

Sabine Windmann\* and Grit Hein

# Altruism from the Perspective of the Social Neurosciences

<https://doi.org/10.1515/nf-2017-A047>

**Abstract:** Altruism is a puzzling phenomenon, especially for Biology and Economics. Why do individuals reduce their chances to provide some of the resources they own to others? The answer to this question can be sought at ultimate or proximate levels of explanation. The Social Neurosciences attempt to specify the brain mechanisms that drive humans to act altruistically, in assuming that overtly identical behaviours can be driven by different motives. The research has shown that activations and functional connectivities of the Anterior Insula and the Temporoparietal Junction play specific roles in empathetic versus strategic forms of altruism, whereas the dorsolateral prefrontal cortex, among other regions, is involved in norm-oriented punitive forms of altruism. Future research studies could focus on the processing of ambiguity and conflict in pursuit of altruistic intentions.

**Keywords:** human; cognitive; brain imaging; evolution; motivation

## Conceptualizations of Altruism

Altruism (lat. *alter*, the other) can be defined as a behaviour that increases the welfare of other/s at a cost to oneself. It serves the prosocial motivation to cooperate in order to maximize the welfare of all (Nowak, 2006). Altruism and cooperation can be contrasted with egoism/individualism (the motivation to maximize one's own gain), competition (maximizing the difference between one's own gain and that of others), and spite (the motivation to minimize the resources of others; c.f. Murphy and Ackermann, 2014). Over the last one or two decades, altruism

and cooperation have become an intensely researched topic in both theoretical and empirical studies by various disciplines including Economics, Biology, Psychology, and the Neurosciences. One of the reasons for the great interest is that the concept cannot be easily reconciled with classical conjectures. For example, from a Darwinian point of view, if altruism is selfless and costly, how can it have evolved and sustained during evolution? Likewise, from an economic perspective expecting humans to maximize their subjective expected utility, why should anyone ever choose to bear any costs to benefit other/s?

Such fundamental theoretical conflict arises particularly when altruism is defined in terms of input-output relations (cost-benefit relations), or in terms of its effects on reproductive fitness (Wilson, 1992). This perspective is requested by natural scientists as they wish to maintain a purely objective stance on the concept of interest. Any behaviour that is costly to oneself but increases the benefit of others counts as altruistic, whether it is shown by humans (Becker and Eagly, 2004), fish (Daniels, 1981), bacteria (Lee et al., 2010), or plants (G. P. Murphy and Dudley, 2009).

Some researchers, mostly Psychologists, tend to find such a “behaviouristic” approach to altruism intuitively unsatisfactory, and instead refer to inner motivational states such as the *intention* to benefit others in defining altruism. For example, Batson and Shaw (1991) define altruism as “a motivational state with the ultimate goal of increasing another's welfare” (p. 108). Hence, these authors refer to proximate conditions, requiring researchers to determine the actor's active goal-states when determining the presence or absence of altruism. Notably, researchers who work with such definitions also tend to accept proximate *consequences* in evaluating the costs and benefits of the actor. For example, feelings of warm-glow that can accompany successful helping are viewed as beneficial for the helper and are therefore counted against the presence of “true” altruism (Batson and Shaw, 1991; Harbaugh et al., 2007).

The present review will not try to discuss or resolve differences between the disciplines in conceptualizing altruism, although we do stress that these differences are important to keep in mind when studying the literature. Instead, we will review ultimate and proximate explana-

\*Corresponding author: Sabine Windmann, Psychologisches Institut der Goethe-Universität Frankfurt, Theodor-W.-Adorno Platz 6, 60323 Frankfurt/Main, Germany, Mail: s.windmann@psych.uni-frankfurt.de

Grit Hein, Translational Social Neuroscience Unit, Julius-Maximilians-University Würzburg, Clinic for Psychiatry and Psychosomatics, Margarete-Höppel-Platz 1, D-97080 Würzburg, Germany; Institute for Medical Psychology, Johann Wolfgang Goethe-University Frankfurt, Heinrich Hoffmann Straße 10, D-60528 Frankfurt/Main, Germany, Mail: Hein\_G@ukw.de

tions of altruism, with a focus on past and potential contributions of the Social Neurosciences. What will become apparent is that in altruism research, some of the traditional boundaries between objective and subjective, biological and cultural, proximate and ultimate, appear questionable (Laland et al., 2011).

## Ultimate explanations

It has long been recognized that natural selection can favour altruism if the donor and the recipient of an altruistic act are genetic relatives, namely, when the costs for the altruist are lower than the degree of relatedness to the recipient. This is what John B. S. Haldane meant when he famously joked: “I will jump into the river to save two of my brothers or eight of my cousins” (Nowak, 2006).

To explain altruism among nonrelatives, however, evolutionary perspectives refer to some other indirect benefit of altruism for the actor, or for the group or network in which the actor lives (Nowak, 2006, 2012). The idea is, in brief, that by investing into the welfare of their social environment, altruistic individuals will “profit back” in the long run. Nowak speaks of a “snuggle for survival” (Nowak, 2012, p. 34).

Two key strategies are particularly noteworthy for their possible proximate implications. One is reciprocity, a mechanism ensuring that the equivalent of the altruistic actor’s investment will eventually be returned (Milinski, 2016). In “direct reciprocity”, the recipient of the returned benefit is the actor himself, so that the two engage in a give-and-take kind of cooperation. By contrast, in “indirect reciprocity”, more than two parties are involved in one of two forms: First, when indirect reciprocity runs “downstream”, bystanders who have observed altruistic behaviours will remember and communicate the actor’s behaviours favourably later on, thereby forming a positive reputation for the actor that will increase his/her future chances to find co-operators and partners, including mating partners (Iredale and Vugt, 2009). Secondly, when indirect reciprocity runs “upstream”, altruistic benefits are passed on indiscriminately, e.g., to other group members, as in Kant’s imperative “Do unto others as you would have others do unto you”.

By showing how altruism can indirectly involve reproductive benefits, the various forms of reciprocity in connection with group and network selection theories render altruism compatible with Darwin’s theory of evolution, even when it occurs among nonrelatives. However, they provide one major drawback, especially for large groups

in which the contribution of each individual to the welfare of the group cannot be easily tracked: the problem of *free riding*. Individuals who profit from the altruism of other group members but who don’t contribute their fair share in return should be better off than those who cooperate. They keep more of their resources in the first place, and benefit from the cooperators’ contributions like all other group members, so in total, their reproductive fitness is higher. This will render them likely to produce more offspring, and, under selection pressures, will drive altruists towards extinction. The parameters determining such evolutionary dynamics are investigated in computational modelling studies (Le and Boyd, 2007; Nowak and Sigmund, 1998; Riolo et al., 2001).

So free riding provides a huge dilemma to altruism theories. However, there is help, and it comes from the economic side of the research field. Ernst Fehr and his colleagues showed in economic games played in the laboratory that group members who observe free riding of other group members are willing to bear costs to punish those defectors (Fehr and Fischbacher, 2003; Fehr and Gächter, 2002). The punishment would reduce these individual’s payoffs, and deter them from free riding in the future. Without the punishment, group cooperation breaks down, but under the threat of punishment, cooperation is maintained (Fehr and Fischbacher, 2003; Fehr and Gächter, 2002).

Punishing defectors is the second key mechanism we wish to discuss here. It is selfless from an objective point of view as it is costly to the punisher but benefits the group by enforcing fairness. It is particularly effective and persuasive when exerted by group-endorsed authorities or institutions. At least on that large scale, it is unlikely to be driven by genetic factors alone, but is shaped by genetic-cultural co-evolution (Bowles et al., 2003; Fehr and Fischbacher, 2003; Henrich et al., 2006; Nowak, 2006).

## Proximate explanations

Any evolved behaviour must include a proximate mechanism that causes the organism to show the behaviour. Wilson (1992, p. 62) asks: “Are behaviors that are altruistic in the evolutionary sense necessarily caused by proximate mechanisms that are altruistic in the psychological sense?” This is where Social Neuroscience comes into play. With the means of brain imaging and other methods, we can specify and dissociate the proximate mechanisms mediating altruistic behaviours, and relate them to the underlying psychological motives (see table 1) quite

Table 1: Possible ultimate and proximate explanations of altruism.

Type of altruism	Helping (rewarding, fostering cooperation)	Costly Punishment (confrontative, prompted by delinquency and fairness violations)
Ultimate Explanation	Genetic relatedness Reciprocity, direct or indirect Group selection Network selection	Group selection Genetic-cultural coevolution
Proximate Explanation	Empathy Perspective taking Expectation of mutuality Unfairness aversion	Unfairness aversion Normative thinking Anger, need for revenge or dominance Moral convictions

independently of introspection and self-report (Camerer, 2008; Hein et al., 2016). Identifying objective traces for subjective motives can be quite important in researching concepts of strong social desirability such as altruism, because self-reports may be biased. Naturally, one needs to remain cautious about the idea of “mind reading” for methodological (Poldrack, 2011) as well as ethical reasons (Evers and Sigman, 2013). That is, from single studies alone (let alone from brain data of single individuals), it is usually not possible to infer subjective and motivational states, mostly because of high individual variability in structural and functional brain anatomy, and because of high regional interconnectedness and functional overlap. For that reason, Social Neuroscience studies do not simply collect and interpret brain data, but carefully select, manipulate or otherwise determine the motives of the participants before linking it with brain measures in order to identify the associated neural correlates and dynamics.

In the case of experiments on altruism, participants typically make decisions that benefit (an)other person(s) at costs to themselves under specified and to-be-compared experimental conditions. In the majority of the studies, the cost of an altruistic decision is monetary, i.e., the participants allocate more money to another person than to themselves (Hein et al., 2016; Morishima et al., 2012; Hutcherson, 2015; for classical reviews of specific paradigms see Sanfey, 2007; Fehr and Camerer, 2007). Some studies have also used physical costs, asking the participants to endure pain in order to benefit another person (Hein et al., 2010, 2011). Correlates of their brain activation, interregional connectivity measures, hormone status, and genetic parameters, among other measures, are being taken and linked to participants’ decisions and preference ratings later on, while considering their experimental treatment condition. The overall data pattern can inform about and sometimes dissociate the psychological motives behind evolutionarily defined altruistic behaviors or behavioral patterns.

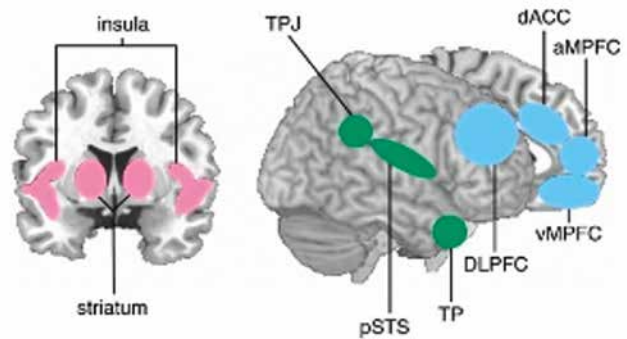
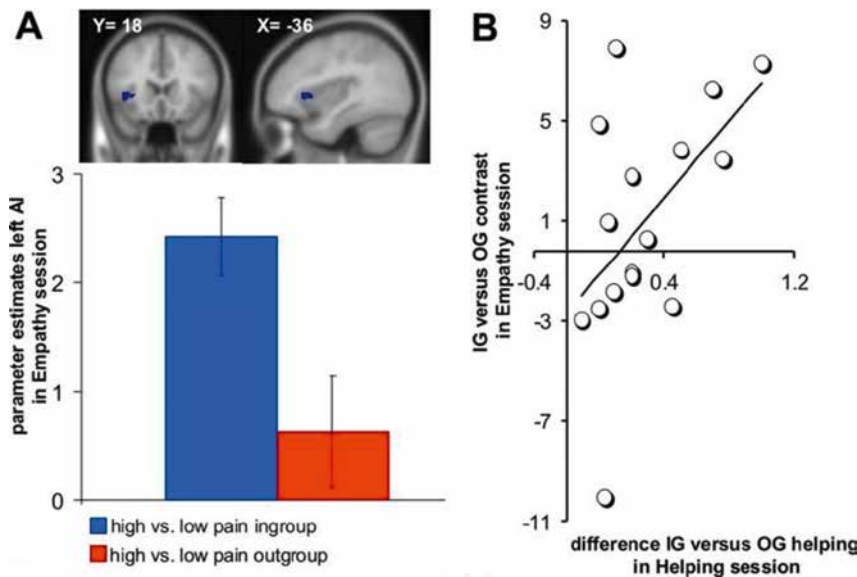


Fig. 1: Brain regions that are commonly involved in social processes. TPJ = Temporal Parietal Junction, dACC = dorsal anterior cingulate cortex, DLPFC = dorsolateral prefrontal cortex, pSTS = posterior part of the superior temporal sulcus, TP = temporal pole, vMPFC = ventral medial prefrontal cortex, aMPFC = anterior medial prefrontal cortex. modified from <https://www.universiteitleiden.nl/en/research/research-projects/social-and-behavioural-sciences/the-social-brain-in-adolescence>

The existing research has shown altruistic decisions to modulate the neural responses or neural connectivities of brain regions that are known to correlate with social processes, sometimes referred to as “the social brain” (Rilling and Sanfey, 2011; Adolphs, 2009; Van Overwalle, 2009). Among others, these brain regions include the insula, the striatum, the anterior cingulate cortex (ACC), the temporoparietal junction (TPJ), and the dorsolateral prefrontal cortex (DLPFC, Figure 1). It is important to note that these brain regions are also found in a variety of studies outside the social domain. This indicates that they correlate with overarching functions that play also a role in social settings. For example, activation in the insula correlates with interoceptive (Craig, 2009) and viscerosensory (Critchley and Harrison, 2014) inputs (e.g. pain) that in the social contexts are used to simulate sensory and emotional states of others, for example to empathize with the pain of another person (Zaki et al., 2016). The striatum is rich



**Fig. 2:** Results of an exemplary study that investigated the motivational basis of ingroup favoritism in altruistic decisions. A) The results revealed significant stronger activation in left anterior insula (AI) when seeing the ingroup member in pain as compared to the outgroup member. B) The individual differences in empathy-related brain activation when seeing the ingroup member and outgroup member in pain predicted individual differences in later helping behavior. The stronger the activation in AI for the ingroup member and not the outgroup member, the more likely was the person to help the ingroup member and not the outgroup member. IG = ingroup, OG = outgroup. Figure taken from Hein et al. (2010), *Neuron* 68, S. 149–160.

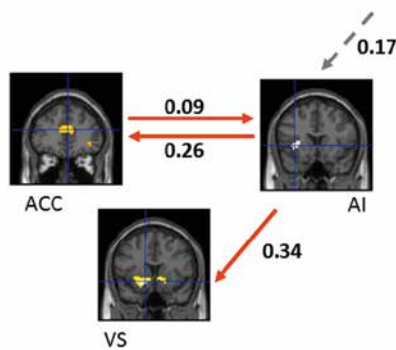
of dopamine, i.e., the neurotransmitter that is involved in transmitting reward signals. Based on that it plays a central role in the processing of rewarding inputs, inside and outside the social domain (Schultz, 2017). The anterior cingulate cortex processes information that is necessary for the flexible regulation of behavior such as errors, rewards, conflicts (Kolling et al., 2016). Such signals are integrated to initiate motivated behaviors in social and nonsocial situations (Shackman et al., 2011). In the following we will discuss two exemplary studies that have used brain activations and neural connectivities to investigate the altruistic behavior and its underlying motives.

One example study (Hein et al., 2010) investigated the motivational basis of ingroup favoritism in altruistic decision-making, i.e., the well-known fact that people behave more altruistically towards members of their own social group (ingroup members) as compared to members of a different social group (outgroup members). Seeing either an ingroup member or a outgroup member receiving a painful shock, the participant could volunteer to receive half of that person's pain himself, thus reducing the intensity of that person's pain stimulation by half. Such helping behaviour was costly for the participant because it resulted in a painful shock. The behavioural results showed an ingroup bias in helping, with participants choosing the more costly helping decisions if the suffering other was an ingroup member. Interestingly, the extent to which a person favoured the ingroup member was predicted by the individual difference in empathy-related brain activation in the anterior part of the insula (AI) when seeing the ingroup member and the outgroup member in pain (as measured in an independent part of the study). The stronger the

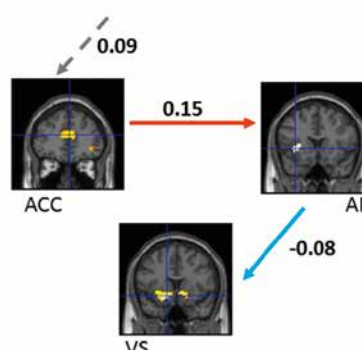
difference in empathy-related AI activation in a direction favouring the ingroup member, the more likely the person was to help the ingroup member and not the outgroup member (Figure 2). Moreover, activation in the ventral striatum when observing the outgroup member suffering predicted a lack of helping towards the outgroup member. The stronger the activation of this reward-related region a person experienced when seeing the outgroup member suffering in the first part of the study, the less likely he was to help the outgroup member later on. Additional analysis (commonality analyses) tested the contribution of brain activations (AI, ventral striatum) and self report measures (ingroup/ outgroup empathy and impression) to explaining variance in helping behaviour towards the ingroup and the outgroup member. The results indicated that behaviour towards a preferred other (ingroup member) is predicted equally well by brain measures and self report, whereas brain measures explain additional variance when it comes to behaviour towards a non-preferred other. Taken together, these results indicate that differences in empathy can motivate differences in altruistic decisions between ingroup and outgroup members. Moreover, they highlight the importance of brain measures for predicting actual behaviour in socially sensitive situations, for example regarding the lack of altruistic decisions towards non-preferred others.

Another recent study (Hein et al., 2016) used a monetary cost paradigm from behavioural economics with the aim to distinguish between two different motives that drove the identical altruistic decision. In the first part of the study, participants underwent a motive induction procedure that activated either an empathy motive or a reci-

### ▪ Reciprocity



### ▪ Empathy



**Fig. 3:** Results of an exemplary study that revealed different motives (empathy and reciprocity) based functional brain connectivity. Solid arrows indicate how much the level of activation in one brain region changes the rate of activation in the respective other brain region of the network (effective connectivities). Dashed arrows indicate changes within brain regions as a result of the experimental manipulation (inputs). Numbers indicate average model parameters that reflect the strength of effective connectivities or inputs. ACC = anterior cingulate cortex, AI = anterior insula, VS = ventral striatum.

procuity motive. To activate empathy, participants observed another person receiving pain. To activate reciprocity, participants received a favour from the other person. Based on these different motives, in the second part of the study, all participants were confronted with the same economic decision task in which they allocated points to themselves or another person (that were later transferred into money). Participants showed a similar number of altruistic decisions (i.e., allocations in favour of the other person that reduced the participant's payout) in the empathy and reciprocity condition, indicating that the two motives could not be differentiated based on behavioural measures. Moreover, analyses that simply looked at the functional activity of specific regions of the brain could not reveal the motive underlying the decisions. Broadly speaking, the same areas in the brain lit up in both settings, including the AI, the ventral striatum, and the anterior cingulate cortex (ACC). In a next step, the authors used Dynamic Causal Modeling (DCM), a method that estimates the direction of functional neuronal connectivities based on assumptions about the generation of the measured signal, to investigate the interplay between these brain regions, and found marked differences between empathy-based and reciprocity-based decisions (Figure 3). The impact of the motives on the interplay between the brain regions was so fundamentally different that it could be used to statistically classify the motive of a person with high accuracy. A further important result was that motives are processed differently in selfish and prosocial people. In selfish people, the empathy but not the reciprocity motive increased the number of altruistic decisions. After activating the empathy motive, selfish individual resembled persons with prosocial preferences in terms of brain connectivity and altruistic behavior. In contrast, prosocial people behaved even more altruistically after activating the reciprocity, but not the empathy motive.

The studies by Hein et al. (2010 and 2016) exemplarily illustrate the paradigms and methods used in the relatively young research field of the Social Neurosciences. These methods need to be consolidated and optimized in future studies, and combined with other approaches and developments in future studies.

## Rewarding versus punitive forms of altruism

The hitherto described research has identified a diversity of motives behind costly helping, a rewarding behavior that promotes cooperation, thereby maximizing the welfare of all. Yet how do these findings relate to altruistic punishment, the second key mechanism that sustains cooperation by retaliating upon defectors? Interestingly, behavioral experiments with economic games have shown the two, costly helping and costly punishment, to be essentially uncorrelated (Peysakhovich et al., 2014; Yamagishi et al., 2012). Once again, Social Neuroscience can help to elucidate the differences in terms of the mediating proximate mechanisms, specifically, in terms of the neuronal basis underlying the motives.

With regards to emotional processes, a number of studies have reported that costly helping is typically accompanied by warm-glow and other positive feeling states (Harbaugh et al., 2007; Hu et al., 2016a, Rand et al., 2015), whereas costly punishment is fueled by anger and the spiteful desire for revenge (Crockett et al., 2014; Fehr and Gächter, 2002; Seip et al., 2009; Singer et al., 2006). This makes the latter appear competitive and confrontational at the proximate level, quite far away from a caring intent to benefit others.

Neuroimaging research confirms that the two types of behaviours are associated with different brain mechanisms. Although both involve ventral and/or dorsal parts of the striatum as part of the elementary reward circuitry of the brain (for helping/sharing, (Genevsky et al., 2013; Harbaugh et al., 2007; Hein et al., 2010; Kuss et al., 2013), for punishment (Buckholtz et al., 2008; de Quervain et al., 2004; Hu et al., 2015; Strobel et al., 2011)), punishment more consistently involves the (typically right) dorsolateral prefrontal cortex (DLPFC), as shown with functional MRI (Buckholtz et al., 2008; Buckholtz et al., 2015; Sanfey et al., 2003), repetitive transcranial brain stimulation (Buckholtz et al., 2015; Knoch et al., 2008; Strang et al., 2015), and resting-state electroencephalography (Knoch et al., 2010). Because the DLPFC is generally thought to goal-dependently modulate automated choice valuation and action selection, its role in the context of costly punishment may be to provide information on social norms, moral values, and other abstract cognitions, in order to prevent reflexes and impulses from automatically driving the behavior (Feng et al., 2015). This contrasts with proximately other-concerned motives underlying helping and sharing which are driven either by affective empathy towards the needy person, involving anterior insula and medial prefrontal cortex, or cognitive empathy and perspective-taking involving the TPJ, among other regions (Haas et al., 2015; Hein et al., 2010; Morelli et al., 2014; Strombach et al., 2015; Tusche et al., 2016).

In conclusion, helping and costly punishment, two types of behaviors that both appear altruistic from an evolutionary point of view, seem to be based proximately on qualitatively different mechanisms from a Social Neuroscience perspective. Although specific interpretations about the functional role of the involved structures are often provisional and speculative, and sometimes appear simplified, the goal is to link these with more established findings and theories of the Cognitive Neurosciences. Over time, a picture shall emerge that shows the diverse facets of altruism as well as the underlying principles and processes at the neuronal level.

## Outlook

In summary, by examining the neural underpinnings, the Social Neurosciences can help to show the different proximate motives underlying various altruistic behaviours which uniformly meet the evolutionary definition of altruism. Some of these mechanisms involve affective care and empathy, whereas others recruit strategical thinking

in pursuit of the interest to maximize one's own return via cooperation. Yet other behaviours may be rooted in the motivation to impose norms on others, in pursuit of abstract goals such as fairness, as opposed to being driven by direct reward expectations. In this sense, the Social Neurosciences can provide clues as to the interpretation of the proximate motives of altruism in terms of more basic cognitive, motivational, and emotional processes. For large-scale coordination of cooperation among strangers, a diversity of proximate motives seems required, only some of which impress as altruistic from a psychological perspective.

Future work could directly contrast the proximate causes of rewarding and punitive altruism (experimental approaches are provided by Hu et al., 2015; Hu et al., 2016b). Are the motives underlying altruistic punishment, involving the DLPFC, truly rooted in *prosocial* moral values and cognitions, or does the drive for social dominance perhaps play a role (e.g., to keep the ingroup homogeneous)? How are acts of moral courage to be evaluated (e.g., whistleblowing) that go beyond normative thinking as is characteristic of altruistic punishment? What is the role of empathy and affect regulation in altruistic acts with other-concerned, caring intentions that can only be achieved by means of confrontation and the induction of aversive feeling states in the recipient? These questions aim at motive-inherent conflicts that need to be resolved by the social brain. They illustrate the multifaceted nature of the phenomenon of altruism, and highlight the importance of interdisciplinary research approaches to which the Social Neurosciences can contribute.

**Funding:** This work was supported by the German Research Foundation (DFG, HE 4566/5-1 to GH).

## References

- Adolphs, R. (2009). [The social brain: neural basis of social knowledge](#). *Annu Rev Psychol* 60, 693–716.
- Batson, C. D., Shaw, L. L. (1991). [Evidence for altruism: toward a pluralism of prosocial motives](#). *Psychol. Inq.* 2(2), 107–122.
- Becker, S. W. & Eagly, A. H. (2004). [The heroism of women and men](#). *Am. Psychol.* 59, 163–78.
- Bowles, S., Choi, J. K. & Hopfensitz, A. (2003). The co-evolution of individual behaviors and social institutions. *J. Theor. Biol.* 223, 135–47.
- Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D. & Marois, R. (2008). [The neural correlates of third-party punishment](#). *Neuron* 60, 930–40.
- Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O. & Marois, R. (2015). From Blame to Punishment:



- Disrupting Prefrontal Cortex Activity Reveals Norm Enforcement Mechanisms. *Neuron* 87, 1369–80.
- Camerer, C. F. (2008). Neuroeconomics: opening the gray box. *Neuron* 60, 416–9.
- Craig, A. D. (2009). How do you feel – now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70.
- Critchley, H. D. & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron* 77, 624–38.
- Crockett, M. J., Ozdemir, Y. & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *J. Exp. Psychol. Gen.* 143, 2279–86.
- Daniels, R. A. (1981). Altruism in an antarctic fish. *Science* 213, 1281.
- de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–8.
- Evers, K. & Sigman, M. (2013). Possibilities and limits of mind-reading: a neurophilosophical perspective. *Conscious. Cogn.* 22, 887–97.
- Fehr, E. & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cognit. Sci.* 11, 419–27.
- Fehr, E. & Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–91.
- Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–40.
- Feng, C., Luo, Y. J. & Krueger, F. (2015). Neural signatures of fairness-related normative decision making in the ultimatum game: a coordinate-based meta-analysis. *Hum. Brain Mapp.* 36, 591–602.
- Genevsky, A., Vastfjall, D., Slovic, P. & Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: affect shifts preferences for giving. *J. Neurosci.* 33, 17188–96.
- Haas, B. W., Brook, M., Remillard, L., Ishak, A., Anderson, I. W. & Filkowski, M. M. (2015). I know how you feel: the warm-altruistic personality profile and the empathic brain. *PLoS One* 10, e0120639.
- Harbaugh, W. T., Mayr, U. & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–5.
- Hein, G., Lamm, C., Brodbeck, C. & Singer, T. (2011). Skin conductance response to the pain of others predicts later costly helping. *PLoS One* 6, e22759.
- Hein, G., Morishima, Y., Leiberg, S., Sul, S. & Fehr, E. (2016). The brain's functional network architecture reveals human motives. *Science* 351, 1074–8.
- Hein, G., Silani, G., Preuschoff, K., Batson, C. D. & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68, 149–60.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. & Ziker, J. (2006). Costly punishment across human societies. *Science* 312, 1767–70.
- Hu, T. Y., Li, J., Jia, H. & Xie, X. (2016a). Helping Others, Warming Yourself: Altruistic Behaviors Increase Warmth Feelings of the Ambient Environment. *Front. Psychol.* 7, 1349.
- Hu, Y., Scheele, D., Becker, B., Voos, G., David, B., Hurlmann, R. & Weber, B. (2016b). The Effect of Oxytocin on Third-Party Altruistic Decisions in Unfair Situations: An fMRI Study. *Sci. Rep.* 6, 20236.
- 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. *Front. Behav. Neurosci.* 9, 24.
- Hutcherson, C. A., Bushong, B. & Rangel, A. (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron* 87, 451–62.
- Iredale, W. & Vugt, M. V. (2009). The peacock's tail of altruism. *The Psychologist* 22, 4.
- Knoch, D., Gianotti, L. R., Baumgartner, T. & Fehr, E. (2010). A neural marker of costly punishment behavior. *Psychol. Sci.* 21, 337–42.
- Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A. & Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation – the example of punishing unfairness. *Cereb. Cortex* 18, 1987–90.
- Kolling, N., Wittmann, M. K., Behrens, T. E., Boorman, E. D., Mars, R. B. & Rushworth, M. F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* 19, 1280–5.
- Kuss, K., Falk, A., Trautner, P., Elger, C. E., Weber, B. & Fließbach, K. (2013). A reward prediction error for charitable donations reveals outcome orientation of donors. *Soc. Cogn. Affect. Neurosci.* 8, 216–23.
- Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W. & Uller, T. (2011). Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? *Science* 334, 1512–6.
- Le, S. & Boyd, R. (2007). Evolutionary dynamics of the continuous iterated prisoner's dilemma. *J. Theor. Biol.* 245, 258–67.
- Lee, H. H., Molla, M. N., Cantor, C. R. & Collins, J. J. (2010). Bacterial charity work leads to population-wide resistance. *Nature* 467, 82–5.
- Lopez-Perez, B., Howells, L. & Gummerum, M. (2017). Cruel to Be Kind: Factors Underlying Altruistic Efforts to Worsen Another Person's Mood. *Psychol Sci* 28, 862–871.
- Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philos. Trans. R. Soc., B.* 371, 20150100.
- Morelli, S. A., Rameson, L. T. & Lieberman, M. D. (2014). The neural components of empathy: predicting daily prosocial behavior. *Soc Cogn Affect Neurosci* 9, 39–47.
- Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C. & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75, 73–9.
- Murphy, G. P. & Dudley, S. A. (2009). Kin recognition: Competition and cooperation in *Impatiens* (Balsaminaceae). *Am. J. Bot.* 96, 1990–6.
- Murphy, R. O. & Ackermann, K. A. (2014). Social value orientation: theoretical and measurement issues in the study of social preferences. *Pers. Soc. Psychol. Rev.* 18, 13–41.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science* 314, 1560–3.
- Nowak, M. A. (2012). Why we help: Far from being a nagging exception to the rule of evolution, cooperation has been one of its primary architects. *Sci. Am.* 307, 34–9.
- Nowak, M. A. & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–7.

- Peysakhovich, A., Nowak, M. A. & Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nat. Commun.* 5, 4939.
- Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron* 72, 692–7.
- Rand, D. G., Kraft-Todd, G. & Gruber, J. (2015). The collective benefits of feeling good and letting go: positive emotion and (dis)inhibition interact to predict cooperative behavior. *PLoS One* 10, e0117426.
- Rilling, J. K. & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48.
- Riolo, R. L., Cohen, M. D. & Axelrod, R. (2001). Evolution of cooperation without reciprocity. *Nature* 414, 441–3.
- Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science* 318, 598–602.
- 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, 1755–8.
- Schultz, W. (2017). Reward prediction error. *Curr. Biol.* 27, R369–R371.
- Seip, E. C., van Dijk, W. W. & Rotteveel, M. (2009). On hotheads and Dirty Harries: the primacy of anger in altruistic punishment. *Ann. N. Y. Acad. Sci.* 1167, 190–6.
- Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J. & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* 12, 154–67.
- 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, 466–9.
- Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B. & Sack, A. T. (2015). Be nice if you have to – the neurobiological roots of strategic fairness. *Soc. Cogn. Affect. Neurosci.* 10, 790–6.
- Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S. & Kirsch, P. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *Neuroimage* 54, 671–80.
- Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I., Tobler, P. N. & Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1619–24.
- Tusche, A., Bockler, A., Kanske, P., Trautwein, F. M. & Singer, T. (2016). Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. *J. Neurosci.* 36, 4719–32.
- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum Brain Mapp* 30, 829–58.
- Wilson, D. S. (1992). On the relationship between evolutionary and psychological definitions of altruism and selfishness. *Biol. Philos.* 7, 61–68.
- Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H. & Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20364–8.
- Zaki, J., Wager, T. D., Singer, T., Keyser, C. & Gazzola, V. (2016). The Anatomy of Suffering: Understanding the Relationship between Nociceptive and Empathic Pain. *Trends Cognit. Sci.* 20, 249–59.

**Article note:** German version available at <https://doi.org/10.1515/nf-2017-0047>

## Bionotes



### Prof Dr Sabine Windmann

Psychologisches Institut der Goethe-Universität Frankfurt, Theodor-W.-Adorno Platz 6, 60323 Frankfurt/Main, Germany  
Phone: +4969-798-35313  
Mail: [s.windmann@psych.uni-frankfurt.de](mailto:s.windmann@psych.uni-frankfurt.de)

Sabine Windmann is a professor of Cognitive Psychology at the Johann-Wolfgang-Goethe University Frankfurt/Main since 2006. She studied Psychology in Gießen and Bonn, obtained her PhD at the University of Trier and her "Habilitation" at the University of Bochum. She spent two years of her post-doctoral research at the University of California San Diego and two more at the University of Plymouth (UK). Her research interests refer to the question of how brain and cognition resolve uncertainty, ambiguity, and conflict, especially in social contexts. She lives in Weinheim/Bergstraße with her husband and her two kids.



### Prof Dr Grit Hein

Translational Social Neuroscience Unit, Julius-Maximilians-University Würzburg, Clinic for Psychiatry and Psychosomatics, Margarete-Höppel-Platz 1, D-97080 Würzburg, Germany;  
Institute for Medical Psychology, Johann Wolfgang Goethe-University Frankfurt, Heinrich Hoffmann Straße 10, D-60528 Frankfurt/Main, Germany  
Mail: [Hein\\_G@ukw.de](mailto:Hein_G@ukw.de)

Grit Hein is a psychologist, neuroscientist and professor of Translational Social Neuroscience at the University of Würzburg. She studied Psychology at the Humboldt-University in Berlin and the City College of New York, and obtained her PhD at the Max-Planck-Institute for Human Cognitive and Brain Sciences in Leipzig. After research visits to Cambridge (England), Frankfurt/Main and Berkeley (USA), she became researcher at the University of Zürich and lecturer at the University of Bern (Switzerland). Her research focuses on the neuronal basis of social motives and its impact on behaviour, the interplay between social motivation and learning, and the motivational deficits of patients with psychological disorders. Grit Hein is married and mother of two sons.