# Differential Roles of Fairness- and Compassion-Based Motivations for Cooperation, Defection, and Punishment

**Tania Singer and Nikolaus Steinbeis**

*Laboratory for Social and Neural Systems Research, University of Zürich, Zürich, Switzerland*

**The present paper briefly describes and contrasts two different motivations crucially involved in decision making and cooperation, namely fairness-based and compassion-based motivation. Whereas both can lead to cooperation in comparable social situations, we suggest that they are driven by fundamentally different mechanisms and, overall, predict different behavioral outcomes. First, we provide a brief definition of each and discuss the relevant behavioral and neuroscientific literature with regards to cooperation in the context of economic games. We suggest that, whereas both fairness- and compassion-based motivation can support cooperation, fairness-based motivation leads to punishment in cases of norm violation, while compassion-based motivation can, in cases of defection, counteract a desire for revenge and buffer the decline into iterative noncooperation. However, those with compassion-based motivation alone may get exploited. Finally, we argue that the affective states underlying fairness-based and compassion-based motivation are fundamentally different, the former driven by anger or fear of being punished and the latter by a wish for the other person's well-being.**

*Key words:* **fairness; compassion; cooperation**

## Introduction

In the present paper we will explore the possible similarities and differences of two motivational states—fairness-based and compassion-based motivation—which have mostly been discussed separately and by different fields. Whereas the concept of fairness has been the focus of multiple investigations in the field of experimental economics (for reviews, see Refs. 1–3), empirical research on empathy and compassion has traditionally been embedded in the field of psychology (e.g., see Refs. 4–8) and recently also social neurosciences (for overviews, see Refs. 9–14). Even though they are not frequently linked, both concepts refer to humans as altruistic beings who care about others' welfare as well as their own. Furthermore, recent research in the field of social neuroscience and neuroeconomics suggests that our sense of fairness and our ability to empathize even share common underlying brain circuitries.[15] In the present paper, we propose, however, that even though fairness-based and compassion-based motivation may share superficial features, they differ in fundamental aspects both with respect to their underlying motivation and to their predicted behavioral consequences.

The world is rife with everyday examples, both in one's private and professional life, of situations in which one encounters unfair behavior and has to decide on a course of action. Imagine someone falsely accuses you of a major or minor transgression. Given the unfounded nature of the claim, a desire for revenge may seem natural, especially if malicious intentions may have been the driving motivational force behind the accusations. Fairness-based

Address for correspondence: Tania Singer, Laboratory for Social and Neural Systems Research, University of Zürich, Blümlisalpstr. 10, CH-8006 Zürich, Switzerland. Voice: +41 44 634 52 20. singer@iew.uzh.ch

Values, Empathy, and Fairness across Social Barriers: Ann. N.Y. Acad. Sci. 1167: 41–50 (2009).
doi: 10.1111/j.1749-6632.2009.04733.x 

attitudes would lead you to engage in retaliatory actions (e.g., counteraccusation, defamation), which would most likely result in a bout of mud-slinging with negative consequences for everyone involved. Alternatively, a compassionate attitude may lead you to try to understand what the possible cause for the driving motivation may have been and in doing so prevent you from taking drastic punitive actions. In fact, you may rather try to seek dialogue with the goal of understanding the other person and reaching an agreement.

In the following sections, we will outline our use of the two crucial concepts and then move to a short overview of the relevant literature on fairness and cooperation from behavioral economics and discuss neuroscientific evidence relevant to punishment. This will be followed by a brief summary of psychological and neuroscientific studies focusing on the effect of empathy and compassion on cooperation and their underlying neural correlates. Finally, we will discuss the similarities and differences of both motivations. Rather than offering an exhaustive literature review on compassion and fairness, we will make a first attempt to combine two concepts that have traditionally been investigated by separate research traditions.

## Definition of Concepts

The Collins English Dictionary equates fair behavior with that which is reasonable, right, and just. Fairness seems to express individual conceptions of justice, which are not set down by law. Violations of justice, which would be considered unfair, can range widely in behavior from slandering to short-changing others. The use of the concept of fairness in the present paper is tightly linked to the one outlined in economic theories, which have formalized notions of fairness and mostly refer to fairness in terms of a just distribution of money or other goods. Within economics, definitions of fairness have to be seen in light of theories of social preferences. Fairness preferences fall into two broad categories: outcome- and intention-based fairness preferences. Outcome-based fairness preferences refer to the strong concern shown for the allocation of resources and outcomes of decisions within one's reference group (e.g., comparing what I get to what you get). In the domain of outcome-based theories, two major motives are discussed, one being equity as a standard of fairness[16] and the other being help offered to the poorest members of the group.[17] Intention-based fairness preferences, on the other hand, refer to the importance of not only considering the final outcome but also others' intentions when goods are being allocated (e.g., the difference between receiving an offer from someone who was forced to make the offer versus someone who has chosen to make the offer).[18,19] Particularly the latter class of theories on fairness provide insight into the role of psychological states in economic decisions. As will be shown below, there is some evidence showing that it matters to people whether someone defected accidentally or intentionally.[20]

Whereas fairness research has been especially prevalent in economics research (but see Refs.13, 20–24), concepts such as empathy, sympathy, or compassion have mostly been reserved for empirical investigations within the fields of psychology and social neuroscience.

Empathy and compassion refer generally to broad, folk, psychological concepts for which there are no single generally accepted definitions available. Therefore, it is essential to briefly review how we define the relevant key concepts in the present paper (see also Refs. 8, 10, 25–28). Very broadly, these concepts denote an affective response to the directly perceived, imagined, or inferred feeling state of another being. More specifically, a distinction can be made between emotional contagion, empathy, sympathy, and compassion, although these different phenomena frequently occur in concert. The most "primitive" state, which is also present in other species and already observed in newborns and young infants, is emotional contagion (primitive empathy[29]). When emotional

contagion arises, a person is affected by another person's affective state but is not even necessarily aware of it (e.g., contagious yawning). Unlike emotional contagion, empathy involves a distinction between oneself and others and an awareness that one is vicariously feeling with someone but that this is not one's own emotion ("I share your pain but I know that it is not my own pain"; for more details, see Ref. 10). Sympathy differs from empathy in that one feels *for* and not *with* someone; that is, the emotion one feels is different from that which the other feels. If you feel pity for someone who is angry, you do not share that person's anger but nevertheless you feel for them. Even though this is not necessarily explicit in the concept itself, sympathy has a condescending quality to it, where the sympathetic other feels that he/she is *above* the one feeling the emotion. Empathic concern, or compassion, usually refers to a feeling *and* a motivation to help the other. Thus, compassion has been defined as a "deep awareness of the suffering of another coupled with the wish to relieve it" (American Heritage Dictionary) or as the "human quality of understanding the suffering of others and wanting to do something about it" (Merriam–Webster). Contemplative traditions typically refer to "loving-kindness" as the wish for happiness for others and of "compassion" as the wish to relieve others' suffering. In the present paper, we use "compassion" both in terms of an emotional as well as a motivational state, whereby the former can be characterized by feelings of warmth, love, and concern for the welfare of others and the latter by the desire to help and promote others' welfare.

Generally, it is assumed that emotional contagion is antecedent to empathy, which in turn precedes sympathy and compassion, which in turn may be followed by prosocial behavior (e.g., see Ref. 14). Empathy is thought to be a necessary but not sufficient condition for compassion to arise because too much empathy may also result in personal distress and avoidance of the suffering other. The lack of empathy as observed, for example, in psychopaths is associated with a lack of compassionate motivation and the occurrence of antisocial behavior[30]).

In the next section, we will explore the relation between fairness-based and compassion-based motivation for cooperation. It is obvious that both an understanding of what is fair and the desire for the other's well-being will most probably lead to reciprocal cooperation in social exchange, also known as *positive reciprocity* (motivation to reciprocate kind acts with kindness). However, based on empirical evidence mostly derived from economics, we suggest that radically different predictions can be derived from the two motivations with regard to what happens once cooperation breaks down and its effect on what is called *negative reciprocity* (the motivation to reciprocate hostile acts with hostility).

## Fairness and Cooperation from the View of Experimental Economics

Human cooperation between nonkin has been an evolutionary puzzle,[3,31] particularly when interactions are not repeated and the ability to form a reputation is limited (i.e., why help a stranger in a large city?). The use of game theory has been able to shed light on how such instances of cooperation may arise, and, by means of simulations and experimental tests, it has become apparent that the use and threat of punishment is a key variable in bringing this about.

### Economic Games as Measures of Cooperation

Two games are of particular importance in understanding the occurrence of cooperation: public goods (PG) and prisoner's dilemma (PD) games. In PG games, all members of a group can share a public good to which each member can contribute. In typical PG experiments, groups of more than two individuals are formed and each individual is given an endowment.

Group members decide simultaneously and for themselves how much of this endowment they would like to keep and how much they would like to contribute to the common good. After all contributions are made, the net amount of contributions by all group members is multiplied by a factor greater than one but smaller than the number of members in the group. This multiplied sum is then redistributed equally among the group members, which means that each member earns, apart from what he or she did not contribute, the multiplied sum divided by the number of members in the group. Seeing that the number of members is greater than the factor of endowment multiplication, the return for each invested monetary unit is less than one. As such, in a one-shot experiment, selfish group members would not be expected to contribute to the public good because their return would be lower than what they invested. However, if everyone was to contribute everything, the sum of contributions would yield an outcome larger than the individual endowment of group members (e.g., if the endowment is 20, group size four, and the factor of multiplication 1.6, then subjects earn 20 if nobody contributes and 32 if everyone contributes everything).

The PD game is essentially a variant of the PG game, the difference being that the interaction consists of only two individuals and two possible actions: contribute everything (cooperate) or nothing (defect). The dilemma arises out of the payoffs, which are realized by the different behavioral combinations of the two individuals. The highest payoff occurs for Player A when he defects and Player B cooperates; mutual cooperation is rewarded a little less; mutual defection results in a low payoff for both players; and the lowest payoff (viz. nothing) occurs for Player A when he cooperates and Player B defects (suckers payoff). In a one-shot interaction, defection is the best strategy regardless of what the other player does, but this strategy is less efficient in iterated interactions because rewards for mutual cooperation are higher that those for mutual defection. Simulations have shown that an unrewarding strategy known as tit-for-tat (TFT; start with cooperation and then copy the other player's previous move) is one of the most effective behavioral strategies for cooperation to evolve in a system in which future interactions occur with a high probability.[31]

Thus, in both dilemmas, cooperation is a distinct and desirable possibility, but selfish interests can override this and lead to a breakdown in cooperation. How people behave in these kinds of situations has been extensively studied. Fischbacher, Gächter, and Fehr[32] investigated subjects' willingness to contribute to a public good, depending on the average contribution of the other group members. They found that over 50% of the participants only contributed under the condition that other members also did so. Free riders made up 30% of the sample. The combination of players who cooperate only if others also do so and a considerable number of free riders would result in cooperation breaking down over time. In fact, there is empirical evidence that cooperation decreases considerably over time, which is very likely the result of conditional cooperators responding negatively to free riders.[33] Other studies have shown that stable cooperation is hardly ever attained in repeated PG experiments where selfish choice is full defection.[34]

However, recent experiments have shown that introducing the possibility of punishing others into PG games drastically increases the rate at which group members cooperate. In games, punishment usually takes the form of a third party observing an interaction between two other players (such as a PD game) and subsequently being able to sanction one, both, or none of the players. Punishment comes at a cost to the punisher, and each time a decision to punish is made it has to be paid for with part of the punisher's endowment, varying payment size with the degree of punishment to be inflicted. Empirical evidence has shown that over 50% of third parties are willing to punish defection in the PD game, particularly when one of the two players cooperated.[35] Thus, punishment is something that occurs irrespective

of any direct material benefit and seems to serve the establishment and enforcement of social norms. However, it can also have direct consequences for one's personal material advantage.

In a study by Fehr and Gächter,[36] groups of four played one-shot rounds of a PG game. Under threat of punishment, 94.2% of the subjects invested more than when others could not punish. Seeing that each round was played with a different combination of group members, this cannot be explained by the deterring effects of future interactions with the same group. When questioned indirectly about the motivation of such punishment, subjects replied that negative emotions, such as anger and annoyance, toward the free rider primarily drove the punishment. The more the free riders' contribution deviated from the average, the angrier the punisher was and in turn the greater the punishment meted out was (for more data on the role of emotions in enforcing normative behavior, see Ref. 37). Likewise, subjects' responses when asked about the expectation of anger and annoyance from others when they have been free riders suggest that fear of these emotions in others was a primary motivation for not free riding in the punishment conditions. Thus, it seems that to maintain stable cooperation in a society, the possibility for punishment has to be available constantly, which in turn means that the permanent threat of such punishment as a result of even accidental transgressions against social norms is highly prevalent.[38–42] Polemically, one could argue that stable cooperation resulting from fairness norms comes at the cost of persistent anger over others' violation of fairness norms and the fear of being punished for one's own violations. On the other hand, recent evidence from the field of neuroeconomics suggests that punishment is experienced as rewarding, which possibly explains why people are willing to endure the cost of punishment. Before summarizing evidence for the role of compassion in cooperation, we will therefore briefly review this new stream of research.

## The Neuronal Basis of Punishment

In a study on third-party punishment, de Quervain and colleagues[43] investigated the neural correlates of decisions on punishment when male subjects observed the abuse of trust between two individuals. They found that the decision to punish a defecting individual activates the dorsal striatum and that the magnitude of punishment correlated with the strength of this activation. Based on the fact that the striatum is frequently involved in the processing of rewards,[44,45] these findings were taken to support the idea that the subjects found it pleasurable to punish those who defected. Further empirical support for this comes from a more recent study[24] in which subjects watched fair and unfair interactions between two others who then received brief electric shocks. In male subjects who were observing this, there was increased activation in the nucleus accumbens when defectors were punished with shocks as compared to when cooperators were punished. Likewise, the strength of this activation correlated with the subjects' individual desire for revenge. Thus there seems to be evidence that punishment of unfair individuals is pleasurable, regardless of whether this is initiated by third parties or not.

Other evidence concerning the neural correlates of punishment has been generated using two-party games, such as the ultimatum game, in which Player A makes Player B an offer about how his endowment can be split. Based on this offer, Player B can either accept or reject the offer. If Player B rejects the offer, the endowment disappears and neither player receives anything, which effectively constitutes an act of punishment on the part of Player B. In a functional (f)MRI study,[47] it was shown that unfair offers elicited greater activation in the left and right anterior insula (AI) as well as the right dorsolateral prefrontal cortex (DLPFC). In a subsequent transcranial magnetic stimulation (TMS) study, Knoch and colleagues[48] showed that a disruption of the right DLPFC led to a greater acceptance of unfair offers even

though subjects were perfectly aware that these offers were unfair. A more recent study has shown that the right DLPFC appears to be critically involved in determining the culpability of other parties depending on their responsibility for certain crimes[49] but not necessarily in determining the appropriate punishment for the crimes.

In sum, recent neuroscientific evidence suggests that punishment of unfair people is associated with brain activation related to reward processing, which may explain why humans are motivated and willing to endure costs and risks to punish unfair people even if they will never see these people again. Moreover, these findings suggest that the implementation of social norm-based behavior, such as rejecting unfair offers and therefore punishing the other or deciding on whether to punish others for crimes perpetrated, requires active cognitive control, a function subserved by the right DLPFC.

## Empathy, Compassion, Generosity, and Cooperation from the View of Psychology

Even though we would argue that compassion, being a prosocial motivation toward the well-being of others, has an influence on prosocial behavior and cooperation, so far the empirical evidence stems mostly from studies focusing on the effect of empathy or generosity on cooperation.

Defection in real-life situations can occur for a variety of reasons, such as deliberate intentions or just by accident from unforeseeable circumstances. The latter occurs particularly frequently and is considered negative noise. This constitutes the discrepancy between an intended outcome (e.g., intending to meet a friend at a prearranged meeting point) and an actual outcome (e.g., not being able to meet the friend because one's car broke down). A TFT strategy would involve responding to defection with defection, regardless of the reasons behind the initial defection. This in turn would lead to a decline in cooperation as a result of mutual defection.[50] As negative noise can be a frequent cause for defection, Van Lange and colleagues[20] conducted an empirical study of the effects of generosity on defection caused by negative noise. Subjects were observed while ostensibly interacting with another subject (but actually a computerized partner). The computerized partner was programmed to adopt one of two strategies throughout the entire game: a TFT strategy or a generous TFT (GTFT) strategy. A GTFT strategy entails always responding to the partner's previous move with the addition of an extra monetary unit (MU). Thus, following the TFT strategy, one would give the partner only two out of a possible 10 MU if one's partner had done so in the previous round; following the GTFT strategy, one would give two plus one out of a possible 10 MU. Also, to introduce noise into the interaction, defection occurred on every sixth trial. They found that, when the computer partner adopted the TFT strategy as compared to the GTFT strategy, participants defected more often after noise. So, adopting a GTFT strategy can prevent the potential decline in cooperation which can ensue from defections caused for whatever reason. This behavioral evidence indicates that a GTFT—as compared to a TFT—strategy ultimately possesses greater promise with respect to the establishment of long-term cooperation, which in turn has implications for the evolution of altruism.[51] One can think of countless contexts in which declines in cooperation take place, from small-scale neighborhood relations to larger scale regional relations resulting in conflict and war (e.g., the conflicts between Indians and Pakistanis following the partitioning of India). It remains open whether a generous tendency would prevail if one knew that the other had intentionally defected.

Rumble, Van Lange, and Parks[23] extended their previous paradigm to study the effects of empathy on cooperation in noisy environments. They used a social dilemma task

adapted from Van Lange and colleagues[20] in which participants played several rounds of the game with the same person. They employed Batson's[21,22] empathy induction (described below) with high-empathy, low-empathy, and no-empathy conditions. They found that the negative effects of noise could only be counteracted in the high-empathy condition. In addition, levels of cooperation were significantly higher in the high-empathy than in the low- or no-empathy conditions. However, when subjects believed that their partner's previous behavior was fully intentional, they were equally less cooperative in all three conditions.

These results confirm findings from previous studies by Batson and colleagues that have studied the effect of an empathy induction procedure on cooperation using a one-shot PD game as a proxy for prosocial behavior.[21,22] Following this procedure, participants were to read the story of a woman, ostensibly their partner in the subsequent game, which revealed her need to be cheered up as a result of her just having broken up with her boyfriend and not being particularly happy at the moment. In one condition, participants were to read the story and imagine how the woman felt (high-empathy condition); in a second condition, participants were to assume an objective perspective (low-empathy condition); in a third condition, no information about the woman was given (no-empathy condition). They found that significantly more subjects chose to cooperate in the high-empathy condition (80%) than in the low-empathy (50%) or no-empathy (40%) conditions.[21] More interestingly, they found that, even when participants knew their partner had defected during the PD game, significantly more participants cooperated in the high-empathy condition (45%) than in the low-empathy (10%) or no-empathy (0%) conditions.[22] In addition, scores on an index of self-reported empathy were significantly positively correlated with cooperation. Particularly the findings from the latter study highlight the potential of prosocial motivation toward others to counteract one's own tendency to defect and thus likely prevent further defection by the other, which would result in the decline in cooperation described above.

In sum, these results suggest that a generous and forgiving behavioral tendency or the induction of an empathic motivational state before engaging in social exchange with other individuals can prevent the sort of decline in cooperation that would typically arise from accidental defection. They also suggest that a preference that takes into account the other person (by wishing him well) as opposed to the norm (one ought to cooperate) may produce behavior, which in the long run is more beneficial for cooperation, especially in noisy and uncertain environments.

## The Neural Basis of Empathy, Compassion, and Cooperation

In the emerging field of social neuroscience, recent attempts to study the neural correlates of empathy and its modulation have been encouraging (for overviews, see Refs. 9–12,52,53). Thus, it has repeatedly been shown that a common neural network of the anterior cingulate cortex (ACC) and the AI is activated when participants experience pain themselves and when they observe others in pain.[13] In addition, the extent of this activation can be modulated by contextual factors, such as whether the other person behaved fairly or unfairly in a previously played economic trust game like those explained above.[24] Thus, when participants observed fair players suffering pain (electrical shocks), they showed an empathic response (activation in ACC and AI). However, at least when male participants observed an unfair player suffering pain, their empathic response was overridden by feelings of "schadenfreude" and revenge. An interesting future experiment could investigate whether this modulation of empathic brain response by the observation of fairness/unfairness could be affected by a prior compassion induction procedure.

As much as empathy research has advanced in the field of neuroscience in the last few years, the investigation of compassion is still in its infancy. A study by Lutz and colleagues[54] looked at the effects of generating a specific state of compassion (through loving-kindness meditation) in novice and expert meditation practitioners when listening to affective and neutral vocalizations. The generation of this particular state was induced by instructing subjects to think about someone they care about and with respect to whom they were open to feelings of altruistic love (wishing well-being) or compassion (wishing freedom from suffering).

The most relevant findings included increased activity of the insula and ACC when listening to affective sounds during periods of meditation (versus rest) as well as an increase in insula activation in response to negative emotional sounds (e.g., distress) compared to positive or neutral sounds for expert as compared to novice practitioners. These data provide further support for the role of ACC and insula in empathic as well as compassionate states and speak to the fact that activity in these brain areas can be enhanced by the practice of generating loving-kindness and compassion for fellow human beings.

## Summary and Conclusion

In the present paper, we have reviewed empirical evidence suggesting that fairness norms may support cooperation but that fairness-based cooperation can also easily break down if social norms are violated. Thus, people who are initially cooperative start defecting when confronted with free riders in order to avoid getting exploited. After cooperation breaks down, there is only a small chance of restoring it. This slippery slope of defection can be prevented by introducing punishment options. Altruistic punishment has been shown to be an efficient way of enforcing social norms and sustaining cooperation in large-scale societies. However, the emotional states underlying fairness- and punishment-based cooperation seem to be predominantly negative as it was shown that violation of fairness is mostly associated with anger and a desire to retaliate and norm reinforcement through punishment is based on a fear of negative outcomes. In contrast to fairness-based motivation, compassion-based motivation seems to buffer the effect of defection in a noisy environment and prevent cooperation from breaking down when participants are confronted with others' defections. Furthermore, compassion is mostly associated with positive feelings of love and concern for the other and its cultivation has been found to have beneficial effects on health, such as improved immune system and stress responses.[55] In contrast, negative feelings, such as anger, have been shown to have significant negative effects on the immune system with long-term consequences.[56,57] It may seem trivial at this point to make such a crass distinction between the polar opposites of the positive and negative emotional states associated with fairness and compassion. Further empirical work is required to provide solid support for the complex interplay between these motivational and emotional states and their long-term effects. Nonetheless, we predict, on the one hand, from a purely health-based perspective, that adopting a compassion-based rather than a fairness-based stance on life is the better option as it should increase one's well-being. On the other hand, the above-mentioned data on cooperation in public good games illustrate quite clearly the downside of purely compassion-based motivation. Whereas the adoption of a compassion-based stance can buffer the downward spiral of defection in noisy environments over short-term interactions, it would make one vulnerable to exploitation and thus be highly maladaptive if interactions with persistent defectors continued over the long term.

In sum, we agree with the Dalai Lama when he writes: "Love and compassion are necessities, not luxuries. Without them humanity cannot survive." However, it seems important to know when to switch from a compassion to a

fairness stance in order to avoid exploitation and to act in the name of justice.

### Acknowledgments

We would like to thank Ernst Fehr for helpful comments on a previous version of the manuscript.

### Conflicts of Interest

The authors declare no conflicts of interest.

## References

1. Camerer, C.F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press. Princeton, NJ.
2. Kagel, J. & A.E. Roth. 1995. *Handbook of Experimental Economics*. Princeton University Press. Princeton, NJ.
3. Fehr, E. & U. Fischbacher. 2004. Social norms and human cooperation. *Trends in Cognitive Sciences* **8:** 185–190.
4. Batson, C.D. 1991. *The Altruism Question: Toward a Social-Psychological Answer*. Erlbaum. Hillsdale, NJ.
5. Batson, C.D. 2009. *These Things Called Empathy. In Social Neuroscience of Empathy*. J. Decety & W. Ickes, Eds.: MIT. Cambridge, MA.
6. Eisenberg, N. & J. Strayer. 1987. Critical issues in the study of empathy. In *Empathy and its Development*. N. Eisenberg & J. Strayer, Eds.: 3–13. Cambridge University Press. New York.
7. Eisenberg, N. 2000. Emotion, regulation, and moral development. *Annual Review in Psychology* **51:** 665–697.
8. Hoffman, M.L. 2000. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge University Press. Cambridge, MA.
9. Decety, J. & P.L. Jackson. 2004. The functional architecture of human empathy. *Behavioural and Cognitive Neuroscience Reviews* **3:** 71–100.
10. De Vignemont, F. & T. Singer. 2006. The empathic brain: How, when and why? *Trends in Cognitive Sciences* **10:** 435–441.
11. Gallese, V. 2003. The manifold nature of interpersonal relations: The quest for a common mechanism Philosophical transactions of the royal society of London, series B. *Biological Sciences* **358:** 517–528.
12. Keysers, C. & V. Gazzola. 2006. Towards a unifying neural theory of social cognition. *Progress in Brain Research* **156:** 379–403.
13. Singer, T., S.J. Kiebel, J.S. Winston, R.J. Dolan & C.D. Frith. 2004. Brain responses to the acquired moral status of faces. *Neuron* **41:** 653–662.
14. Singer, T. & C. Lamm. 2009. The social neuroscience of empathy. *Annals of the New York Academy of Sciences*: in press.
15. Singer, T. 2007. The neuronal basis of empathy and fairness. In *Empathy and Fairness*. G. Bock & J. Goode, Eds.: 20–30. Wiley. Chichester, UK.
16. Fehr, E. & K.M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* **114:** 817–868.
17. Charness, G. & M. Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* **117:** 817–869.
18. Rabin, M. 1993. Incorporating fairness into game theory and economics. *American Economic Review* **83:** 1281–1302.
19. Falk, A. & U. Fischbacher. 2006. A theory of reciprocity. *Games and Economic Behavior* **54:** 293–315.
20. Van Lange, P.A., J.W. Ouwerkerk & M.J. Tazelaar. 2002. How to overcome the detrimental effects of noise in social interaction: The benefits of generosity. *J Pers Soc Psychol* **82:** 768–780.
21. Batson, C.D. & T. Moran. 1999. Empathy-induced altruism in a prisoner's dilemma European. *Journal of Social Psychology* **29:** 909–924.
22. Batson, C.D. & N. Ahmad. 2001. Empathy-induced altruism in a prisoner's dilemma II: What if the target of empathy has defected? *European Journal of Social Psychology* **31:** 25–36.
23. Rumble, A.C., P.A.M. van Lange & C.D. Parks. 2009. The benefits of empathy: When empathy may sustain cooperation in social dilemmas. *European Journal of Social Psychology*: In press.
24. Singer, T., B. Seymour, J.P. O'doherty, K.E. Stephan, R.J. Dolan & C.D. Frith. 2006. Empathic neural responses are modulated by the perceived fairness of others. *Nature* **439:** 466–469.
25. Batson, C.D. 2009. These things called empathy. In *Social Neuroscience of Empathy*. Decety, J. & W. Ickes, Eds.: MIT. Cambridge, MA.
26. Eisenberg, N. & R.A. Fabes. 1990. Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion* **14:** 131–149.
27. Preston, S.D. & F.B.M. de Waal. 2002. Empathy: Its ultimate and proximate bases. *Behavioral and Brain Science* **25:** 1–72.
28. Wispé, L. 1986. The distinction between sympathy and empathy: To call forth a concept, a word is needed. *J. Pers. Soc. Psychol.* **50:** 314–321.
29. Hatfield, E., R.L. Rapson & Y.L. Le. 2009. Emotional contagion and empathy. In *The Social Neuroscience of Empathy*. J. Decety & W. Ickes, Eds.: MIT. Cambridge, MA.
30. Blair, J., D.R. Mitchell & K. Blair. 2005. *The Psychopath: Emotion and the Brain*. Blackwell. Oxford, UK.

31. Axelrod, R. & W.D. Hamilton. 1981. The evolution of cooperation. *Science* **211:** 1390–1396.
32. Fischbacher, U., S. Gachter & E. Fehr. 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters* **71:** 397–404.
33. Fehr, E. & S. Gachter. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* **90:** 980–994.
34. Dawes, R.M. & R.H. Thaler. 1988. Cooperation. *Journal of Economic Perspectives* **2:** 187–197.
35. Fehr, E. & U. Fischbacher. 2004. Third-party punishment and social norms. *Evolution and Human Behavior* **25:** 63–87.
36. Fehr, E. & S. Gachter. 2002. Altruistic punishment in humans. *Nature* **415:** 137–140.
37. Hopfensitz, A. & E. Reuben. 2009. The importance of emotions for the effectiveness of social punishment. *The Economic Journal*: In press.
38. Anderson, C.M. & L. Putterman. 2006. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior* **54:** 1–24.
39. Carlsmith, K.M., J.M. Darley & P.H. Robinson. 2002. Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of personality and social psychology* **83:** 284–299.
40. Carpenter, J.P. 2007. The demand for punishment. *Journal of Economic Behavior & Organization* **62:** 522–542.
41. Fowler, J.H., T. Johnson & O. Smirnov. 2005. Egalitarian motive and altruistic punishment. *Nature* **433:**.
42. Kocher, M.G., P. Martinsson & M. Visser. 2008. Does stake size matter for cooperation and punishment? *Economics Letters* **99:** 508–511.
43. de Quervain, D.J.F., U. Fischbacher, V. Treyer, M. Schelthammer, U. Schnyder, A. Buck & E. Fehr. 2004. The neural basis of altruistic punishment. *Science* **305:** 1254–1258.
44. Schultz, W. 2000. Multiple reward signals in the brain Nature reviews. *Neuroscience* **1:** 199–207.
45. O'Doherty, J.P. 2004. Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Curr Opin Neurobiol* **14:** 769–776.
46. Singer, T. 2006. The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews* **30:** 855–863.
47. Sanfey, A.G., J.K. Rilling, J.A. Aronson, L.E. Nystrom & J.D. Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* **300:** 1755–1758.
48. Knoch, D., A. Pascual-Leone, K. Meyer, V. Treyer & E. Fehr. 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* **314:** 829–832.
49. Buckholtz, J.W., C.L. Asplund, P.E. Dux, D.H. Zald, J.C. Gore, O.D. Jones & R. Marois. 2008. The neural correlates of third-party punishment. *Neuron* **60:** 930–940.
50. Tetlock, P.E., C.B. Mcguire & G. Mitchell. 1991. Psychological perspectives on nuclear deterrence. *Annual Review of Psychology* **42:** 239–276.
51. Nowak, M.A. & K. Sigmund. 1992. Tit for tat in heterogeneous populations. *Nature* **355:** 250–253.
52. Hein, G. & T. Singer. 2008. I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology* **18:** 153–158.
53. Singer, T. & S. Leiberg. 2009. Sharing the emotions of others: The neural bases of empathy. In *The Cognitive Neurosciences IV*. M.S. Gazzaniga, Ed.
54. Lutz, A., J. Brefczynski-Lewis, T. Johnstone & R.J. Davidson. 2008. Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise. *PLoS ONE* **3:** e1897.
55. Pace, T.W., L.T. Negi, D.D. Adame, S.P. Cole, T.I. Sivilli, T.D. Brown, M.J. Issa & C.L. Raison. 2009. Effect of compassion meditation on neuroendocrine, innate immune and behavioral responses to psychosocial stress. *Psychoneuroendocrinology* **34:** 87–98.
56. Kiecolt-Glaser, J.K., L. Mcguire, T.F. Robles & R. Glaser. 2002. Emotions, morbidity, and mortality: A new perspective from psychoneuroimmunology. *Annual Review of Psychology* **53:** 83–107.
57. Van Oyen Witvliet, C., T.E. Ludwig & K.L. Vander Laan. 2001. Granting Forgiveness or Harboring Grudges: Implications for Emotion, Physiology, and Health. *Psychological Science* **12:** 117–123.