

events by a pathway that is distinct from—and upstream of—the caspase inhibition that is responsible for its survival-promoting effect. The fact that ITA-protected neurons have extensive neurites suggests that ITA may be involved in these effects. It will be interesting to examine this more closely, and in particular to determine whether ITA-protected neurons maintain normal levels of protein synthesis and fail to show induction of c-Fos and c-Jun following NGF withdrawal—if so, this might suggest that ITA can have effects in addition to caspase inhibition.

Although ITA is likely to work by inhibiting caspases, the identities of the specific caspases affected are not yet clear. Other IAPs such as XIAP, IAP1, and IAP2 bind to and inhibit caspases-3, -7, and -9, whereas they have no effect on caspases-1, -6, -8, and -10 (ref. 13). NGF withdrawal induces the activation of caspase-2 but not caspase-3 (ref. 12). Antisense RNA targeted to caspase-2 transcripts inhibits the death of sympathetic neurons induced by NGF withdrawal¹⁴, although sympathetic neurons from caspase-2 knockout mice are still sensitive to NGF deprivation¹⁵ (perhaps reflecting developmental compensation by another caspase). Thus, although caspase-2 may be involved, it is unclear how many caspases contribute to sympathetic neuronal cell death. It will be interesting to determine whether ITA or other members of the IAP family can inhibit caspase-2, and whether other IAPs can inhibit the death of sympathetic and sensory neurons induced by NGF withdrawal.

There is evidence that NGF withdrawal activates two parallel pathways that together lead to death. One pathway requires Bax, a pro-apoptotic member of the Bcl-2 family². Sympathetic neurons from mice lacking Bax do not undergo apoptosis in response to NGF deprivation (although they do exhibit somatic atrophy and a reduced number of neurites)¹⁰. Bax appears to act by promoting the leakage of cytochrome c from the mitochondria into the cytoplasm, a critical late step in the apoptotic pathway⁵. Cytochrome c alone, when injected into the cytoplasm, is not able to kill neurons that are maintained in the presence of NGF. But following withdrawal of NGF, the neurons become highly sensitive to cytochrome c injection, a state that has been termed 'competence-to-die'¹⁰. A similar degree of sensitization is seen in wild-type or Bax-deficient neurons, implying that NGF

withdrawal is acting through a Bax-independent pathway, in addition to its Bax-dependent effect on cytochrome c release. The new findings of Wiese *et al.* raise the possibility that this pathway may correspond to the downregulation of ITA.

When sympathetic neurons mature, they cease to be sensitive to NGF withdrawal, perhaps because they express a survival-promoting gene (or genes) in a constitutive manner. However, ITA does not appear to be involved in this process; Wiese *et al.* showed that ITA mRNA is detectable *in vivo* only during embryogenesis, when neurons are still dependent on NGF for survival.

Finally, several other recent studies have implicated IAPs in neuronal survival⁵. For example, mutation of the gene for the IAP family member NIAP was shown to contribute to the pathogenesis of spinal muscular atrophy, a devastating genetic disorder of childhood that is characterized by the progressive degeneration of motor neurons. There is also evidence that the induction of NIAP may contribute to neuronal survival after ischemic brain damage. Thus, the IAP family of proteins, in addition to playing a role in normal development, may also be key

determinants of survival or death under pathological conditions.

1. Burek, M.J. & Oppenheim, R.W. *Brain Pathol.* **6**, 427–446 (1996).
2. Deshmukh, M. & Johnson, E.M. *Mol. Pharmacol.* **51**, 897–906 (1997).
3. Wiese, S. *et al. Nat. Neurosci.* **2** 978–983 (1999).
4. Birnbaum, M.J., Clem, R.J. & Miller, L.K. *J. Virol.* **68**, 2521–2528 (1994).
5. Deveraux, Q.L. & Reed, J.C. *Genes Dev.* **13**, 239–252 (1999).
6. Crowder, J.C. & Freeman, R.S. *J. Neurosci.* **18**, 2933–2943 (1998).
7. Ozes, O.N. *et al. Nature* **401**, 82–85 (1999).
8. Romashkova, J.A. & Makarov, S.S. *Nature* **401**, 86–90 (1999).
9. Maggirwar, S.B., Sarmiere, P.D., Dewhurst, S. & Freeman, R.S. *J. Neurosci.* **18**, 10356–10365 (1998).
10. Deshmukh, M. & Johnson, E.M. *Neuron* **21**, 695–705 (1998).
11. Gagliardini, V. *et al. Science* **263**, 826–828 (1994).
12. Deshmukh, M. *et al. J. Cell Biol.* **135**, 1341–1354 (1996).
13. Roy, N., Deveraux, Q.L., Takahashi, R., Salvesen, G.S. & Reed, J.C. *EMBO J.* **16**, 6914–6925 (1997).
14. Troy, C.M., Stefanis, L., Greene, L.A. & Shelanski, M.L. *J. Neurosci.* **17**, 1911–1918 (1997).
15. Bergeron, L. *et al. Genes Dev.* **12**, 1304–1314 (1998).

News On Views: Pandemonium Revisited

Michael J. Tarr

How do we recognize objects from different viewpoints? A new model, based on the known properties of cortical neurons, may help resolve this long-standing debate.

How does the brain recognize objects? Consider that the visual information that falls on our retinae is a 2D projection of the actual 3D world. Moreover, these projections are rather unstable in that they vary, sometimes dramatically, with changes in an object's orientation, position, distance, illumination and configuration. Thus, what we mentally remember about an object at one moment in time may not match what we see at some later moment. To make matters even worse, our visual memories must generalize not only over changes in viewing conditions, but over different instances of an object class

—for instance, when looking at cars, we must recognize a new model as a member of the class, even though we have never seen it before.

The mechanisms by which object recognition is accomplished has been a popular question in the brain sciences almost since their inception. Almost everyone agrees that neural representations of previously-seen objects — visual memories — are compared to newly-created neural representations of visual inputs. On the other hand, it often seems that those studying object recognition can agree about little else. The differing stances between researchers can be effectively divided into two camps: 'view-based' theorists and 'structural-description' theorists^{1,2}. The former hypothesize that the neural representations used in the recog-

Michael J. Tarr is in the Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, Rhode Island 02912, USA.
e-mail: michael_tarr@brown.edu

niton process are tied to the appearance of objects as originally viewed — that is, from a specific viewpoint in which an object has actually appeared. In contrast, the latter hold that the neural representations that form the basis of the recognition process are organized into hierarchies of features or parts that are either partially or completely independent of any particular viewpoint. Which approach better captures the body of extant results has been the subject of heated debate over the past several years^{1,2}. Resolving this debate, however, has been rather difficult, in that the best-specified theory, a structural-description model in which objects are represented as collections of 3D volumes³, is not particularly consistent with either neural⁴ or behavioral⁵ data. On the other hand, the strongest point in favor of view-based models has been a set of experimentally-generated phenomena^{4,5}, rather than a detailed theory that can account for these data. Indeed, view-based models have remained relatively simplistic entities which do not generalize very well from familiar to novel viewing conditions⁶. As such, the current debate has reached a bit of stalemate.

In this issue, Riesenhuber and Poggio⁷ present a new model that has the potential to remove the impasse. (Because their term for the model — “hierarchical model of object recognition” is rather cumbersome, I have dubbed it ‘HMAX’ which roughly stands for ‘Hierarchical Model And X’ — where X is a highly non-linear maximum operation.) What they present is a computational implementation of a view-based theory of object recognition that rests heavily on the functional architecture of the cortical temporal lobe stream — the part of the brain that is believed to mediate visual object recognition. The functional properties of this neural pathway were first described by Hubel and Wiesel⁸, who demonstrated that information processing in this part of visual cortex proceeds in a hierarchical fashion. Specifically (see Fig. 1), the cortical temporal lobe stream progresses from local responses driven by simple stimulus properties — for example, oriented lines — to more global responses driven by more complex stimulus properties — for example, bars of particular lengths and widths⁸. More recently, it has been demonstrated that this hierarchy continues into inferotemporal cortex (IT), where cells presumably combine the responses of earlier cortical areas into highly specific pattern detectors — for example, neurons that respond most strongly to

complex shapes⁹, individual faces¹⁰, or objects⁴. Similarly, in HMAX, Riesenhuber and Poggio implement a hierarchy of conjunctions and disjunctions of progressively more and more complex feature combinations, culminating in object-specific units that are ‘view-tuned’ — that is, object representations that respond most strongly to a single viewpoint. Correspondingly, the same sensitivity to viewpoint is found in the neurophysiology of IT, where the vast majority of neurons that are object-specific appear to respond preferentially to a particular viewpoint (although there are also some neurons that respond equally well to any view)^{4,10}. At the same time, view-tuned cells typically respond in an invariant manner to changes in size or distance. Thus, the challenge is to develop a theory that predicts viewpoint-dependent performance for recognizing known objects, but with little or no scale- or position-dependence.

Riesenhuber and Poggio’s model shows precisely this sort of response pattern, being robust over changes in scale or position — yet, as already mentioned, the units coding for specific objects within HMAX are highly viewpoint-dependent. The primary reason for this behavior is that HMAX relies on a non-linear maximum operation (‘MAX’) for combining feature responses at one stage in order to create more complex feature detectors at a subsequent stage. In the model, the use of the MAX operator means that the strongest signal among features feeding into a unit at the next layer will determine the response of this unit. This method for pooling responses also allows Riesenhuber and Poggio’s model to perform well even with images containing more than one object. As with the pattern of responses for feature detectors in HMAX, the non-linear MAX mechanism for pooling afferents seems to have an analog in neurophysiology, possibly arising from lateral inhibition between cells at each processing stage.

In contrast to the wide explanatory power of HMAX, this group’s earlier model of visual recognition⁶ dealt almost exclusively with techniques for using viewpoint-specific object representations to achieve viewpoint-invariant recognition. Their solution, typical of nearly all instantiations of the view-based approach^{1,5}, was to encode multiple views of each known object, so that almost any new view would likely be near to a familiar view. Consequently, recognition performance would be relatively viewpoint-invariant. Adding additional

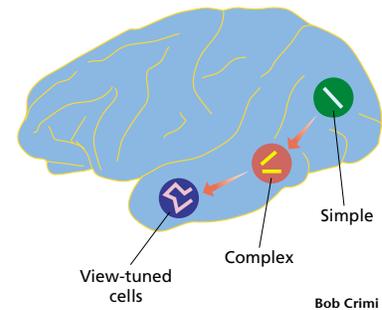


Fig. 1. The temporal stream within visual cortex processes information in a hierarchical fashion. Earlier visual areas are most responsive to simple stimulus patterns such as oriented lines. In contrast, later visual areas such as inferotemporal cortex (IT) have recently been shown to be sensitive to complex shapes or specific objects in specific views. It is thought that these more complex object representations are constructed out of progressively more and more complex feature detectors.

power to such systems, known views of a given object or object class are not treated completely independently of one another; rather they are ‘pooled’ to form ‘multiple-views’ object representations. For example, in Poggio and Edelman’s implementation⁶, a computational network learned specific views of novel stimuli, but then was able to accurately recognize the same stimuli in new viewpoints by interpolating between the appearance of two or more known views for a particular object. Other view-based models have proposed similar normalization mechanisms, for instance, aligning a description of the input image with a known view¹¹ or accumulating evidence across a set of viewpoint-specific feature detectors¹². For all of these models, the critical prediction is progressively poorer generalization — in the form of either weaker neural responses or diminished recognition performance — with increasing distance between a test view and any known view of an object. At the same time, whereas view-based models have tacitly acknowledged that scale- and position-invariance are desirable properties that are suggested by the extant data, they have not actually proposed mechanisms for achieving either type of invariance. Even worse, implementing such invariances would be difficult given the kinds of features typically used to construct viewpoint-specific representations, for example, the (X,Y) image coordinates of an object’s vertices⁶. Indeed, the use of such highly specific coding schemes led to some of the strongest criticisms of

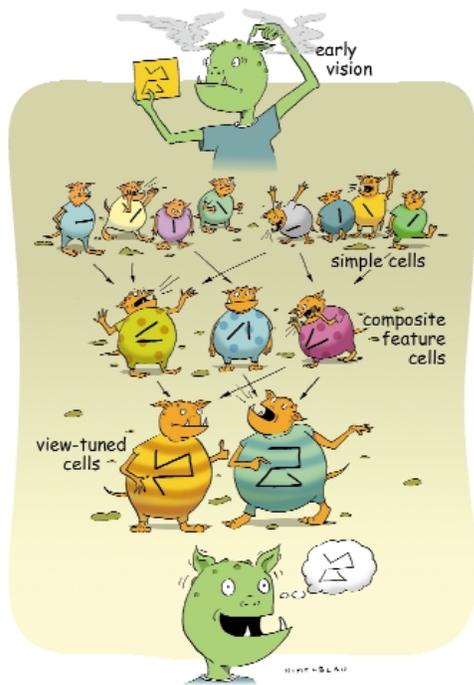


Fig. 2. Pandemonium revisited. The HMAX model introduced by Riesenhuber and Poggio bears some resemblance to Selfridge's 'Pandemonium' model. Although both models pool information over progressively more complex detectors, the detectors in HMAX are purely visual and are based on what we now know regarding the response properties of successive layers in temporal cortex.

view-based models, for example, that they would fail to generalize across instances of a class².

By way of comparison, structural-description models appear to be inherently more robust across changes in the image. This is because they commonly assume that object representations are the end-result of a reconstruction process whereby the 3D structure of the object is recovered from the 2D image. For example, in one of the earliest structural-description theories, object structure was captured by 3D parts that are spatially related to one another in an object-centered coordinate system¹³. In more recent instantiations, 3D parts are qualitative entities whereby a consistent description is arrived at so long as the same configuration of parts can be identified in the image³. In both cases, the key point is that the resulting structural description is no longer tied to the appearance of the object in the image, but rather describes the object independently of viewing parameters such as scale, position or viewpoint. Thus, the critical prediction is robust generalization — in the form of stable neural responses or uniform recognition

performance — with changes in test views relative to known views of objects. In principle, strong invariance across image transformations is a desirable characteristic. To this point, however, neural and behavioral studies seem to suggest a different conclusion: generalization across view-points in primate vision does not appear to be strongly invariant in terms of either neural responses⁴ or behavior⁵. Thus, structural-description models have been criticized as inconsistent with the extant data¹.

The HMAX model of object vision appears to be a significant advance beyond these tried and true approaches. Why? First, HMAX generally builds on what is actually known regarding the architecture of temporal stream processing in the primate visual cortex (Fig. 1). As mentioned earlier, the organization

of this pathway is believed to be hierarchical, building progressively more and more complex pattern detectors based on the inputs of earlier visual layers. At the end of this pathway, the responses of neurons in IT are markedly specific — often being tuned to single objects in specific viewpoints. HMAX follows a similar hierarchical progression in which inputs are combined from one level to the next, ending in view-tuned units that are quite similar to the cells found in IT. Consequently, HMAX stands a reasonable chance of being neurally plausible. Moreover, as our knowledge regarding the specifics of temporal stream processing increases, new findings can be readily incorporated into HMAX and the resulting predictions can then be compared to actual data (for instance, see the supplementary material provided at http://neurosci.nature.com/web_specials/).

Second, HMAX's behavior appears to be highly consistent with neural and behavioral findings regarding sensitivity to changes in scale, position, and viewpoint. For instance, Riesenhuber and Poggio find that a specific view-tuned unit in HMAX responds in an invariant

manner across changes in object size or object position. On the other hand, the same unit exhibits a highly view-dependent response with increasingly poorer generalization as its preferred object is rotated further from the training viewpoint. This latter pattern of performance is characteristic of both single-unit recordings in IT^{4,10} and psychophysical studies of human object recognition⁵ — results that strongly suggest the existence of view-based object representations. At the same time, HMAX is scale- and position-invariant, as in structural-description approaches. Thus, unlike earlier view-based models, HMAX is able to retain specificity regarding the appearance of an object in terms of viewpoint, but not at the expense of being overly sensitive to translation or changes in size. Beyond these standard image transformations, HMAX can provide a good account of neural responses across a wide range of more subtle stimulus conditions. For example, view-tuned units in HMAX sometimes show a strong response to 'pseudo-mirror' test views, that is, 180° depth rotations where the 2D silhouette of the object in the test view is simply a mirror-reflection of the 2D silhouette of the object in the known view. Again, the same generalization pattern has been obtained in single-unit recordings in IT⁴. Similarly, the HMAX architecture is such that robust recognition in scenes containing visual noise or more than one object is a real possibility (something not addressed in any previous model).

Third, HMAX incorporates certain properties of structural-description models that are highly desirable in any theory of pattern recognition. Specifically, the object representations used in HMAX are compositional¹⁴, in that they are built up from conjunctions and disjunctions of simpler features according to the MAX-like operation. The advantage of this approach is that the input image is divided into much more stable elements that are far more likely to be detected when there are changes in the image; that is, there is a unique interpretation for each known object because the lower-level elements that are used to represent them can be reliably extracted from the image. Riesenhuber and Poggio's model demonstrates that compositionality is not the exclusive domain of the structural-description approach; HMAX is a view-based implementation that is compositional — that is, a hierarchy in which the complexity of features increases with each stage in the model. There are,

however, some important differences in the way HMAX realizes the composition of features. For instance, the features used in HMAX are not encoded relative to one another in terms of their positions in space — they are combined, but without reference to where individual features fall in the image. This type of coding may seem problematic at first glance — one reaction is that HMAX could be ‘fooled’ into identifying a scrambled image with the appropriate features as an instance of a known object. Luckily, the set of features forming each view-tuned unit is ‘over-complete’, that is, HMAX represents each object with an extremely large collection of different features that is highly redundant. Because each object is extremely over-specified, the odds of generating a scrambling in which all of the features describing an object are preserved is essentially zero. Moreover, this same coding redundancy allows HMAX to perform well at recognizing objects across configural deformations or across class instances. In both instances, although some of the features describing an object or class might be different, there are still many features that are likely to be consistent between what was learned and what is now being recognized.

Finally, HMAX offers an advance in how to pool over feature responses at each stage of processing. Earlier models that have proposed summing over differently weighted afferents to create progressively more complex detectors have typically relied on linear summation¹². HMAX, in contrast, relies on the non-linear MAX operation. The MAX approach provides much of the robust generalization HMAX exhibits across changes in size and with images filled with visual clutter — for example, other objects in the scene. Importantly, MAX-like mechanisms are reflected in recent neurophysiological results — specifically, it has been observed that the response of a neuron to two features is related to its strongest activation to each of the features alone. Other results also support the idea that the responses of IT cells are often highly non-linear.

Interestingly, HMAX is in many ways a throwback to one of the earliest models of object recognition, a system referred to as “Pandemonium” that was proposed by Selfridge¹⁵ in 1959. Pandemonium relied on a series of local feature detectors (‘demons’) that scanned each image for the presence of specific patterns, for example, a particular configuration of lines or an oriented curve. The responses of these feature detectors

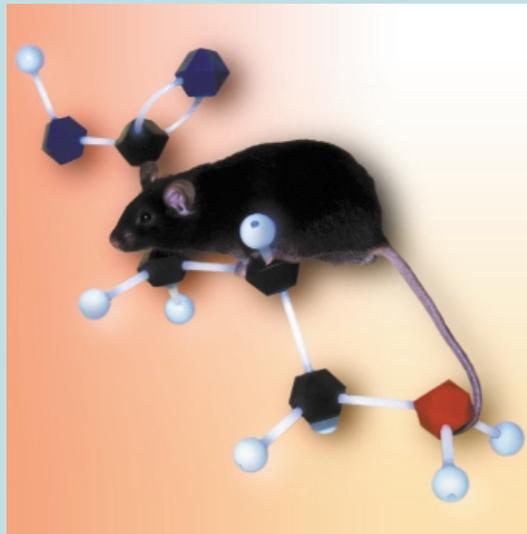
were then pooled to create more complex feature detectors at the next stage within the model. As with most theories that followed, summation across each layer of feature detectors was linear, with stronger responses at one layer being proportional to the number of appropriate features responding at the previous stage. Like Pandemonium, HMAX pools over progressively more complex detectors. However, unlike Pandemonium, these detectors are purely visual. Moreover, as illustrated in Figure 2, HMAX builds on what we now know regarding the response properties of successive layers in temporal cortex. Additionally, HMAX relies on the non-linear pooling of responses from one stage to the next. Even so, the general approach in Pandemonium and HMAX is markedly similar. Thus, everything old is new again. Yet because it so successfully addresses many of the perceived shortcomings of view-based models, HMAX truly is “News on Views.”

1. Tarr, M. J. & Bülthoff, H. H. *J. Exp. Psychol.: Hum. Percept. Perform.* 21, 1494–1505 (1995).

2. Biederman, I. & Gerhardstein, P. C. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1506–1514 (1995).
3. Biederman, I. *Psychol. Rev.* 94, 115–147 (1987).
4. Logothetis, N. K. & Pauls, J. *Cereb. Cortex* 3, 270–288 (1995).
5. Tarr, M. J., Williams, P., Hayward, W. G. & Gauthier, I. *Nat. Neurosci.* 1, 275–277 (1998).
6. Poggio, T. & Edelman, S. *Nature* 343, 263–266 (1990).
7. Riesenhuber, M. & Poggio, T. *Nat. Neurosci.* 2, 1019–1025 (1999).
8. Hubel, D. H. & Wiesel, T. *J. Physiol.* 160, 106–154 (1962).
9. Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. *J. Neurophysiol.* 66, 170–189 (1991).
10. Perrett, D. I., Rolls, E. T. & Caan, W. *Exp. Brain Res.* 47, 329–342 (1982).
11. Ullman, S. *Cognition* 32, 193–254 (1989).
12. Perrett, D. I., Oram, M. W. & Wachsmuth, E. *Cognition* 67, 111–145 (1998).
13. Marr, D. & Nishihara, H. K. *Proc. R. Soc. Lond. B Biol. Sci.* 200, 269–294 (1978).
14. Bienenstock, E., Geman, S. & Potter, D. in *Advances in Neural Information Processing Systems 9* (eds. Mozer, M. C., Jordan, M. I. & Petsche, T.) 838 (MIT Press, Cambridge, Massachusetts, 1997).
15. Selfridge, O. G. in *Symposium on the Mechanisation of Thought Processes* 513–526 (HM Stationery Office, London, 1959).

Creating teetotaler mice

The GABA_A receptor mediates the sedative and anxiety-reducing effects of ethanol. Although ethanol can potentiate GABA_A receptor function, the effects are variable, and it has been suggested that the receptor’s sensitivity to ethanol may depend on its phosphorylation state. In this issue (pages 997–1002), Hodge and colleagues show that one isoform of protein kinase C (PKC ϵ) mediates both the behavioral and biochemical response to ethanol by phosphorylation of the GABA_A receptor. The authors found that mutant mice lacking PKC ϵ show markedly reduced ethanol self-administration compared to wild-type mice, and are much more sensitive to the acute behavioral effects of ethanol. Biochemically, these effects are associated with an increased GABA_A receptor sensitivity. Because the mutant mice consume less ethanol, the authors suggest that inhibitors of PKC ϵ may be useful for treating alcoholism. Interestingly, the mutants also showed normal locomotor activity and did not appear to be sleepy or sedated. Inhibitors of PKC ϵ could therefore also be a non-sedating alternative to enhance GABA_A function when treating disorders such as anxiety or epilepsy.



Kalyani Narasimhan