

playing a role in VWM or claim that online perceptual processing exclusively involves early sensory cortex; neither does it make any specific claim about the nature of VWM contents stored in the PFC and PPC. It simply rejects sensory regions as being essential in VWM storage due to a lack of sufficient supporting evidence. The small behavioral distractor inference effect reported is consistent with VWM representation tracking perceptual representation but is agnostic to where it is stored in the brain.

Scimeca *et al.* further argue that the fact that top-down signals can sustain VWM representation in sensory cortex supports a sensory account of VWM storage [4]. It is unclear how the presence of such an interaction supports the necessity of sensory region in VWM storage, especially in situations in which a decreased VWM representation in sensory cortex does not correlate with a decrement in behavioral performance [7]. Similar evidence can also be found in monkey neurophysiology research [2]. The finding that top-down signals reinstate VWM representations in sensory cortex after they are disrupted further shows that sensory cortex relies on distractor-resilient VWM content stored elsewhere for its representation, making sensory representation and its interaction with top-down signals unnecessary in supporting VWM storage. Although much remains to be learned about the significance of the top-down signals on sensory regions, it likely aids probe detection at the end of the delay period, rather than directly participates in VWM storage [1].

Advantages of VWM Storage in Nonsensory Cortex

Although an overlap between perception and VWM storage in the sensory cortex can save cortical resources [3], it would subject the content of VWM to greater distraction by incoming visual input and make VWM less useful in real world vision

[1]. Moreover, a sensory storage would result in VWM content being scattered across different cortical regions, making an integrated representation less accessible and less able to be manipulated. Having a storage away from the sensory cortex enables better protection, amplification, integration, and manipulation of the VWM content, while freeing up sensory resources and allowing them to be continuously engaged in the processing of incoming visual input [1]. A nonsensory storage account does require PFC and PPC to be directly involved in the representation of visual information initially processed in sensory cortex. Indeed, PFC and in particular PPC have been shown to be capable of representing and holding a diverse array of visual information [10–12].

Concluding Remarks

While both commentaries bring forth additional data and arguments, they do not provide adequate support for the sensory account of VWM storage. Together, the evidence and critiques against the sensory account of VWM storage stand as originally presented [1,2]. That said, the continuous dialogs and debates among researchers are valuable, healthy and welcome forces that will continuously move the field forward.

¹Harvard University, 33 Kirkland Street, Room 780, Cambridge, MA, 02138, USA

*Correspondence: yaodaxu@wjh.harvard.edu (Y. Xu).
<http://dx.doi.org/10.1016/j.tics.2017.12.008>

References

- Xu, Y. (2017) Reevaluating the sensory account of visual working memory storage. *Trends Cogn. Sci.* 21, 794–815
- Leavitt, M.L. *et al.* (2017) Sustained activity encoding working memories: not fully distributed. *Trends Neurosci.* 40, 328–346
- Gayet, S. (2018) Visual working memory storage recruits sensory processing areas. *Trends Cogn. Sci.* 22, 189–190
- Scimeca, J.M. (2018) Reaffirming the sensory recruitment account of working memory. *Trends Cogn. Sci.* 22, 190–192
- Pierrot-Deselligny, C. *et al.* (2004) Eye movement control by the cerebral cortex. *Curr. Opin. Neurol.* 17, 17–25
- Beck, D.M. *et al.* (2001) Neural correlates of change detection and change blindness. *Nat. Neurosci.* 4, 645–650
- Bettencourt, K.C. and Xu, Y. (2016) Decoding under distraction reveals distinct occipital and parietal contributions

to visual short-term memory representation. *Nat. Neurosci.* 19, 150–157

- Vogel, E.K. and Machizawa, M.G. (2004) Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748–751
- Xu, Y. and Chun, M.M. (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95
- Christophel, T.B. *et al.* (2017) The distributed nature of working memory. *Trends Cogn. Sci.* 21, 111–124
- Jeong, S.K. and Xu, Y. (2016) Behaviorally relevant abstract object identity representation in the human parietal cortex. *J. Neurosci.* 36, 1607–1619
- Vaziri-Pashkam, M. and Xu, Y. (2017) Goal-directed visual processing differentially impacts human ventral and dorsal visual representations. *J. Neurosci.* 37, 8767–8782

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

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.

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 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.

¹Department 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).
<http://dx.doi.org/10.1016/j.tics.2018.01.003>

References

1. Tamir, D.I. and Thornton, M.A. (2018) Modeling the predictive social mind. *Trends Cogn. Sci.* 22, 201–212
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

Spotlight

How Primate Brains Vary and Evolve

Aida Gómez-Robles^{1,2,*}

Studies of brain evolution tend to focus on differences across species rather than on variation within

species. A new study measures and compares intraspecific variation in macaque and human brain anatomy to explore the effect that short-term diversity has on long-term evolution.

Evolutionary neuroscience focuses on diverse aspects of variation that are particularly relevant to human and nonhuman primate brain evolution. Plenty of studies have compared the evolution of total brain size and of the size of certain brain regions in different species. Very few studies, however, have measured and compared variation within species, which is commonly considered the raw material for evolution: the more variable a population is, the more options natural selection will have to choose from. Therefore, we can expect that increased variation within species will go hand in hand with rapid evolution. Conversely, there can be reasons to expect the opposite relationship: perhaps a high level of variation will hamper evolution because many individuals will show trait values that will substantially differ from optimal values. Of course, selection can influence intraspecific variation as well: selection can reduce variation when certain traits are substantially more advantageous than others or increase it when environments are also variable. Because of the lack of relevant comparative data, the exact relationship between human brain variation and evolution, certain aspects of which seem to have occurred at a particularly fast pace [1], is currently unknown.

In a recent publication in *Cerebral Cortex*, Croxson and colleagues [2] address this topic, testing the hypothesis that the human brain is more variable than the brains of other primates and that this has facilitated and accelerated brain evolution in our species. Their study quantifies and compares variability in human and macaque brains using structural

MRI scans. To do so, they produce species-specific templates and measure the deformations required to match each individual to the corresponding template. They separate their analyses into those corresponding to gray matter and white matter and use several analytical approaches and visualization tools to compare cortical regions, assessing differences between hemispheres.

As compelling as the reasoning laid out in their paper is, Croxson and colleagues' results are unexpectedly unclear. Once size variation is removed from analyses, overall human brains are not particularly more variable than macaque brains. It seems, however, that patterns of variation may differ between the two species, with evolutionarily recent association areas being more variable in humans. These results, however, are also less clear than I would have expected: while a positive association between variability and evolutionary expansion is observed in the temporal lobe, such an association is not found (or is extremely weak and not consistent across hemispheres) in the parietal and frontal lobes, which are also involved in higher cognitive functions (Figure 1).

So, how can we build on these results? First, increasing sample size, which amounts to only ten individuals per species in Croxson and colleagues' study, would be important to more accurately reflect intraspecific variation. Second, the comparison of only two species is necessarily very limited because it does not allow unequivocal clarification of the traits that are human-specific versus those that are macaque-specific. Assembling relatively large samples of *in vivo* MRI scans of nonhuman primates is painfully difficult, but an effort to increase these samples may pay off and show a clearer-cut scenario. Indeed, there are some resources that can be easily added to this study and to other, similar ones.