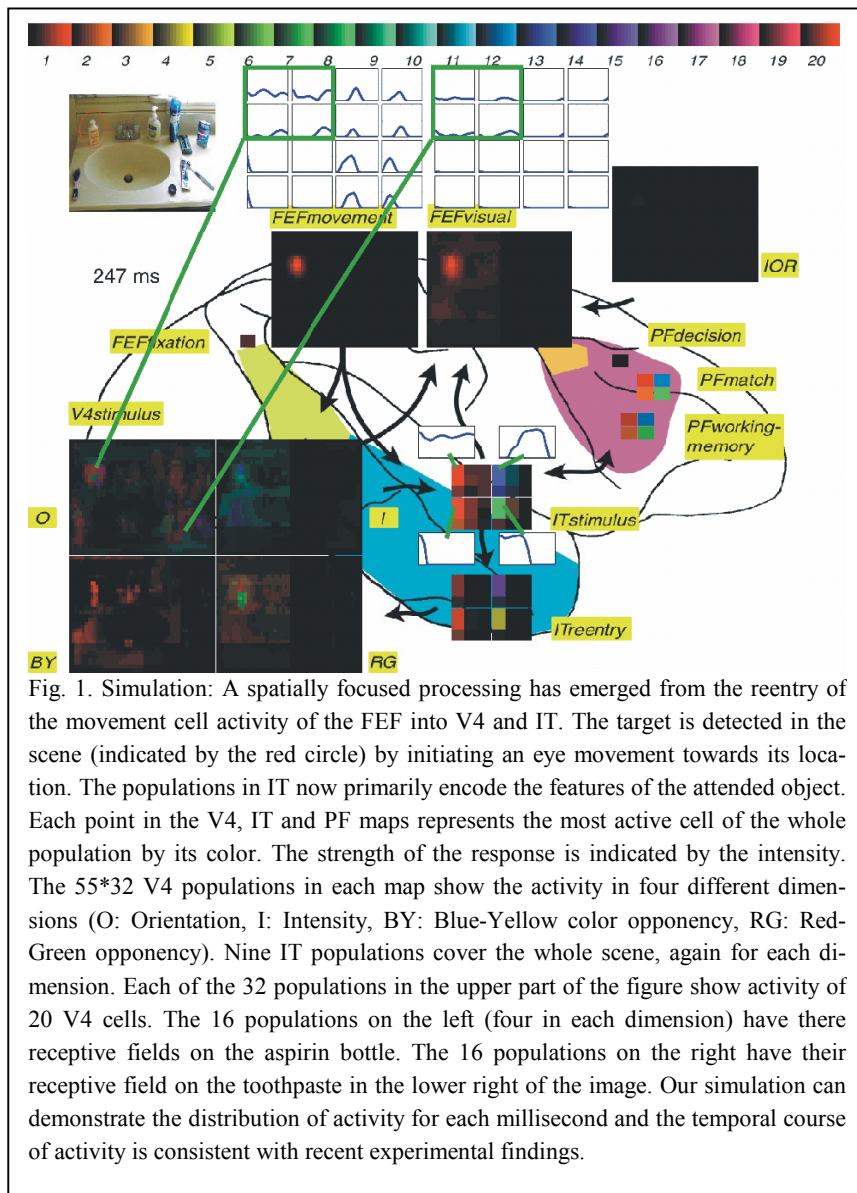


How does attention facilitate object recognition?

Fred H. Hamker
CALIFORNIA INSTITUTE OF TECHNOLOGY
Division of Biology 139-74
Pasadena, CA 91125

Object recognition generally implemented in a hierarchical bottom-up process leads to a strong overlapping of populations encoding features belonging to different objects. Although experiments indicate that some scenes allow the detection of categories during very brief presentations, ambiguities in cell populations encoding features within the same receptive field limits recognition in non-segmented scenes like natural images. We show how attention resolves these ambiguities over time and facilitates object recognition.



According to our approach, object recognition is a dynamic, constructive process in which attention occurs as an emergent result of feedback from later areas into earlier areas (Fig. 1). We have previously shown that this model can simulate the results of various attention experiments. The areas simulated in our model, their interactions and the temporal activity of the cells are grounded in neuroanatomy and neurophysiology. The model consists of the following key aspects. Features encoding the scene are modeled by a population code. Stimulus-driven saliency is implemented by center-surround operations, as suggested by Itti and Koch (2000). However, instead of constructing an explicit saliency map, V4 cells are modulated according to the saliency. Populations encoding features within the same receptive field compete with each other as suggested by the biased competition framework (Desimone and Duncan, 1995). Feedback from higher areas implements an input gain control. Areas responsible for action preparation feed back into extrastriate visual areas and implement a spatially organized gain control.

The task the model is tested with, is to detect a previously shown object within a natural scene by selecting it for an eye movement. We predict that such a task is solved by the brain as follows. Cells in early stages (V1 – V4) construct a set of features from the scene within spatially organized maps and encode the visual saliency by the strength of their firing rate. Due to inherent competition salient stimuli are given preferential processing. As soon as the signals enter IT, the top-down knowledge about the object starts to modulate the activity. If the target sufficiently discriminates

from the background it can be reported without any specific selection. However, if the overlap of feature detectors in the same receptive field is too strong, the second phase is essential. Feedback from IT to V4 emphasizes the features of interest by parallel feature-based attention. Areas responsible for eye-movement preparation like the FEF, which are first driven by the saliency of the stimuli, now receive information about the task relevance of a stimulus. Pre-motor cells feed back into extrastriate areas and a focused processing known as spatial attention emerges. Now the full detail of the object of interest is encoded by IT cells and can be reported.

