Sound Neuroscience: An Undergraduate Neuroscience Journal

Volume 2
Issue 1 *Women in Neuroscience*

Article 18

2015

Rebecca Saxe: Investigating Theory of Mind using Multi-Voxel Pattern Analysis in Autistic Adults

Olivia G. Cadwell
Olivia Cadwell, ocadwell@pugetsound.edu

Olivia Cadwell

Neuroscience 201

Rebecca Saxe: Investigating Theory of Mind using Multi-Voxel Pattern Analysis in Autistic Adults

Rebecca Saxe is a neuroscientist in the field of cognitive science and research. Her lab is located at Massachusetts Institute of Technology and conducts studies to relate cognitive states and mental state reasoning to moral judgment. Saxe uses multi-voxel pattern analysis (MVPA) in her research to study spatial neural responses in the right temporo-parietal junction (RTPJ) of the human brain. The RTPJ brain region of her interest is thought to comprise the "theory of mind network," which contains the neural basis for thinking about and understanding others' beliefs, intentions, and mental states (6). Through her studies, Saxe seeks to discover if there are neural differences in the brain that allow non-autism spectrum disorder (neurotypical) adults to ascribe intentionality behind the actions of other individuals. The ability to assign whether or not a harmful action was intentional is often deficit in high functioning adults with autism spectrum disorder (ASD). This particular study investigates the Theory of Mind (ToM) system using MVPA in order to learn whether distinct neural responses are associated with the ability to differentiate mental states such as intentional versus accidental acts of harm

In order for humans to distinguish a harmful situation as intentional or on accident, adults rely on moral cognitive abilities to consider another's intentions, beliefs, and emotions as a way of determining in what mindset a harmful action was committed (6). The ability to understand, relate to, and accurately interpret another person's behavior is referred by psychologists as Theory of Mind (ToM). According to Saxe, the theory of mind network is located primarily in the RTPJ brain region where there are clusters of neural populations that program this theory, though the exact mechanisms are unknown.

Individual's deficit in ToM abilities are typically diagnosed with social cognitive disorders such as ASD and social anxiety disorders and may exhibit atypical activity in this area, according to Saxes' hypothesis. Yet, while adults with high functioning ASD still struggle with ToM skills, Saxe discovered in an earlier study that these individuals do not consistently judge intentional and accidental harms as morally dissimilar (4). Instead, these individuals appear to exhibit ignorance about any information regarding a person's unintended actions, and subsequently an over emphasis on the action's negative implications. Succeeding research has

found MVPA to decode properties of perceived stimulus, (such as objects or sentences) and has made it possible for researchers to detect fluctuations in cognitive states. In turn, MVPA can be correlated to an individual's ongoing behavior, and is therefore suggested to be as close to 'mind reading' as currently possible in science (5). In order to interpret a participant's cognitive state, Saxe applied this sophisticated pattern-classification algorithm to fMRI data from a few seconds of brain activity she measured the existence or absence of mental states (5).

Following these studies and preliminary results, Saxe's conducted another study published "Decoding moral judgments from neural representations of intentions", (3) using subjects who are high functioning adults with ASD. She hypothesized that when the participants read about accidental or intentional harmful acts, the spatial patterns created from brain activity in the RTPJ could be decoded to infer whether the individual interpreted the harm as accidental or intentional. Through this research, Saxe inspected the sub layers of neurons that organize into clusters over the cortex to form reliable spatial patterns of activity measurable via fMRI. To distinguish cognitive features, she applied MVPA to these patterns and related the fMRI representations to the ASD participants' behavioral performance (3).

From these analyses, Saxe found evidential traits from mental state reasoning represented in a brain area involving ToM (3). These results indicated discrimination from MVPA of intentionality through neural pathways in the RTPJ of neurotypical adults. They also specified that a difference of magnitude in neural responses exists while the neurotypical adults read about the intentions and while they made ethical judgments. These participants exhibited more RTPJ activity for accidental harms than intentional harms, which Saxe explained coincides with prior studies to suggest that RTPJ is activated while individuals are justifying mental states. Therefore, when given information about the mind of the offender, the greater the degree of RTPJ in response to accidental harms indicates that neurotypical adults incorporate ToM reasoning to their moral judgments (3).

Alternatively, the high functioning ASD participants showed atypical neural activity in the RTPJ. Saxe found no distinction between intentional and accidental harms in the theory of mind RTPJ or any other relative region. Upon reading both harm related narrative scenarios, the subjects showed equal activation, which implies that accidental harms do not elicit a greater contemplation of mental states than intentional harms. These results correlate with the behavioral description of ASD adults. Namely, they are efficient in accurately detecting moral judgment but

cannot create a distinction between innocent transgressions and intentionally hurtful actions. Saxe stated, while taking in consideration many prior studies, the current results from her research advocates that ASD effects the organization and voxel pattern of information in the ToM brain areas. Thus, the result of Saxe's research suggests that individuals with ASD have consistent and reliable responses to in the RTPJ in reaction to all harmful acts, despite of whether the root of the action was accidental or

Individuals who experience difficulties with social interaction and communication extend beyond ASD, and these results present a glance at the neural mechanisms where these issues may originate. Other research on the topic of ToM has been conducted to identity regions in the brain that may contain the unknown encoding equipment for ToM skills (2). One interrelated region is the mirror neuron system and is "in line with introspection that a non-intended action does not produce attempts to attribute an intention," (2) and is supposed to be an integral member of social readabilities. Mirror neurons have been found to interpret intentions behind the actions of others' and appear to be the basic infrastructure to understanding intentionality. Another recent fMRI study found that there is a contrast of non-intended versus intended actions that exhibit activation in several brain regions, one of which is the RTPJ (2).

The Theory of Mind system has been discovered to exhibit trainability which suggests a continued elasticity and malleable nature (1) and the possibility of manipulation if correct cognitive therapy is given. There is great hope that that studies like Rebecca Saxe's will continue to be explored regarding ASD and other social cognitive disorders, and that its results will encourage new methods and interventions for therapy for an increased quality of life.

References

1. Bretherton, I., & Beeghly, M., (1982). Talking about internal states: The acquisition of anexplicit theory of mind. *Developmental Psychology, 18*, 906-921.
2. Fabbri-Destro, M., & Rizzolatti. G., (2008). Mirror Neurons and Mirror Systems in Monkeys and Humans. *Psychology 23:* 171-179; doi:10.1152/phisiol.00004.2008.
3. Saxe, R., J. Koster-Hale., & Young, L. L., (2013). Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences USA, 110(14):* 5648-5653.
4. Moran, J. M., Young, L. L., Saxe, R., Lee, S. Me., O'Young, D., Mavros, P. L., & Gabrieli, J. D., (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences USA, 108(7):* 2688-2692.
5. Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V., (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences,* doi:10.1016/j.tics.2006.07.005.
6. Uddin, L. Q., Iacobini, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends in Cognitive Science*, *11*(4), 153-157.