GROUNDING OBJECT CONCEPTS IN PERCEPTION AND ACTION: EVIDENCE FROM fMRI STUDIES OF TOOLS

Michael S. Beauchamp and Alex Martin
(Section on Cognitive Neuropsychology, Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, USA)

ABSTRACT

Studies of categories of objects, including tools, have spurred the development of the sensory-motor model of object concept representation. According to this model, information about objects is represented in the same neural subsystems that are active when we perceive and use them. In turn, this model has provided insight into the brain mechanisms of tool use. For tools, three types of information are especially important for identification: the characteristic motion with which they move (such as the up and down motion of a hammer), their visual form, and the way that they are manipulated. Evidence from neuropsychological, non-human primates, and neuroimaging studies suggest a mapping between specific brain regions and these fundamental identifying properties of tools. We focus on neuroimaging studies of the left posterior middle temporal gyrus. This brain region is active both when subjects perceive moving tools and when they answer questions about tools, and is responsive to the type of visual motion characteristic of tools: rigid, unarticulated motion. We describe a simple model that explains how low-level receptive field properties like those known to exist in area MT/V5 could give rise to the high-level category-related representations observed in functional imaging experiments.

Key words: manipulable objects, human visual cortex, conceptual representation, middle temporal gyrus, functional magnetic resonance imaging

GROUNDING OBJECT CONCEPTS IN PERCEPTION AND ACTION

We begin with a brief introduction to the sensory/motor model of object concept representation (Barsalou et al., 2003; Martin, 1998; Martin et al., 2000). For any given individual, the concept of a specific object (such as a hammer) consists of everything the individual knows about the object, such as that hammers have handles, are usually made of wood or metal, are useful to pound nails, and that Maxwell had a silver hammer in the Beatles song of that name. This level of representation of an object is often referred to as associative or encyclopedic knowledge, and is characterized by its explicit expression (usually via language), unlimited size and scope, and individual idiosyncrasies. A different level of object concept representation is referred to as “semantic primitives” (Martin, 1998). In contrast to encyclopedic knowledge, semantic primitives are accessed implicitly and automatically in the service of comprehension, are highly constrained in number, and are universal. This level of representation allows us to quickly and efficiently identify objects and understand words, and forms the foundation for our vast stores of encyclopedic knowledge about objects. While the sensory/motor model does not address the organization of encyclopedic knowledge, it makes strong claims about the organization of semantic primitives with regard to both their representational content and organization in the brain.

In the sensory/motor model, the concept of an object is composed of semantic primitives that represent those properties of the object that allow for fast and efficient recognition. For tools, these properties include stored representations of what they look like, how they move when used, and how we manipulate them. These primitives are stored within the same neural systems active when we learned about those properties. Specifically, they are...
stored within visual processing systems for perceiving object form and object motion, and action systems responsible for visuomotor transformations and for grasping and manipulating objects.

ROLE OF TEMPORAL LOBES IN REPRESENTING VISUAL PROPERTIES OF TOOLS: FORM AND MOTION

In the human brain, tools elicit activity in a distributed network of brain areas, concentrated in the left hemisphere. This network is active across presentation format (pictures, visual words or auditory words) and processing task (viewing objects, naming objects, retrieving information about tools and their action-related properties, performance of the object-appropriate action). The brain regions in this network can be divided into four anatomical regions (Figure 1A). Like all object categories, visual presentation of tools evokes activity in posterior ventral temporal cortex.

Tool-related activity in this region is strongest in the medial portion of the fusiform gyrus, located in between face-preferring regions more laterally and house/place-preferring regions more medially (Chao et al., 1999). Greater activity to tools than to other object categories in the medial fusiform gyrus has been replicated a number of times using both pictures and words (Beauchamp et al., 2002, 2003; Chao et al., 2002; Devlin et al., 2005; Whatmough et al., 2002). A second focus of tool-related activity is found in left posterior lateral temporal cortex. This region is located on the lateral surface in the middle temporal gyrus and inferior temporal sulcus (abbreviated MTG). The remaining tool-responsive regions (also in the left hemisphere) consist of an anterior portion of the intraparietal sulcus (IPS) in the inferior parietal lobule and ventral premotor cortex (VPM) (Binkofski et al., 1999; Chao and Martin, 2000; Handy et al., 2003; Johnson-Frey et al., 2004; Kellenbach et al., 2003; Gerlach et al., 2002; Grabowski et al., 1998).

There is general consensus on the importance of the frontal-parietal circuit for the grasping and manipulation of objects, and computation of the required sensorimotor transformations (Jeannerod et al., 1995; Johnson-Frey, 2003; Rizzolatti and Luppino, 2001). In addition to action performance (such as the manual manipulation of tools) this circuit, especially in the left hemisphere, is also important for perceiving the actions of others, imagining, imitating or pantomiming actions, and knowing about them (e.g., Buxbaum and Saffran, 2002; Gainotti et al., 1995; Goldenberg et al., 2003; Grezes and Decety, 2001; Jeannerod, 2001; Johnson-Frey, 2004; Martin, 2001; Rumiati et al., 2004). While the importance of parietal and premotor regions in human interactions with tools has received considerable attention, the function of left posterior MTG has remained relatively obscure.

One hint to the role of left MTG in knowing about tools and their functional characteristics came from a functional neuroimaging study using property production tasks. Subjects were required to verbally generate object properties in response to object pictures or their written names (Martin et al., 1995). When subjects generated a word denoting an action associated with the object (e.g., ‘write’ in response to pencil) activity in the left MTG was greater than when subjects generated the name of a color associated with the object (e.g., ‘yellow’ in response to pencil). The reverse contrast (color > action word generation) elicited activity in ventral temporal cortex, anterior to regions associated with color sensation (e.g., Beauchamp et al., 1999; Hadjikhani et al., 1998; Lueck et al., 1989). For action word generation, the active portion of MTG was located just anterior to area MT/V5, known to be an important locus of cortical visual motion processing (e.g., Beauchamp et al., 1997; Tootell et al., 1995; Watson et al., 1993).

Because of the selectivity of the left MTG response to action word generation, and its proximity to the visual motion area MT/V5, we suggested that information about the motion properties associated with object actions may be stored in this region (Martin et al., 1995). This speculation received additional support from an object naming study that found greater activity for naming tools than animals in the same region of left MTG active during action word generation (Martin et al., 1996). These reports were quickly followed by neuropsychological evidence linking impaired tool recognition to left MTG damage (Tranel et al., 1997). The association between action word generation, tool naming, and activation of the posterior region of the left MTG has been replicated numerous times using different imaging methods and experimental paradigms (e.g., Devlin et al., 2002; Phillips et al., 2002; for extensive review, see Martin, 2001) as has the association between tool and action knowledge deficits and left MTG lesions (Tranel et al., 2003). The consistency of these findings suggests that the left MTG is a critical node in the neural circuitry underlying knowledge of tools and their associated actions.

Several lines of evidence support the idea that the motion of real-world object is processed in lateral temporal cortex. Viewing static images of objects that imply motion – such as a picture of a cup in mid-fall – evoke greater activity in lateral temporal cortex than static images without implied motion (Kourtzi and Kanwisher, 2000; Senior et al., 2000). Extensive evidence from single unit recording studies in non-human primates and imaging studies in humans indicate that part of lateral temporal cortex in an around the superior temporal sulcus (STS) is specialized for processing biological motion (reviewed in Puce and Perrett, 2003). These findings raised the possibility that information about object-associated motion was represented in lateral temporal cortex, with
biological motion in STS and non-biological motion typical of tools in the MTG.

Direct tests of this motion property hypothesis required new functional magnetic resonance (fMRI) experiments. Although previous studies had shown that viewing static images of tools evokes activity in left MTG, no information was available about brain responses to moving tools. If motion properties associated with tools are represented in MTG, this region should not only show a stronger response to tools than animate, biological objects, but should respond more strongly to moving tools than to static images of these objects. In contrast, if information about the form of an object, and form-related properties like color, is represented in ventral temporal cortex, then ventral temporal areas, such as fusiform gyrus, should show category-related responses, but no advantage for moving over static object images.

In order to test these ideas, it was important to use objects of clear ethological relevance, depicting familiar, easily discriminable motions. Therefore, we used short video clips of familiar tools moving in their characteristic fashion (e.g., saws sawing,
hamme hammering) and human actors performing common whole body movements (e.g., jumping jacks, sitting down). Digital editing techniques were used to create tool videos that showed tools moving naturally without a visible hand or arm manipulating them. This permitted an experiment that purely dissociated the response to either non-biological or biological moving stimuli (for details, see Beauchamp et al., 2002). One difficulty with complex stimuli is that it is difficult, if not impossible, to control all of the differences in low-level visual features (such as speed of motion) between stimulus categories. To circumvent this problem, many exemplars from each category were used (28 different tool and human motions) filmed from five different viewpoints on an equiluminant gray background to yield 140 unique stimuli per object category. Because each stimulus category contained a variety of movements of different speeds and directions (hammers hammer quickly while water pitchers pour slowly; jumping jacks are rapid while sitting motions are slower) between-category differences would be unlikely to reflect simple differences in the speed or direction of movement.

A series of experiments were performed, using a variety of designs (block and rapid event-related) and tasks (delayed march-to-sample, object categorization, motion-type classification). In each experiment, category-related dissociations were observed in lateral and ventral temporal regions (Figure 1A). Consistent with the studies reviewed above, in lateral temporal cortex, MTG preferred tool stimuli while STS preferred human stimuli. In ventral temporal cortex, medial fusiform preferred tools while lateral fusiform preferred humans. Visual area MT straddled the boundary between tool and human preferring cortex, and showed on average no preference for either category of objects (Figure 1B). Interestingly, MT preferred both types of meaningful stimuli (humans and tools) to simple moving gratings (the standard stimulus used to localize MT) emphasizing the potency of object motion in evoking neuronal responses from lateral temporal cortex.

Having replicated the expected category-related differences in lateral and ventral temporal cortex, we then addressed the central question: is MTG selective for tool motion? As predicted, MTG showed significantly larger responses to moving than static images of tools (Figure 1C). Importantly, medial fusiform showed no difference in response to moving and static tools (Figure 1C). This serves as a negative control, demonstrating that the larger response observed in MTG for moving stimuli was not due to a non-specific effect of visual attention or arousal on the entire visual cortex. These results provide strong support for the claim that MTG codes some aspect of tool motion, while medial fusiform codes some aspect of tool form. While these studies used behavioral tasks to equate difficulty across conditions, previous studies had demonstrated that activation is observed in MTG and fusiform even for passive viewing of stimuli (Chao et al., 1999), suggesting that motion and form information is automatically accessed whenever tools are viewed.

A critical issue was to characterize the motion properties to which MTG was responsive. Assuming that access to knowledge of its up-and-down motion is important for identifying a hammer, it follows that neurons in MTG should be sensitive to this type of motion. A simple quality shared by most tool motions is that they are simple translations that take place without articulation. That is, hammers, saws, pencils, utensils, and other tools move with different trajectories, but do not change shape as they move. This is in sharp contrast to animate biological objects, which move with many degrees of articulation as limbs and digits shift about their joints.

To study whether specific motion properties give rise to tool-related activity in MTG, it is necessary to separate the motion pattern of the stimulus from the object form. One approach is simply to create artificial unarticulated motion vectors and apply them to tools and bodies. If MTG prefers unarticulated motion, then it should respond to artificial unarticulated motion, while if STS prefers articulated motions, it should show weaker responses to biological forms moving in an unarticulated way. This was exactly the result that was observed (Figure 2A) (for details, see Beauchamp et al., 2003). The stimuli were equally novel or artificial – neither saws nor bodies are normally observed rotating about their center of mass – but MTG responded similarly to typical and artificial unarticulated motions of tools, while STS preferred articulated to unarticulated motions of bodies.

One confound in this experiment is that in both cases, the visual form of the preferred category was present (tools for MTG and bodies for STS). Therefore, in the next experiment, Johansson (1977) point light displays were created (Figure 2B). In these displays the complex form of the original object was replaced with a much sparser display containing 5-10 points whose trajectories matched the motion of the corresponding real object. The visual system is able to extract a remarkable amount of information from these displays (Johansson, 1977). The advantage of point-light displays is that they contain none of the color or form cues of natural stimuli, allowing an assessment of the importance of motion in isolation. Using these stimuli, we found that MTG preferred the unarticulated motion found in tool point light displays, while STS preferred the articulated motion found in human point light displays (Figure 2B). Interestingly, the response of MTG to point light displays was as large as the response to real tool stimuli, highlighting the importance of motion properties to MTG responses. In contrast, the medial fusiform showed a strong preference for real
compared to point light displays (Figure 2B),
supporting the idea that this region encodes the
form and color of tools, but not their motion.
Nevertheless, although ventral temporal regions
responded weakly to the impoverished point light
displays, category-related responses were observed.
Specifically, the medial fusiform showed a greater
response to point light displays of moving tools than
moving humans. This finding underscores the
critical point that responses in these regions are
associated with what the stimulus is interpreted to
be, not solely with its physical characteristics (for
evidence of category-related responses in lateral and
ventral temporal cortex in response to moving
geometric forms interpreted as depicting either
social or mechanical events and interactions, see
Martin and Weisberg, 2003).

These findings suggest a simple neuronal
architecture that could give rise to the observed
fMRI results (Figure 3). Motion-sensitive neurons
in V1 and early stages of MT respond to specific
parts of the visual field. These neurons have
specific preferred directions which are oriented in a
columnar fashion (Figure 3A; Albright et al.,
1984). If neurons that encode similar directions in
adjacent parts of the visual field projected to MTG,
this would produce maximal responses for objects
that move in an unarticulated way (Figure 3B).
Conversely, biological movements in which
different limbs move in different directions would
evoke only weak activity in MTG (Figure 3C).
While this speculative proposal is only a cartoon
sketch that does not capture the complex feed-
forward and feedback organization of visual cortex

---

Fig. 2 – Average responses to different tool and human stimuli. Locations of active regions and time series conventions shown in
Figure 1. (A) Average responses in MTG (left) and STS (right) to single 2.5 second presentations (indicated by gray bars) of
unarticulated tools, humans performing articulated whole body movements, artificial unarticulated tools, and artificial unarticulated
humans. (B) Responses in MTG (left) and medial fusiform (right) to video clips of moving humans performing whole body movements,
Johansson (1977) point-light displays of humans performing the same movements, video tools and point-light tools.
(as well as the many other receptive field properties of MT neurons) it provides an illustration of how relatively low-level receptive field properties could produce the high-level category related activity observed in visual cortex.

**FUTURE AVENUES OF EXPLORATION**

The findings reviewed here add to an increasing body of literature on the neural systems that mediate perceiving and knowing about object categories. Functional neuroimaging reveals that, like all categories of objects tested to date, information about tools is represented in a distributed network. The sensory-motor model provides a framework for understanding these distributed activations. Our results reveal that in temporal lobe, MTG is particularly responsive to the type of motion – rigid, unarticulated motion – that is typical of manmade, manipulable objects. More generally, MTG, especially in the left hemisphere, is also activated
by a variety of conceptual processing tasks involving tools and their associated actions, and damage to this region results in impaired naming of tools and impaired knowledge of their actions. These findings provide compelling evidence that this region is critical for both perceiving and knowing about the motion characteristics of tools.

A number of important questions remain to be addressed. First, although we have highlighted the specialized processing and representational characteristics of the brain regions engaged by tools (especially MTG) little is known about how these disparate representations are integrated. A promising avenue of exploration is to study how closely intertwined representations from different modalities – such as the visual and tactile domains – are integrated (Binkofski et al., 2004; Goldenberg et al., 2004; Holmes et al., 2004; Lädavas and Farnè, 2004). Our attempt to address this issue has revealed that auditory-visual multisensory integration about tools may be dependent on an associative region located in lateral temporal cortex superior and anterior to MTG (Beauchamp et al., 2004a, 2004b). Multisensory responses in lateral temporal cortex appear to be organized by modality, but not necessarily by conceptual category. The precise relationship between category-related and modality-related organization remain to be determined. A related question is the role of non-visual responses to tools in MTG. If motion properties are critical for identifying tools, and if activity in MTG is the neural substrate that represents these motion properties, then activity in MTG may be necessary for perceiving tools, whether they are presented in the visual or auditory modality. A direct test of this prediction suggests that, in fact, MTG does respond to tool-associated sounds (Lewis et al., 2004; Yang et al., 2005). Another important question is the extent to which the motion properties represented in MTG are selective for tools because their motion is a critical property for determining their function, or responds to any manmade object that can be manipulated, regardless of the relationship between its associated motion and its function. For example, motion and manual manipulation is involved in tying a necktie or using a toaster, but their movement and motions characteristics are not relevant for the concept of a tie or a toaster. On-going studies in our laboratory are exploring this issue.

Acknowledgments. This work was supported by the National Institute of Mental Health Intramural Research Program. Michael Beauchamp’s current address is Department of Neurobiology and Anatomy, University of Texas Health Science Center at Houston, Houston, TX, USA.

References


Devlin JT, Rushworth MF and Matthews PM. Category-related activation for written words in the posterior fusiform is task specific. Neuropsychologia, 43: 69-74, 2005.


LISSAUER H. A case of visual agnosia with a contribution to theory. Archiv für Psychiatrie, 21: 222-270, 1890.


Michael S. Beauchamp and Alex Martin

(Received 8 February 2005; accepted 7 April 2005)