Dick, Frederic and Krishnan, S. (2018) What underlies the emergence of stimulus- and domain-specific neural responses? Commentary on Hernandez, Claussenius-Kalman, Ronderos, Castilla-Earls, Sun, Weiss, & Young (2018). Journal of Neurolinguistics 49 , pp. 235-236. ISSN 0911-6044.

What underlies the emergence of stimulus- and domain-specific neural responses?

Commentary on

Hernandez, Claussenius-Kalman, Ronderos, Castilla-Earls, Sun, Weiss, & Young (2018)

Frederic Dick[1,2] & Saloni Krishnan[3]

[1] Department of Psychological Sciences, Birkbeck, University of London

[2] Department of Experimental Psychology, UCL

[3] Department of Experimental Psychology, Oxford University

Hernandez et al (2018) provide a welcome historical perspective and synthesis of emergentist theories over the last decades. Particularly useful is the highlighting of theoretical differences regarding the etiology and character of seemingly domain- or category-specific neural response preferences. In an ideal world, competing theoretical accounts win or lose influence based on their ability to predict existing and new empirical results with the greatest accuracy, and the fewest tweaks to the model. It is equally important to make sure the empirical results the theories address are themselves robust, and that sources of variability in findings are accounted for within the theory.

One of the most-discussed cases of domain-specificity in the language literature is the putative 'Visual Word Form Area' in ventral occipitotemporal (VWFA), claimed to be a focal brain region that responds most strongly to written words versus all other stimuli (including part words and word strings), and is invariant to written form and retinal position (Cohen & Dehaene, 2003). However, there is considerable counter-evidence to each claim (many of these are summarized in Price & Devlin, 2003, 2004). The work of Vogel and colleagues (Vogel et al., 2011, 2012) is instructive here: in a series of fMRI studies, they show that a) activation for a two-word comparison task in the VWFA of English speakers was *weaker,* not stronger, than for performing the same task with two Amharic letter strings, line drawings, or consonants; b) activation in and directly around the VWFA coordinates is driven by stimulus complexity regardless of domain; c) activation in the putative VWFA is modulated by how similar (or 'groupable') the Amharic and consonant string pairs were, and d) there is weak connectivity between the putative VWFA and multiple cortical regions known to be involved in or necessary for reading. These and many other results (synthesized by Vogel, Petersen, & Schlaggar, 2014) suggest that the seeming domain-specificity of ventral occipitotemporal cortex initially reported for written words may reflect visual characteristics and processing demands that are associated with, but not limited to, word reading.

Such a process-based approach - also as mentioned by Hernandez et al - can shed light even on the archetype of category-specific neural responses: the so-called fusiform face area (Kanwisher, McDermott, & Chun, 1997). An elegant yet under-appreciated study by Haist, Lee, &

Stiles (2010) first replicated the strong preference in FFA for faces compared to diverse objects or wristwatches under passive viewing. They then showed that this neural face preference in the FFA *completely vanishes* when the same subjects see the same objects, but are asked to individuate them - activation for all three visual categories is equally strong. This finding suggests that the increased fMRI response to passively-viewed faces in this region may reflect the emergence of an automatized neural process of individuating complex, highly similar faces - perhaps the primary goal of 'real-world' face processing. These results from the reading and vision literature demonstrate that even the 'superstar' examples of domain-specific neural responses are much more complex - and less philosophically tidy - than the phenomena being accounted by at least some emergentist-style accounts.

A final case of category- or domain-specific neural responses is that of multiple putative 'speech-sensitive' regions in the left and right superior temporal gyrus and sulcus (STG/STS, reviewed by Price, Thierry, & Griffiths, 2005). There is now little serious discussion of a region dedicated to speech processing, but there is also little question that passive listening to speech (even backward speech) tends to drive STS activation in particular much more strongly than other stimulus classes, such as diverse environmental sounds, even those containing non-linguistic vocalizations (review and analysis in Leech & Saygin, 2011). What drives the emergence of such strong and seemingly stimulus-class-specific response preferences? Dick, Lee, Nusbaum, & Price (2011) suggested that such dramatic response selectivity might emerge as a result of long-term, intensive experience in both perceiving *and producing* a sound class. They tested this hypothesis by comparing passive neural responses in professional violinists and actors when they listened to solo violin and dramatic speech excerpts, with the expectation that violinists would show substantially increased fMRI activation for violin music in multiple 'speech-selective' regions when compared to actors. This was indeed the case, even in posterior STS regions found to be strongly speech-selective in previous studies (Tervaniemi et al., 2006). Moreover, Dick et al. (2011) found multiple audiomotor regions (right STG, bilateral speech-sensitive premotor cortex, right cerebellum) with strong soundclass—selective responses that were driven by subjects' experience, where actors showed greater activation for dramatic speech than violin, but violinists showed greater activation for violin than speech.

Parallel visuomotor expertise effects have also been observed: a classic experiment by Calvo-Merino et al., (2005) showed that long-term experience with a particular style of dance movement, ballet or capoiera, was associated with recruitment of premotor and inferior parietal cortex during the perception of these actions. Krishnan et al (under review) demonstrate that brain activity for auditory perception is similarly modulated by instrument-specific long-term experience. Non-musicians, beatboxers, and guitarists passively listened to novel beatbox and guitar music pieces, which evoked neural activity in dorsal stream regions such as the left inferior frontal gyrus and left inferior parietal cortex only in the performers, *and only when they listened to a style of music they had experience producing.* The most plausible explanation for such instrument-selective activity (see also Margulis, Mlsna, Uppunda, Parrish, & Wong, 2009) is that internal motor models are spontaneously produced when participants listen to sounds they are expert at producing, reflecting a consequence of the unique long-term associations between perception and production systems that develop in visuo- and audio-motor experts (Cook et al., 2014).

The evidence from neuroimaging studies favors theories that can organically account for how an organism develops graded, process-sensitive neural response patterns to 'special' stimulus categories or domains - but does not ignore the fact that the same 'stimulus-sensitive' regions also keep their neural day jobs.

References

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. Cerebral Cortex 15(8), 1243–1249.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. NeuroImage, 22(1), 466–476.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. Behavioral and Brain Sciences, 37(02), 177–192.

Dick, F., Lee, H. L., Nusbaum, H., & Price, C. J. (2011). Auditory-motor expertise alters “speech selectivity” in professional musicians and actors. Cerebral Cortex, 21(4), 938–948.

Haist, F., Lee, K., & Stiles, J. (2010). Individuating faces and common objects produces equal responses in putative face-processing areas in the ventral occipitotemporal cortex. Frontiers in Human Neuroscience, 4, 181.

Hernandez, A.E., Claussenius-Kalman, H. L. Ronderos, J. Castilla-Earls, A. P., Sun, L., Weiss, S. D. and Young, D. R. (in press). Neuroemergentism: A Framework for Studying Cognition and the Brain. Journal of Neurolinguistics.

Krishnan, S., Lima, C., Evans, S., Chen, S., Guldner, S., Manly, T., & Scott, S. (in review). Beatboxers and guitarists engage sensorimotor regions selectively when listening to the instruments they can play.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17(11), 4302–4311.

Leech, R., & Saygin, A. P. (2011). Distributed processing and cortical specialization for speech and environmental sounds in human temporal cortex. Brain and Language, 116(2), 83–90.

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. M. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. Human Brain Mapping, 30(1), 267–275.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. NeuroImage, 19(3), 473–481.

Price, C. J., & Devlin, J. T. (2004). The pro and cons of labelling a left occipitotemporal region: "the visual word form area". NeuroImage, 22(1), 477–479.

Tervaniemi, M., Szameitat, A. J., Kruck, S., Schröger, E., Alter, K., De Baene, W., & Friederici, A. D. (2006). From air oscillations to music and speech: functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. Journal of Neuroscience, 26(34), 8647–8652.

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2012). The left occipitotemporal cortex does not show preferential activity for words. Cerebral Cortex, 22(12), 2715-2732.

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2014). The VWFA: it's not just for words anymore. Frontiers in human neuroscience, 8, 88.