



## Philosophy of Mind

# The Depictive Nature of Visual Mental Imagery

*Norman Yujen Teng \**  
*National Chung Cheng University*  
[pyyjt@ccunix.ccu.edu.tw](mailto:pyyjt@ccunix.ccu.edu.tw)

**ABSTRACT:** Tye argues that visual mental images have their contents encoded in topographically organized regions of the visual cortex, which support depictive representations; therefore, visual mental images rely at least in part on depictive representations. This argument, I contend, does not support its conclusion. I propose that we divide the problem about the depictive nature of mental imagery into two parts: one concerns the format of image representation and the other the conditions by virtue of which a representation becomes a depictive representation. Regarding the first part of the question, I argue that there exists a topographic format in the brain but that does not imply that there exists a depictive format of image representation. My answer to the second part of the question is that one needs a content analysis of a certain sort of topographic representations in order to make sense of depictive mental representations, and a topographic representation becomes a depictive representation by virtue of its content rather than its form.

## Problem Setting

Tye (1991) argues that visual mental images rely at least in part on depictive representations. His argument, in its barest and crudest form, goes like this. We know that visual experiences have their contents encoded in topographically organized regions of the visual cortex. We also know that visual mental images are constructed by activating previously stored visual information. These two empirical findings suggest that imagistic experiences also have their contents encoded in the topographically mapped areas. Now, it seems clear that the topographically organized regions of the visual cortex support depictive representations. Therefore, visual mental images rely at least in part on depictive representations (see also Kosslyn, 1994, pp. 12-21).

Though the argument just described makes its point with cogency and force, I think that it is not sufficient to support the view that visual mental images rely on depictive representations. The problem, as I see it, is this. It is indeed interesting to note that in visual perception, the information coming in through our eyes goes through a series of stages of processing, with the retinal image being reconstructed in the visual cortex, so that in a quite literal sense adjacent parts of the visual cortex represent adjacent parts of the retinal image (see also Churchland & Sejnowski, 1992, pp. 31-34). In this topographically representational scheme, it is impossible to represent an object without also representing the spatial order of adjacent parts of the represented object. This result makes topographic representations very unlike propositional representations, where the information about the

spatial order of adjacent parts of the represented object needs not be encoded. That topographic representations are not propositional does not imply that they are depictive representations, however. Notice that the topographic mapping preserves only order, but not size and not really shape. (Colors and other secondary qualities surely are never preserved.) Which makes a topographic representation a bit unlike a depictive representation (Guzeldere, 1995, p. 353, footnote 19). For depictive representations seem to require something more than just order-preserving. Moreover, topographic maps are found not only in the visual cortex. A similar topographic relation exists in other sensory modalities. For example, in auditory perception the anterior portion of the auditory cortex responds to high-frequency tones and posterior regions respond to progressively lower frequencies. (Stilling et al., 1995, p. 298) The map is for frequency, not spatial order of adjacent parts of a depicted scene, and is thus at a considerably further remove from what we might take to be a depictive representation. So, if topographic mapping in auditory perception does not support depictive representation, then topographic mapping in visual perception by itself does not support depictive representation, either. Thus, one might conclude, visual mental images does not depend on depictive representations.

The problem just described hinges on two assumptions: (A1) Depictive representations require something more than just order-preserving, and (A2) topographic representations in the auditory cortex are not depictive representations. In what follows I first elaborate these two assumptions, arguing that one should accept them and take the above problem seriously.

Next, I examine Tye's theory, arguing that his theory only partly explains the depictive nature of visual mental images. This paper concludes with my proposal that we need to divide the problem about the depictive nature of mental imagery into two parts:

(Q1) The format problem: What is the format of mental imagery?

(Q2) The representation problem: What are the conditions by virtue of which a representation become a depictive representation?

Regarding the first question, I argue that there exists a topographic format in the brain, and one should abandon the talk about depictive format of image representation. My answer to the second question is that one needs a content analysis of a certain sort of topographic representations to make sense of depictive mental representations, and a topographic representation becomes a depictive representation by virtue of its content rather than its form.

## **Two Assumptions**

That depictive representations require something more than just order-preserving is, as I take it, quite uncontroversial. Nevertheless, certain remarks are needed to clarify what is at issue and what is not. First, what is at issue is not about what qualifies as a representation, but whether or not there are depictive mental representations. Second, for the present purposes (A1) is strong enough to support the view that topographic representations are not depictive representations, though we still do not have a fully developed and complete theory of depictive representations. And I will not attempt one in the following discussion. That would be far beyond the scope of this paper. That being said, let me elucidate the distinction between depictive and topographic representations by the following example. Consider a square. The depiction of this shape normally places its sides parallel to the gravitational frame, that is, with its top side and bottom side parallel to the horizontal ground and the other two sides aligned with the vertical. Now rotate the square 45 degrees. Observe that the shape now looks like a diamond. This simple transformation preserves any

spatial order of adjacent parts of the original shape, yet the resultant configuration depicts a diamond, rather than a square (Arnheim, 1974, pp. 98-103; Leyton, 1992, pp. 342-346; Mach, 1897/1959). This square-diamond phenomenon shows that depiction cannot be merely topographic representation.

Consider now the second assumption that topographic representations in the auditory cortex are not depictive representations. In defending this assumption it will be helpful to make it clear at the outset that depiction is primarily a visual phenomenon. (1) If representing visual qualities is a necessary feature of depiction, then representations of sound frequencies cannot be depictive representations. However, the argument for (A2) cannot be so straightforward. We know that in synesthesia, music or voices may be perceived to have shapes, textures, and colors. We also know that people normally judge that low-frequency sounds express larger visual size than do high-frequency sounds, and may describe a soprano's tone as sharper than an alto's (Marks, 1996; Marks, Hammeal & Bornstein, 1987). Both synesthesia and the perceptual correspondence across visual and auditory modalities seem to suggest that, perhaps, an auditory representation can be a depictive representation via its associated link to visual perception. Notice, however, that the associated link between auditory and visual qualities is asymmetric in the sense that sounds are perceived to have, or judge to express, certain visual qualities, but the perceptual dimensions of visual experience are not perceived to have, or judged to express, auditory qualities. Which suggests that an auditory representation can become a depictive representation only with reference to its associated visual qualities. So, any auditory representation that can be realized in some medium with its representational character independent of visual representations is not a depictive representation. The topographic representations in the auditory cortex clearly satisfy this condition. Thus, I conclude, they are not depictive representations. So (A2) should be accepted.

With (A1) and (A2) in hand, it is clear that one should take seriously the problems described in the last section. I turn now to Tye's theory and examine how he can respond to the problems (Q1) and (Q2).

### **Tye's Theory**

Tye (1991, pp. 90-91) proposes the following account of what a visual mental image is:

A mental image of an F (though no one F in particular) is a symbol-filled array to which a sentential interpretation having the content "This represents an F" is affixed. The array itself, which is very like Marr's 2 1/2-D sketch, occurs in a fixed medium resembling Kosslyn's visual buffer, and is generated from information in long-term memory that consists in part of viewer-centered information about the visual appearances of Fs and in part of information about the spatial structure of Fs.

Let me unpack this formulation by analyzing its four components in the following order: (1) the visual buffer, (2) the viewer-centered representations, (3) Marr's 2 1/2-D sketch, and (4) mental images as interpreted symbol-filled arrays. I shall restrict my exposition to a number of basic points which concern us here.

### **The Visual Buffer**

To understand what the visual buffer is, it will be helpful to start with the following analogy Kosslyn proposed. Visual mental images are conceived of on the model of displays on a cathode-ray tube screen attached to a computer; such displays are generated on the screen by a computer program (Kosslyn, 1980, pp. 5-9). In this model, the internal representations that underlie our experience of having an image are functionally analogous

to the displays on the screen. And the visual buffer in which the internal representations are encoded is functionally analogous to the screen. The visual buffer in this functional analysis has at least the following two properties. (V1) Its components are individual cells capable of either being activated or not. Those cells, when activated, represent single spatial points on the surface of the imaged object.

(V2) Those cells are structured into an array or matrix, with adjacent cells representing adjacent parts of the surface of the imaged object. (2) The properties (V1) and (V2) clearly make the visual buffer capable of supporting topographic representations discussed above. This result, buttressed by the empirical discovery that human visual cortex includes topographically mapped areas, shows that the visual buffer is indeed psychologically real (Kosslyn, 1994, pp. 12-20).

### **Viewer-Centered Representations**

Topographic mapping in the visual system, as we have noted, preserves only spatial order of adjacent parts, which is not sufficient to support depictive representations. To remedy this situation, it is important to note that in visual perception the mapping is from the retinal image onto the brain. When one sees an object, one always sees it from a particular point of view. Which means that patterns of activation in retinotopic maps are viewer-centered. One can then postulate that after mapping the viewer-centered information is preserved. The information clearly cannot be preserved merely in terms of topographic relations. Yet the information can be preserved in the intentional contents expressed in the topographic maps. And the constraint of preserving viewer-centered information in visual information processing explains why a topographic representation can be a

depictive representation. If this indeed is a case, we have the following scenario. For visual perception, the visual buffer is activated by processes that operate on information contained in the light striking the eyes. For mental images, the visual buffer is activated by generational processes that act on the viewer-centered information stored in long-term memory about the appearances of objects and their spatial structures. Note that this distinction between visual perception and visual mental imagery does not exclude the possibility that imagery is an integral part of how perception operates. To make the above points clearer, we need to examine the third component in Tye's formulation.

### **Marr's 2 1/2-D Sketch**

Marr's (1982) notion of the 2 1/2-D sketch is very complicated, but the basic points which concern us here are not. I shall restrict my discussion to those basic points. According to Marr, a 2 1/2-D sketch is a representation of visible surfaces in three dimension. This notion can be illustrated in the following way. Suppose I am looking at a circular object, say a coin on my desk, which is tilted away from me. The appearance of this object as it presents to me is dependent upon the point of view from which I see it. And I see the object as being circular and as being tilted away from me. But let us specify the appearance of the object without mentioning the fact that it seems to be tilted away from me, that is, without any reference to its outward distance and orientation in my visual field. The apparent shape of this object then looks elliptical. The information of this apparent shape, presumably, can be encoded in an image-based representation in the form of an array in my visual system, with adjacent cells in the array representing the adjacent parts of the apparent shape.

How this can be done needs not concern us here. The important point is that the function of the 2 1/2-D sketch is to assign portions of the image to the surfaces in the environment and to specify the distance and orientation of those surfaces relative to the viewer. This can be done by letting the cells in the array contain symbols representing such features as orientation and depth of the patch of surface relative to the viewer (Tye, 1991, pp. 81-83).

Having said this, it is clear now how the viewer-centered information can be preserved in the intentional contents expressed in the topographic maps. Activity in any given cell in the maps can be conceived of as containing descriptive labels, which represent the following local features: presence of a tiny patch of surface, orientation and depth of the patch of surface, determinate shade of color, texture, and so on. This result, in my view, can in principle explain why a topographic representation can become a depictive representation, though Tye didn't directly address this issue.

### **Mental Images As Interpreted Symbol-Filled Arrays**

Tye is quite aware of the fact that mental images are pre-interpreted and are composed of organized units rather than arbitrary parts. That is why he introduces the notion that descriptive labels that provide a more specific content are appended to the array at different levels of groupings. For example, some labels are attached to single cells representing orientation of the patch of surface and so on, as described above. Further labels are attached to groups of cells representing nonlocal features, such as a circle or a square, and in more complex cases, a duck or a rabbit. He argues that the labels are linguistic in form (Tye, 1991, pp. 90-102; see also Raffman, 1997, p. 190). But he also argues that the groupings and how the descriptive labels are appended to the array can be done at a nonconceptual level (Tye, 1995, pp. 122-123). A problem immediately arises concerning the linguistic nature of the labels and its relation to the nonconceptual activity of appending the labels to the array. I will return to this problem in the next section.

### **The Imagery Format and Depictive Representation**

Now is the time to answer the two problems set out at the beginning of this inquiry. Consider first the problem (Q1). The format of image representation specifies what types of representational elements can be used as its components and the way in which these components can be arranged and processed. The format can be conceived of as constituted by topographically structured arrays or matrices to which labels of certain types can be attached. The groupings of the cells and the labelling activities form an overlapping and nested hierarchy, depending on how the patterns of activation of the cells are formed. In the case of visual mental imagery, the labels are of the type which can be interpreted as having the intentional contents which make the labels stand for certain determinate surface features from a particular point of view. In the case of auditory mental imagery, the labels are of the type which can be interpreted as having the intentional contents which make the labels stand for certain sound properties. We don't need the so-called depictive format of mental imagery, which is beset by problems of the sort we have discussed (cf. Kosslyn, 1980, pp. 31-35; 1983, p. 32-37; 1994, pp. 12-20).

Consider the problem (Q2). It should now be clear that topographic representations can become depictive representations by virtue of the intentional contents of certain appropriate labels that are attached to them. Tye argues that the labels are linguistic in form, but the grouping and labelling activity can be done at a nonconceptual level. This is possible once we note that the internal resources that are deployed when one is using linguistic entities need not be parts of a linguistic system (Bechtel & Abrahamsen, 1991, Ch. 7; Clark, 1993; Rumelhart, 1992; Rumelhart et al., 1986; Teng, forthcoming). It seems to me, however, that Tye glosses over the different levels of grouping and labelling activity by saying that all the labels are linguistic in nature. If the image is a rabbit, perhaps a linguistic label, say, 'This is a rabbit' will be attached to it. The linguistic labels, presumably, constitute a representational scheme, independent of the topographically organized format of image representations. But for labels that indicate presence of a tiny patch of surface, orientation of the patch of surface, and so on, it is the topographic medium, rather than some linguistic system, that accounts for their organization and how they can be grouped into consistent parts of an image representation. Those labels do not constitute a representational scheme

independent of the activity of the cells in the topographically organized array. They are embodied in the activity. It is under certain functional analyses that we talk of the activities and the groupings of the cells as if some descriptive labels were attached to them. And what happens when one attaches a descriptive label to an image representation is that one has the representation whose content is then brought under the given concept expressed by the descriptive label. Consider an image of a square, for example. The image has its content encoded in the topographically organized array; the pattern of the activities of the cells indicates that a square is represented. The pattern and the activity of each cell, presumably, are nonconceptual. One can further bring the pattern under the concept expressed by a linguistic label, say, 'This is a square', for further information processing.

## Notes

\* This research was supported by a grant from National Science Council, ROC, NSC87-2411-H-194-009.

(1) See Budd, 1993; Hopkins, 1995; Peacocke, 1987. It is also interesting to note that depictive experience is possible through touch as well as sight; see Kennedy, 1993.

(2) A number of empirical studies conducted by Kosslyn and his collaborators show that the visual buffer also has a limited extent and a specific shape, with a limited resolution which is sharpest near the center of the medium and more degraded toward the periphery, and so on. These empirical results are very interesting, but for the present purposes I restrict my discussion to the properties (V1) and (V2) of the visual buffer (Kosslyn, 1980, pp. 139-141; 1994, pp. 85-87).

## References

Arnheim, R. (1974). *Art and visual perception: A psychology of the creative eye, the new version*. Berkeley, CA: University of California Press.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Cambridge, MA: Basil Blackwell.

Budd, M. (1993). How pictures look. In D. Knowles & J. Skorupski (Eds.) *Virtue and taste: Essays on politics, ethics and aesthetics, in memory of Flint Schier*. Oxford: Blackwell.

Churchland, R. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Clark, A. (1993). *Associative engines: Connectionism, concepts, and representational change*. Cambridge, MA: MIT Press.

Guzeldere, G. (1995). Is consciousness the perception of what passes in one's own mind. In T. Metzinger (Ed.) *Conscious experience* (pp. 335-357). Thorverton, UK: Imprint Academic.

Hopkins, R. (1995). Explaining depiction. *Philosophical Review*, 104, 425-455.

Kennedy, J. M. (1993). *Drawing & the blind: Pictures to touch*. New Haven: Yale University Press.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge: Harvard University Press.

- Kosslyn, S. M. (1983). *Ghosts in the mind's machine: Creating and using images in the brain*. New York: Norton.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge: MIT Press.
- Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA: MIT Press.
- Mach, E. (1959). *The analysis of sensations*. New York: Dover. (Original work published 1897).
- Marks, L. E. (1996). On perceptual metaphors. *Metaphor and Symbolic Activity*, 11, 39-66.
- Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, 52 (1, Serial No. 215).
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.
- Peacocke, C. (1987). Depiction. *Philosophical Review*, 3, 383-410.
- Raffman, D. (1997). Review of M. Tye, *Ten problems of consciousness*. *Journal of Consciousness Studies*, 4, 188-190.
- Rumelhart, D. E. (1992). Toward a microstructural account of human reasoning. In S. Davis (Ed.) *Connectionism: Theory and practice*. Oxford: Oxford University Press.
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart and the PDP Research Group (Eds.) *Parallel distributed processing (Vol. 2): Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L. and Rissland E. L. (1995). *Cognitive science: An introduction*. Cambridge, MA: MIT Press.
- Teng, N. Y. *The language of thought and the embodied nature of language use*. *Philosophical Studies*, forthcoming.
- Tye, M. (1991). *The imagery debate*. Cambridge, MA: MIT Press.
- Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.