Computational Cognitive Neuroscience of the Visual System

Stephen A. Engel

University of Minnesota

ABSTRACT—*Should psychologists care about functional magnetic resonance imaging (fMRI)? Within the field of visual perception, the answer to this question is an emphatic "yes." There is a long history of close interactions between psychology and neuroscience in the study of vision. In the 1980s, vision researchers developed a general framework for combining the two fields, and that framework still supports much current research. This article will briefly cover this general approach and then will illustrate how neuroimaging allows the framework to be applied to human perception. In short, fMRI can measure how information is represented in sequential stages of processing. These same representations can also be measured behaviorally. Theories of vision specify how the sequential representations participate in well-defined computations that underlie perception.*

KEYWORDS—*vision; fMRI; color*

One argument about why cognitive psychologists should *not* care about fMRI can be traced back at least as far as Neisser's *Cognitive Psychology* (Neisser, 1967, p. 6):

> The task of a psychologist trying to understand human cognition is analogous to that of a man trying to discover how a computer has been programmed. . . . Given this purpose, he will not care much whether his particular computer stores information in magnetic cores or in thin films; he wants to understand the program, not the hardware.

For psychologists, Neisser argues, the uninteresting hardware is the brain.

The flaw in this argument is that the brain is not a general-purpose computer. Its specialized hardware may strongly influence what software can run. Neural hardware does, in fact, greatly influence the cognitive software underlying perception.

Examples of this influence abound in vision science. One of the oldest arises in Helmholtz's trichromatic theory of color vision. A psychological fact—that the color of any single light can be matched perceptually by a combination of three independent lights—is entirely accounted for by a neural constraint, that the initial neural encoding of light in the retina is by three classes of cone photoreceptors. Similarly, Herring's observation of color opponency is explained by neurons whose responses reflect simple linear combinations of cone signals. For example, the psychological fact that no light appears simultaneously reddish and greenish (while many appear simultaneously bluish and greenish) is accounted for by color-opponent neurons that encode a difference between cones that respond to reddish and greenish light. These neurons can signal the relative redness or greenness of a stimulus but cannot signal the presence of both redness and greenness; because they encode a difference, simultaneous signals from both reddish- and greenish-responsive cones cancel each other.

By the 1970s, many vision scientists expressed their theories using information-processing terminology. In this language, the initial *representation* of color information is the responses of the cones, and the color-opponent representation is a second stage of a *processing stream*, along which sequential representations are *computed*. The power of this abstract language is that evidence for the same representations can be gathered using either methods from psychology ("psychological representations") or methods from neuroscience ("neural representations").

Current theories of visual perception contain an additional level of explanation: *computational theory*, named by David Marr in the early 1980s. This level contains a precise specification of the goals, inputs, and outputs of the processing stream.

Also by the 1980s, neuroscientists had systematically characterized the many separable parts of the brain that are involved in visual perception. To date, over 30 distinct regions, called

Address correspondence to Stephen Engel, Psychology Department, University of Minnesota, Elliott Hall S253, 75 East River Road, Minneapolis, MN 55455; e-mail: engel@umn.edu.

visual areas, have been identified. Anatomical connections between the areas link them into streams, along which information flows from initial, "earlier" areas to "later" ones.

Armed with these developments, the linked goals of contemporary research in visual perception are to (in no specific order)

- Identify psychological representations that are important for a perceptual phenomenon
- Identify the neural representations in different visual areas that are important for a perceptual phenomenon
- Test for correspondence between the two representations (e.g., are the neural representations the implementation of the psychological ones?)
- Develop formal theories of how the representations participate in a computation that is the basis of the perceptual phenomenon

A well-worked-out example of this approach comes from the study of visual motion, in which behavioral and electrophysiological data have been combined with computational analyses (for a review, see, e.g., Simoncelli, 2004). To summarize the main results of this relatively complete account: One goal of motion perception is to infer the true motion of objects in the world from the changing pattern of retinal stimulation. The primary visual cortex (V1) computes an initial step in making that inference. Direction-selective neurons in that part of the brain encode very local measurements of image motion. A later visual area, designated MT, contains a second-stage representation. MT neurons receive input from V1, and a subset of them combines their inputs in such a way as to encode more global "pattern-" or object-based motion. The neural representation in MT corresponds closely to the psychological representation of motion; altering the neural representation by electrically stimulating small populations of neurons (microstimulation) can change animals' percepts of motion. Reasonably complete quantitative theories express formally what the local and global representations in V1 and MT are and how neurons compute each of them.

IMAGING RESULTS

Results from human imaging studies fit squarely within the tradition of computational cognitive neuroscience outlined above. This section reviews some recent findings. Early reviews of the use of neuroimaging data in computational studies of perception are presented in Engel (1996) and Wandell (1999).

Identifying Visual Areas

A major assumption of the computational neuroscience approach is that different cortical areas contain different neural representations. Accordingly, a first and crucial step was to identify visual areas in human subjects. Standard functional magnetic resonance imaging (fMRI) methods for mapping visual areas were developed over 10 years ago and are now used routinely in individual subjects.

Early visual areas can be identified with fMRI by measuring how the visual field maps to the visual cortex (DeYoe et al., 1996; Engel et al., 1994; Sereno et al., 1995). Each visual area contains a distinct spatial representation, or map, of a quarter or half of the visual field. These maps are measured by stimulating portions of the visual field and noting correlations between stimulus location and the location of neural activity in the cortex—that is, determining which visual-field location generates the highest level of activity at each location in the cortex. Plotting each cortical location's preferred visual-field location as color on a graphical model of the brain anatomy reveals many distinct visual-field representations, which are visible as clusters of color that span the range of a quarter or half of the visual field. Each of these visual-field representations corresponds to a distinct visual area.

Later visual areas are identified using a variety of different techniques. Some later areas have subtle visual-field maps that can be identified as described above. Motion selective areas can be identified by comparing responses to moving and stationary stimuli, and areas important for object recognition can be identified by comparing responses to objects with those to nonobjects such as texture patterns (see Grill-Spector & Sayres, 2008, this issue). Dorsal areas involved in visual attention and motor programming can also be identified (see Yantis, 2008, this issue).

Given maps of visual areas, one can then examine those areas' responses to stimuli that are carefully chosen to test for the presence of specific neural representations important for visual perception. In the next section, I review one such effort from my laboratory.

Color Perception

There exists broad agreement about the earliest computations that are involved in color perception (some of these were described previously). A surprising amount of debate, however, surrounds how neurons in V1 represent color information. The majority of neurons in V1 respond selectively to orientation, firing robustly only to stimuli at a narrow range of orientations. Most of these are light–dark color-opponent neurons, which respond most strongly to stimuli that are close to black and white. Very conflicting conclusions have been reported about how many red–green color-opponent neurons—which prefer patterns of reddish and greenish light—also show orientation selectivity. One view holds that almost all color-opponent neurons fail to show selectivity. In this view, color is represented separately from form, as orientation selectivity is a first stage in making explicit information about object shape. An opposing view is that many or most color-opponent neurons do show orientation selectivity, jointly representing color and form (for a more complete discussion, see Engel, 2005).

Psychophysical results suggest that color and orientation are represented jointly at some stage in the visual system. For example, prolonged exposure to a red–green pattern with a particular spatial orientation greatly reduces the visibility of red–green patterns that are oriented similarly (Bradley, Switkes, & De Valois, 1988). Orthogonally oriented patterns are affected less by this adaptation, and differently colored patterns even less. These results are usually interpreted to mean that exposure to a colored pattern selectively reduces responses of neurons that jointly represent a particular combination of color and orientation.

We used this adaptation paradigm, in conjunction with fMRI, to measure the orientation selectivity of responses to stimuli that drive red–green and light–dark color-opponent neurons (Fig. 1; Engel, 2005). We recorded the response of visual areas to horizontal and vertical red–green and light–dark patterns both prior to and following prolonged exposure to one of the patterns. Adaptation scans began with subjects viewing a high-contrast adapting pattern for 1 minute. Low-contrast "test" patterns were then presented in alternation with the high-contrast adapters. Because of their lower contrast, the test patterns were decrements in contrast and so generated decreases in neural activity that were visible as troughs in the fMRI timecourse (see Fig. 1). Critically, deeper troughs indicated weaker response to the corresponding test, as the fMRI signal fell to lower levels from the high level established by the adapting pattern.

Our results showed strong effects of adaptation, whose pattern indicated that large numbers of neurons responding to the red–green patterns were selective for orientation. Prior to adaptation, responses to the test patterns were roughly equal for all conditions. Following adaptation, the test pattern that was the same color and orientation as the adapter generated reliably weaker responses in V1 than did the other three test patterns; adaptation was selective. The joint selectivity of the adaptation was visible for both red–green and light–dark adapters and was most likely due to many of the adapted neurons jointly encoding color and orientation.

Additionally, for red–green adaptation, the red–green test that differed in orientation from the adapter gave an intermediate response, weaker than the response to either light–dark test. This suggested that some of the adaptation was due to decreased responsiveness in red–green neurons that respond to many orientations. Overall, then, the fMRI data indicated that two populations of neurons, one selective for orientation and one unselective, encode color in V1.

To test whether these neural representations corresponded to representations identified behaviorally, we measured perceptual effects of adaptation using the same stimuli in the same subjects. Subjects adjusted the appearance of an unadapted test stimulus to match the appearance of an adapted one. Overall, the behavioral measurements closely paralleled the fMRI data. Effects of adaptation were jointly selective; tests that were the same color and orientation as the adapter appeared greatly weakened following adaptation. For red–green adaptation, the test that
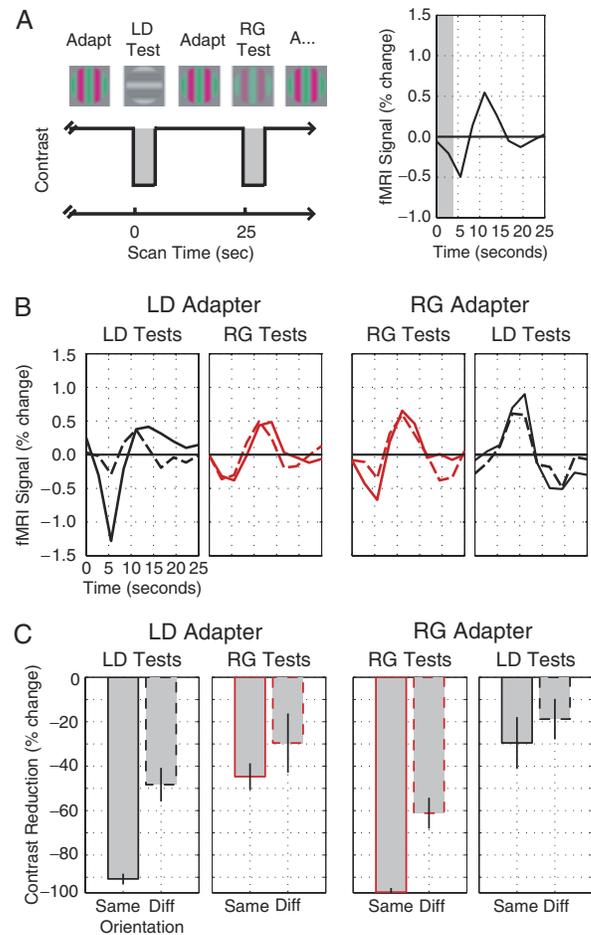
Fig. 1. Study of color- and orientation-selective adaptation in the human visual cortex (V1) using functional magnetic resonance imaging (fMRI). Subjects viewed low-contrast test patterns in alternation with high-contrast adapters (A, left, shows a sample stimulus sequence with a red–green [RG] adapter and both light–dark [LD] and red–green tests). The tests generated troughs in the fMRI signal, followed by peaks generated by the adapter reappearing (A, right, shows average fMRI response to the test, whose presentation time is indicated in gray). The depth of the trough reflected the strength of response to the test. For LD adaptation, the LD test that was the same orientation as the adapter generated a very weak response, a deep trough (B, left graph, black solid line; solid lines indicate tests that were the same orientation as the adapter and dashed lines indicate tests that differed in orientation; black lines indicate LD tests and red lines indicate RG tests). For LD adaptation, tests in which either the color or orientation differed from the adapter generated equal and relatively strong responses. For RG adaptation, the RG test that was the same orientation as the adapter also produced the weakest response (red solid line) and the RG test that differed in orientation produced an intermediate response (red dashed line). In perceptual measurements (C), adapting to either LD or RG greatly reduced the apparent contrast of the test that had the same color and orientation as the adapter and had smaller effects on other stimuli. Adapting to RG produced an intermediate effect on the RG pattern that differed in orientation from the adapter.

differed in orientation from the adapter showed an intermediate effect of adaptation.

The correspondence between the behavioral and fMRI data suggests that both oriented and unoriented red–green neurons in V1 are the neural basis of representations that are important for

color perception. This conclusion fits within the computational cognitive-neuroscience approach, placing corresponding neural and behavioral representations within a sequence of computations. From the viewpoint of purely cognitive theory, our results suggest that color has a dual representation in early stages of processing, one that is separate from form and one that is bound together with form.

Other Results

Many other imaging studies of vision have measured neural representations and established their correspondence with perceptual representations. While reviewing all such work is beyond the scope of this review, two additional examples illustrate the approach.

Zenger-Landolt and Heeger (2003) studied a phenomenon called surround suppression, in which the appearance of a small patch of stripes can be greatly affected by surrounding it with a larger patch of stripes. This phenomenon likely reflects a computation in which the neural responses to the central patch are roughly divided by the responses to the surrounding patch. This computation, termed *contrast normalization*, may serve to place different neural responses on a similar scale, facilitating comparisons between them at later stages of visual processing. Zenger-Landolt and Heeger found evidence of surround suppression in V1 using fMRI and also established a close quantitative correspondence between suppression measured behaviorally and the neural suppression. These results identify a critical transformation of visual representation that occurs in V1.

Meng, Remus, and Tong (2005) examined whether early visual areas contain representations of visual phantoms. Visual phantoms can appear when relatively faint bars of light move together on either side of an empty gap, and the bars are aligned so that if they continued through the gap they would connect to form larger bars. When viewing such displays, humans see ghostly phantoms moving in the gap, as though the visual system extends the bars through it. Meng et al., measured effects of this phantom in V1 with fMRI; V1 response was high in conditions in which phantoms were present and low in control conditions that did not generate a phantom. The researchers also used binocular rivalry (presenting two different stimuli to the two eyes) to create conditions in which subjects' perception of the phantom varied widely over time with an unchanging stimulus. These perceptual variations were visible in the fMRI timecourse. V1 activity was higher when subjects perceived the phantom and lower when they did not. Together, these results suggest that neurons in V1 represent the phantom. This representation is likely part of a larger computation in which the visual system infers the presence of stimuli from data that are interrupted spatially.

CHALLENGES AND FUTURE DIRECTIONS

Understanding the visual system remains an enormous task, even with the relatively well-worked-out paradigm described above. This section reviews some ongoing challenges for neuroimaging research.

Measuring Representations With fMRI

The low spatial resolution of fMRI data makes it difficult to measure neural representations directly. For example, V1 contains neurons that are selective to various orientations but neurons with different preferred orientations are mixed within any given localized measurement (voxel) of fMRI data.

Several methods have been developed to solve this problem. Adaptation is a classic psychophysical method for revealing multiple representations from univariate data. Long-term adaptation was used, for example, to demonstrate specificity for color and for orientation in the color-vision study discussed above. Short-term or "fMRI" adaptation, in which effects of relatively brief prior exposures are measured, may also be used, although this method can sometimes fail to find known representations—for example, orientation specificity in V1 (Boynton & Finney, 2003). Parametric stimulus variation (systematically changing quantitative aspects of the stimulus), coupled with formal modeling, is also a powerful method for inferring representations within visual areas (Engel, Zhang, & Wandell, 1997).

Another, recently developed method tests whether subtle aspects of the spatial pattern of activity in an area correlate with the presented stimulus (e.g., Kamitani & Tong, 2005). For example, small changes in the spatial pattern of V1 activity can predict the orientation of a set of bars that a subject is viewing. This ability to predict the stimulus implies that some distinct representation of each orientation is present in V1.

Establishing Correspondence Between fMRI and Behavioral Data

Comparing fMRI and behavioral data is a challenge because they are measured in different units: fMRI activity is usually measured by percent change of fMRI signal strength. How can this be related to behavior as measured in percent correct in a task? A useful first step is to show simple qualitative agreement across multiple conditions, as was done for the color data above.

A stronger method is to keep the stimulus constant and to test whether the fMRI data correlate with residual changes in behavior. This approach increases confidence that behavioral changes are due to neural changes, since the behavioral changes cannot be attributed to the (constant) external stimulus. This approach was used in the visual-phantom study described above.

A third, even stronger method is to use parametric stimulus variations to relate fMRI and behavioral data in a quantitative fashion. With such data it is often possible to build a single mathematical model that can simultaneously explain the parametric changes in behavior and those in the fMRI signal. For example, relatively simple models exist that can predict how subjects' ability to see a small increase in stimulus contrast depends upon the base level of contrast. The behavior can be modeled using an underlying contrast-response function, which

can be simultaneously fit to fMRI data (Boynton, Demb, Glover, & Heeger, 1999; Zenger-Landolt & Heeger, 2003).

Technical Advances

Finally, two rapidly advancing lines of work promise to improve the contributions of fMRI to computational cognitive neuroscience. First, uncertainty about the neural sources of the fMRI signal make it difficult to compare fMRI and single-unit data. Such comparisons will become easier in the future, as the relation between the fMRI signal and various neural processes is clarified. Second, in humans it is currently difficult to identify the anatomical connections between visual areas that trace successive stages of computation. Diffusion tensor imaging, which measures movement of water in tissue, may provide a method for identifying connections between functionally identified areas in the human visual cortex.

CONCLUSION: BEYOND LOCALIZATION

There is little doubt that fMRI is already making a contribution to cognitive theories of visual function. Importantly, the approach outlined here is not limited to vision. Human neuroimaging has established itself as one of many useful tools in the general framework of computational cognitive neuroscience.

The committed skeptic might note that most of the examples discussed here simply localized perceptual representations that had already been identified behaviorally. While localization results do constrain cognitive theory, they are a relatively limited role for neuroimaging. I believe that this is a historical accident. Because imaging is a relatively new and expensive technique, it makes sense to work conservatively and start from solid behavioral findings. But already, some results run in the opposite direction. For example, our study of color perception identified adaptation in red–green neurons that generalized across orientation. We only later measured corresponding behavioral effects and found a similar pattern that had not been reported in the existing behavioral literature. As imaging matures as a field, we should see more examples of novel representations uncovered initially with fMRI. These will then be measured behaviorally and incorporated into expanding theories of visual perception.

Recommended Reading

Boynton, G.M., Demb, J.B., Glover, G.H., & Heeger, D.J. (1999). (See References). A classic example of combining behavioral and fMRI data through mathematical modeling.

Engel, S.A. (2005). (See References). A recent paper that identifies neural representations using adaptation.

Kamitani, Y., & Tong, F. (2005). (See References). A recent paper that identifies neural representations using correlations between the spatial pattern of activity and stimuli.

REFERENCES

Boynton, G.M., Demb, J.B., Glover, G.H., & Heeger, D.J. (1999). Neuronal basis of contrast discrimination. *Vision Research*, *39*, 257–269.

Boynton, G.M., & Finney, E.M. (2003). Orientation-specific adaptation in human visual cortex. *Journal of Neuroscience*, *23*, 8781–8787.

Bradley, A., Switkes, E., & De Valois, K. (1988). Orientation and spatial frequency selectivity of adaptation to color and luminance gratings. *Vision Research*, *28*, 841–856.

DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences, USA*, *93*, 2382–2386.

Engel, S.A., Zhang, X., & Wandell, B. (1997). Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, *388*, 68–71.

Engel, S.A. (1996). Looking into the black box: New directions in neuroimaging. *Neuron*, *17*, 375–378.

Engel, S.A. (2005). Adaptation of oriented and unoriented color-selective neurons in human visual areas. *Neuron*, *45*, 613–623.

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., & Shadlen, M.N. (1994). fMRI of human visual cortex. *Nature*, *369*, 525.

Grill-Spector, K., & Sayres, R. (2008). Object recognition: Insights from advances in fMRI methods. *Current Directions in Psychological Science*, *17*, 73–79.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*, 679–685.

Meng, M., Remus, D.A., & Tong, F. (2005). Filling-in of visual phantoms in the human brain. *Nature Neuroscience*, *8*, 1248–1254.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*, 889–893.

Simoncelli, E. (2004). Local analysis of visual motion. In L.M. Chalupa & J.S. Werner (Eds.), *The visual neurosciences* (pp. 1616–1623). Cambridge, MA: MIT Press.

Wandell, B.A. (1999). Computational neuroimaging of human visual cortex. *Annual Review of Neuroscience*, *22*, 145–173.

Yantis, S. (2008). The neural basis of selective attention: Cortical sources and targets of attentional modulation. *Current Directions in Psychological Science*, *17*, 86–90.

Zenger-Landolt, B., & Heeger, D.J. (2003). Response suppression in v1 agrees with psychophysics of surround masking. *Journal of Neuroscience*, *23*, 6884–6893.