Naturalistic stimuli reveal a dominant role for agentic action in visual representation

James V. Haxby [a,*], M. Ida Gobbini [b,c], Samuel A. Nastase [d]

[a] Center for Cognitive Neuroscience, Dartmouth College, Hanover, NH, USA
[b] Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale, University of Bologna, Italy
[c] Cognitive Science Program, Dartmouth College, Hanover, NH, USA
[d] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA

A B S T R A C T

Naturalistic, dynamic movies evoke strong, consistent, and information-rich patterns of activity over a broad expanse of cortex and engage multiple perceptual and cognitive systems in parallel. The use of naturalistic stimuli enables functional brain imaging research to explore cognitive domains that are poorly sampled in highly-controlled experiments. These domains include perception and understanding of agentic action, which plays a larger role in visual representation than was appreciated from experiments using static, controlled stimuli.

1. Introduction

The advent of digital video and programs for controlling presentation of digital video made it feasible to present dynamic, naturalistic movies in functional brain imaging experiments. Pioneering studies in 2004 (Bartels and Zeki, 2004; Hasson et al., 2004) have led to widespread use of such stimuli, producing clear evidence of their utility for evoking reliable, information-rich patterns of brain activity over a larger extent of cortex than is activated by more controlled experiments. Movies also provide a rich context with narrative structure (Hasson et al., 2008a, 2015; Chen et al., 2017), better hold attention (Hasson et al., 2008b), and enhance subject compliance (Vanderwal et al., 2015).

Whereas controlled experiments minimize extraneous information in stimuli and tasks, experiments with naturalistic stimuli better simulate the full richness of natural visual and auditory experience. Movies sample a broad range of brain states and engage multiple perceptual and cognitive systems in parallel. Even within a sensory modality, such as vision, different types of information are layered and simultaneously present in natural movies. The rich, layered information in movies has allowed concurrent modeling of multiple stages of perceptual processing (Nishimoto et al., 2011; Huth et al., 2012; Güçlü and van Gerven, 2017). Broader sampling and efficient engagement of multiple systems motivated the use of movie-viewing data as the basis for developing algorithms for aligning functional anatomy (Conroy et al., 2009, 2013; Sabuncu et al., 2010; Haxby et al., 2011; Chen et al., 2015; Guntupalli et al., 2016, 2018). Researchers have begun to dissociate overlapping

neural representations in naturalistic paradigms using analytic methods such as multivariate pattern classification (MVPC; Haxby et al., 2001, 2014), representational similarity analysis (Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013), and forward encoding models (Naselaris et al., 2010; Nunez-Elizalde et al., 2019). Critically, the relative contributions of different types of information are more faithfully represented in naturalistic stimuli than in controlled experimental manipulations.

We argue here that studies of brain activity evoked by viewing a naturalistic, dynamic movie better reflect the statistics of natural viewing in a complex, cluttered, changing, and continuous visual environment. Using naturalistic, dynamic stimuli has the potential to lead to surprising new insights about how neural resources for visual perception are structured and allocated. We refer here to naturalistic, dynamic stimuli as those that present visual episodes, sometimes accompanied by a soundtrack, with the complexity of natural scenes. We examine the use of naturalistic audiovisual clips or movies and distinguish these from still images and from highly-controlled or schematic videos (e.g. point-light displays or videos of isolated body parts performing simple actions with no context), which have an intermediate status between non-naturalistic stimuli and naturalistic, complex videos.

We focus here on one cognitive domain – the representation of action by animate agents. We present evidence that agentic actions in naturalistic movies are represented in patterns of neural activity across a surprisingly wide expanse of cortex. Representation of agentic action appears to play a more dominant role in the neural representation of vision than perception of form, perhaps even in the ventral visual

---

pathway, which was previously thought to be dedicated to form recognition. We contend that the relative importance of action representation was underappreciated due to the field's reliance on tightly-controlled static stimuli, which were often deliberately engineered to remove any such agentic information. We review some earlier evidence for agentic action representation in ventral temporal (VT) cortex that is often overlooked or minimized and discuss how the apparent conflict with evidence for the central role of the ventral visual pathway in form recognition can be reconciled. We go on to discuss how the use of naturalistic, dynamic stimuli can shed new light on the representation of agentic actions.

2. Representation of agentic action in natural vision

Nearly all movies depict people or other agents, such as animals and robots, performing actions. In the following, we walk through a series of studies providing evidence—sometimes incidentally—for the primacy of action representation in natural vision. In a recent study (Nastase et al.,

2017), we used RSA to study the geometry of representations evoked by viewing naturalistic movie clips of behaving animals. We analyzed the data with separate predictors of representational geometry based on the similarities of action categories (eating, fighting, running, swimming), independent of the species of animal performing the actions, and on similarities of taxonomic categories (primates, ungulates, birds, reptiles, insects), independent of the action being performed. The results showed that the geometry of representations evoked by videos of behaving animals was dominated by species-invariant behaviors across a remarkably broad expanse of cortex, including most of extrastriate visual cortex in both the dorsal and ventral pathways, as well as parietal, motor, and premotor cortices (Fig. 1A). By contrast, representational geometry that correlated with behavior-invariant taxonomic categories was more restricted to the lateral and ventral occipitotemporal cortices of the ventral visual pathway. Surprisingly, even the representational geometry in VT cortex was dominated by animal behavior rather than by taxonomic form, accounting for 2.5 times more variance (Fig. 1B). This was
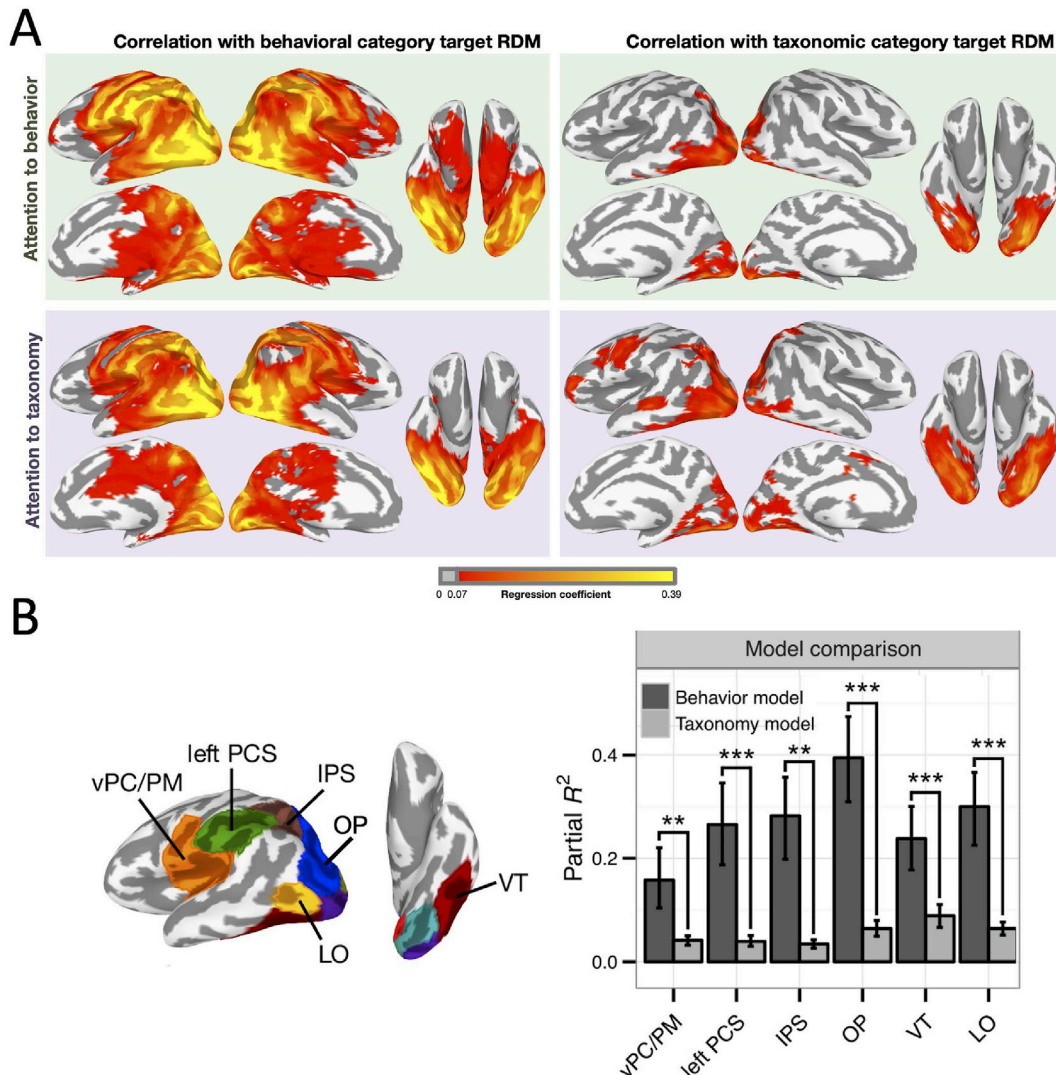

Fig. 1. The dominant role of action in neural representation of movie clips of behaving animals. A. In a human fMRI study of representation of animal behavior and taxonomy (Nastase et al., 2017), the representational geometry of responses to video clips of behaving animals was dominated by geometry that reflected action categories, as compared to geometry that reflected taxonomic categories. A regression model was constructed to evaluate the relative contributions of representational geometries capturing action categories and taxonomic categories. This model was applied using a searchlight analysis, revealing the broader extent of cortex with stronger representation of actions, including occipital, ventral and lateral temporal, posterior parietal, and premotor cortices. This dominance was evident whether the participants attended to behavior or to taxonomic classes. B. Variance accounted for by models of representational geometry capturing action categories and taxonomic categories, averaged across tasks. Action representation overwhelmingly dominated taxonomic form representation in all dorsal and ventral extrastriate, posterior parietal, and premotor cortical ROIs (vPC/PM: ventral pericentral and premotor; PCS: left postcentral sulcus; IPS: intraparietal sulcus; OP: occipitoparietal; VT: ventral temporal, LO: lateral occipital). Adapted from Nastase et al. (2017).

true even using a cross-classification analysis where action classification was cross-validated across taxonomic categories (and stimuli; Nastase et al., 2017, Supplementary Figure 3). Note that the dominant role of action in the geometry of responses to these video clips was independent of the task performed by participants (one-back repetition detection based on action or taxonomic category). The dominance of action representation was evident for both tasks individually (Fig. 1A), and the results in Fig. 1B were averaged over tasks. In other words, differences were found in analyses that distinguished the action- and taxonomy-related representational content in the same response patterns to the same trials and, therefore, cannot be attributed to differences in task or eye movements. This unexpected result contrasts with previous work based on experiments using still images that emphasized the role of VT cortex in the perception of form-related properties (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Ishai et al., 1999; Haxby et al., 1994, 2001, 2011; Connolly et al., 2012, 2016; Konkle and Oliva, 2012; Kravitz et al., 2013; Grill-Spector and Weiner, 2014; Sha et al., 2015; among many others).

In an fMRI study of monkeys viewing naturalistic movies, Russ and Leopold (2015) used a forward encoding analysis to predict voxelwise responses based on low-level visual features, namely luminance, contrast, and motion energy, and high-level features, namely the presence of faces, extremities, and animals. The results showed that motion energy was the strongest predictor of activity in lateral inferior temporal (IT), anterior superior temporal sulcal (STS), and lateral prefrontal cortices, in addition to visual motion areas such as MT, MST, and FST (Fig. 2A). This effect was not due to simple motion energy but, more specifically, to biological motion, as motion only weakly predicted responses to movies without animals. Surprisingly, motion energy was a stronger predictor of responses in the temporal face patches (Tsao et al., 2008) than was the presence of faces, suggesting that even these areas may be more involved in the representation of biological motion than the representation of facial form in natural viewing.

Related movie-viewing fMRI studies in humans (Huth et al., 2012) also demonstrated that the principal component that captures the largest share of response variance distinguishes between semantic labels for entities and actions with high and low movement energy (Fig. 2). This principal component was a strong, positive predictor of neural responses across mostly the same cortices that Nastase et al. (2017) found to contain representation of taxonomy-invariant animal behaviors. Even

after regressing out this dimension, semantic qualities such as animacy and social interaction still dominate the residual response space. Even in the FFA, the semantic qualities of animacy, social interaction, and mobility play a surprisingly important role in the response space (Çukur et al., 2013).

3. Convergent evidence from non-naturalistic paradigms

Earlier evidence also supports the hypothesis that VT cortex plays a role in the representation of perceived agentic action (Grosbas et al., 2012). Controlled experiments dissociated responses to agentic action from responses to animate form using stimuli that were dynamic but not naturalistic and depict biological motion, social interaction, or goal-directed action in the absence of animate form. These studies consistently found that activity in VT cortex is modulated by perception of agentic action. Point-light displays of human body and human hand motion (Bonda et al., 1996; Grossman and Blake, 2002; Beauchamp et al., 2003; Peelen et al., 2006; Gobbini et al., 2007) consistently evoke stronger responses in VT cortex than point-light displays of scrambled motion and tool motion (Fig. 3A-D). Animations showing social interactions depicted with moving, simple geometric figures (Heider and Simmel, 1944) consistently evoke stronger responses in VT cortex than the same geometric figures with movement driven by external physical forces (Castelli et al., 2000; Schultz et al., 2003; Gobbini et al., 2007, Fig. 3E-H). Other studies have used videos of moving robots performing meaningful actions (Gobbini et al., 2011; Shultz and McCarthy, 2011) and show that the representation of agentic action in VT cortex does not require the agent to have a naturalistic animate form. Shultz and McCarthy (2011) showed that goal-directed actions performed by inanimate robots, as compared to non-goal-directed actions, evoke strong activity in both ventral and lateral temporal areas, as well as parietal and premotor cortices, even when the robots performing the goal-directed actions bear little resemblance to naturalistic animate forms. These results suggest that action representation in VT may be related to the meaning and goals that motivate agentic actions, analogous to VT form representation that is related to the semantics of categories and animacy (Connolly et al., 2012, 2016; Grill-Spector and Weiner, 2014; Sha et al., 2015; Thorat et al., 2019).

More recent studies (Wurm and Lingnau, 2015; Wurm et al., 2016) have controlled for the effect of animate form by keeping the animate
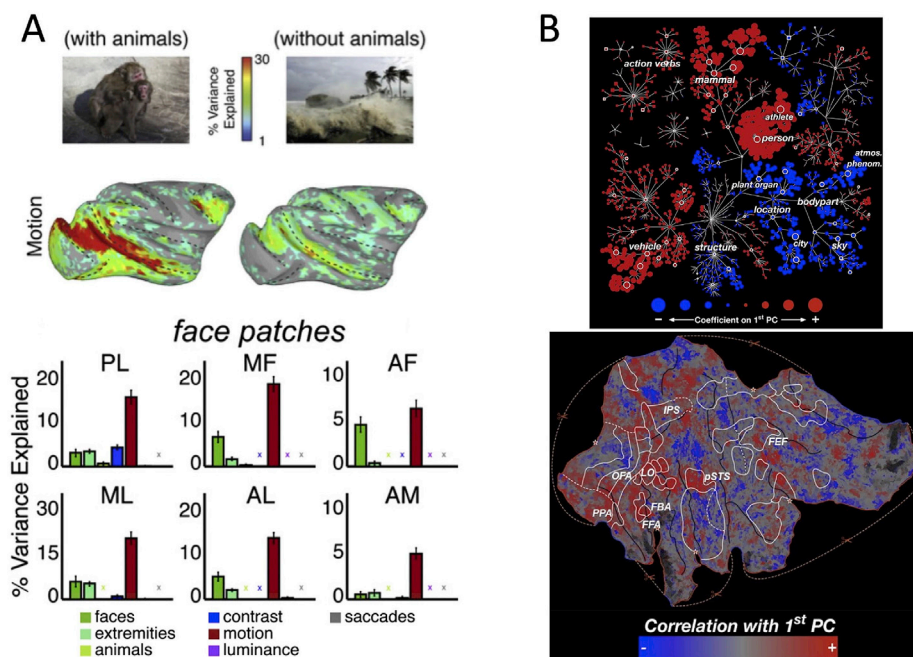

Fig. 2. A. In an fMRI study of monkeys viewing movies, motion energy was the best predictor for widespread neural responses, and this effect was limited to scenes with animal motion (Russ and Leopold, 2015). The stronger response to motion as compared to the presence of animate form was even evident in temporal face patches, cortical fields previously thought to be dominated by responses to the presence of faces (from Russ and Leopold, 2015). B. In an analysis of the semantic features that predict responses to natural movies (Huth et al., 2012), the principal component that explained the most variance (1st PC) reflected a contrast between responses to frames labeled with agents that move (e.g. people, animals), vehicles, and action verbs, and to frames labeled with names of stationary objects. This action-related semantic principal component was a positive predictor of responses in a broad extent of cortex that included dorsal and ventral extrastriate, posterior parietal, and premotor cortices (IPS: intraparietal sulcus, FEF: frontal eye field, FBA: fusiform body area, FFA: fusiform face area, OFA: occipital face area, PPA: parahippocampal place area, LO: lateral occipital, pSTS: posterior superior temporal sulcus). Adapted from Huth et al. (2012).
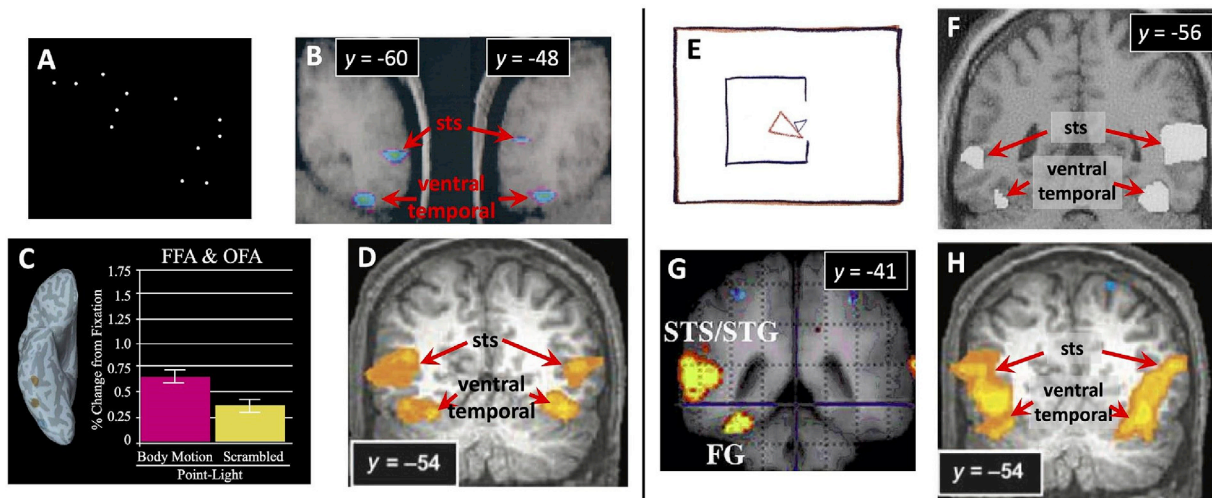
Fig. 3. Stronger responses in VT cortex and the STS to dynamic stimuli depicting agentic action in the absence of naturalistic animate form. Point-light displays of body actions (A, mid-point in a handspring sequence) evoke stronger responses to point-light displays of scrambled motion. Converging results from studies by (B) Bonda et al. (1996; PET-rCBF; these images are for hand motion, viewing body motion produced similar results; Copyright 1996 Society for Neuroscience), (C) Grossman and Blake (2002; fMRI), and (D) Gobbini et al. (2007; fMRI). See also Beauchamp et al. (2003) and Peelen et al. (2006). Responses measured with fMRI to social interactions depicted by moving, rigid geometric shapes (E, frame from a sequence depicting the big triangle coaxing the little triangle to go outside) evoke stronger responses in VT and STS cortices than sequences showing movements driven by physical forces. Converging results from studies by (F) Castelli et al., 2000, (G) Schultz et al. (2003), and (H) Gobbini et al. (2007).

form constant across a limited set of different actions performed in a single, controlled environment. The type of hand action could be decoded from patterns of brain activity in a set of cortices similar to those found in Nastase et al. (2017). The authors emphasize the role of lateral occipitotemporal cortices in action representation (Lingnau and Downing, 2015), along with parietal and premotor cortices, but their results also show decoding of perceived hand actions in VT cortex.

Studies of the effects of brain lesions in the ventral pathway show that VT lesions can impair perception of motion coherence (Gilaie-Dolan et al., 2013), structure from motion (Cowey and Vaina, 2000; Blanke et al., 2007), and biological motion in point-light displays (Cowey and Vaina, 2000; Gilaie-Dolan et al., 2015). Unilateral VT lesions can cause mild impairment of biological motion perception, especially in patients with prosopagnosia, and patients with lesions in the posterior superior sulcus and premotor cortex show more severe impairments (Gilaie-Dolan et al., 2015). Individuals with congenital prosopagnosia often show severe impairment of perception of biological motion in point-light displays (Lange et al., 2009). These studies suggest that action representation observed in VT using neuroimaging likely plays a causal, albeit complex, role in the distributed system for perception of agentic action.

4. Reconciling action and form representation

The representation of action in VT cortex appears at first glance to run counter to studies that show that moving animate stimuli generally do not evoke stronger responses than do still images of animate stimuli in VT cortex, whereas such augmentation of activity is seen in lateral temporal areas in the superior temporal sulcus and middle temporal gyrus (Beauchamp et al., 2002; Fox et al., 2009; Haxby and Gobbini, 2012; Pitcher et al., 2011). These studies all used still and video images of animate forms, whereas the studies cited above either factored out the activity evoked by animate form by using dynamic stimuli that are stripped of animate form or investigated the representation of information conveyed by different actions with equivalent animate form. The studies that showed no difference between responses in VT cortex to moving versus still images generally use videos that add little action information that isn't implied by the photographic still images of the same animate forms. By contrast, the studies that dissociate the

representation of form from the representation of action all use stimuli in which information about action is revealed or specified only by the movement. For example, the still images used in Beauchamp et al. (2002), Pitcher et al. (2011)), and Haxby and Gobbini (2012) were all taken from the video stimuli used in the moving faces and bodies conditions. For point-light and social interaction stimuli, on the other hand, the presence and nature of agentic action is only evident in the dynamic stimuli. In the richer naturalistic stimuli used in Nastase et al. (2017), the videos convey information about types of different agentic behaviors that are independent of taxonomic categories. In more controlled videos of hand actions (Wurm and Lingnau, 2015; Wurm et al., 2016), the same animate forms (hands) in the same controlled environments perform different actions that can be decoded from activity evoked in VT cortex. The studies that find no increase in response when movement is added to images of animate stimuli, therefore, mostly show that VT areas are not sensitive to simple motion, unlike motion vision areas in hMT+ and pSTS, but an extensive literature consistently shows that VT areas, nonetheless, represent information about agentic action.

It is firmly established that VT cortex plays a central role in the perception of object and animate form. Indeed, the original work on two visual cortical pathways dissociated perception of spatial relations and action in the dorsal spatial vision pathway from perception of static form in the ventral object vision pathway (Ungerleider and Mishkin, 1982; Mishkin et al., 1983). This distinction has been repeated in the initial imaging studies of this hypothesis (Haxby et al., 1991, 1994; Ungerleider and Haxby, 1994), in Goodale and Milner's reformulation of the two visual systems hypothesis (Goodale & Milner, 1992), and in the application of this framework to the face perception system (Haxby et al., 2000; Gobbini and Haxby, 2007; Haxby and Gobbini, 2012). Neuroimaging studies of the functional organization of the ventral visual pathway have similarly emphasized the representation of form, as exemplified in category-selective areas (Puce et al., 1996; Kanwisher et al., 1997; McCarthy et al. 1997; Epstein and Kanwisher, 1998; Haxby et al., 1999; Ishai et al., 1999; Downing et al., 2001; Grill-Spector and Weiner, 2014), the distinction between animate and inanimate forms (Kiani et al., 2007; Kriegeskorte et al., 2008; Grill-Spector and Weiner, 2014), a continuum for animacy or agency (Connolly et al., 2012, 2016; Sha et al., 2015; Thorat et al., 2019), a continuum for real world size (Konkle and Oliva, 2012), and high-dimensional models of the

representational space (Haxby et al., 2001, 2011; Guntupalli et al., 2016).

The evidence that neural activity in VT cortex also represents information about agentic action doesn't contradict evidence for the central role of VT cortex in form perception. The evidence reviewed here indicates that VT cortex represents information about agentic action concurrently with information about form. In fact, concurrent representations of form and action in VT cortex appear to be related. The topographic organization for the representation of form in VT cortex is strongly related to representation of potential for agentic action. This dimension of the topography for representation of still images has been labeled the "animacy continuum," with a lateral-to-medial gradient for agency and a posterior-to-anterior gradient for a transition from a representation of this continuum based on static visual features to one based on semantic features (Connolly et al., 2012, 2016; Sha et al., 2015; Thorat et al., 2019).

The neuroimaging evidence for agentic action representation in VT cortex is not new but, rather, begins (Bonda et al., 1996) at the same time as the discovery of a face-selective cortical field in the fusiform gyrus (Puce et al., 1996; McCarthy et al. 1997; Kanwisher et al., 1997). The significance of this extensive and consistent literature has been overshadowed by attention to the role of the pSTS and middle temporal gyrus in the same studies. Pre-existing assumptions based on knowledge about motion vision and form vision in the dorsal and ventral pathways were reinforced by the studies that showed no increase in activity for moving relative to static animate stimuli. The emphasis on the role of VT cortex in the representation of animate and object form appears to have effectively excluded consideration of the possibility that it may also participate in other visual functions. Concomitantly, evidence for the roles played by the STS/MTG, parietal, and premotor cortices in action perception steered many away from considering that VT cortex may also be part of a broadly distributed system for action perception.

Studies using naturalistic, dynamic stimuli can serve to broaden our understanding of how the visual system disentangles and interprets different types of information that are layered in complex, natural stimuli. By using analytic methods that can disentangle the roles the same cortical field plays in the representation of information about form and agentic action, brain activity evoked by naturalistic, dynamic stimuli can more fully reveal the complexity of visual function and how it is adapted to the statistics of the natural vision. Multivariate pattern classification (Haxby et al., 2001, 2011, 2014; Haxby, 2012) can decode information about action that is invariant across variations in the form of the agent and vice versa. Representational similarity analysis (Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013) can examine the representational geometry of population responses to different actions, factoring out the geometry of responses to different forms, and vice versa. Forward encoding analysis (Naselaris et al., 2010; Huth et al., 2012; Nunez-Elizalde et al., 2019) can estimate the extent to which action features and form features independently predict responses.

Existing results using naturalistic, dynamic movies suggest that the representation of agentic action is more extensive and stronger than the representation of form in natural vision, even in VT cortex (Russ and Leopold, 2015; Nastase et al., 2017). This unexpected result requires further corroboration and elaboration with different stimuli, different imaging modalities, and converging analytic methods. For example, studies of the geometry of representations for a rich variety of human actions in naturalistic contexts may produce fresh insights. Perception of action may play a disproportionate role in adaptive behavior where rapid and accurate assessment of the actions of other people and animals can be critical for achieving goals and avoiding harm. The dominant role of action perception also may be a function of the statistics of natural viewing and the sensitivity of neural systems to change. Features that are invariant to movement and support recognition of form don't change, by definition, whereas action perception is driven by dynamic features that change continuously. Consequently, whereas representations of form – objects, animals, and people – can be established quickly and thereafter remain constant, requiring no further updating, perception of actions unfolds over time, is continuously changing, and requires constant processing. Imaging methods with higher temporal resolution, such as MEG (cf. Cichy et al., 2014, 2017; Isik et al., 2018) may be able to resolve the roles of transient and sustained population responses for actions and forms, although the study of brain activity evoked by naturalistic stimuli with high temporal resolution electrophysiology presents many, as yet unresolved, challenges (Park et al., 2017).

5. Summary and conclusion

Naturalistic, dynamic stimuli evoke strong neural responses over a large extent of cortex. The information in a natural audiovisual movie engages multiple neural systems in parallel to process different types of information that are layered and simultaneously present. Studies using naturalistic, dynamic movies reveal that agentic action plays a surprisingly dominant role in visual representation. Previous studies with still images concentrated on the perception of the form of objects, animals, and people, but the perception of action appears to play a decidedly larger role in neural representation during natural vision.

Paradigm shifts in science are often driven by technical developments, such as the introduction of naturalistic, dynamic stimuli in cognitive neuroscience. These developments not only change how experiments are conducted but also reveal gaps in our collective knowledge and motivate fundamentally novel questions. Visual experiments with still images revealed mechanisms for perception of object and animate form but were not well-suited for study of the perception of action. The use of naturalistic, dynamic stimuli, in which form and action are simultaneously present, and analytic methods that can distinguish their overlapping representations, will allow us to better appreciate the powerful role of observed action in vision.

CRediT authorship contribution statement

James V. Haxby: Conceptualization, Writing - original draft. **M. Ida Gobbini:** Writing - review & editing. **Samuel A. Nastase:** Writing - review & editing.

References

Bartels, A., Zeki, S., 2004. Functional brain mapping during free viewing of natural scenes. Hum. Brain Mapp. 21, 75–85. https://doi.org/10.1002/hbm.10153.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34, 149–159. https://doi.org/10.1016/S0896-6273(02)00642-6.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2003. FMRI responses to video and point-light displays of moving humans and manipulable objects. J. Cogn. Neurosci. 15, 991–1001. https://doi.org/10.1162/089892903770007380.

Blanke, O., Brooks, A., Mercier, M., Spinelli, L., Adriani, M., Lavanchy, L., Safran, A.B., Landis, T., 2007. Distinct mechanisms of form-from-motion perception in human extrastriate cortex. Neuropsychologia 45, 644–653. https://doi.org/10.1016/j.neuropsychologia.2006.07.019.

Bonda, E., Petrides, M., Ostry, D., Evans, A., 1996. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J. Neurosci. 16, 3737–3744. https://doi.org/10.1523/jneurosci.16-11-03737.1996.

Castelli, F., Happé, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12 (3), 314–325. https://doi.org/10.1006/nimg.2000.0612.

Chen, J., Leong, Y.C., Honey, C.J., Yong, C.H., Norman, K.A., Hasson, U., 2017. Shared memories reveal shared structure in neural activity across individuals. Nat. Neurosci. 20, 115–125. https://doi.org/10.1038/nn.4450.

Chen, P.-H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., Ramadge, J., 2015. A reduced dimension fMRI shared response model. In: Advances in Neural Information Processing Systems 28, pp. 460–468. http://papers.nips.cc/paper/5855-a-reduced-dimension-fmri-shared-response-model.

Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17, 455–462. https://doi.org/10.1038/nn.3635.

Cichy, R.M., Khosla, A., Pantazis, D., Oliva, A., 2017. Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. Neuroimage 153, 346–358. https://doi.org/10.1016/j.neuroimage.2016.03.063.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32, 2608–2618. https://doi.org/10.1523/jneurosci.5547-11.2012.

Connolly, A.C., Sha, L., Guntupalli, J.S., Oosterhof, N.N., Halchenko, Y.O., Nastase, S.A., Visconti di Oleggio Castello, M., Abdi, H., Jobst, B., Gobbini, M.I., Haxby, J.V., 2016. How the human brain represents perceived dangerousness or "predacity" of animals. J. Neurosci. 36, 5373–5384. https://doi.org/10.1523/jneurosci.3395-15.2016.

Conroy, B.R., Singer, B.D., Haxby, J.V., Ramadge, P.R., 2009. MRI-based inter-subject cortical alignment using functional connectivity. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (Eds.), Advances in Neural Information Processing Systems, vol. 22, pp. 378–386. https://papers.nips.cc/paper/3764-fmri-based-inter-subject-cortical-alignment-using-functional-connectivity.

Conroy, B.R., Singer, B.D., Guntupalli, J.S., Ramadge, P.R., Haxby, J.V., 2013. Inter-subject alignment of human cortical anatomy using functional connectivity. Neuroimage 81, 400–411. https://doi.org/10.1016/j.neuroimage.2013.05.009.

Cowey, A., Vaina, L.M., 2000. Blindness to form from motion despite intact static form perception and motion detection. Neuropsychologia 38, 566–578. https://doi.org/10.1016/S0028-3932(99)00117-7.

Çukur, T., Huth, A.G., Nishimoto, S., Gallant, J.L., 2013. Functional subdomains within human FFA. J. Neurosci. 33, 16748–16766. https://doi.org/10.1523/jneurosci.1259-13.2013.

Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473. https://doi.org/10.1126/science.1063414.

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392, 598–601. https://doi.org/10.1038/33402.

Fox, C.J., Iaria, G., Barton, J.J.S., 2009. Defining the face processing network: optimization of the functional localizer in fMRI. Hum. Brain Mapp. 30, 1637–1651. https://doi.org/10.1002/hbm.20630.

Gilaie-Dolan, S., Saygin, A.P., Lorenzi, L.J., Egan, R., Rees, G., Behrmann, M., 2013. The role of human ventral visual cortex in motion perception. Brain 136, 2784–2798. https://doi.org/10.1093/brain/awt214.

Gilaie-Dolan, S., Saygin, A.P., Lorenzi, L.J., Rees, G., Behrmann, M., 2015. Ventral aspect of the visual form pathway is not critical for the perception of biological motion. Proc. Natl. Acad. Sci. U.S.A. E361–E370. https://doi.org/10.1073/pnas.1414974112.

Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: a comparison of theory of mind tasks. J. Cogn. Neurosci. 19, 1803–1814. https://doi.org/10.1162/jocn.2007.19.11.1803.

Gobbini, M.I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M., Pietrini, P., 2011. Distinct neural systems involved in agency and animacy detection. J. Cogn. Neurosci. 23, 1911–1920. https://doi.org/10.1162/jocn.2010.21574.

Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. 15, 536–548. https://doi.org/10.1038/nrn3747.

Grosbas, M.-H., Beaton, S., Eickhoff, S.B., 2012. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. Hum. Brain Mapp. 33, 431–454. https://doi.org/10.1002/hbm.21222.

Grossman, E.D., Blake, R., 2002. Brain areas active during visual perception of biological motion. Neuron 35, 1167–1175. https://doi.org/10.1016/S0896-6273(02)00897-8.

Güçlü, U., van Gerven, M.A.J., 2017. Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. Neuroimage 145, 329–336. https://doi.org/10.1016/j.neuroimage.2015.12.036.

Guntupalli, J.S., Feilong, M., Haxby, J.V., 2018. A computational model of shared fine-scale structure in the human connectome. PLoS Comput. Biol. 14, e1006120. https://doi.org/10.1371/journal.pcbi.1006120.

Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Ramadge, P.J., Haxby, J.V., 2016. A model of representational spaces in human cortex. Cereb. Cortex 36, 5373–5384. https://doi.org/10.1093/cercor/bhw068.

Hasson, U., Yang, E., Vallines, I., Heeger, D.J., Rubin, N., 2008a. A hierarchy of temporal receptive windows in human cortex. J. Neurosci. 28, 2539–2550. https://doi.org/10.1523/jneurosci.5487-07.2008.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., Heeger, D.J., 2008b. Neurocinematics: the neuroscience of film. Projections 2, 1–26. https://doi.org/10.3167/proj.2008.020102.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640. https://doi.org/10.1126/science.1089506.

Hasson, U., Chen, J., Honey, C.J., 2015. Hierarchical process memory: memory as an integral component of information processing. Trends Cogn. Sci. 19, 304–313. https://doi.org/10.1016/j.tics.2015.04.006.

Haxby, J.V., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., Rapoport, S.I., 1991. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc. Natl. Acad. Sci. U.S.A. 88, 1621–1625. https://doi.org/10.1073/pnas.88.5.1621.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends Cogn. Sci. 4, 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0.

Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P., Grady, C.L., 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J. Neurosci. 14, 6336–6353. https://doi.org/10.1523/jneurosci.14-11-06336.1994, 1994.

Haxby, J.V., Gobbini, M.I., 2012. Distributed neural systems for face perception. In: Calder, A.J., Rhodes, G., Johnson, M.H., Haxby, J.V. (Eds.), Oxford Handbook of Face Perception. Oxford University Press, Oxford, pp. 93–110. https://doi.org/10.1093/oxfordhb/9780199559053.001.0001.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430. https://doi.org/10.1126/science.1063736.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72, 404–416. https://doi.org/10.1016/j.neuron.2011.08.026.

Haxby, J.V., 2012. Multivariate pattern analysis of fMRI: the early beginnings. Neuroimage 62, 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016.

Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37, 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., Martin, A., 1999. The effect of face inversion on activity in human neural systems for face and object perception. Neuron 22, 189–199. https://doi.org/10.1016/S0896-6273(00)80690-X.

Heider, F., Simmel, M., 1944. An experimental study of apparent behavior. Am. J. Psychol. 57, 243–259. https://doi.org/10.2307/1416950.

Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76, 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. U.S.A. 96, 9379–9384. https://doi.org/10.1073/pnas.96.16.9379.

Isik, L., Tacchetti, A., Poggio, T., 2018. A fast, invariant representation for human action in the visual system. J. Neurophysiol. 119, 631–640. https://doi.org/10.1152/jn.00642.2017.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311. https://doi.org/10.1523/jneurosci.17-11-04302.1997.

Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. 97, 4296–4309. https://doi.org/10.1152/jn.00024.2007.

Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. Neuron 74, 1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036.

Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., Mishkin, M., 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends Cogn. Sci. 17, 26–49. https://doi.org/10.1016/j.tics.2012.10.011.

Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry: integrating cognition, computation, and the brain. Trends Cogn. Sci. 17, 401–412. https://doi.org/10.1016/j.tics.2013.06.007.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4. https://doi.org/10.3389/neuro.06.004.2008.

Lange, J., de Lussanet, M., Kuhlmann, S., Zimmermann, A., Lappe, M., Zwitserlood, P., Dobel, C., 2009. Impairments of biological motion perception in congenital prosopagnosia. PLoS One 4, e7414. https://doi.org/10.1371/journal.pone.0007414.

Lingnau, A., Downing, P.E., 2015. The lateral occipitotemporal cortex in action. Trends Cogn. Sci. 19, 268–277. https://doi.org/10.1016/j.tics.2015.03.006.

McCarthy, G., Puce, A., Gore, J.C., Allison, T., 1997. Face-specific processing in the human fusiform gyrus. J. Cognit. Neurosci. 9, 605–610. https://doi.org/10.1162/jocn.1997.9.5.605.

Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983. Object vision and spatial vision: two cortical pathways. Trends Neurosci. 6, 414–417. https://doi.org/10.1016/0166-2236(83)90190-X.

Naselaris, T., Kay, K.N., Nishimoto, S., Gallant, J.L., 2010. Encoding and decoding in fMRI. Neuroimage 56, 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073.

Nastase, S.A., Connolly, A.C., Oosterhof, N.N., Halchenko, Y.O., Guntupalli, J.S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M.I., Haxby, J.V., 2017. Attention selectively reshapes the geometry of distributed semantic representation. Cereb. Cortex 27, 4277–4291. https://doi.org/10.1093/cercor/bhx138.

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 21, 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031.

Nunez-Elizalde, A.O., Huth, A.G., Gallant, J.L., 2019. Voxelwise encoding models with non-spherical multivariate normal priors. Neuroimage 197, 482–492. https://doi.org/10.1016/j.neuroimage.2019.04.012.

Park, J., Russ, B.E., McMahon, D.B.T., Koyano, K.W., Berman, R.A., Leopold, D.A., 2017. Functional subpopulations of neurons in a macaque face patch revealed by single-unit fMRI mapping. Neuron 95, 971–981. https://doi.org/10.1016/j.neuron.2017.07.014.

Peelen, M.V., Wiggett, A.J., Downing, P., 2006. Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. Neuron 49, 815–822. https://doi.org/10.1016/j.neuron.2006.02.004.

Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., Kanwisher, N., 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage 56, 2356–2363. https://doi.org/10.1016/j.neuroimage.2011.03.067.

Puce, A., Allison, T., Asgori, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. 16, 5205–5215. https://doi.org/10.1523/jneurosci.16-16-05205.1996.

Russ, B.E., Leopold, D.A., 2015. Functional MRI mapping of dynamic visual features during natural viewing in the macaque. Neuroimage 109, 85–94. https://doi.org/10.1016/j.neuroimage.2015.01.012.

Sabuncu, M., Singer, B.D., Conroy, B., Bryan, R.E., Ramadge, P.J., Haxby, J.V., 2010. Function-based intersubject alignment of human cortical anatomy. Cereb. Cortex 20, 130–140. https://doi.org/10.1093/cercor/bhp085.

Schultz, R.T., Grelotti, D.J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., Skudlarski, P., 2003. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 415–427. https://doi.org/10.1098/rstb.2002.1208.

Sha, L., Haxby, J.V., Abdi, H., Guntupalli, J.S., Oosterhof, N.N., Halchenko, Y.O., Connolly, A.C., 2015. The animacy continuum in the human ventral visual pathway. J. Cogn. Neurosci. 27, 665–678. https://doi.org/10.1162/jocn_a_00733.

Shultz, S., McCarthy, G., 2011. Goal-directed actions activate the face-sensitive posterior temporal sulcus and fusiform gyrus in the absence of human-like perceptual cues. Cereb. Cortex 22, 1098–1106. https://doi.org/10.1093/cercor/bhr180.

Thorat, S., Proklova, D., Peelen, M.V., 2019. The nature of the animacy organization in human ventral temporal cortex. eLife 8, e47142. https://doi.org/10.7554/eLife.47142.

Tsao, D.Y., Moeller, S., Freiwald, W., 2008. Comparing face patch systems in macaques and humans. Proc. Natl. Acad. Sci. U.S.A. 105, 19514–19519. https://doi.org/10.1073/pnas.0809662105.

Ungerleider, L.G., Mishkin, M., 1982. Two cortical visual systems. In: Ingle, D.J., Goodale, M.A., Mansfield, R.J.W. (Eds.), Analysis of Visual Behavior. MIT Press, Cambridge, MA, pp. 549–586.

Ungerleider, L.G., Haxby, J.V., 1994. 'What' and 'where' in the human brain. Curr. Opin. Neurobiol. 4, 157–165. https://doi.org/10.1016/0959-4388(94)90066-3.

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., Castellano, F.X., 2015. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122, 222–232. https://doi.org/10.1016/j.neuroimage.2015.07.069.

Wurm, M.F., Ariani, G., Greenlee, M.W., Lingnau, A., 2016. Decoding concrete and abstract action representations during explicit and implicit conceptual processing. Cereb. Cortex 26, 3390–3401. https://doi.org/10.1093/cercor/bhv169.

Wurm, M.F., Lingnau, A., 2015. Decoding actions at different levels of abstraction. J. Neurosci. 35, 7727–7735. https://doi.org/10.1523/jneurosci.0188-15.2015.