Trends in Cognitive Sciences

Letter

Seeing Other Minds in 3D

Rebecca Saxe[1,*]

To be kind or compassionate, or cruel or condescending, a social agent must understand what others are likely to want, feel, and choose. In this issue Tamir and Thornton [1] offer a powerful and parsimonious account of human social knowledge (how we represent the internal states of one another) and social prediction (how we anticipate dynamics in those internal states).

Tamir and Thornton [1] have identified three key dimensions that organize our understanding of other minds. These dimensions (glossed as valence, social impact, and rationality) can capture the similarities and differences between concepts of internal experiences (anger, loneliness, gratitude), and also between concepts of personalities (aggressive, introverted, agreeable). Most impressively, the three dimensions explain the patterns of hemodynamic activity in our brains as we consider these experiences [2] (Box 1). States such as anger and gratitude are invisible, but the patterns evoked in our brain as we think about them are as predictable by the model of Tamir and Thornton as the patterns evoked in our visual cortex when we look at chairs, bicycles, or pineapples are predictable by models of high-level vision [3]. Human social prediction follows the same dimensions: observers predict that transitions are more likely between states that are 'nearby' in this abstract 3D space [4]. Thus, we expect that a friend now feeling 'anxious' will be more likely to feel 'sluggish' than 'energetic' later.

In contrast to the traditional focus on basic emotions, beliefs, and desires, Tamir and Thornton [2] measured a broad set of 60 concepts ranging from 'drunkenness' and 'disarray' to 'skepticism' and 'self-pity'. Any future attempt to capture the structure of intuitive social knowledge should follow this lead. It is likely that even 60 concepts under-sample the space. For example, the mind versus body dimension, which drove robust neural activity in prior work [5,6], is the fourth and least important dimension in the current analysis. This dimension may reappear in future studies that include concepts of body states such as 'hunger', 'thirst', and 'pain'.

The success of the Tamir and Thornton model raises three challenges for future research: how to incorporate (i) context, (ii) content, and (iii) cultural origins into formal models of intuitive social knowledge.

First, the current model captures the structure and dynamics of internal experiences, in the absence of any external context. Thus, if we know only that a friend is feeling anxious, we predict that she will later feel more sluggish than energetic. However, if we know that the friend was anxious before giving a public lecture, and the lecture was a roaring success, then we predict that she will leave the stage feeling more energetic than sluggish. Nor is this example unusual. Emotions are evaluative perceptions of events that lead to actions – the temporal dynamics of emotions depend on what happened next, including what the person herself did (or could not do), what others did, and merely what befell [7].

Second, the current model captures concepts of 'states of mind', without any specific content. This approach contrasts with traditional 'mental states', which are composed of an attitude (or evaluative perception) towards a proposition (or content). In the classic Sally-Anne false belief task [8], for example, children are tested on whether they can track the content of Sally's belief (that her ball is in the box), and combine that with the content of her desire (to get the ball) so as to predict her action (looking in the box). All action predictions pose this problem. We cannot ask how a person's belief will influence her next action without knowing: her belief about what?

The challenge for future research will thus be to incorporate context and content into formal models of intuitive social knowledge and prediction. Unfortunately, the models that parsimoniously

---

Box 1. Social and Developmental Psychology Converge

Both social and developmental psychologists study mind perception. Social psychologists test how people form impressions and categorize others into groups. Developmental psychologists focus on a seemingly different problem: how children learn to infer an individual's specific desires, beliefs, and emotions in a particular context.

Tamir and Thornton [1,2] are social psychologists. Their model captures how we make computationally efficient judgments of a person, knowing only that she is (temporarily or persistently) in a state of contemplation, drunkenness, lust, or self-pity.

Inspired by developmental psychology, we recently [6,12] investigated neural patterns evoked when people infer specific beliefs, desires, and emotions from context. These independent approaches produced highly convergent results. The same cortical regions have been implicated in intuitive social knowledge, including the temporoparietal junction and medial prefrontal cortex. The same dimensions were reliably decoded from patterns of activity in those regions, including valence and rationality or justification [12]. Reproducible results from different laboratories, using different stimuli and derived from distinct traditions, are very encouraging. Intuitive social knowledge is apparently robustly organized in cortex at a spatial scale that is accessible by fMRI. For researchers interested in developing computational models of complex human cognition that can be validated with neural measurements, this topic is fertile ground.

Trends in Cognitive Sciences

CellPress
REVIEWS

capture personalities and states of mind likely cannot simply be extended. Vector space models cannot naturally encode logical or causal structure (context), and lack compositionally (content). The difference between feeling 'playful' versus 'serious' can be measured as a distance along one continuous dimension, but the difference between 'wanting the ball' versus 'wanting to go to the ball' cannot. Different formal structures will likely be required [6,9].

Finally, as Tamir and Thornton [2] note, the success of their model in capturing the social knowledge of Harvard undergraduates raises the question of cultural variability. Explicit theories about the mind vary substantially across the different cultures of the world [10]. Are dimensions of valence, social impact, and rationality implicit in descriptions of human experience in other cultures? Either way, these three dimensions leave out concerns with loyalty and divinity – that are both likely to be highly salient aspects of both

personality and states of mind for many humans [11].

In sum, we perceive other humans to have rich internal lives. When we consider a friend's feelings of anxiety, excitement, skepticism, or curiosity, diagnostic patterns of activity are evoked in a specific set of cortical regions in our brains [2]. These patterns of brain activity encode a core 3D space of internal experiences, as discovered by Tamir and Thornton [1]. How the more complex, causal, and compositional inferences we make about other minds can likewise be implemented in a neural system remains to be seen.

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 46-4019 MIT, 43 Vassar Street, Cambridge, MA 02139, USA

*Correspondence: saxe@mit.edu (R. Saxe).

References
1. Tamir, D.I. and Thornton, M.A. (2018) From social knowledge to social prediction: modeling the layers of social cognition. *Trends Cogn. Sci.*
2. Tamir, D.I. *et al.* (2016) Neural evidence that three dimensions organize mental state representation: rationality, social impact, and valence. *Proc. Natl. Acad. Sci.* 113, 194–199
3. Nishimoto, S. *et al.* (2011) Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646
4. Thornton, M.A. and Tamir, D.I. (2017) Mental models accurately predict emotion transitions. *Proc. Natl. Acad. Sci.* 114, 5982–5987
5. Bruneau, E. *et al.* (2013) How we know it hurts: item analysis of written narratives reveals distinct neural responses to others' physical pain and emotional suffering. *PLoS One* 8, e63085
6. Skerry, A.E. and Saxe, R. (2015) Neural representations of emotion are organized around abstract event features. *Curr. Biol.* 25, 1945–1954
7. Fontaine, J.J. *et al.*, eds (2013) *Components of Emotional Meaning: A Sourcebook*, Oxford University Press
8. Baron-Cohen, S. *et al.* (1985) Does the autistic child have a 'theory of mind'? *Cognition* 21, 37–46
9. Baker, C.L. *et al.* (2017) Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nat. Hum. Behav.* 1, 0064
10. Barrett, H.C. *et al.* (2016) Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proc. Natl. Acad. Sci.* 113, 4688–4693
11. Graham, J. *et al.* (2009) Liberals and conservatives rely on different sets of moral foundations. *J. Pers. Soc. Psychol.* 96, 1029
12. Koster-Hale, J. *et al.* (2017) Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *Neuroimage* 161, 9–18