

Online Supplemental Box 3: Implications for Theories of Emotion

Our integration of the active inference framework^{1,2} and the structural model of corticocortical connections^{3,4} provides the key anatomical and functional basis for an evolutionary, constructionist theory of emotion⁵⁻⁸. This theoretical approach proposes that emotions emerge when people make situated meaning of the physical changes in their own bodies or in the actions they observe in others. The theory solves the emotion paradox⁸: studies of brain circuitry in non-human animals^{7,9-13}, electrical stimulation in humans^{13,14}, brain imaging^{13,15,16}, lesions in humans^{17,18}, autonomic function in humans^{8,13,19,20}, facial expressions in infants and adults²¹⁻²³ and vocal acoustics^{22,24} all fail to find specific and consistent biological and behavioral “fingerprints” that type each emotion category, such as anger, sadness, fear, and so on, and yet humans have no difficulty distinguishing instances of these categories in their own experiences and in their perceptions of others (in other humans, in non-human animals, and even in moving geometric figures²⁵). In the past, constructionist theories have appealed to socialization, culture, and context to explain how people make physical changes meaningful as emotional episodes, but have failed to provide a set of hypotheses that are consistent with the principles of evolutionary biology that can be tested at biological, psychological and social levels of analysis. Recently, a family of constructionist theories have developed that do just that²⁶. These theories build on older, less well-developed constructionist theories of William James²⁷ and Schachter & Singer²⁸ to hypothesize how the experience and perception of emotion emerges from the interplay of core brain systems that are not themselves specific to emotion.

According to the Conceptual Act Theory, for example, each instance of emotion is constructed within the brain’s functional architecture for creating situated

concepts, involving domain-general core systems for interoception, exteroception (vision, audition, etc.), categorization, attention and executive control, and so on^{5-8, 29}. The theory hypothesizes that the brain is a generative model of the world. Using knowledge from past experience and learning stored in its vast network of connections, the default mode network and interoceptive system initiate predictions that cascade through the rest of the brain to anticipate, categorize, and therefore explain incoming sensory input. This is the brain's default mode of interacting with the world. Thus, the brain is using stored knowledge of past experiences to predict incoming sensory inputs as a way of categorizing and explaining them, thereby constructing perception and action.

The brain constructs an instance of the emotion concept (e.g., an instance of anger) that best fits the immediate situation based on some Bayesian estimation of past experience, allowing an individual to prepare for situated action. This is called a situated conceptualization. Every waking moment of life, the brain is estimating the body's homeostatic requirements, weighing them against immediate resources, and making interoceptive predictions that become constituents of these situated conceptualizations. Your experience – be it a feeling of pent up, burning anger, defeated, anguished anger, or exuberant, energetic anger – has several properties: affective feelings of valence and arousal that derive from your brain's best guess about how your body should be responding in a particular situation, plus what these feelings mean relation to the rest of the situation at hand. The situated instance of an emotion concept that provides this meaning (as a prediction) itself is the homeostatic prediction that allow the body to act, as well as the interoceptive prediction that allow the brain to make sense of the sensory world it inhabits.

Without concepts, people are experientially blind. Sounds are just noise instead of words or music. Visual inputs are just blobs of light instead of objects and images. And interoceptive inputs are just bodily sensations or affective feelings of pleasure and displeasure instead of emotions like anger, sadness, fear, happiness, gratitude, etc. Indeed, making sense of the sensory world is a key feature of the active inference framework. If interoceptive predictions are a constituent part of an instance of an emotion concept, then when incoming interoceptive inputs are accounted for by the interoceptive prediction, the inputs are, in effect, ‘explained’ by the concept. And the goal of predictive coding, after all, is to allow the brain to ‘explain away’ (i.e., infer) the causes of the sensations it receives only from their effects^{1, 30}.

In the Conceptual Act Theory, emotions are constructions of the world, not responses to it. The brain’s default mode of interacting with the world is via continuous, intrinsic predictive activity that is more or less constrained by sensations coming from the world and the body. Categorization is achieved when sensory input is adequately predicted. And some of these mental events are experienced as emotions.

References

1. Friston, K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* **11**, 127-38 (2010).
2. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* **36**, 181-204 (2013).
3. Barbas, H. & Rempel-Clower, N. Cortical structure predicts the pattern of corticocortical connections. *Cereb Cortex* **7**, 635-46 (1997).
4. Barbas, H. General Cortical and Special Prefrontal Connections: Principles from Structure to Function. *Annu Rev Neurosci* (2015).
5. Barrett, L.F., Wilson-Mendenhall, C.D. & Barsalou, L.W. in *The psychological construction of emotion* (eds. Barrett, L.F. & Russell, J.A.) 83-110 (Guilford, New York, 2015).
6. Barrett, L.F. Psychological Construction: The Darwinian Approach to the Science of Emotion. *Emotion Review* **5**, 379-89 (2013).
7. Barrett, L.F. Emotions are real. *Emotion* **12**, 413-29 (2012).
8. Barrett, L.F. Solving the emotion paradox: categorization and the experience of emotion. *Pers Soc Psychol Rev* **10**, 20-46 (2006).
9. Barrett, L.F., Mesquita, B., Ochsner, K.N. & Gross, J.J. The experience of emotion. *Annu Rev Psychol* **58**, 373-403 (2007).
10. LeDoux, J. Rethinking the emotional brain. *Neuron* **73**, 653-76 (2012).
11. LeDoux, J.E. The slippery slope of fear. *Trends Cogn Sci* **17**, 155-6 (2013).
12. Gross, C.T. & Canteras, N.S. The many paths to fear. *Nat Rev Neurosci* **13**, 651-8 (2012).
13. Barrett, L.F. et al. Of Mice and Men: Natural Kinds of Emotions in the Mammalian Brain? A Response to Panksepp and Izard. *Perspect Psychol Sci* **2**, 297-311 (2007).
14. Guillory, S.A. & Bujarski, K.A. Exploring emotions using invasive methods: review of 60 years of human intracranial electrophysiology. *Soc Cogn Affect Neurosci* (2014).
15. Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E. & Barrett, L.F. The brain basis of emotion: a meta-analytic review. *Behav Brain Sci* **35**, 121-43 (2012).
16. Wager, T.D. et al. A bayesian model of category-specific emotional brain responses. *PLoS Comput Biol* **11**, e1004066 (2015).
17. Hurlmann, R. et al. Reduced 5-HT(2A) receptor signaling following selective bilateral amygdala damage. *Soc Cogn Affect Neurosci* **4**, 79-84 (2009).
18. Feinstein, J.S. et al. Fear and panic in humans with bilateral amygdala damage. *Nat Neurosci* **16**, 270-2 (2013).
19. Cacioppo, J.T., Berntson, G.G. & Larsen, J.T. The psychophysiology of emotion. *The handbook of emotion* (2000).
20. Kreibig, S.D. Autonomic nervous system activity in emotion: a review. *Biol Psychol* **84**, 394-421 (2010).
21. Gendron, M., Roberson, D., van der Vyver, J.M. & Barrett, L.F. Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion* **14**, 251-62 (2014).

22. Russell, J.A., Bachorowski, J.A. & Fernandez-Dols, J.M. Facial and vocal expressions of emotion. *Annu Rev Psychol* **54**, 329-49 (2003).
23. Barrett, L.F. Was Darwin Wrong About Emotional Expressions? *Current Directions in Psychological Science* **20**, 400-6 (2011).
24. Gendron, M., Roberson, D., van der Vyver, J.M. & Barrett, L.F. Cultural relativity in perceiving emotion from vocalizations. *Psychol Sci* **25**, 911-20 (2014).
25. Heider, F. & Simmel, M. An Experimental Study of Apparent Behavior. *The American Journal of Psychology* **57**, 243-59 (1944).
26. The psychological construction of emotion (Guilford Publications, 2015).
27. James, W. The Principles of Psychology: In Two Volumes (Martino Publishing, 1890).
28. Schachter, S. & Singer, J.E. Cognitive, social, and physiological determinants of emotional state. *Psychol Rev* **69**, 379-99 (1962).
29. Oosterwijk, S., Touroutoglou, A. & Lindquist, K.A. in *The Psychological Construction of Emotion* (eds. Barrett, L.F. & Russell, J.A.) (Guilford, New York, 2015).
30. Shipp, S., Adams, R.A. & Friston, K.J. Reflections on agranular architecture: predictive coding in the motor cortex. *Trends Neurosci* **36**, 706-16 (2013).