues are available (e.g., discussing strategies with a partner before starting the game), the likelihood of reciprocity rises by as much as 40% (Sally, 1995). In contrast, when there is uncertainty within the environment (e.g., ambiguity around the size of the resource being split, or how many members are using the resource), it reduces an individual's willingness to cooperate (Budescu, Rapoport, & Suleiman, 1990, 1992; Budescu, Suleiman, & Rapoport, 1995).

NORM COMPLIANCE IS REWARDING

From a decision-making perspective, an individual who chooses to comply with moral norms demonstrates that the value of norm compliance is greater than the value of selfishly maximizing one's self-benefit. In this section, we evaluate evidence from the neuroimaging literature that demonstrates how norm compliance and reciprocal behaviors systematically engage the brain's reward network.

The clearest neural evidence that people value reciprocity comes from studies on trust, cooperation, and fairness. In the domain of trust, neuroimaging experiments utilizing the Trust Game illustrate that the caudate, a region critical for indexing reward, computes information

about the intention to reciprocate trusting acts (King-Casas et al., 2005), and that other regions within the reward network—most notably, the ventral tegmental area and ventral striatum—subserve reciprocal exchanges of trust between two players (Krueger et al., 2007; Phan, Sripada, Angstadt, & McCabe, 2010). A recent meta-analysis further reveals that these value signals are likely to be linked to aspects of reciprocity rather than to trust itself. When deciding to trust in repeated games (where direct reciprocity has the opportunity to manifest), there is a high likelihood that ventral striatum is recruited, but not in one-shot games, where direct reciprocity is impossible (Bellucci, Chernyak Sergey, Goodyear, Eickhoff Simon, & Krueger, 2016). Indeed, reciprocation of trust can be experimentally increased by stimulating right orbitofrontal cortex using transcranial direct current stimulation, suggesting that reward regions contribute critically to reciprocal action (Wang, Li, Yin, Li, & Wei, 2016). The value associated with reciprocal trust also appears to be conditional on social distance. Individuals trust close friends with more money than strangers (even when friends and strangers reciprocate at the same rate), which is associated with greater ventral striatum activity (Fareri et al., 2015). Thus, not only does reciprocity appear to depend on immediate observations (i.e., did my partner just behave in a way that reciprocated my trust?), but it also seems linked to previously learned expectations (i.e., is my partner generally someone who would reciprocate my trust?).

Studies of reciprocal cooperation demonstrate a similar engagement of reward-processing regions. An early experiment using the Prisoner's Dilemma observed that mutual cooperation was reported to be highly satisfying, and that these cooperative decisions were associated with enhanced Blood Oxygen Level Dependent (BOLD) activity within in the nucleus accumbens, caudate, and orbitofrontal cortex (Rilling et al., 2002). Subsequent work contrasting neural responses in cooperative and competitive variants of a coordination game found that mutual cooperation recruits orbitofrontal cortex (Decety, Jackson, Sommerville, Chaminade, & Meltzoff, 2004), even when coordination does not increase monetary reward. This suggests that even when monetary benefits to the self are not maximized, the act of cooperating is in itself rewarding.

There is also an abundance of evidence illustrating that reciprocal actions are valued in the wake of a fairness violation. In these cases, however, violating fairness norms characteristically engenders behaviors that are construed as negative reciprocity, such as punishing the perpetrator (Fehr & Fischbacher, 2004). Nearly two decades of work demonstrates that receiving unfair offers in the Ultimatum Game is associated with increased anterior insula and anterior cingulate cortex activity, regions associated with negative emotional experiences and conflict (Chang & Sanfey, 2011; Sanfey et al., 2003; Xiang, Lohrenz, & Montague, 2013). In contrast, receiving fair offers recruits the reward network, including ventral striatum and orbitofrontal cortex (Tabibnia, Satpute, & Lieberman, 2008). These reward regions also become engaged when punishment is levied on the transgressor, suggesting that people highly value enforcing fairness norms, even when punishment comes with a monetary cost (de Quervain, Fischbacher, Treyer, & Schellhammer, 2004; Hu, Strang, & Weber, 2015; Singer et al., 2006).

Neural evidence for the value of reciprocity in altruism is less straightforward and less abundant, largely due to the fact that altruism appears to draw upon an indirect reciprocity mechanism. This tautologically requires that any expected returns from norm compliance be abstracted from the altruistic action itself (e.g., in the form of 'social capital'). Accordingly, identifying the neural underpinnings of reciprocity in the domain of altruism requires observing how indirect reciprocity manifests over time, or at least between multiple individuals in an iterated task. These features make it relatively difficult to study the neural value of reciprocity within an altruistic context, and thus there is limited work on the topic. However, in the few cases in which researchers have fruitfully examined the BOLD signal underpinning the effects of indirect reciprocity during altruistic social exchanges, evidence dovetails with work on trust, fairness, and cooperation: altruistic decisions are influenced by indirect reciprocity motivations, which is subserved, in part, by increased caudate activity (Watanabe et al., 2014). In other words, even an indirect reciprocity mechanism that manifests across multiple individuals behaving altruistically appears to rely on regions that process reward.

In contrast, those who have broken a moral norm (oftentimes to selfishly enhance their own monetary benefit) demonstrate a different pattern of neural activity that does not reliably include reward regions. Several neuroimaging studies across multiple different social domains illustrate that the dorsolateral prefrontal cortex (dlPFC) is recruited when selfishly violating a moral norm (Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; De Neys, Novitskiy, Geeraerts, Ramautar, & Wagemans, 2011; FeldmanHall, Dalgleish, et al., 2012; Ruff, Ugazio, & Fehr, 2013; Yamagishi et al., 2016). Given the role of the dlPFC in cognitive control (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Mansouri, Tanaka, & Buckley, 2009; Ochsner & Gross, 2005), activation of this region during selfish decisions suggests that it may be difficult for individuals to adjudicate between options when a selfish opportunity is sufficiently tempting. Neural activity in dlPFC may therefore reflect the deployment of cognitive control to overcome concern for another's welfare (Rilling et al., 2007). These neural data paint an emerging picture that cognitive control appears to be required to resolve self-other conflicts that ultimately favor the self.

EMOTIONS FACILITATE RECIPROCAL BEHAVIOR

Although emotion has historically been regarded as an irrational and dangerous threat to our moral calculus (Plato, 1955), the last few decades have fruitfully illustrated how emotion can play a special role in the establishment of response-dependent values and norm compliance (D'Arms & Jacobson, 1994; Phelps, Lempert, & Sokol-Hessner, 2014). Take, for instance, a situation where you contemplate cheating on your spouse. You might feel a pang of disapproval or shame upon considering such behavior. These moral emotions moderate moral standards (is it wrong if you are in an unhappy marriage?), and by extension, moral behavior (do you decide to have the affair?). In essence, the link between norm compliance and moral behavior is thought to be influenced by moral emotions (Tangney, Stuewig, & Mashek, 2007), insofar that emotional experiences can sustain one's own compliance with moral norms and motivate enforcement of norm compliance in others (Dunning, Fetchenhauer, & Schlösser, 2012; Fehr & Gächter, 2002).

Self-directed emotions

Guilt and shame are emotions that are explicitly linked to promoting the interests of society rather than one's own interests (Pizarro, 2000). These moral emotions emerge early in childhood (Vaish, 2018), and are negative evaluations of one's own morally transgressive behavior (Eisenberg, 2000). Guilt appears to be a particularly salient motivator of reparative behavior, as it encourages people to make amends for violating moral norms, and can thus enhance how positively the transgressing person is perceived (Stearns & Parrott, 2012). Guilt-proneness consistently correlates with measures of perspective-taking and is inversely related to antisocial and criminal behavior (Tangney et al., 2007). Aligning with these findings, several neuroimaging studies have found that when describing moral transgressions, feelings of guilt are associated with neural activity in a network that corresponds with thinking about other people (Basile et al., 2011; Shin et al., 2000; Takahashi et al., 2004), which may reflect that a key function of guilt is to promote perspective-taking. In these cases, it is likely that individuals are thinking about their partner's expectations, and thus guilt seems to exert the greatest influence on reciprocal moral actions. Because guilt is associated with breaches of moral norms and social standards, the existence of guilt (or even the anticipation of guilt) is a potent motivator for upholding moral norms (Battigalli & Dufwenberg, 2007; Chang et al., 2011).

While emotions such as guilt encourage people to avoid breaking norms, other emotions motivate people to actively comply with norms. For example, some theories propose that empathy sensitizes people to value altruism (Batson et al., 1991; Preston, 2013; Zaki, 2014). To the extent that the interplay between norms enables flexible moral action, it may therefore be the case that empathy's primary contribution is the promotion of altruism above other norms (Rumble Ann, Van Lange Paul, & Parks Craig, 2009; Zaki & Mitchell, 2011), which can be amplified by warm glow motives (Andreoni, 1990; Ashar, Andrews-Hanna, Dimidjian, & Wager, 2017; FeldmanHall et al., 2015). Indeed, recent work on extraordinary altruists demonstrates that these individuals maintain atypically high concern for the welfare of distant others (Vekaria, Brethel-Haurwitz, Cardinale, Stoycos, & Marsh, 2017), a finding that is mirrored by experimental inductions of empathy in normative populations (Klimecki, Mayer, Jusyte, Scheeff , & Schönenberg, 2016). In addition to warm glow motives, other positive emotions (such as

happiness) can also actively facilitate prosocial behaviors through a reward reinforcement mechanism (Aknin, Van de Vondervoort, & Hamlin, 2018).

Other-directed emotions

Negative emotions such as anger and disgust arise from being treated unfairly, and are believed to motivate punishment (Pillutla & Murnighan, 1996; Srivastava, Espinoza, & Fedorikhin, 2009; Van't Wout, Kahn, Sanfey, & Aleman, 2006). Recent work reveals that the act of punishing can alleviate the onslaught of these negative emotional experiences (Hétu, Luo, D'Ardenne, Lohrenz, & Montague, 2017). Unsurprisingly, watching people break moral norms that target other individuals can also give rise to a similar set of moral emotions, including righteous anger, indignation, contempt, and disgust (Dubreuil, 2010; Moll et al., 2002; Rozin, Lowery, Imada, & Haidt, 1999).

Contempt (the moral denunciation of others) is often expressed in response to the violation of communal codes, and is therefore a negative social evaluation of others (Tangney et al., 2007). Contempt is most often expressed by those not directly harmed by the violation, and thus deals with norm compliance from a third-party perspective. Bystanders observing an injustice can express contempt to ostracize the agent causing harm. For example, people feel contempt towards those who violate social hierarchy norms (Rozin et al., 1999). When presented with an angry, contemptuous face criticizing a norm violation, individuals report greater feelings of guilt (Giner-Sorolla & Espinosa, 2011), which can affect rates of future norm compliance. In essence, these third-party emotions are used for social policing, with the aim to minimize morally offensive behavior (Tangney, Miller, Flicker, & Barlow, 1996).

Atypical emotion processing

Emotion's critical role in guiding norm compliance is even more evident when considering populations whose processing of emotions is atypical. Individuals who fail to generate an emotional arousal response before approving harmful, immoral actions illustrate how a lack of anticipatory emotional response results in behavior that is insensitive to moral norms and future

consequences (Blair, 1996; Harenski, Harenski, Shane, & Kiehl, 2010; Moretto, Làdavas, Mattioli, & Di Pellegrino, 2010; Rilling et al., 2007; Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). For example, lesions to the medial frontal cortex (mPFC) typically lead to blunted emotional responding (Bechara, Damasio, & Damasio, 2000), and accumulating evidence indicates that this region is critical for evaluating emotional states and integrating this information within the context of current goal states, such as adhering to relevant social norms (Forbes & Grafman, 2010). In other words, the mPFC likely processes internal emotional signals alongside cues about social norms to help guide successful moral behavior.

Individuals diagnosed with psychopathy and conduct disorder also provide a compelling case for the intimate link between disrupted emotional responses and patterns of aberrant moral behaviors. For example, when watching others in pain, adult psychopaths, adolescents who exhibit psychopathic traits, and adolescents diagnosed with conduct disorder all show attenuated engagement of brain regions known to respond to another's pain (Decety, Chen, Harenski, & Kiehl, 2013; Decety, Michalska, Akitsuki, & Lahey, 2009; Marsh et al., 2013). For psychopaths, these failures in appreciating the emotional aspects of a victim's suffering has been explicitly linked to both abnormal (i.e., immoral) judgments (Young, Koenigs, Kruepke, & Newman, 2012), and an insensitivity to norms of generosity (Koenigs, Kruepke, & Newman, 2010). There is emerging research, however, that suggests these aberrant moral behaviors may also be a product of failures in processing value (Baskin-Sommers, Stuppy-Sullivan, & Buckholtz, 2016; Hosking et al., 2017; Mitchell et al., 2006). Individuals with higher levels of psychopathy cooperate less and exhibit reduced activity in orbitofrontal cortex when cooperating (Rilling et al., 2007), hinting at a causal role of reward in motivating cooperative behavior. However, given the intimate link between emotion and reward (Adolphs, 2002; Berridge & Robinson, 2003; Knutson, Adams, Fong, & Hommer, 2001; Kringelbach, 2005; Murray, 2007; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; Phelps & LeDoux, 2005), it is likely that perturbed representations of value manifest because of failures in generating an emotional response (Bechara, 2004; Bechara, Damasio, Damasio, & Anderson, 1994; Bechara, Damasio, Tranel, & Damasio, 1997), which can subsequently result in immoral behavior.

LEARNING MORAL NORMS THROUGH RECIPROCITY

Moral norms develop and are transmitted through social interactions and relationships (Ho, MacGlashan, Littman, & Cushman, 2017). The frequency with which these norms are attended and adhered to suggests that they are indoctrinated at an early age (House, 2018). Children as young as three years old can distinguish between legal and social violations (Smetana, 1983). Recent developmental research further reveals that children begin to obey norms after an authority figure illustrates they should be followed (Hardecker & Tomasello, 2017; Schmidt, Butler, Heinz, & Tomasello, 2016). Such vicarious learning appears early in the developmental trajectory, and can help facilitate the distinction between social and moral norms in young children (e.g., wearing pajamas to school versus hitting another; (Turiel, 1983). As a child's moral calculus develops further, they begin to consider contextual factors, such as intent, provocation, and duty when evaluating which moral norms might be appropriate for the situation (Engelmann, Herrmann, & Tomasello, 2017).

Once learned, moral norms seem to be sustained through reward and punishment contingencies that are based on expectations of reciprocity (Göckeritz, Schmidt, & Tomasello, 2014; Hardecker, Schmidt, & Tomasello, 2017; Leimgruber, 2018). These expectations can be expressed both directly (e.g., monetary benefit) and indirectly (e.g., gaining social capital; (Hackel & Zaki, 2018). For example, breaking certain social norms, such as wearing the wrong outfit to school, can evoke scorn and mockery from peers, and if the transgression is particularly egregious, it may even induce social rejection. Accordingly, the feedback received from others acts as a reinforcement mechanism that can dictate the adherence to (or deviance from) moral norms (Aknin et al., 2018). Over the last few years, researchers have begun to successfully apply reinforcement learning frameworks to explain a mechanism for how social learning unfolds. For example, prediction errors—when an actual outcome deviates from an expected outcome—allow individuals to update their expectations about the social world to align with reality (Behrens, Hunt, Woolrich, & Rushworth, 2008; FeldmanHall, Otto, & Phelps, 2018; Joiner, Piva, Turrin, & Chang, 2017; Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009; Montague & Lohrenz, 2007). These

prediction errors, which are largely generated by the midbrain dopaminergic system and the structures it innervates (Haber & Knutson, 2010; Ruff & Fehr, 2014; Schultz, Dayan, & Montague, 1997), may drive moral learning by encoding norm violations.

In one of the first studies illustrating that norm violations generate prediction errors, researchers found that subjects in a Trust Game transferred less money to partners who violated trust (King-Casas et al., 2005). This behavior was underpinned by prediction error signals in the caudate, such that the magnitude of neural activity in response to a partner's reciprocation (or lack thereof) tracked decisions to trust a partner with more money on the next round. Though the prediction error signal was initially observed after subjects saw feedback about whether a partner upheld a trust norm, it began to shift backward in time as subjects learned more about a partner's trustworthiness, suggesting that subjects were developing a stable impression of their partner's moral traits (i.e., their willingness to reciprocate). Subsequent work further decoupled monetary reward from learning about moral traits (e.g., generosity), revealing that activity in a key learning hub—the ventral striatum—indexes dissociable prediction errors when learning about money and stable moral characteristics such as generosity (Hackel, Doll, & Amodio, 2015). Moreover, prediction errors associated with learning about another's generosity correlated with activity in a network of brain regions implicated in impression updating (including ventrolateral prefrontal cortex and right temporoparietal junction), illustrating that people find norm violations especially diagnostic in helping to form a stable impression of another's personality (Mende-Siedlecki, Baron, & Todorov, 2013).

However, an individual's ability to glean information about their social world (and to subsequently adaptively update their behavior) may depend on the social context and the relevant moral norm. To probe whether prediction errors are contextually modulated, researchers have dynamically manipulated moral expectations using the Ultimatum Game. When led to believe that unfair offers are ubiquitous, subjects were less willing to punish partners who break fairness norms (Sanfey, 2009), which provides compelling evidence that people adjust their behaviors according to the prevailing norms of a specific social environment. These context-

sensitive decisions to punish were supported by prediction errors in canonical learning regions such as ventral striatum, substantia nigra, and VTA (Hétu et al., 2017; Xiang et al., 2013). The notion that stable impressions about another's moral traits are dependent on moral expectations is also supported by memory research. In a study from our own lab examining how decision-making is influenced by episodic memory, we observed that people adaptively play with past partners if accurate impressions of the partner's norm compliance have been fully encoded by rich episodic memories (Murty, FeldmanHall, Hunter, Phelps, & Davachi, 2016). Together, these results suggest that people use learned impressions of others' moral traits to guide adaptive decision-making.

The tight coupling between norms, moral learning, and adaptive decision-making demonstrates that people use knowledge of norm violations to form impressions of others' moral traits. Direct experience of another person's failure to comply with norms produces prediction errors, and these errors drive fast and flexible learning about others' moral traits, such as generosity and trustworthiness (Hackel et al., 2015; King-Casas et al., 2005). Once moral impressions stabilize, learning regions cease to track deviations from expected normative behavior (Delgado, Frank, & Phelps, 2005). Because norm violations provide diagnostic information about others' traits (Mende-Siedlecki et al., 2013), stable impressions can guide optimal choices by enabling people to affiliate with those who are likely to be rewarding social partners and to avoid those who are likely to be unrewarding (Murty et al., 2016). In fact, these moral impressions can weigh so heavily on social decisions that people choose not to cooperate with a stranger if they know that the stranger is friends with a norm violator (Martinez, Mack, Gelman, & Preston, 2016).

INTEGRATION WITH OTHER THEORIES

While we posit that moral decision-making is largely motivated by four fundamental norms, other prominent theories have argued that a number of additional norms are critical for successful socialization (Moral Foundations Theory; (Haidt, 2007), or that all moral behaviors can be reduced to a single motivation—the desire to reduce harm (Theory of Dyadic Morality; (Schein & Gray, 2017). Here, we have tried to strike a balance between parsimony and explanatory power.

For example, Moral Foundations Theory may place undue weight on certain norms (e.g., purity) that are less represented in many everyday moral quandaries. On the other hand, the Theory of Dyadic Morality is parsimonious by its very nature. Although we would agree that many moral situations can be perceived through a lens of harm, such an account can be overly restrictive when trying to explain the wide range of findings in psychology, economics, and neuroscience.

By allowing the findings from psychology and neuroeconomics to guide us, we have highlighted reciprocity as a common mechanism that motivates adherence to a discrete suite of moral norms. The idea that reciprocity provides a unifying principle for social behavior is not new (Berg et al., 1995; Bolton & Ockenfels, 2000; de Waal & Luttrell, 1988; Fehr & Gächter, 2000b; Gouldner, 1960). Early models such as Social Exchange Theory suggested that reciprocity is a universally-held principle (Gouldner, 1960), and that high-quality relationships can emerge and flourish through reciprocal actions (Cropanzano & Mitchell, 2005; Thibaut & Kelley, 1959). We build on this work, examining how this one simple mechanism can explain people's adherence to a set of specific moral norms, and how these moral norms collectively provide an overarching framework for understanding moral behavior across a variety of domains.

CONCLUSIONS

Moral norms facilitate harmonious interpersonal exchanges by providing people with a set of common expectations. Here we highlight four norms—fairness, altruism, trust, and cooperation—that we believe to be the most foundational for successful social living. By discussing the ways in which these norms can shape behavior, we offer an account for the proximate psychological mechanisms motivating moral norm compliance: reciprocity. People comply with moral norms because they have the direct or indirect expectation that others will also adhere to these norms, and because they believe that norm violations may have negative repercussions for the future well-being of both specific individuals and entire societies. Activation in the brain's reward network supports active adherence to moral norms, suggesting that people find value in complying with norms and engaging in reciprocal behaviors with others. In addition, we examine how aversive moral emotions such as contempt and guilt facilitate norm

enforcement by devaluing selfish, norm-violating actions. Finally, we review evidence that learning about norm violators depends on a network of brain regions that encode for reward and violated expectations of receiving reward, suggesting that people learn about others' social value through a reinforcement learning mechanism.

The degree to which humans act fairly, help, trust, and cooperate is often viewed as a puzzle across an array of disciplines. Some of the deepest thinkers in human history, including Adam Smith, Jean-Jacques Rousseau, and Charles Darwin, have attempted to provide accounts of how social norms dictate appropriate behaviors in nearly every aspect of human life, from the trivial (e.g. wearing the correct attire to a wedding) to the deeply consequential (e.g. punishing a criminal with the death sentence). However, few accounts have successfully reconciled two seemingly contradictory features of norm compliance: although social norms are pervasive and often perceived as inflexible in nature, the degree to which an individual adheres to these norms produces malleable and context-specific behaviors. Emerging research in moral psychology and neuroscience elucidates how norms are supported by the simple cognitive mechanism of reciprocity. Reciprocal behavior is stable enough to support interpersonal exchanges between strangers, yet flexible enough to accommodate adaptive behavior across a range of social environments. We provide a unifying framework for understanding how a wide variety of putatively unrelated moral behavior—helping a homeless person, getting angry at a fraudster, asking a stranger at the library to look after your computer while you take a call—are supported by expectations of reciprocity and the associated neural encoding of reward.

CONFLICT OF INTEREST

None declared.

REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*(2), 169-177. doi:https://doi.org/10.1016/S0959-4388(02)00301-X

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. doi:https://doi.org/10.1016/0749-5978(91)90020-T

Aknin, L. B., Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Positive feelings reward and promote prosocial behavior. *Current Opinion in Psychology, 20*, 55-59. doi:https://doi.org/10.1016/j.copsyc.2017.08.017

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal, 100*(401), 464-477. doi:https://doi.org/10.2307/2234133

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal, 103*(418), 570-585. doi:https://doi.org/10.2307/2234532

Arrow, K. J. (1974). *The limits of organization*: WW Norton & Company.

Ashar, Y. K., Andrews-Hanna, J. R., Dimidjian, S., & Wager, T. D. (2017). Empathic Care and Distress: Predictive Brain Markers and Dissociable Brain Systems. *Neuron, 94*(6), 1263-1273.e1264. doi:https://doi.org/10.1016/j.neuron.2017.05.014

Azar, O. H., Lahav, Y., & Voslinsky, A. (2015). Beliefs and social behavior in a multi-period ultimatum game. *Frontiers in Behavioral Neuroscience, 9*. doi:https://doi.org/10.3389/fnbeh.2015.00029

Barr, N. (2012). *Economics of the welfare state*: Oxford University Press.

Bartke, S., Friedl, A., Gelhaar, F., & Reh, L. (2017). Social comparison nudges—Guessing the norm increases charitable giving. *Economics Letters, 152*, 73-75. doi:https://doi.org/10.1016/j.econlet.2016.12.023

Basile, B., Mancini, F., Macaluso, E., Caltagirone, C., Frackowiak, R. S. J., & Bozzali, M. (2011). Deontological and altruistic guilt: Evidence for distinct neurobiological substrates. *Human Brain Mapping, 32*(2), 229-239. doi:https://doi.org/10.1002/hbm.21009

Baskin-Sommers, A., Stuppy-Sullivan, A. M., & Buckholtz, J. W. (2016). Psychopathic individuals exhibit but do not avoid regret during counterfactual decision making. *Proceedings of the National Academy of Sciences, 113*(50), 14438. doi:https://doi.org/10.1073/pnas.1609985113

Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology, 61*(3), 413-426. doi:https://doi.org/10.1037/0022-3514.61.3.413

Battigalli, P., & Dufwenberg, M. (2007). Guilt in Games. *The American Economic Review, 97*(2), 170-176. doi:https://doi.org/10.1257/aer.97.2.170

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron, 64*(5), 756-770. doi:https://doi.org/10.1016/j.neuron.2009.11.017

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience, 14*(11), 1468-1474. doi:https://doi.org/10.1038/nn.2933

Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition, 55*(1), 30-40. doi:https://doi.org/10.1016/j.bandc.2003.04.001

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1), 7-15. doi:https://doi.org/10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebral Cortex, 10*(3), 295-307. doi:https://doi.org/10.1093/cercor/10.3.295

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding Advantageously Before Knowing the Advantageous Strategy. *Science, 275*(5304), 1293. doi:https://doi.org/10.1126/science.275.5304.1293

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy, 82*(6), 1063-1093. doi:https://doi.org/10.1086/260265

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*(7219), 245-249. doi:https://doi.org/10.1038/nature07538

Bellucci, G., Chernyak Sergey, V., Goodyear, K., Eickhoff Simon, B., & Krueger, F. (2016). Neural signatures of trust in reciprocity: A coordinate-based meta-analysis. *Human Brain Mapping, 38*(3), 1233-1248. doi:https://doi.org/10.1002/hbm.23451

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?—evidence from donations. *Experimental Economics, 11*(3), 268-281. doi:https://doi.org/10.1007/s10683-007-9192-y

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122-142. doi:https://doi.org/10.1006/game.1995.1027

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences, 26*(9), 507-513. doi:https://doi.org/10.1016/S0166-2236(03)00233-9

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*: Cambridge University Press.

Bicchieri, C., & Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making, 23*(2), 161-178. doi:https://doi.org/10.1002/bdm.648

Blair, R. J. R. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders, 26*(5), 571-579. doi:https://doi.org/10.1007/BF02172277

Blake, P. R., Rand, D. G., Tingley, D., & Warneken, F. (2015). The shadow of the future promotes cooperation in a repeated prisoner's dilemma for children. *Scientific Reports, 5*. doi:https://doi.org/10.1038/srep14559

Bohnet, I., & Frey, B. S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior & Organization, 38*(1), 43-57. doi:https://doi.org/10.1016/S0167-2681(98)00121-8

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review, 90*(1), 166-193. doi:https://doi.org/10.1257/aer.90.1.166

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior, 10*(1), 95-121. doi:https://doi.org/10.1006/game.1995.1026

Brañas-Garza, P., Rodríguez-Lara, I., & Sánchez, A. (2017). Humans expect generosity. *Scientific Reports, 7*, 42446. doi:https://doi.org/10.1038/srep42446

Brosnan, S. F., & de Waal, F. B. M. (2014). Evolution of responses to (un)fairness. *Science, 346*(6207), 1251776. doi:https://doi.org/10.1126/science.1251776

Budescu, D. V., Rapoport, A., & Suleiman, R. (1990). Resource dilemmas with environmental uncertainty and asymmetric players. *European Journal of Social Psychology, 20*(6), 475-487. doi:https://doi.org/10.1002/ejsp.2420200603

Budescu, D. V., Rapoport, A., & Suleiman, R. (1992). Simultaneous vs. sequential requests in resource dilemmas with incomplete information. *Acta Psychologica, 80*(1), 297-310. doi:https://doi.org/10.1016/0001-6918(92)90052-F

Budescu, D. V., Suleiman, R., & Rapoport, A. (1995). Positional order and group size effects in resource dilemmas with uncertain resources. *Organizational Behavior and Human Decision Processes, 61*(3), 225-238. doi:https://doi.org/10.1006/obhd.1995.1018

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*: Princeton University Press.

Camerer, C., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives, 9*(2), 209-219. doi:https://doi.org/10.1257/jep.9.2.209

Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology, 100*(1), 1-15. doi:https://doi.org/10.1037/a0021643

Capraro, V., Smyth, C., Mylona, K., & Niblo, G. A. (2014). Benevolent characteristics promote cooperative behaviour among humans. *PLoS One, 9*(8), e102881. doi:https://doi.org/10.1371/journal.pone.0102881

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish?: Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*(2), 284. doi:https://doi.org/10.1037/0022-3514.83.2.284

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of Sciences, 105*(10), 3721-3726. doi:https://doi.org/10.1073/pnas.0710069105

Chang, L. J., & Sanfey, A. G. (2011). Great expectations: neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience, 8*(3), 277-284. doi:https://doi.org/10.1093/scan/nsr094

Chang, L. J., & Smith, A. (2015). Social emotions and psychological games. *Current Opinion in Behavioral Sciences, 5*, 133-140. doi:https://doi.org/10.1016/j.cobeha.2015.09.010

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron, 70*(3), 560-572. doi:https://doi.org/10.1016/j.neuron.2011.02.056

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117*(3), 817-869. doi:https://doi.org/10.1162/003355302760193904

Cialdini, R. B. (2003). Crafting Normative Messages to Protect the Environment. *Current Directions in Psychological Science, 12*(4), 105-109. doi:https://doi.org/10.1111/1467-8721.01242

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*(6), 1015. doi:http://dx.doi.org/10.1037/0022-3514.58.6.1015

Cottrell, C. A., Neuberg, S. L., & Li, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology, 92*(2), 208-231. doi:https://doi.org/10.1037/0022-3514.92.2.208

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior, 46*(2), 260-281. doi:https://doi.org/10.1016/S0899-8256(03)00119-2
Cronk, L., Chagnon, N., & Irons, W. (2000). *Adaptation and human behavior: an anthropological perspective*: Routledge.
Cropanzano, R., & Mitchell, M. S. (2005). Social Exchange Theory: An Interdisciplinary Review. *Journal of Management, 31*(6), 874-900. doi:https://doi.org/10.1177/0149206305279602
D'Arms, J., & Jacobson, D. (1994). Expressivism, morality, and the emotions. *Ethics, 104*(4), 739-763. doi:https://doi.org/10.1086/293653
Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology, 27*(1), 100-108. doi:https://doi.org/10.1037/h0034449
Darwin, C. (1859). *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. London: J. Murray.
Dawkins, R. (1989). *The selfish gene*. New York, NY: Oxford University Press.
De Neys, W., Novitskiy, N., Geeraerts, L., Ramautar, J., & Wagemans, J. (2011). Cognitive control and individual differences in economic ultimatum decision-making. *PLoS One, 6*(11), e27107. doi:https://doi.org/10.1371/journal.pone.0027107
de Quervain, D. J. F., Fischbacher, U., Treyer, V., & Schellhammer, M. (2004). The neural basis of altruistic punishment. *Science, 305*(5688), 1254. doi:https://doi.org/10.1126/science.1100735
de Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology, 59*, 279-300. doi:https://doi.org/10.1146/annurev.psych.59.103006.093625
de Waal, F. B. M. (2009). *Primates and philosophers: How morality evolved*: Princeton University Press.
de Waal, F. B. M., & Luttrell, L. M. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology, 9*(2), 101-118. doi:https://doi.org/10.1016/0162-3095(88)90016-7
Decety, J., Chen, C., Harenski, C., & Kiehl, K. A. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Frontiers in Human Neuroscience, 7*, 489. doi:https://doi.org/10.3389/fnhum.2013.00489
Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *NeuroImage, 23*(2), 744-751. doi:https://doi.org/10.1016/j.neuroimage.2004.05.025
Decety, J., Michalska, K. J., Akitsuki, Y., & Lahey, B. B. (2009). Atypical empathic responses in adolescents with aggressive conduct disorder: A functional MRI investigation. *Biological Psychology, 80*(2), 203-211. doi:https://doi.org/10.1016/j.biopsycho.2008.09.004
Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8*(11), 1611-1618. doi:https://doi.org/10.1038/nn1575
Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences, 108*(32), 13335. doi:https://doi.org/10.1073/pnas.1102131108

Dubreuil, B. (2010). Punitive emotions and norm violations. *Philosophical Explorations, 13*(1), 35-50. doi:https://doi.org/10.1080/13869790903486776
Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47*(2), 268-298. doi:https://doi.org/10.1016/j.geb.2003.06.003
Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122-141. doi:https://doi.org/10.1037/a0036673
Dunning, D., Fetchenhauer, D., & Schlösser, T. M. (2012). Trust as a social and emotional act: Noneconomic considerations in trust behavior. *Journal of Economic Psychology, 33*(3), 686-694. doi:https://doi.org/10.1016/j.joep.2011.09.005
Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51*(1), 665-697.
Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who takes the most revenge? Individual differences in negative reciprocity norm endorsement. *Personality and Social Psychology Bulletin, 30*(6), 787-799. doi:https://doi.org/10.1177/0146167204264047
Engel, C. (2011). Dictator games: a meta study. *Experimental Economics, 14*(4), 583-610. doi:https://doi.org/10.1007/s10683-011-9283-7
Engelmann, J. M., Herrmann, E., & Tomasello, M. (2017). Concern for Group Reputation Increases Prosociality in Young Children. *Psychological Science*, 0956797617733830. doi:https://doi.org/10.1177/0956797617733830
Fareri, D. S., Chang, L. J., & Delgado, M. R. (2012). Effects of Direct Social Experience on Trust Decisions and Neural Reward Circuitry. *Frontiers in Neuroscience, 6*, 148. doi:https://doi.org/10.3389/fnins.2012.00148
Fareri, D. S., Chang, L. J., & Delgado, M. R. (2015). Computational Substrates of Social Value in Interpersonal Collaboration. *The Journal of Neuroscience, 35*(21), 8170. doi:https://doi.org/10.1523/JNEUROSCI.4775-14.2015
Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences, 11*(10), 419-427. doi:https://doi.org/10.1016/j.tics.2007.09.002
Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785-791. doi:https://doi.org/10.1038/nature02043
Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63-87. doi:https://doi.org/10.1016/S1090-5138(04)00005-4
Fehr, E., & Gächter, S. (2000a). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review, 90*(4), 980-994. doi:https://doi.org/10.1257/aer.90.4.980
Fehr, E., & Gächter, S. (2000b). Fairness and Retaliation: The Economics of Reciprocity. *The Journal of Economic Perspectives, 14*(3), 159-181. doi:https://doi.org/10.1257/jep.14.3.159
Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137-140. doi:https://doi.org/10.1038/415137a
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817-868. doi:https://doi.org/10.1162/003355399556151

FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. *NeuroImage, 105*, 347-356. doi:https://doi.org/10.1016/j.neuroimage.2014.10.043

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience, 7*(7), 743-751. doi:https://doi.org/10.1093/scan/nss069

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition, 123*(3), 434-441. doi:https://doi.org/10.1016/j.cognition.2012.02.001

FeldmanHall, O., Otto, A. R., & Phelps, E. A. (2018). Learning another's preference to punish enhances one's own punitive behavior. *Journal of Experimental Psychology: General*.

Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience, 33*, 299-324. doi:https://doi.org/10.1146/annurev-neuro-060909-153230

Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences, 107*(12), 5334-5338. doi:https://doi.org/10.1073/pnas.0913149107

Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World. *The American Economic Review, 102*(2), 720-749. doi:https://doi.org/10.1257/aer.102.2.720

Gächter, S., & Fehr, E. (1999). Collective action as a social exchange. *Journal of Economic Behavior & Organization, 39*(4), 341-369. doi:https://doi.org/10.1016/S0167-2681(99)00045-1

Giner-Sorolla, R., & Espinosa, P. (2011). Social cuing of guilt by anger and of shame by disgust. *Psychological Science, 22*(1), 49-53. doi:https://doi.org/10.1177/0956797610392925

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology, 206*(2), 169-179. doi:https://doi.org/10.1006/jtbi.2000.2111

Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong Reciprocity and the Roots of Human Morality. *Social Justice Research, 21*(2), 241-253. doi:https://doi.org/10.1007/s11211-008-0067-y

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology, 213*(1), 103-119. doi:https://doi.org/10.1006/jtbi.2001.2406

Göckeritz, S., Schmidt, M. F. H., & Tomasello, M. (2014). Young children's creation and transmission of social norms. *Cognitive Development, 30*, 81-95. doi:https://doi.org/10.1016/j.cogdev.2014.01.003

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472-482. doi:https://doi.org/10.1086/586910

Gottlieb, J., & Carver, C. S. (1980). Anticipation of future interaction and the bystander effect. *Journal of Experimental Social Psychology, 16*(3), 253-260. doi:https://doi.org/10.1016/0022-1031(80)90068-2

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 161-178. doi:https://doi.org/10.2307/2092623

Gray, K., Ward, A. F., & Norton, M. I. (2014). Paying it forward: Generalized reciprocity and the limits of generosity. *Journal of Experimental Psychology: General, 143*(1), 247-254. doi:https://doi.org/10.1037/a0031047

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron, 44*(2), 389-400. doi:https://doi.org/10.1016/j.neuron.2004.09.027

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The Competitive Advantage of Sanctioning Institutions. *Science, 312*(5770), 108. doi:https://doi.org/10.1126/science.1123633

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*(4), 367-388. doi:https://doi.org/10.1016/0167-2681(82)90011-7

Haber, S. N., & Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human Imaging. *Neuropsychopharmacology, 35*(1), 4-26. doi:https://doi.org/10.1038/npp.2009.129

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nature Neuroscience, 18*(9), 1233-1235. doi:https://doi.org/10.1038/nn.4080

Hackel, L. M., & Zaki, J. (2018). Propagation of Economic Inequality Through Reciprocity and Reputation. *Psychological Science, 29*(4), 604-613. doi:https://doi.org/10.1177/0956797617741720

Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science, 316*(5827), 998.

Hamilton, W. D., & Axelrod, R. (1981). The evolution of cooperation. *Science, 211*(27), 1390-1396. doi:https://doi.org/10.1126/science.7466396

Hardecker, S., Schmidt, M. F. H., & Tomasello, M. (2017). Children's Developing Understanding of the Conventionality of Rules. *Journal of Cognition and Development, 18*(2), 163-188. doi:https://doi.org/10.1080/15248372.2016.1255624

Hardecker, S., & Tomasello, M. (2017). From imitation to implementation: How two- and three-year-old children learn to enforce social norms. *British Journal of Developmental Psychology, 35*(2), 237-248. doi:https://doi.org/10.1111/bjdp.12159

Harenski, C. L., Harenski, K. A., Shane, M. S., & Kiehl, K. A. (2010). Aberrant neural processing of moral violations in criminal psychopaths. *Journal of Abnormal Psychology, 119*(4), 863-874. doi:https://doi.org/10.1037/a0020979

Harrington, J. E. (1995). Cooperation in a one-shot Prisoners' Dilemma. *Games and Economic Behavior, 8*(2), 364-377. doi:https://doi.org/10.1016/S0899-8256(05)80006-5

Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*: Ecco/HarperCollins Publishers.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., . . . Ensminger, J. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences, 28*(6), 795-815. doi:https://doi.org/10.1017/S0140525X05000142

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., . . . Henrich, N. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*(5972), 1480-1484. doi:https://doi.org/10.1126/science.1182238

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly Punishment Across Human Societies. *Science, 312*(5781), 1767. doi:https://doi.org/10.1126/science.1127333

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial Punishment Across Societies. *Science, 319*(5868), 1362. doi:https://doi.org/10.1126/science.1153808

Hétu, S., Luo, Y., D'Ardenne, K., Lohrenz, T., & Montague, P. R. (2017). Human substantia nigra and ventral tegmental area involvement in computing social error signals during the ultimatum game. *Social Cognitive and Affective Neuroscience*, nsx097-nsx097. doi:https://doi.org/10.1093/scan/nsx097

Hilbe, C., Chatterjee, K., & Nowak, M. A. (2018). Partners and rivals in direct reciprocity. *Nature Human Behaviour*. doi:https://doi.org/10.1038/s41562-018-0320-9

Ho, M. K., MacGlashan, J., Littman, M. L., & Cushman, F. (2017). Social is special: A normative framework for teaching with and learning from evaluative feedback. *Cognition*. doi:https://doi.org/10.1016/j.cognition.2017.03.006

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: SAGE Publications.

Hosking, J. G., Kastman, E. K., Dorfman, H. M., Samanez-Larkin, G. R., Baskin-Sommers, A., Kiehl, K. A., . . . Buckholtz, J. W. (2017). Disrupted Prefrontal Regulation of Striatal Subjective Value Signals in Psychopathy. *Neuron, 95*(1), 221-231.e224. doi:https://doi.org/10.1016/j.neuron.2017.06.030

House, B. R. (2018). How do social norms influence prosocial development? *Current Opinion in Psychology, 20*, 87-91. doi:https://doi.org/10.1016/j.copsyc.2017.08.011

Hu, Y., Strang, S., & Weber, B. (2015). Helping or punishing strangers: neural correlates of altruistic decisions as third-party and of its relation to empathic concern. *Frontiers in Behavioral Neuroscience, 9*. doi:https://doi.org/10.3389/fnbeh.2015.00024

Isaac, R. M., Walker, J. M., & Williams, A. W. (1994). Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of Public Economics, 54*(1), 1-36. doi:https://doi.org/10.1016/0047-2727(94)90068-X

Joiner, J., Piva, M., Turrin, C., & Chang, S. W. C. (2017). Social learning through prediction error in the brain. *npj Science of Learning, 2*(1), 8. doi:https://doi.org/10.1038/s41539-017-0009-2

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530*(7591), 473-476. doi:https://doi.org/10.1038/nature16981

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, S285-S300.

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies, 59*(1), 63-80. doi:https://doi.org/10.2307/2297925

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science, 308*(5718), 78. doi:https://doi.org/10.1126/science.1108062

Klimecki, O. M., Mayer, S. V., Jusyte, A., Scheeff, J., & Schönenberg, M. (2016). Empathy promotes altruistic behavior in economic interactions. *Scientific Reports, 6*, 31961. doi:https://doi.org/10.1038/srep31961

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement Learning Signal Predicts Social Conformity. *Neuron, 61*(1), 140-151. doi:https://doi.org/10.1016/j.neuron.2008.11.027

Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics, 112*(4), 1251-1288. doi:https://doi.org/10.1162/003355300555475

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of Increasing Monetary Reward Selectively Recruits Nucleus Accumbens. *The Journal of Neuroscience, 21*(16), RC159. doi:https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001

Koenigs, M., Kruepke, M., & Newman, J. P. (2010). Economic decision-making in psychopathy: A comparison with ventromedial prefrontal lesion patients. *Neuropsychologia, 48*(7), 2198-2204. doi:https://doi.org/10.1016/j.neuropsychologia.2010.04.012

Kohlberg, L., & Hersh, R. H. (1977). Moral Development: A Review of the Theory. *Theory Into Practice, 16*(2), 53-59.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676. doi:https://doi.org/10.1038/nature03701

Kringelbach, M. L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience, 6*, 691. doi:https://doi.org/10.1038/nrn1747

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., . . . Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences, 104*(50), 20084-20089. doi:https://doi.org/10.1073/pnas.0710103104

Leimgruber, K. L. (2018). The developmental emergence of direct reciprocity and its influence on prosocial behavior. *Current Opinion in Psychology, 20*, 122-126. doi:https://doi.org/10.1016/j.copsyc.2018.01.006

Lerner, M. J. (1980). The belief in a just world *The Belief in a Just World* (pp. 9-30): Springer.

Leung, K. (1997). Negotiation and reward allocations across cultures *New perspectives on international industrial/organizational psychology.* (pp. 640-675). San Francisco, CA, US: The New Lexington Press/Jossey-Bass Publishers.

Ma, Q., Meng, L., & Shen, Q. (2015). You Have My Word: Reciprocity Expectation Modulates Feedback-Related Negativity in the Trust Game. *PLoS One, 10*(2), e0119129. doi:https://doi.org/10.1371/journal.pone.0119129

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience, 10*, 141. doi:https://doi.org/10.1038/nrn2538

Mao, A., Dworkin, L., Suri, S., & Watts, D. J. (2017). Resilient cooperators stabilize long-run cooperation in the finitely repeated Prisoner's Dilemma. *Nature Communications, 8*, 13800. doi:https://doi.org/10.1038/ncomms13800

Marsh, A. A., Finger, E. C., Fowler, K. A., Adalio, C. J., Jurkowitz, I. T. N., Schechter, J. C., . . . Blair, R. J. R. (2013). Empathic responsiveness in amygdala and anterior cingulate cortex in youths with psychopathic traits. *Journal of Child Psychology and Psychiatry, 54*(8), 900-910. doi:https://doi.org/10.1111/jcpp.12063

Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., & Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences, 111*(42), 15036-15041. doi:https://doi.org/10.1073/pnas.1408440111

Martinez, J. E., Mack, M. L., Gelman, B. D., & Preston, A. R. (2016). Knowledge of Social Affiliations Biases Economic Decisions. *PLoS One, 11*(7), e0159918. doi:https://doi.org/10.1371/journal.pone.0159918

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization, 52*(2), 267-275. doi:https://doi.org/10.1016/S0167-2681(03)00003-9

Meer, J., Miller, D. H., & Wulfsberg, E. (2016). The Great Recession and Charitable Giving. *National Bureau of Economic Research Working Paper Series, No. 22902*. doi:https://doi.org/10.3386/w22902

Melnikoff, D. E., & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. *Proceedings of the National Academy of Sciences*. doi:https://doi.org/10.1073/pnas.1714945115

Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *The Journal of Neuroscience, 33*(50), 19406-19415. doi:https://doi.org/10.1523/JNEUROSCI.2334-13.2013

Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas: A review. *Review of Personality and Social Psychology, 4*(1), 11-44.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11*(4), 143-152. doi:https://doi.org/10.1016/j.tics.2006.12.007

Mikhail, J. (2008). The Poverty of the Moral Stimulus. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Evolution of Morality: Adaptations and Innateness* (Vol. 1, pp. 353): MIT Press.

Mitchell, D. G. V., Fine, C., Richell, R. A., Newman, C., Lumsden, J., Blair, K. S., & Blair, R. J. R. (2006). Instrumental learning and relearning in individuals with psychopathy and in patients with lesions involving the amygdala or orbitofrontal cortex. *Neuropsychology, 20*(3), 280-289. doi:https://doi.org/10.1037/0894-4105.20.3.280

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J. n., Andreiuolo, P. A., & Pessoa, L. (2002). The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions. *The Journal of Neuroscience, 22*(7), 2730. doi:https://doi.org/10.1523/JNEUROSCI.22-07-02730.2002

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. doi:https://doi.org/10.1073/pnas.0604475103

Montague, P. R., & Lohrenz, T. (2007). To detect and correct: norm violations and their enforcement. *Neuron, 56*(1), 14-18. doi:https://doi.org/10.1016/j.neuron.2007.09.020

Moretto, G., Làdavas, E., Mattioli, F., & Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience, 22*(8), 1888-1899. doi:https://doi.org/10.1162/jocn.2009.21367

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends in Cognitive Sciences, 11*(11), 489-497. doi:https://doi.org/10.1016/j.tics.2007.08.013

Murty, V., FeldmanHall, O., Hunter, L. E., Phelps, E. A., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. *Journal of Experimental Psychology: General, 145*(5), 548-558. doi:https://doi.org/10.1037/xge0000158

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*: Oxford University Press.
Nishi, A., Shirado, H., Rand, D. G., & Christakis, N. A. (2015). Inequality and visibility of wealth in experimental social networks. *Nature, 526*, 426. doi:https://doi.org/10.1038/nature15392
Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. *Science, 314*(5805), 1560. doi:https://doi.org/10.1126/science.1133755
Nowak, M. A., Sasaki, A., Taylor, C., & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature, 428*(6983), 646-650. doi:https://doi.org/10.1038/nature02414
Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*, 573. doi:https://doi.org/10.1038/31225
Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291. doi:https://doi.org/10.1038/nature04131
Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature, 466*(7310), 1057-1062. doi:https://doi.org/10.1038/nature09205
O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*, 95. doi:https://doi.org/10.1038/82959
Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences, 9*(5), 242-249. doi:https://doi.org/10.1016/j.tics.2005.03.010
Olson, M. (1965). *Logic of Collective Action: Public Goods and the Theory of Groups (Harvard economic studies. v. 124)*: Harvard University Press.
Ostrom, E. (2014). Collective action and the evolution of social norms. *Journal of Natural Resources Policy Research, 6*(4), 235-252. doi:https://doi.org/10.1257/jep.14.3.137
Ostrom, E., & Walker, J. (2003). *Trust and reciprocity: Interdisciplinary lessons for experimental research*: Russell Sage Foundation.
Peysakhovich, A., & Rand, D. G. (2015). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science, 62*(3), 631-647. doi:https://doi.org/10.1287/mnsc.2015.2168
Pfeiffer, T., Tran, L., Krumme, C., & Rand, D. G. (2012). The value of reputation. *Journal of the Royal Society Interface, 9*(76), 2791-2797. doi:https://doi.org/10.1098/rsif.2012.0332
Phan, K. L., Sripada, C. S., Angstadt, M., & McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences, 107*(29), 13099. doi:https://doi.org/10.1073/pnas.1008137107
Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron, 48*(2), 175-187. doi:https://doi.org/10.1016/j.neuron.2005.09.025
Phelps, E. A., Lempert, K. M., & Sokol-Hessner, P. (2014). Emotion and Decision Making: Multiple Modulatory Neural Circuits. *Annual Review of Neuroscience, 37*(1), 263-287. doi:https://doi.org/10.1146/annurev-neuro-071013-014119
Piketty, T. (2017). *Capital in the twenty-first century*: Harvard University Press.
Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*(3), 208-224. doi:https://doi.org/10.1006/obhd.1996.0100

Pitman, R. L., Deecke, V. B., Gabriele, C. M., Srinivasan, M., Black, N., Denkinger, J., . . . Neilson, J. L. (2017). Humpback whales interfering when mammal-eating killer whales attack other species: Mobbing behavior and interspecific altruism? *Marine Mammal Science, 33*(1), 7-58. doi:https://doi.org/10.1111/mms.12343

Pizarro, D. (2000). Nothing More than Feelings? The Role of Emotions in Moral Judgment. *Journal for the Theory of Social Behaviour, 30*(4), 355-375. doi:https://doi.org/10.1111/1468-5914.00135

Plato. (1955). *Plato: Collected Dialogues* (G. M. Grube & C. D. C. Reeve, Trans.). Princeton: Princeton University Press.

Preston, S. D. (2013). The origins of altruism in offspring care. *Psychological Bulletin, 139*(6), 1305. doi:https://doi.org/10.1037/a0031755

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*(1), 1-20. doi:https://doi.org/10.1017/S0140525X02000018

Quervel-Chaumette, M., Dale, R., Marshall-Pescini, S., & Range, F. (2015). Familiarity affects other-regarding preferences in pet dogs. *Scientific Reports, 5*, 18102. doi:https://doi.org/10.1038/srep18102

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review, 83*(5), 1281-1302.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences, 17*(8), 413-425. doi:https://doi.org/10.1016/j.tics.2013.06.003

Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., & Lilienfeld, S. O. (2007). Neural Correlates of Social Cooperation and Non-Cooperation as a Function of Psychopathy. *Biological Psychiatry, 61*(11), 1260-1271. doi:https://doi.org/10.1016/j.biopsych.2006.07.021

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. doi:https://doi.org/10.1016/S0896-6273(02)00755-9

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *The American Economic Review, 81*(5), 1068-1095.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*(4), 574. doi:http://dx.doi.org/10.1037/0022-3514.76.4.574

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience, 15*, 549. doi:https://doi.org/10.1038/nrn3776

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science, 342*(6157), 482-484. doi:https://doi.org/10.1126/science.1241399

Rumble Ann, C., Van Lange Paul, A. M., & Parks Craig, D. (2009). The benefits of empathy: When empathy may sustain cooperation in social dilemmas. *European Journal of Social Psychology, 40*(5), 856-866. doi:https://doi.org/10.1002/ejsp.659

Sally, D. (1995). Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Rationality and Society, 7*(1), 58-92. doi:https://doi.org/10.1177/1043463195007001004

Sanfey, A. G. (2009). Expectations and social decision-making: biasing effects of prior knowledge on Ultimatum responses. *Mind & Society*, *8*(1), 93-107. doi:https://doi.org/10.1007/s11299-009-0053-6
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science, 300*(5626), 1755.
Schein, C., & Gray, K. (2017). The theory of dyadic morality: reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 1088868317698288. doi:https://doi.org/10.1177/1088868317698288
Schmidt, M. F. H., Butler, L. P., Heinz, J., & Tomasello, M. (2016). Young Children See a Single Action and Infer a Social Norm: Promiscuous Normativity in 3-Year-Olds. *Psychological Science, 27*(10), 1360-1370. doi:https://doi.org/10.1177/0956797616661182
Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science, 275*(5306), 1593. doi:https://doi.org/10.1126/science.275.5306.1593
Schwartz, S. H. (1977). Normative influences on altruism. *Advances in Experimental Social Psychology, 10*, 221-279.
Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex, 46*(5), 668-677. doi:https://doi.org/10.1016/j.cortex.2009.04.008
Sherif, M. (1936). The psychology of social norms.
Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., . . . Rauch, S. L. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological Psychiatry, 48*(1), 43-50. doi:https://doi.org/10.1016/S0006-3223(00)00251-1
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. doi:https://doi.org/10.1177/019027250807100106
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*, 466. doi:https://doi.org/10.1038/nature04271
Smetana, J. G. (1983). Social-cognitive development: Domain distinctions and coordinations. *Developmental Review, 3*(2), 131-147. doi:https://doi.org/10.1016/0273-2297(83)90027-8
Sommerville, J. A., & Enright, E. A. (2018). The origins of infants' fairness concerns and links to prosocial behavior. *Current Opinion in Psychology, 20*, 117-121. doi:https://doi.org/10.1016/j.copsyc.2018.01.005
Sripada, C. S. (2008). Nativism and moral psychology: Three models of the innate structure that shapes the contents of moral norms *Moral Psychology* (Vol. 1, pp. 319-343): MIT Press Cambridge.
Srivastava, J., Espinoza, F., & Fedorikhin, A. (2009). Coupling and decoupling of unfairness and anger in ultimatum bargaining. *Journal of Behavioral Decision Making, 22*(5), 475-489. doi:https://doi.org/10.1002/bdm.631
Stearns, D. C., & Parrott, W. G. (2012). When feeling bad makes you look good: Guilt, shame, and person perception. *Cognition & Emotion, 26*(3), 407-430. doi:https://doi.org/10.1080/02699931.2012.675879

Straub, P. G., & Murnighan, J. K. (1995). An experimental investigation of ultimatum games: Information, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior & Organization, 27*(3), 345-364. doi:https://doi.org/10.1016/0167-2681(94)00072-M

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. doi:https://doi.org/10.1111/j.1467-9280.2008.02091.x

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *NeuroImage, 23*(3), 967-974. doi:https://doi.org/10.1016/j.neuroimage.2004.07.054

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology, 70*(6), 1256. doi:http://dx.doi.org/10.1037/0022-3514.70.6.1256

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual Review of Psychology, 58*(1), 345-372. doi:https://doi.org/10.1146/annurev.psych.56.091103.070145

Tavoni, A., Dannenberg, A., Kallis, G., & Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences, 108*(29), 11825. doi:https://doi.org/10.1073/pnas.1102493108

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. Oxford, England: John Wiley.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*(1), 35-57.

Tsui, A., & Wang, D. (2002). Employment Relationships from the Employer's Perspective: Current Research and Future Directions *International Review of Industrial and Organizational Psychology 2002* (pp. 77-114): John Wiley & Sons Ltd.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*: Cambridge University Press.

Ullmann-Margalit, E. (1978). *The emergence of norms*. New York, NY: Oxford University Press.

Vaish, A. (2018). The prosocial functions of early social emotions: the case of guilt. *Current Opinion in Psychology, 20*, 25-29. doi:https://doi.org/10.1016/j.copsyc.2017.08.008

Van de Vondervoort, J. W., & Hamlin, J. K. (2018). The early emergence of sociomoral evaluation: infants prefer prosocial others. *Current Opinion in Psychology, 20*, 77-81. doi:https://doi.org/10.1016/j.copsyc.2017.08.014

Van Den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., & Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the Trust Game. *Social Cognitive and Affective Neuroscience, 4*(3), 294-304. doi:https://doi.org/10.1093/scan/nsp009

Van Lange, P. A. M., Liebrand, W. B. G., Messick, D. M., & Wilke, H. A. M. (1992). Social dilemmas: The state of the art *Social dilemmas: Theoretical issues and research findings* (pp. 3-28).

Van Vugt, M., & Samuelson, C. D. (1999). The impact of personal metering in the management of a natural resource crisis: A social dilemma analysis. *Personality and Social Psychology Bulletin, 25*(6), 735-750. doi:http://dx.doi.org/10.1177/0146167299025006008

Van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Experimental Brain Research, 169*(4), 564-568. doi:https://doi.org/10.1007/s00221-006-0346-5

Vekaria, K. M., Brethel-Haurwitz, K. M., Cardinale, E. M., Stoycos, S. A., & Marsh, A. A. (2017). Social discounting and distance perceptions in costly altruism. *Nature Human Behaviour, 1*, 0100. doi:https://doi.org/10.1038/s41562-017-0100

Von Neumann, J., & Morgenstern, O. (1945). *Theory of games and economic behavior*: Princeton University Press Princeton, NJ.

Wang, G., Li, J., Yin, X., Li, S., & Wei, M. (2016). Modulating activity in the orbitofrontal cortex changes trustees' cooperation: A transcranial direct current stimulation study. *Behavioural Brain Research, 303*, 71-75. doi:https://doi.org/10.1016/j.bbr.2016.01.047

Warneken, F., Hare, B., Melis, A. P., Hanus, D., & Tomasello, M. (2007). Spontaneous altruism by chimpanzees and young children. *PLoS Biology, 5*(7), e184. doi:https://doi.org/10.1371/journal.pbio.0050184

Watanabe, T., Takezawa, M., Nakawake, Y., Kunimatsu, A., Yamasue, H., Nakamura, M., . . . Masuda, N. (2014). Two distinct neural mechanisms underlying indirect reciprocity. *Proceedings of the National Academy of Sciences, 111*(11), 3990. doi:https://doi.org/10.1073/pnas.1318570111

Wilensky, H. L. (1974). *The welfare state and equality: Structural and ideological roots of public expenditures* (Vol. 140): Univ of California Press.

Wilson, E. O. (2000). *Sociobiology*: Harvard University Press.

Xiang, T., Lohrenz, T., & Montague, P. R. (2013). Computational Substrates of Norms and Their Violations during Social Exchange. *The Journal of Neuroscience, 33*(3), 1099. doi:https://doi.org/10.1523/JNEUROSCI.1642-12.2013

Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly*, 265-271. doi:https://doi.org/10.2307/2786924

Yamagishi, T., Takagishi, H., Fermin, A. d. S. R., Kanai, R., Li, Y., & Matsumoto, Y. (2016). Cortical thickness of the dorsolateral prefrontal cortex predicts strategic choices in economic games. *Proceedings of the National Academy of Sciences, 113*(20), 5582-5587. doi:https://doi.org/10.1073/pnas.1523940113

Young, L., Koenigs, M., Kruepke, M., & Newman, J. P. (2012). Psychopathy increases perceived moral permissibility of accidents. *Journal of Abnormal Psychology, 121*(3), 659. doi:https://doi.org/10.1037/a0027489

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin, 140*(6), 1608-1647. doi:https://doi.org/10.1037/a0037679

Zaki, J., & Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences, 108*(49), 19761. doi:https://doi.org/10.1073/pnas.1112324108

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics, 6*(3), 299-310. doi:https://doi.org/10.1023/A:1026277420119

Zisis, I., Di Guida, S., Han, T. A., Kirchsteiger, G., & Lenaerts, T. (2015). Generosity motivated by acceptance - evolutionary analysis of an anticipation game. *Scientific Reports, 5*, 18076. doi:https://doi.org/10.1038/srep18076