Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see http://jneurosci.org/content/preparing-manuscript#journalclub.

An Eye for an Eye: Neural Correlates of the Preference for Punishment-Based Justice

Samantha J. Fede, Joshua L. Gowin, and Peter Manza
National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland 20892
Review of Stallen et al.

Justice is a fundamental part of organized social behavior; punishing social violations and compensating for unfair losses are rationally consistent with, and even required by, social contract. Yet whether human justice behaviors are rational is still an open question. In many cases, retributive justice (i.e., punishment for transgressions) is not the most effective method for reducing antisocial behavior (Latimer et al., 2005), yet it remains popular, so understanding how the brain processes experiences of injustice and decisions regarding punishment could help inform solutions that are both widely supported and broadly effective. For example, restorative justice focuses on repairing harm caused by antisocial behavior, such as by having a thief work to pay back stolen money. Much research has identified brain networks associated with processing unfairness, particularly in laboratory paradigms where a social "relationship" can be developed over repeated interactions (Feng et al., 2015). Stallen et al. (2018) pushed that effort forward by investigating the neural underpinnings of decisions regarding justice, where bystanders may choose how to allocate their resources. Participants must decide how much effort to devote to punishing the perpetrator for unfair behavior and how much effort to spend compensating victims for their losses. These conditions may correspond to retributive and restorative justice, respectively.

The authors used a novel "Justice Game" to examine how the brain decides between punishment and compensation. Participants were given 200 chips that could be redeemed for money at the end of the experiment. They then watched as a partner, who also had 200 chips, decided to accept their allotment or instead steal chips from a victim, who could be either the participant or a third person. The partner and third person were computers, but participants were told they were peers. On half of the trials, the partner stole up to 100 chips from the victim. When the participant was the victim, they were given the opportunity to punish the perpetrator by taking chips away from him at a cost of 1 chip paid per 3 chips taken. When the participant observed the perpetrator taking from a third party, they could either give chips to the victim from their own supply or take chips from the perpetrator; that is, they could decide to compensate or punish. Participants performed this task during an fMRI scan.

There were several important findings from the study. First, trials where the partner did not take any chips compared with those where he did were associated with medial prefrontal, posterior insula, posterior cingulate, and temporoparietal junction (TPJ) activations. The reverse contrast (partner taking chips vs not taking chips) was associated with lateral prefrontal, anterior cingulate, anterior insula, and precuneus activations. Second, participants were more likely to punish the perpetrator than compensate the victim, and decisions to punish rather than compensate were associated with higher ventral striatum activation. The greater the transgression (i.e., the more chips taken), the more severe the punishment participants doled out, particularly when the participant was the victim.

As another component of the study, some participants were assigned to receive either intranasal oxytocin or placebo spray. Oxytocin is an important hormone regulating social behaviors, such as promoting social bond development (Young and Wang, 2004) and trust during social interactions (Kosfeld et al., 2005). However, the function of oxytocin is complex and context-dependent. For instance, oxytocin increases aggressive behavior associated with maternal protectiveness (Ferris et al., 1992). It was not known how oxytocin would affect human response to injustice. Surprisingly, Stallen et al. (2018) found that oxytocin administration was associated with more frequent, but less severe, punishment, particularly when the participant was the victim. Individuals who received oxytocin also reported hav-

Received May 21, 2018; revised July 11, 2018; accepted July 19, 2018.
The authors declare no competing financial interests.
Correspondence should be addressed to Dr. Samantha J. Fede, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 301-594-7416, 10 Center Drive, 1-5417, MSC 1108, Bethesda, MD 20892. E-mail: fedesj@nih.gov.
DOI:10.1523/JNEUROSCI.1282-18.2018
 0270-6474/18/387559-03$15.00/0

ing a greater expectation of fairness than the nonoxytocin group. This effect corresponded to greater activation of the right anterior insula. Based on these findings, the authors suggest that oxytocin may influence the decision to punish based on cognitive judgments rather than emotion. They operationalize cognitive decisions to punish as those driven by the best interest of the group; however, meta-analytic work suggests that restorative justice, rather than punishment-based retributive justice, leads to better societal outcomes (e.g., less reoffending) (Latimer et al., 2005). Moreover, interpreting the anterior insula response to fairness violations as a cognitive process overlooks the diversity of functions attributed to this brain region, such as uncertainty, salience, pain, empathy, and interoception (Craig, 2009; Singer et al., 2009). Thus, insula function might contribute to processing during this task in both cognitive and emotional ways. For example, oxytocin may increase awareness of social rules, leading to a greater expectation of social conformity. When these higher expectations of justice are violated (i.e., when witnessing unfair behavior), these individuals may have a greater affective response in the insula corresponding to the greater difference between expected and witnessed behavior. This interpretation is consistent with the finding by Stallen et al. (2018) that individuals receiving oxytocin reported greater expectations of fairness than the placebo group and previous work that has shown oxytocin increases time spent on social feature processing and ability to use those features to interpret intentionality (Guastella et al., 2008).

Another region found to be involved in processing just and unjust decisions was the TPJ. This region is thought to support social cognitive processes, such as integrating semantic and social rule information into decision making, as well as theory of mind (i.e., thinking about what others are thinking) (Carter and Huettel, 2013). In the Stallen et al. (2018) study, greater TPJ engagement was associated with the choice to refrain from punishment. The TPJ is thought to integrate intentionality into judgments of blameworthiness. For example, inhibition of right TPJ using transcranial magnetic stimulation led participants to disregard intention in judging the permissibility of an action, such as judging an instance of accidental harm as "unacceptable" but attempted murder as "okay" because no one actually got hurt (Young et al., 2010). Therefore, when deciding whether to punish perpetrators or compensate victims, TPJ engagement may reflect more consideration of the possible reasons for which the perpetrator stole chips. Oxytocin is also related to engagement of the TPJ and use of intentionality in social cognition (Walter et al., 2012). For example, oxytocin administration increased left TPJ engagement when observing altruism (Hu et al., 2016). However, in this study (Stallen et al., 2018), oxytocin did not significantly modulate TPJ engagement in either compensation or punishment conditions. There are several possible explanations for the discrepant findings. There may be a ceiling to the utility of oxytocin in theory of mind processing in the TPJ. In other words, typical levels of oxytocin may have sufficiently influenced the activity of the TPJ in the context of this task such that additional oxytocin did not increase activation. Additionally, there were several differences in study design. Hu et al. (2016) reported a relationship between oxytocin and TPJ activity during third party observation of altruism, but here, participants made decisions about whether to punish or compensate. Finally, given a smaller sample size and that the TPJ was not included as an ROI *a priori*, the Stallen et al. (2018) study may have needed more participants to detect these effects.

Stallen et al. (2018) also provide insight into the neural mechanisms behind our desire to see perpetrators punished. Participants spent more chips and had greater activation in the ventral striatum when choosing to punish perpetrators rather than compensate victims. The authors suggested that these findings may indicate that participants found punishing the perpetrator more rewarding than helping the victim. This would be consistent with previous work that found increased ventral striatum activation when individuals chose to administer a shock to a person perceived as unfair (e.g., Singer et al., 2006). However, this is not the only explanation for these findings because higher ventral striatal activation during punishment relative to compensation may relate to cognitive dissonance. Individuals often paradoxically overvalue a choice that is seen as costly, as in the case of punishment in this task, where value is lost by both parties. Presumably, this increased motivation or valuation reflects a desire to alleviate dissonance, and these types of behaviors have been associated with ventral striatum activation in previous studies (Inzlicht et al., 2018). However, an important limitation of this finding is that this study only used male participants. There are established gender differences in the field of social cognition; for example, previous work has indicated that males, but not females, have ventral striatum activity associated with punishing for unfair behavior (Singer et al., 2006) and during cooperative social activity (Feng et al., 2015). Therefore, it is possible that the current results may be specific to males. Of note, oxytocin administration did not alter ventral striatum activation. This is not surprising, given that there are few to no oxytocin receptors in human ventral striatum (Boccia et al., 2013). Therefore, future research should investigate whether pharmacological manipulations of neurotransmitters with high ventral striatum receptor density (e.g., dopamine or opioids) affect punishment versus compensation decisions more potently.

The results of this work by Stallen et al. (2018) might be applied to better understand substance use disorders. Incarcerated individuals with a substance use disorder have abnormal neural responses when making moral decisions (Fede et al., 2016). In addition, laboratory experiments have found that, in response to having their money stolen, individuals with substance use disorders are more likely than healthy controls to punish a perpetrator by stealing money (Kose et al., 2015). Some theories of addiction suggest that oxytocin deficits may impel preferences for substances over social relationships (McGregor et al., 2008). Accordingly, this task offers an opportunity to explore differences in neural processing of justice that might explain the relationship between substance use and antisocial behavior. Finally, it offers an opportunity to test whether any potential differences might be corrected through oxytocin administration. Thus, this study represents a starting point to explore new aspects of neural function that may lead to exciting new therapeutic approaches.

References

Boccia ML, Petrusz P, Suzuki K, Marson L, Pedersen CA (2013) Immunohistochemical localization of oxytocin receptors in human brain. Neuroscience 253:155–164. CrossRef Medline

Carter RM, Huettel SA (2013) A nexus model of the temporal-parietal junction. Trends Cogn Sci 17:328–336. CrossRef Medline

Craig AD (2009) How do you feel–now? The anterior insula and human awareness. Nat Rev Neurosci 10:59–70. CrossRef Medline

Fede SJ, Harenski CL, Borg JS, Sinnott-Armstrong W, Rao V, Caldwell BM, Calhoun VD (2016) Abnormal fronto-limbic engagement in incarcerated stimulant users during moral process-

ing. Psychopharmacology (Berl) 233:3077–3087. CrossRef Medline

Feng C, Hackett PD, DeMarco AC, Chen X, Stair S, Haroon E, Ditzen B, Pagnoni G, Rilling JK (2015) Oxytocin and vasopressin effects on the neural response to social cooperation are modulated by sex in humans. Brain Imaging Behav 9:754–764. CrossRef Medline

Feng C, Luo YJ, 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. CrossRef Medline

Ferris C, Foote K, Meltser H, Plenby M, Smith K, Insel T (1992) Oxytocin in the amygdala facilitates maternal aggression. Ann N Y Acad Sci 652:456–457. CrossRef Medline

Guastella AJ, Mitchell PB, Dadds MR (2008) Oxytocin increases gaze to the eye region of human faces. Biol Psychiatry 63:3–5. CrossRef Medline

Hu Y, Scheele D, Becker B, Voos G, David B, Hurlemann R, Weber B (2016) The effect of oxytocin on third-party altruistic decisions in unfair situations: an fMRI study. Sci Rep 6:20236. CrossRef Medline

Inzlicht M, Shenhav A, Olivola CY (2018) The effort paradox: effort is both costly and valued. Trends Cogn Sci 22:337–349. CrossRef Medline

Kose S, Steinberg JL, Moeller FG, Gowin JL, Zuniga E, Kamdar ZN, Schmitz JM,Lane SD (2015) Neural correlates of impulsive aggressive behavior in subjects with a history of alcohol dependence. Behav Neurosci 129:183–196. CrossRef Medline

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. Nature 435:673–676. CrossRef Medline

Latimer J, Dowden C, Muise D (2005) The effectiveness of restorative justice practices: a meta-analysis. Prison J 85:127–144. CrossRef

McGregor IS, Callaghan PD, Hunt GE (2008) From ultrasocial to antisocial: a role for oxytocin in the acute reinforcing effects and long-term adverse consequences of drug use? Br J Pharmacol 154:358–368. CrossRef Medline

Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD (2006) Empathic neural responses are modulated by the perceived fairness of others. Nature 439:466–469. CrossRef Medline

Singer T, Critchley HD, Preuschoff K (2009) A common role of insula in feelings, empathy and uncertainty. Trends Cogn Sci 13:334–340. CrossRef Medline

Stallen M, Rossi F, Heijne A, Smidts A, De Dreu CK, Sanfey AG (2018) Neurobiological mechanisms of responding to injustice. J Neurosci 38:2944–2954. CrossRef Medline

Walter NT, Montag C, Markett S, Felten A, Voigt G, Reuter M (2012) Ignorance is no excuse: moral judgments are influenced by a genetic variation on the oxytocin receptor gene. Brain Cogn 78:268–273. CrossRef Medline

Young LJ, Wang Z (2004) The neurobiology of pair bonding. Nat Neurosci 7:1048–1054. CrossRef Medline

Young L, Camprodon JA, Hauser M, Pascual-Leone A, Saxe R (2010) Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proc Natl Acad Sci U S A 107:6753–6758. CrossRef Medline