



Viewing Another Act as You Would Creates Altruistic Desires towards That Other

Paul Bogdan*

School of Information, University of Michigan, Ann Arbor, MI, United States

There has been growing evidence for the existence of distributed, frequently updating social “indices”, which are related to the reputation of others and predict altruism towards them. However, the means by which the brain modifies an index based on experiences is still unknown. This work utilizes recent insights on the role of the anterior cingulate cortex during perspective taking, dorsolateral prefrontal representations of context, the temporoparietal junctions relationship with understanding another’s background, and dorsomedial prefrontal activation patterns tracking reputation. It aims to show that cognitive empathy causes comparisons between a target’s action and the action one would wish to do in the target’s position. It also suggests that viewing a target perform the same action that one would in the target’s position creates altruistic desires towards the target. By considering these comparisons as central to understanding prosocial and antisocial motivations, a variety of behavioral studies are better explained. This piece seeks to open questions and discussions on the interplay of those brain regions, suggest future approaches to relationship therapy, and establish fundamentals for multi-agent models aimed at normative sociality.

Keywords: social behavior, empathy, altruism, fairness, cingulate cortex, medial prefrontal cortex, temporoparietal junction, action selection

OPEN ACCESS

Edited by:

Xiaolin Zhou,
Peking University, China

Reviewed by:

Jie Hu,
University of Zurich, Switzerland
Chao Liu,
Beijing Normal University, China

***Correspondence:**

Paul Bogdan
pcbogdan@umich.edu

Received: 12 September 2017

Accepted: 22 November 2017

Published: 08 December 2017

Citation:

Bogdan P (2017) Viewing Another Act as You Would Creates Altruistic Desires towards That Other. *Front. Hum. Neurosci.* 11:594. doi: 10.3389/fnhum.2017.00594

INTRODUCTION

Altruism is widely interpreted as taking actions for the benefits of another which confer a cost to oneself. While Charles Darwin initially found such behavior paradoxical, Nowak (2006) mathematically outlined situations where altruism is an evolutionarily beneficial aspect of cooperation. However, there still exists debate on how one’s cognitive processes give rise to such prosocial behavior. A systematic review identified explanations on the origins of altruistic behavior related to empathy, socially learned norms and fairness/reciprocation (Feigin et al., 2014). The concept of fairness/reciprocity are also closely associated with “reputation” (also referred to as status). In this context, the reputation of an individual or group determines the extent to which one values their benefits (Izuma, 2012). A recent review identified both competent and moral behavior as being common means of improving reputation (Bai, 2017). Notably though, members of a community may view the reputation of the same person differently. There does not yet exist a model for sociality which unifies or explains the above routes to altruism.

I argue that comparing how oneself and another would act for a given task is a fundamental aspect of social cognition, an outcome of cognitive empathy (theory of mind), and should be an element of unifying models. If the target would act how the judge (performing cognitive empathy) would wish to act in the target’s position, then the judge feels greater motivations to be altruistic

towards the target. If target does not act how the judge believes would wish to act, particularly if the target is acting less altruistically than expected, then the judge will feel less altruistic desires towards the target. For instance, a judge may become upset that they were [cut off] by a target, while the target was <entering the highway>, the task. The judge's anger towards the target is centered on the belief, that the judge would not [cut off] the target, if the judge was <entering the highway>. However, the target may believe they were justified as their task is <driving to hospital while wife is currently going through labor>. The upset judge would act the same if they understood the circumstances, but by neglecting the healthcare details, the judge believes the target has treated them *unfairly*. Moreover, if they knew the other was a student driver, this element, related to their background, would have made the other's actions more acceptable. These examples show how antisocial desires emerge from seeing a target act not how the judge would. However, by changing the considered contextual behavior, the judge may realize that the target is acting how the judge would, creating prosocial desires.

To support the existence of such computations, I will first show how a view, which considers judges comparing the actions of targets to the fictive actions the judge would perform, as opposed to reciprocation theories (Levine, 1998; Behrens et al., 2009), affords explanations to paradoxical social psychology findings. Notably this includes results on defectors punishing cooperators in economic games, motor synchrony promoting altruism, and a preference for those who treat judges with fairness rather than favoritism.

To support this view's neural plausibility, I will then outline findings on a cognitive empathy network involving the dorsolateral prefrontal cortex (dlPFC), temporoparietal junction (TPJ), anterior cingulate cortex gyrus (ACCg) and dorsomedial prefrontal cortex (dmPFC). The ACCg, which is known to assess the motivations of others (Apps et al., 2016), will be described as performing comparisons between one and another's action plans. I will also show that the context of an action (<entering highway> vs. <entering highway while wife is in labor>) is represented by the dlPFC, that accounting for the differences between the backgrounds of others (driver vs. student driver) utilizes the TPJ, and that indices which represent the extent to which a specific other's benefits are valued are represented by the dmPFC.

REPUTATION INDICES DEFINE THE EXTENT TO WHICH ANOTHER'S BENEFITS ARE VALUED

Nowak (2006) summarized findings showing that altruistic tendencies can evolve if they are directed at others who one expects will eventually reciprocate the altruism. Moreover, the extent of reciprocation expected determines the strength of one's altruistic tendencies, which are defined by the amount of benefits the other receives relative to the cost one is willing to incur. However, direct reciprocity is dependent on repeated encounters with the same individual and does not explain the advantages of

altruistic tendencies towards those who are generally helpful in one's community, indirect reciprocity. Nowak (2006) elaborated that the concept of reputation accounts such intercommunity altruism between donors and recipients who may never directly interact again.

The following formalism is also provided, showing that an individual should perform an altruistic act when another's reputation is greater than the ratio of the costs one incurs relative to the benefits another incurs, $r > c/b$. This view can also be interpreted as one receiving utility equal to $r * b$ when helping another (Levine, 1998).

Notably, people may view the reputation of the same community member differently causing varying altruistic behavior towards that member. Levine (1998) also showed that it is common for people to perceive the reputation of novel others as being greater than zero, $r > 0$, creating light altruistic tendencies towards people one is unfamiliar with. Reputation indices may also be less than zero and promote spiteful actions which cause a large cost to another at a smaller cost to oneself. Spiteful actions are beneficial for one's own interests as they coax defectors into cooperating for the benefits of others (Herrmann et al., 2008; Starmans et al., 2017). Both prosocial and antisocial reputation indices have been found to exist in the brain and are frequently updating and distributed entities (Parkinson et al., 2017). This article proceeds concerned with how reputation indices change.

COGNITIVE EMPATHY ASSESSES DESERVINGNESS

A meta-analysis revealed that higher cognitive empathy scores were correlated with prosocial behavior (Imuta et al., 2016). Although, the correlation with altruism was weak ($r = 0.19$, $p < 0.05$, 37 studies sampled) and a number of the studies found a lack of effect (Imuta et al., 2016). Correlations between cognitive empathy and antisocial behavior have also been found (Cowell et al., 2015).

I interpret these findings as suggesting, that *cognitive* empathy does not necessarily promote altruism, but performs computations to assess the reputation/deservingness of others. This conclusion is supported by cognitive empathy questionnaire scores predicting the extent to which one is sensitive to justice (Decety and Yoder, 2016). Justice sensitivity predicts the extent to which one will reciprocate the altruism of others by modulating the rate of reputation index change following another's prosocial or antisocial actions (Levine, 1998). The expected behavioral outcomes of this relationship between cognitive empathy and changing reputation indices indeed occur. Greater cognitive empathy was found to promote altruistic offers in ultimatum games only when their partner made fair offers previously (Schug et al., 2016). Greater cognitive empathy also affects rejection rates of ultimatum games (spiteful actions; Beadle et al., 2012). Notably, young adults showing greater cognitive empathy will decrease their rejection rate (an altruistic shift) after viewing their partner be selfish, while older adults will increase their rejection rate (a spiteful shift;

Beadle et al., 2012). Such results are understandable given our view of altruistic desires emerging from judging another to be acting similarly to how one would in their position, as a different study found that young adults are more selfish during ultimatum games than older adults (Bailey et al., 2012).

This piece proceeds showing how further social psychology findings are well explained using a view which predicts altruistic desires towards those acting how one would act. Explanations on altruism due to affective empathy (Feigin et al., 2014) utilize different neural processes (Kanske et al., 2016) and are not considered. Similarly, explanations pertaining to social norms (Feigin et al., 2014) also involve different neural circuits than those related to reputations (where the utility of another is treated as one's own; Ruff et al., 2013; Parkinson et al., 2017). The remaining explanations, as identified by recent reviews on altruism and reputation (Feigin et al., 2014; Bai, 2017), are concerned with how performing fair/moral actions increases one's reputation. Such views relate closely to positions stating that one feels altruistic desires towards those that will reciprocate (Levine, 1998; Nowak, 2006). The following section explains how a variety of findings on fairness are equally or better explained using this work's focus on action comparisons and what the judge would do in the other's position.

THE MEANING OF FAIRNESS

Perceiving a partner in an economic game as being fair creates altruistic desires towards that person (Rabin, 1993; Levine, 1998; Herrmann et al., 2008). Similarly, perceiving a partner to be unfair prompts spiteful actions towards them (Nowak et al., 2000; Forber and Smead, 2014). Moreover, viewing a target to be fair or unfair, even when one is an unaffected third party, also affects valuations of the other's utility (Fehr and Fischbacher, 2004). This relationship has been identified in third-party subjects as young as 15-months old (Schmidt and Sommerville, 2011). Fairness promoting altruism also manages to explain altruistic behavior better than purely reciprocity-based theories, as 6–8-year old subjects, when comparing a distributor who shared toys equally among the subject and others with another distributor who showed favoritism towards the subject, were split 50/50 in terms of which distributor they preferred (Shaw et al., 2012).

However, it prompts the question, “what is fairness?”. In ultimatum/dictator games, fairness is straightforward: both players receive similar amounts of money (Tam, 2014). Although outside the laboratory, a consensus will not be so easily found, and even in ultimatum games, receiving subjects deem situations where they are given more money to be fairer than those where their partner receives additional money (Tam, 2014). This suggests fairness is subjective, often self-serving and without consistency (Konow, 2005; Feng et al., 2013).

To account for this, I believe biologically concerned positions on assessing the fairness of others, should deem that another's action is “fair” (promoting of altruistic desires) if it is what one *perceives* one would do in the other's position. The element

of perception (and not having to actually perform) also allows prescriptive biases to enter.

A variety of findings on fairness can be explained utilizing this other-perspective, action comparison view. For instance, studies have shown that participants will utilize self-sacrificing punishments to coax others into cooperating during economic games (Boyd and Richerson, 1992; Herrmann et al., 2008). In these studies, spiteful actions are usually administered by cooperators towards defectors due to unfair defector behavior (the defectors are not acting how the cooperators would). However, Herrmann et al. (2008) paradoxically found that in some cultures, defectors will punish cooperating partners. Such results are inexplicable by reciprocity theories. However, a view which predicts antagonistic intentions towards those who act differently than one would in their position can better explain Herrmann's et al. (2008) findings.

Levine's (1998) model, which proposed that one acts altruistically towards another whom one perceives as altruistic, is an example of a reciprocity model struggling to account for all findings. Such findings also include Nowak's et al. (2000) study, which found that those who accepted unfair offers garnered lower reputation and received more abuse. This conflicts with Levine's (1998) stance that their altruistic, forgiving behavior should beget a higher reputation (reviewed by Bai, 2017). In Nowak's et al. (2000) work, those who received abuse were *not* acting as the judging abuser would have acted in their position (and retaliate via punishment). By understanding the victim's actions in this expectations view, the continuing abuse becomes clearer.

OTHER SOCIAL PSYCHOLOGY FINDINGS IN THIS CONTEXT

The centrality of sharing views may explain why apologies are particularly powerful when the target assures the victim, that the action will not occur again (Scher and Darley, 1997). Such apologies, if believed, result in the victim perceiving themselves and the target to have similar views on the acceptable actions for a situation.

Viewing another as having synchronous motor activity with one is another means of perceiving them to be performing similar actions. This type of action comparison intuitively is unrelated to altruistic motivations, but motor synchrony (walking in-step vs. not or mouthing the words to a song in sync vs. out of sync) was indeed found to promote altruism (Wiltermuth and Heath, 2009). In support of these findings being due to cognitive processes, Wiltermuth and Heath (2009) also showed altruistic motivations not to be associated with emotional responses. Similar relationships between altruism and synchrony were confirmed in 14-month infants (Trainor and Cirelli, 2015).

Further evidence for the behavioral claims emerges from its proposed routes to conflicts (not having similar views on a what action is best for a task) being common issues relationship therapy methods address. For instance, most couples therapy models include focuses on five points: each member understanding the relationship's role in the other's life,

identifying the effects of actions on the other, not avoiding emotions regarding the relationship, effective communication, and vocalizations of appreciations (Benson et al., 2012). The successes of these approaches can all be represented in the proposed similarity view: discussing the role of the relationship aligns context representations/norms (which allow rational conclusions of appropriate actions), identifying effects on others aligns valuations (allowing similar decisions on what actions are appropriate), and avoiding emotions prevents effective communication. Communication is necessary to align tasks and action valuations. Action valuations are also better aligned by emphasizing positives.

These approaches aim to address failures to fully take the other's perspective. Coplan (2011) elaborates on the meaning of other-oriented perspective taking, as opposed to self-oriented perspective taking. The latter only involves one putting themselves in the other's context without accounting for differences in life histories/backgrounds, which may result in different competencies, assets or valuation schemes. Other-oriented perspective, on the other hand, accounts for these but requires greater knowledge of the other and cognitive effort.

The differences Coplan (2011) describe are useful for tying the neural relationships between the TPJ and dLPFC to behavioral outcomes. The following sections elaborate on these and other neural findings from the regions involved in cognitive empathy, the dmPFC, ACCg, TPJ and dLPFC, and how these findings support a view on perspective taking comparisons promoting altruism.

ALTRUISTIC TENDENCIES ARE ENCODED BY dmPFC INDICES

The dmPFC is understood as highly relevant to social cognition. Greater dmPFC activation occurs when subjects are told to think about others, in contrast to vmPFC, dLPFC and left insula activations, which all were greater during self-thinking (Denny et al., 2012). Increased dmPFC activation also wasn't exclusive to direct viewing of others, but also occurs when viewing video clips (Iacoboni et al., 2004) or images (Wagner et al., 2011) of social interactions.

Its activation is expected to be for the assessment of reputations, as dmPFC activation tracks cues for assessing reputations (Muscatell et al., 2012) and changes in self-esteem due to social evaluation (Eisenberger et al., 2011). Related to this role, the dmPFC also shows greater activation when witnessing unfair offers (Cheng et al., 2017), when learning norms or when witnessing the violation of norms (Berthoz et al., 2002), when realizing that a social group does not share your opinion (Klucharev et al., 2009), or when realizing the advice of another is bad (Behrens et al., 2008). Also, supporting the relationship between deservingness and reputation, medial prefrontal cortex (mPFC) activations were found to encode judgments of another's deservingness, features notably not encoded by the TPJ (Tusche et al., 2016).

As expected given the relationship between reputation and altruism, dmPFC activation was also found to be correlated with

both altruistic intentions (Rilling et al., 2008; Waytz et al., 2012) and punishment of others at a cost to oneself (spite; Feng et al., 2016). This has given rise to explanations that the dmPFC's role in altruism is encoding reputation indices, r , and modulating activity in the vmPFC (Izuma, 2012), which encodes the rewards associated with an action, $r * b$ (Weilbacher and Gluth, 2016; Hunt and Hayden, 2017). More recent developments have refined the relationship between the dmPFC and vmPFC, showing that reputation indices of higher valued others are located more ventrally in the mPFC, where purely selfish rewards exist (Sul et al., 2015).

ACCg activation similarly reflects beliefs of another's reputation (King-Casas et al., 2005), and its activation reflected learning of social hierarchies (Qu et al., 2017). These findings, identification of reciprocal projections between the ACCg and dmPFC (Apps et al., 2016), along with projections from the dLPFC and TPJ to the ACCg communicating information necessary for perspective taking (Lockwood et al., 2015) have given rise to explanations that the ACCg is performing computations allowing encoding of reputations in the dmPFC (Behrens et al., 2009; Apps et al., 2016). These findings and their relations to a general ACC role in decision-making are elaborated in the next section.

ACTION COMPARISONS IN THE ACC

The dorsal ACC (dACC; also found referred to as the ACC sulcus) is expected to have a role in comparing possible actions for a task (Hayden et al., 2009; Kolling et al., 2016), as the dACC's activation patterns reflect the valuations of actions (Alexander and Brown, 2011) and non-reward state transformations (Hayden and Platt, 2010) of not only performed but also of fictive decisions (Hayden et al., 2009; Kolling et al., 2016). For these functions, the dACC densely shares reciprocal connections with mPFC and lateral PFC (LPFC) regions (Beckmann et al., 2009; Medalla and Barbas, 2010). Such findings make ACC activation being correlated with regret, cognitive dissonance, and attitude change unsurprising (van Veen et al., 2009). Rangel and Hare (2010) and Cisek (2012) offer richer reviews of ACC comparisons and the existence of valuations in an action-space.

The ACCg has similar properties as the dACC, but for social assessments of another's actions, which I argue occur during cognitive empathy. Such a parallel is understandable in the context of the ACC as an action assessment module if empathy is tempered with regards to perception-action coupling (Preston and de Waal, 2002) and the Empathy Simulation Hypothesis (Gallese and Goldman, 1998). Perception-action coupling involves the same neurons representing an action to be also activated when viewing another perform said action (Preston and de Waal, 2002). Such findings have given rise to simulatory explanations for empathy where empathy is explained as simulating oneself in another's situation (Gallese and Goldman, 1998).

Beyond previously mentioned findings, there exist further results showing that the ACCg is a social parallel to the dACC, where the actions of others are compared to the fictive actions

the judge believes are ideal given the other's position and background. ACCg activations have been found to reflect the rewards, pains, and net value another receives for both the actions one performs (Apps and Ramnani, 2014) and actions the other performs (Apps et al., 2016). Notably, such reward judgments utilize expected valuations from the other agent's reference frame (Apps et al., 2016), which requires information from the TPJ and dLPFC to understand.

CONTEXT AND REFERENCE FRAME ENCODINGS BY THE dLPFC AND TPJ

The dLPFC shows increased activation when assessing the appropriateness of another's action (Buckholz et al., 2008), as it represents the other's context (and the morals/norms/expectations associated with such a context) during action assessment. Support for this extensive. Disruption of the dLPFC reduces norm sensitivity (Buckholz et al., 2015) and impairs the ability to account for the others' context during moral judgment (Tassy et al., 2011). Disruption of the right dLPFC results in lower rejection rates of unfair ultimatum game offers (Knoch et al., 2006). Increased dLPFC activity is also associated with better accounting of context in a variety of non-social paradigms (Ruff et al., 2013; Smittenaar et al., 2013; Rudorf and Hare, 2014; Balderston et al., 2017).

Although among social paradigms, many which identify relationships between dLPFC activation and behavioral outcomes found similar correlations with TPJ activation. For instance, dLPFC lesioned patients and low TPJ gray mass volume subjects both do worse in the eyes test for assessing theory of mind capabilities (Geraci et al., 2010; Sato et al., 2016). dLPFC and TPJ increased activations are also both correlated with better understandings of intentionality (Güroğlu et al., 2011), and flexibility during moral prescription tasks (Tei et al., 2017).

However, there exist experiments where the two regions perform differently, showing that the TPJ plays a role in representing another's unique background. While dLPFC activity will track the behavior of computers (Seo and Lee, 2008), TPJ activity is largely exclusive to interpreting humans and does not change from baseline due to computer responses (Lee and Harris, 2013). Further evidence for this role in mentalizing others is extensive: greater TPJ activation predicts the degree of cognitive other-perspective taking (Tusche et al., 2016), transcranial DC

stimulation of the TPJ also increases one's capabilities to imitate another (Hogeveen et al., 2015) and electrical stimulation of the right TPJ was found to induce out-of-body experiences (Blanke et al., 2002).

Most relevant for action judgment, during assessment of blameworthiness: TPJ activity peaks prior to dLPFC activation, suggesting distinct roles for the two regions or a modulatory/filter role for the TPJ (Buckholz et al., 2008), and increased dLPFC activity was associated with another's circumstances being deemed the cause of another's actions, while TPJ activation correlated with the other's character being the cause (Berthoz et al., 2002). Coplan (2011) provides a review of the behavioral manifestations of accounting for someone's context instead of their context and background.

Together, the above findings on the dLPFC and TPJ show the existence of context and background representations, which provide necessary information for ACCg comparison computations that change the values of reputation indices, r , in the dmPFC.

CONCLUDING REMARKS

I aimed to show that a necessary aspect of cognitive empathy is comparing the actions/thoughts of another to one's own actions/thoughts. By reinterpreting behavioral findings, I highlighted how altruistic desires can be reasonably hypothesized as occurring due to viewing another as acting as one would. A formal model is not proposed here, but I hope future works addresses the importance of comparisons between another's action and one's own action plans in models of sociality.

AUTHOR CONTRIBUTIONS

PB is the sole author and is responsible for the reading, writing and editing.

ACKNOWLEDGMENTS

The author thanks Henry Boldizar, Greg Davis, Matthew Rastovac, and Jonathan Shobrook for providing helpful conversations; Justin Webb and all the members of the Alpha Gamma Club for the inspiration for this project.

REFERENCES

- Alexander, W. H., and Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–1344. doi: 10.1038/nn.2921
- Apps, M. A., and Ramnani, N. (2014). The anterior cingulate gyrus signals the net value of others' rewards. *J. Neurosci.* 34, 6190–6200. doi: 10.1523/JNEUROSCI.2701-13.2014
- Apps, M. A., Rushworth, M. F., and Chang, S. W. (2016). The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron* 90, 692–707. doi: 10.1016/j.neuron.2016.04.018
- Bai, F. (2017). Beyond dominance and competence: a moral virtue theory of status attainment. *Pers. Soc. Psychol. Rev.* 21, 203–227. doi: 10.1177/1088868316649297
- Bailey, P. E., Ruffman, T., and Rendell, P. G. (2012). Age-related differences in social economic decision making: the ultimatum game. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 68, 356–363. doi: 10.1093/geronb/gbs073
- Balderston, N. L., Hsiung, A., Ernst, M., and Grillon, C. (2017). Effect of threat on right dlPFC activity during behavioral pattern separation. *J. Neurosci.* 37, 9160–9171. doi: 10.1523/JNEUROSCI.0717-17.2017

- Beadle, J. N., Paradiso, S., Kovach, C., Polgreen, L., Denburg, N. L., and Tranel, D. (2012). Effects of age-related differences in empathy on social economic decision-making. *Int. Psychogeriatr.* 24, 822–833. doi: 10.1017/S1041610211002547
- Beckmann, M., Johansen-Berg, H., and Rushworth, M. F. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190. doi: 10.1523/JNEUROSCI.3328-08.2009
- Behrens, T. E., Hunt, L. T., and Rushworth, M. F. (2009). The computation of social behavior. *Science* 324, 1160–1164. doi: 10.1126/science.1169694
- Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538
- Benson, L. A., McGinn, M. M., and Christensen, A. (2012). Common principles of couple therapy. *Behav. Ther.* 43, 25–35. doi: 10.1016/j.beth.2010.12.009
- Berthoz, S., Armony, J. L., Blair, R. J. R., and Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain* 125, 1696–1708. doi: 10.1093/brain/awf190
- Blanke, O., Ortigue, S., Landis, T., and Seeck, M. (2002). Neuropsychology: stimulating illusory own-body perceptions. *Nature* 419, 269–270. doi: 10.1038/419269a
- Boyd, R., and Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* 13, 171–195. doi: 10.1016/0162-3095(92)90032-y
- Buckholz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., et al. (2008). The neural correlates of third-party punishment. *Neuron* 60, 930–940. doi: 10.1016/j.neuron.2008.10.016
- Buckholz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., et al. (2015). From blame to punishment: disrupting prefrontal cortex activity reveals norm enforcement mechanisms. *Neuron* 87, 1369–1380. doi: 10.1016/j.neuron.2015.08.023
- Cheng, X., Zheng, L., Li, L., Zheng, Y., Guo, X., and Yang, G. (2017). Anterior insula signals inequalities in a modified Ultimatum Game. *Neuroscience* 348, 126–134. doi: 10.1016/j.neuroscience.2017.02.023
- Cisek, P. (2012). Making decisions through a distributed consensus. *Curr. Opin. Neurobiol.* 22, 927–936. doi: 10.1016/j.conb.2012.05.007
- Coplan, A. (2011). Will the real empathy please stand up? A case for a narrow conceptualization. *South. J. Philos.* 49, 40–65. doi: 10.1111/j.2041-6962.2011.00056.x
- Cowell, J. M., Samek, A., List, J., and Decety, J. (2015). The curious relation between theory of mind and sharing in preschool age children. *PLoS One* 10:e0117947. doi: 10.1371/journal.pone.0117947
- Decety, J., and Yoder, K. J. (2016). Empathy and motivation for justice: cognitive empathy and concern, but not emotional empathy, predict sensitivity to injustice for others. *Soc. Neurosci.* 11, 1–14. doi: 10.1080/17470919.2015.1029593
- Denny, B. T., Kober, H., Wager, T. D., and Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J. Cogn. Neurosci.* 24, 1742–1752. doi: 10.1162/jocn_a_00233
- Eisenberger, N. I., Inagaki, T. K., Muscatell, K. A., Haltom, K. E. B., and Leary, M. R. (2011). The neural sociometer: brain mechanisms underlying state self-esteem. *J. Cogn. Neurosci.* 23, 3448–3455. doi: 10.1162/jocn_a_00027
- Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evol. Hum. Behav.* 25, 63–87. doi: 10.1016/s1090-5138(04)00005-4
- Feigin, S., Owens, G., and Goodyear-Smith, F. (2014). Theories of human altruism: a systematic review. *Ann. Neurosci. Psychol.* 1:1. Available online at: <http://www.vipoa.org/neuropsychol>
- Feng, C., Deshpande, G., Liu, C., Gu, R., Luo, Y. J., and Krueger, F. (2016). Diffusion of responsibility attenuates altruistic punishment: a functional magnetic resonance imaging effective connectivity study. *Hum. Brain Mapp.* 37, 663–677. doi: 10.1002/hbm.23057
- Feng, C., Luo, Y., Gu, R., Broster, L. S., Shen, X., Tian, T., et al. (2013). The flexible fairness: equality, earned entitlement, and self-interest. *PLoS One* 8:e73106. doi: 10.1371/journal.pone.0073106
- Forber, P., and Smead, R. (2014). The evolution of fairness through spite. *Proc. Biol. Sci.* 281:20132439. doi: 10.1098/rspb.2013.2439
- Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/s1364-6613(98)01262-5
- Geraci, A., Surian, L., Ferraro, M., and Cantagallo, A. (2010). Theory of Mind in patients with ventromedial or dorsolateral prefrontal lesions following traumatic brain injury. *Brain Inj.* 24, 978–987. doi: 10.3109/02699052.2010.487477
- Güroğlu, B., van den Bos, W., van Dijk, E., Rombouts, S. A., and Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. *Neuroimage* 57, 634–641. doi: 10.1016/j.neuroimage.2011.04.032
- Hayden, B. Y., and Platt, M. L. (2010). Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* 30, 3339–3346. doi: 10.1523/JNEUROSCI.4874-09.2010
- Hayden, B. Y., Pearson, J. M., and Platt, M. L. (2009). Fictive reward signals in the anterior cingulate cortex. *Science* 324, 948–950. doi: 10.1126/science.1168488
- Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science* 319, 1362–1367. doi: 10.1126/science.1153808
- Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., et al. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Soc. Cogn. Affect. Neurosci.* 10, 1003–1009. doi: 10.1093/scan/nsu148
- Hunt, L. T., and Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nat. Rev. Neurosci.* 18, 172–182. doi: 10.1038/nrn.2017.7
- Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., et al. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal bold fmri signal increases compared to a resting baseline. *Neuroimage* 21, 1167–1173. doi: 10.1016/j.neuroimage.2003.11.013
- Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., and Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: a meta-analytic review. *Dev. Psychol.* 52, 1192–1205. doi: 10.1037/dev0000140
- Izuma, K. (2012). The social neuroscience of reputation. *Neurosci. Res.* 72, 283–288. doi: 10.1016/j.neures.2012.01.003
- Kanske, P., Böckler, A., Trautwein, F. M., Parianen Lesemann, F. H., and Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Soc. Cogn. Affect. Neurosci.* 11, 1383–1392. doi: 10.1093/scan/nsw052
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83. doi: 10.1126/science.11108062
- Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., and Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron* 61, 140–151. doi: 10.1016/j.neuron.2008.11.027
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156
- Kolling, N., Behrens, T., Wittmann, M., and Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. *Curr. Opin. Neurobiol.* 37, 36–43. doi: 10.1016/j.conb.2015.12.007
- Konow, J. (2005). Blind spots: the effects of information and stakes on fairness bias and dispersion. *Soc. Justice Res.* 18, 349–390. doi: 10.1007/s11211-005-8566-6
- Lee, V. K., and Harris, L. T. (2013). How social cognition can inform social decision making. *Front. Neurosci.* 7:259. doi: 10.3389/fnins.2013.00259
- Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Rev. Econ. Dyn.* 1, 593–622. doi: 10.1006/redy.1998.0023
- Lockwood, P. L., Apps, M. A., Roiser, J. P., and Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *J. Neurosci.* 35, 13720–13727. doi: 10.1523/JNEUROSCI.1703-15.2015

- Medalla, M., and Barbas, H. (2010). Anterior cingulate synapses in prefrontal areas 10 and 46 suggest differential influence in cognitive control. *J. Neurosci.* 30, 16068–16081. doi: 10.1523/JNEUROSCI.1773-10.2010
- Muscattell, K. A., Morelli, S. A., Falk, E. B., Way, B. M., Pfeifer, J. H., Galinsky, A. D., et al. (2012). Social status modulates neural activity in the mentalizing network. *Neuroimage* 60, 1771–1777. doi: 10.1016/j.neuroimage.2012.01.080
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science* 314, 1560–1563. doi: 10.1126/science.1133755
- Nowak, M. A., Page, K. M., and Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science* 289, 1773–1775. doi: 10.1126/science.289.5485.1773
- Parkinson, C., Kleinbaum, A. M., and Wheatley, T. (2017). Spontaneous neural encoding of social network position. *Nat. Hum. Behav.* 1:72. doi: 10.1038/s41562-017-0072
- Preston, S. D., and de Waal, F. B. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20. doi: 10.1017/s0140525x02000018
- Qu, C., Ligneul, R., Van der Henst, J. B., and Dreher, J. C. (2017). An integrative interdisciplinary perspective on social dominance hierarchies. *Trends Cogn. Sci.* 21, 893–908. doi: 10.1016/j.tics.2017.08.004
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302.
- Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* 20, 262–270. doi: 10.1016/j.conb.2010.03.001
- Rilling, J. K., Dagenais, J. E., Goldsmith, D. R., Glenn, A. L., and Pagnoni, G. (2008). Social cognitive neural networks during in-group and out-group interactions. *Neuroimage* 41, 1447–1461. doi: 10.1016/j.neuroimage.2008.03.044
- Rudolf, S., and Hare, T. A. (2014). Interactions between dorsolateral and ventromedial prefrontal cortex underlie context-dependent stimulus valuation in goal-directed choice. *J. Neurosci.* 34, 15988–15996. doi: 10.1523/JNEUROSCI.3192-14.2014
- Ruff, C. C., Ugazio, G., and Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science* 342, 482–484. doi: 10.1126/science.1241399
- Sato, W., Kochiyama, T., Uono, S., Sawada, R., Kubota, Y., Yoshimura, S., et al. (2016). Structural neural substrates of reading the Mind in the Eyes. *Front. Hum. Neurosci.* 10:151. doi: 10.3389/fnhum.2016.00151
- Scher, S. J., and Darley, J. M. (1997). How effective are the things people say to apologize? Effects of the realization of the apology speech act. *J. Psycholinguist. Res.* 26, 127–140. doi: 10.1023/A:1025068306386
- Schmidt, M. F., and Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS One* 6:e23223. doi: 10.1371/journal.pone.0023223
- Schug, J., Takagishi, H., Benech, C., and Okada, H. (2016). The development of theory of mind and positive and negative reciprocity in preschool children. *Front. Psychol.* 7:888. doi: 10.3389/fpsyg.2016.00888
- Seo, H., and Lee, D. (2008). Cortical mechanisms for reinforcement learning in competitive games. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3845–3857. doi: 10.1098/rstb.2008.0158
- Shaw, A., DeScioli, P., and Olson, K. R. (2012). Fairness versus favoritism in children. *Evol. Hum. Behav.* 33, 736–745. doi: 10.1016/j.evolhumbehav.2012.06.001
- Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D., and Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* 80, 914–919. doi: 10.1016/j.neuron.2013.08.009
- Starmans, C., Sheskin, M., and Bloom, P. (2017). Why people prefer unequal societies. *Nat. Hum. Behav.* 1:82. doi: 10.1038/s41562-017-0082
- Sul, S., Tobler, P. N., Hein, G., Leiberg, S., Jung, D., Fehr, E., et al. (2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proc. Natl. Acad. Sci. U S A* 112, 7851–7856. doi: 10.1073/pnas.1423895112
- Tam, N. (2014). Quantification of fairness perception by including other-regarding concerns using a relativistic fairness-equity model. *Adv. Soc. Sci. Res. J* 1, 159–168. doi: 10.14738/assrj.14.291
- Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., et al. (2011). Disrupting the right prefrontal cortex alters moral judgement. *Soc. Cogn. Affect. Neurosci.* 7, 282–288. doi: 10.1093/scan/nsr008
- Tei, S., Fujino, J., Kawada, R., Jankowski, K. F., Kauppi, J. P., van den Bos, W., et al. (2017). Collaborative roles of temporoparietal junction and dorsolateral prefrontal cortex in different types of behavioural flexibility. *Sci. Rep.* 7:6415. doi: 10.1038/s41598-017-06662-6
- Trainor, L. J., and Cirelli, L. (2015). Rhythm and interpersonal synchrony in early social development. *Ann. N Y Acad. Sci.* 1337, 45–52. doi: 10.1111/nyas.12649
- Tusche, A., Böckler, A., Kanske, P., Trautwein, F. M., and Singer, T. (2016). Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *J. Neurosci.* 36, 4719–4732. doi: 10.1523/JNEUROSCI.3392-15.2016
- van Veen, V., Krug, M. K., Schooler, J. W., and Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nat. Neurosci.* 12, 1469–1474. doi: 10.1038/nn.2413
- Wagner, D. D., Kelley, W. M., and Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb. Cortex* 21, 2788–2796. doi: 10.1093/cercor/bhr074
- Waytz, A., Zaki, J., and Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *J. Neurosci.* 32, 7646–7650. doi: 10.1523/JNEUROSCI.6193-11.2012
- Weilbacher, R. A., and Gluth, S. (2016). The interplay of hippocampus and ventromedial prefrontal cortex in memory-based decision making. *Brain Sci.* 7:4. doi: 10.3390/brainsci7010004
- Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Bogdan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.