This article was an invited commentary on a reprinted article, "Impaired 'Affective Theory of Mind' Is Associated with Right Ventromedial Prefrontal Damage."

"Affective Theory of Mind" and the Function of the Ventral Medial Prefrontal Cortex

(*Cogn Behav Neurol* 2018;31:36–50)

The hardest problems of our daily lives often involve other people's emotions: Does she respect me? Did I offend him? What did she mean by that look? To resolve these questions, humans use not only our "general intelligence" and problem-solving skills, but also a distinct cognitive ability, sometimes called a "theory of mind." Figuring out what other people are thinking and feeling relies on a distinct group of brain regions, including the temporoparietal junction (TPJ) and medial prefrontal cortex (MPFC). Within theory of mind, the MPFC is more engaged in reasoning about others' feelings, while the TPJ is more engaged in reasoning about others' beliefs (Peelen et al, 2010; Saxe and Kanwisher, 2003). As a result, damage that mostly affects the MPFC, for example, can selectively impair a person's ability to understand others' feelings, while leaving their verbal and problem-solving abilities intact, a striking dissociation that was first proposed by Shamay-Tsoory et al (2005) in their groundbreaking article "Impaired 'Affective Theory of Mind' Is Associated with Right Ventromedial Prefrontal Damage."

Shamay-Tsoory et al (2005) inferred that the right ventral MPFC was necessary for understanding others' feelings. The evidence for this claim was indirect, because the tasks they used implied but did not explicitly measure emotion understanding. Lesions in the right ventral MPFC predicted difficulty understanding the intended meaning of sarcastic utterances (eg, seeing an employee resting, and saying "Don't work too hard!") or detecting who "said something wrong" in a faux pas (eg, insulting someone without knowing they could hear). Subsequently, Shamay-Tsoory and Aharon-Peretz (2007) directly measured emotion understanding in the same patients, and confirmed a selective impairment. Further, a TPJ lesion leads to decrements in the ability to pass a false belief task (Samson et al, 2004). Combined, these early lesions studies hint at a double dissociation between cognitive and affective theory of mind.

Received for publication February 28, 2018; accepted February 28, 2018.
Commentary on: Shamay-Tsoory SG, Tomer R, Berger BD, Goldsher D, Aharon-Peretz J. 2005. Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cogn Behav Neurol.* 18:55–67.
The commentators declare no conflicts of interest.

Many subsequent studies have confirmed the link between the MPFC and understanding others' emotions. For example, neurodegenerative damage to the right ventral MPFC causes abnormally low emotional responses to the suffering of victims of intentional harms, while leaving comprehension of the intentional action intact (Baez et al, 2016). Converging evidence comes from functional magnetic resonance imaging studies. Activity in the MPFC is high when people read stories about characters experiencing intense negative emotions (eg, proposing marriage and being refused), compared to intense physical pain (eg, breaking an ankle and tearing the ligaments) (Jacoby et al, 2016). The pattern of activity in the MPFC differs when observers infer that a character is happy versus sad, whether emotions are perceived in facial expressions, animated cartoons, or verbal narratives (Peelen et al, 2010; Skerry and Saxe, 2014). In contrast, cognitive aspects of the character's intentions, beliefs, and knowledge appear to be represented in the TPJ (Koster-Hale et al, 2017; Koster-Hale and Saxe, 2011).

Thus, Shamay-Tsoory's original insight, that damage to the MPFC could disproportionately impair understanding of others' emotions, has withstood the test of time. Still, many interesting questions remain open. First, regions of the MPFC have been implicated in a range of high-level cognitive abilities, including responding to objects or options that are valued (Kim et al, 2011), reacting to words or descriptions that are self-relevant (eg, one's own name, or self-descriptive traits) (Farb et al, 2007), and making moral judgments that integrate emotions and utilitarian outcomes (Shenhav and Greene, 2014). How are these different functions of the MPFC related to each other, and related to understanding others' emotions? In functional magnetic resonance imaging data, there are at least partially distinct subregions for thinking about self versus others (Jenkins et al, 2008), about values versus emotions (Skerry and Saxe, 2014), and about physical pain versus negative emotions (Kragel et al, 2018). However, both the detailed architecture and the more general computational theory of the MPFC remain open challenges.

A second open question concerns the effect of age at lesion onset. For many cortical regions, damage acquired earlier leads to less severe cognitive outcomes (Mosch et al, 2005; Trauner et al, 2013); for example, children who receive therapeutic hemispherectomies can learn to walk and talk even without their entire left hemisphere (Fritz et al, 2011). It is therefore a striking contrast that despite the migration of new neurons to the ventral MPFC in the first two years of life (Sanai et al, 2011), early damage to that region causes *more* severe impairments than late damage, particularly

in social and moral reasoning (Anderson et al, 1999; Taber-Thomas et al, 2014). Koenigs et al (2007) hypothesized that these results reveal a critical role for the ventral MPFC in emotional learning of social and moral knowledge during childhood. It would be fascinating to test whether understanding of others' emotions follows the same pattern of more severe impairments following early damage, and thus may depend on a similar mechanism of early learning.

Finally, future research should also investigate the interaction of brain anatomy and environmental experience in children's learning about others' emotions. There are initial hints that physically abused children have selectively smaller right ventral MPFC regions (Hanson et al, 2010), and that children raised in institutional settings or adopted in their preschool years tend to show impairments in understanding of emotion (Barone and Lionetti, 2012; Wismer Fries and Pollak, 2017). But the degree to which development of the ventral MPFC mediates the effects of early life adversity on emotion understanding and social and moral learning, and the implications of this mechanism for effective intervention, remain unknown.

In sum, 15 years after Shamay-Tsoory and colleagues' landmark article, this area of research is more active and more exciting than ever.

Heather L. Kosakowski, BA
hlk@mit.edu
Rebecca Saxe, PhD
saxe@mit.edu
Department of Brain and Cognitive Science
Massachusetts Institute of Technology
Cambridge, Massachusetts

REFERENCES

Anderson SW, Bechara A, Damasio H, et al. 1999. Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat Neurosci*. 2:1032–1037.

Baez S, Morales JP, Slachevksy A, et al. 2016. Orbitofrontal and limbic signatures of empathic concern and intentional harm in the behavioral variant frontotemporal dementia. *Cortex*. 75:20–32.

Barone L, Lionetti F. 2012. Attachment and emotional understanding: a study on late-adopted pre-schoolers and their parents. *Child Care Health Dev*. 38:690–696.

Farb NA, Segal ZV, Mayberg H, et al. 2007. Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Soc Cogn Affect Neurosci*. 2:313–322.

Fritz SL, Rivers ED, Merlo AM, et al. 2011. Intensive mobility training postcerebral hemispherectomy: early surgery shows best functional improvements. *Eur J Phys Rehabil Med*. 47:569–577.

Hanson JL, Chung MK, Avants BB, et al. 2010. Early stress is associated with alterations in the orbitofrontal cortex: a tensor-based morphometry investigation of brain structure and behavioral risk. *J Neurosci*. 30: 7466–7472.

Jacoby N, Bruneau E, Koster-Hale J, et al. 2016. Localizing pain matrix and theory of mind networks with both verbal and non-verbal stimuli. *Neuroimage*. 126:39–48.

Jenkins AC, Macrae CN, Mitchell JP. 2008. Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc Natl Acad Sci U S A*. 105:4507–4512.

Kim H, Shimojo S, O'Doherty JP. 2011. Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb Cortex*. 21:769–776.

Koenigs M, Young L, Adolphs R, et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*. 446:908–911.

Koster-Hale J, Richardson H, Velez N, et al. 2017. Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *Neuroimage*. 161:9–18.

Koster-Hale J, Saxe R. 2011. Theory of mind brain regions are sensitive to the content, not the structural complexity, of belief attributions. In: Carlson L, Hoelscher C, Shipley TF, eds. *Proceedings of the 33rd Annual Cognitive Science Society Meeting*. Austin, Texas: Cognitive Science Society; 3356–3361.

Kragel PA, Kano M, Van Oudenhove L, et al. 2018. Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. *Nat Neurosci*. 21:1–7.

Mosch SC, Max JE, Tranel D. 2005. A matched lesion analysis of childhood versus adult-onset brain injury due to unilateral stroke: another perspective on neural plasticity and recovery of social functioning. *Cogn Behav Neurol*. 18:5–17.

Peelen MV, Atkinson AP, Vuilleumier P. 2010. Supramodal representations of perceived emotions in the human brain. *J Neurosci*. 30:10127–10134.

Samson D, Apperly IA, Chiavarino C, et al. 2004. Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci*. 7:499–500.

Sanai N, Nguyen T, Ihrie RA, et al. 2011. Corridors of migrating neurons in the human brain and their decline during infancy. *Nature*. 478:382–386.

Saxe R, Kanwisher N. 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind." *Neuroimage*. 19:1835–1842.

Shamay-Tsoory SG, Aharon-Peretz J. 2007. Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*. 45:3054–3067.

Shamay-Tsoory SG, Tomer R, Berger BD, et al. 2005. Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cogn Behav Neurol*. 18:55–67.

Shenhav A, Greene JD. 2014. Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. *J Neurosci*. 34:4741–4749.

Skerry AE, Saxe R. 2014. A common neural code for perceived and inferred emotion. *J Neurosci*. 34:15997–15998.

Taber-Thomas BC, Asp EW, Koenigs M, et al. 2014. Arrested development: early prefrontal lesions impair the maturation of moral judgement. *Brain*. 137:1254–1261.

Trauner DA, Eshagh K, Ballantyne AO, et al. 2013. Early language development after peri-natal stroke. *Brain Lang*. 127:399–403.

Wismer Fries AB, Pollak SD. 2017. The role of learning in social development: illustrations from neglected children. *Dev Sci*. 20:1–11.