

color. The signal that arrives from the retina is mainly determined not by the local light stimulus, but by the local contrast between a stimulus and its immediate surround [15,16]. This could be invoked as a basis for the constant hue loci (as well as for the illusion of Fig. 1). Such a neural representation captures the differences between different surfaces, but indicates nothing directly about the viewing conditions, such as the overall cast of illumination. The processes by which illumination and other parameters of the viewing conditions are reinstated in perception are not yet well elucidated, but it makes sense that these should require more integration with context than does perception of surface quality. And of course the brain has enough signals to support color experience of a dimensionality far higher than three. This is no heretical departure from the trichromatic theory: all that is necessary is that the signals underlying such perceptual dimensions combine information from different points in the visual field – and indeed the retinal ‘local contrast’ signals already do this.

To understand how the local-contrast signals within the image are integrated in perception is a problem that neurophysiologists are only beginning to address, although it is already clear that long-range interactions do occur, in primary visual cortex [17,18] as well as at later stages of processing [19]. The work of Ekroll *et al.* increases the challenge of this problem, by highlighting neglected subtleties in the way that perception of a surface depends on the viewing conditions. Gratifying correspondences have been revealed between neural firing rates and judgments of surface hue [19] and lightness [17] under varying viewing conditions, but a complete account of surface perception will require in addition some representation of the viewing conditions themselves, as well as of material properties such as glossiness [20], which also depend on fairly complex spatial computations.

References

- 1 Ekroll, V. *et al.* (2002) The natural center of chromaticity space is not always achromatic: a new look at color induction. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13352–13356
- 2 Shevell, S.K. (1978) The dual role of chromatic backgrounds in color perception. *Vis. Res.* 18, 1649–1661

- 3 Walraven, J. (1976) Discounting the background: the missing link in the explanation of chromatic induction. *Vis. Res.* 16, 289–295
- 4 Chichilnisky, E.J. and Wandell, B.A. (1995) Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching. *Vis. Res.* 35, 239–254
- 5 Whittle, P. (in press) Contrast colors. In *Colour Perception: From Light To Object* (Mausfeld, R. and Heyer, D., eds), Oxford University Press
- 6 Katz, D. (1911) *Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung*, Barth, Leipzig
- 7 Katz, D. (1935) *The World of Color*, Kegan, Paul, Trench, Trubner and Co
- 8 Mausfeld, R. (1998) Color perception: from Grassman codes to a dual code for object and illumination colors. In *Color Vision* (Backhaus, W.G.K. *et al.*, eds), pp. 219–250, De Gruyter
- 9 Bühler, K. (1922) *Die Erscheinungsweisen der Farben*, Fischer, Jena
- 10 Hagedorn, J. and D’Zmura, M. (2000) Color appearance of surfaces viewed through fog. *Perception* 29, 1169–1184
- 11 Brown, R.O. and MacLeod, D.I. (1997) Color appearance depends on the variance of surround colors. *Curr. Biol.* 7, 844–849
- 12 D’Zmura, M. *et al.* (2000) The colors seen behind transparent filters. *Perception* 29, 911–926
- 13 Faul, F. and Ekroll, V. (2002) Psychophysical model of chromatic perceptual transparency based on subtractive color mixture. *J. Opt. Soc. Am. Ser. A* 19, 1084–1095
- 14 Arend, L. and Goldstein, R. (1987) Simultaneous constancy, lightness and brightness. *J. Opt. Soc. Am. Ser. A* 4, 2281–2285
- 15 Lee, B.B. *et al.* (1990) Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers. *J. Opt. Soc. Am. Ser. A* 7, 2223–2236
- 16 Shapley, R.M. and Enroth-Cugell, C. (1984) Visual adaptation and retinal gain control. *Prog. Retinal Res.* 3, 263–346
- 17 MacEvoy, S.P. and Paradiso, M.A. (2001) Lightness constancy in primary visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8827–8831
- 18 Wachtler, T. *et al.* (2001) Nonlocal interactions in color perception: nonlinear processing of chromatic signals from remote inducers. *Vis. Res.* 41, 1535–1546
- 19 Zeki, S. (1980) The representation of colours in the cerebral cortex. *Nature* 284, 412–418
- 20 Adelson, E.H. (2001) On seeing stuff: the perception of materials by humans and machines. In *Proc. of the SPIE (Vol. 4299: Human Vision and Electronic Imaging VI)* (Rogowitz, B.E. and Pappas, T.N., eds) pp. 1–12, SPIE

1364-6613/03/\$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S1364-6613(03)00022-6

Taking the MAX from neuronal responses

Guillaume A. Rousselet, Simon J. Thorpe and Michèle Fabre-Thorpe

Centre de Recherche Cerveau & Cognition, CNRS-UPS UMR 5549, Faculté de médecine de Rangueil, 133, route de Narbonne, 31062 Toulouse cedex, France

By taking the MAX from their inputs, neurons in the ventral visual pathway might preserve their selectivity even when stimulated with natural scenes. This computational hypothesis has received recent direct physiological evidence from recordings of V4 neuronal

responses, in a recent study by Gawne and Martin (2002). Object vision might rely more heavily on parallel processing than generally thought.

In monkeys, neurons in infero-temporal cortex (IT) can respond selectively to a large range of complex stimuli as early as 80–120 ms after stimulus onset [1,2]. This fast processing is generally thought to be achieved through a

Corresponding author: Guillaume A. Rousselet
guillaume.rousselet@cerco.ups-tlse.fr

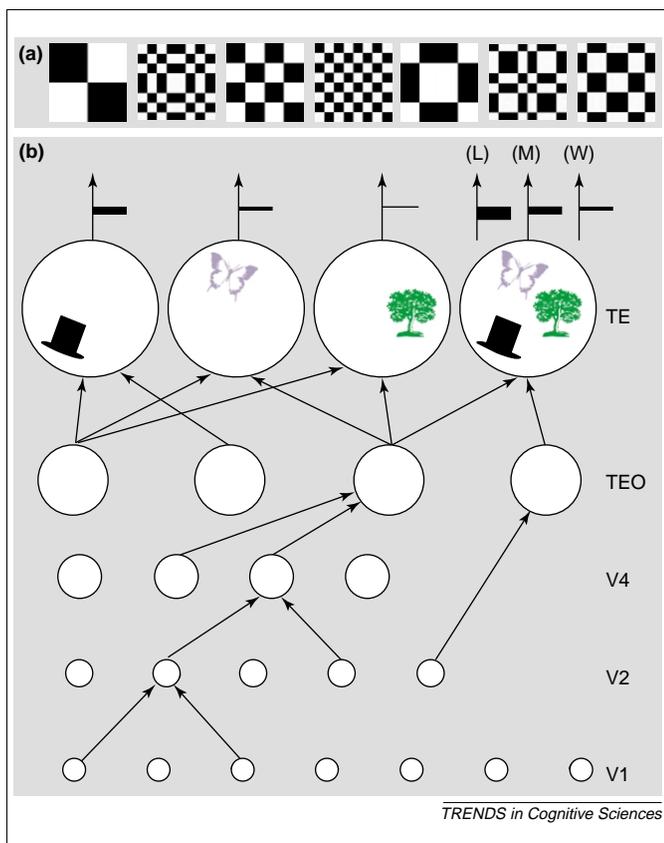


Fig. 1. Visual processing of simultaneously presented stimuli. (a) The seven Walsh patterns used by Gawne and Martin [7] to test stimulus interactions in V4 neuronal receptive fields (RF). (b) Schematic representation of the ventral visual pathway. Through a hierarchy of cortical areas, from V1 through temporo-occipital cortex (TEO) to temporal cortex (TE), complex and invariant object representations are progressively built by integrating inputs from lower levels. As RF size increases (presumably to allow translation and size invariance) neurons at higher levels typically receive inputs from more than one object at a time during natural scene perception. The response output of TE neurons to stimuli falling in their RF is illustrated by a vertical arrow, with the strength of the mean firing represented by a black rectangle. According to a linear model (L) of neuronal integration all the stimuli in its RF influence the response of the TE neuron, leading to 'blurred' representations. This potential problem might be solved if neurons implement a 'MAX' function, responding only to the stronger of their inputs (M). An alternative is that the neurons perform a weighted-average function, here illustrated with an equal participation of each input (W). Gawne and Martin found that V4 neurons appear to follow a MAX function, which would make the ventral pathway less sensitive to visual clutter than generally thought. This promising result needs to be tested in area TE where more complex features are processed.

hierarchy of cortical areas where complex and invariant object representations are progressively built by integrating inputs from lower levels. Although linear integration mechanisms might be sufficient to account for the processing of isolated elements, problems would arise in the case of natural scenes containing multiple objects. There is a progressive loss of spatial information as one progresses through the levels of the hierarchy due to an increase of receptive field size [3]. Consequently, neurons at the highest levels can potentially receive information from many different parts of the visual field, and in this case, a linear mechanism would lead to 'blurred' neuronal responses in the highest levels. Neuronal representations at the top of the hierarchy would contain mixed information from several objects, thus rendering recognition in clutter virtually impossible. However, the human visual system is able to categorize objects in natural scenes in parallel, within 150 ms [4,5].

How is this possible? It has been suggested that such performance might be explained by the use of a non-linear mechanism, the 'MAX' function [6]. The MAX function pools input activation so that the output of a neuron is determined by the strongest afferent; that is, the input with the maximum firing rate. This constitutes a form of a winner-take-all mechanism. By making a selection from various inputs, such a mechanism might potentially explain selective responses to objects in natural scenes.

Neurophysiological evidence for the MAX function

In a recent experiment, Gawne and Martin provided some direct experimental evidence in favour of this powerful theoretical idea [7]. They directly tested the MAX model by recording neurons from visual cortical area V4 of two monkeys. While animals performed a passive fixation task, two artificial visual stimuli (two-dimensional Walsh patterns; Fig. 1a) were flashed simultaneously in a neuron's receptive field. The neuronal responses recorded under simultaneous presentations were compared with those elicited by the two stimuli separately. They found that for a significant number of the recorded neurons, responses to the two stimuli presented simultaneously were well-fitted by a MAX model. Furthermore, the MAX model did much better than a linear model in almost all cases.

Importantly, the authors carefully verified that the capacity of the MAX model to fit the data could not be due to simple explanations like response saturation, or to cases where the neuron failed to respond to one of the stimuli. For some recorded neurons, no simple function could be found to explain their behaviour, suggesting that several non-linear mechanisms might be at work in the visual system. They also compared the MAX model with a weighted-average model that had been found to explain neuronal responses to simultaneous presentations of simple bars in V4 [8]. Although the weighted-average model did a much better job at explaining the data than the linear model, overall it was less successful than the MAX model. Thus, the MAX model seems to provide a good description of the behaviour of neurons in V4 when two stimuli are present in their receptive fields.

Explaining data inconsistent with the MAX model

So why did the data in the study by Reynolds *et al.* indicate a weighted-average model for V4 neuronal responses under similar experimental conditions? Gawne and Martin suggest that the difference might reflect their selection of neurons with large receptive fields, which allowed a maximal separation between the two stimuli – a selection that was not done by Reynolds *et al.* Increasing the separation could reduce the importance of neuronal interactions between the two stimulus representations at lower levels of processing. Note also, that, as outlined by Gawne and Martin, some neurons recorded by Reynolds *et al.* did in fact have responses that were explained by a MAX function.

Furthermore, it is worth examining carefully data reported by Desimone's group, not only when V4 neurons were stimulated by simple bars, but also when V4 and IT neurons were presented with more complex objects in their

receptive fields [8–10]. In most cases, it appeared that the neuronal responses to the simultaneous presentation of two stimuli in V4 or IT neuronal receptive fields could be explained by the weighted sum of the responses to each stimulus presented in isolation. However, it is also clear that the early part of the response to a preferred stimulus (one that elicits a strong response when presented in isolation) seemed to be partially or totally unaffected when this preferred stimulus is flanked by another stimulus associated with a much smaller response when presented alone. It would appear that the effect of a non-optimal stimulus on the response to a preferred stimulus occurs only after some delay. Therefore, even when neuronal responses to simultaneous presentations appear to follow a weighted-sum model, the early part of the response might in fact be explained by a MAX model.

Given that information theory analysis of neuronal responses in IT has shown that most of the information about stimulus properties is conveyed in the early part of the neuronal response [11,12], this result clearly leaves open the possibility that several objects might be encoded in parallel before some form of competition between neuronal responses occurs [13]. In the future, therefore, it will be very important for models of object recognition to determine what part of the neuronal response is used by the visual system to take a decision. It will also be important to determine whether lower firing rates in response to a stimulus flanked by a non-optimal stimulus reflects a degraded representation, or whether this response might instead incorporate additional information about spatial and/or shape relations between the two stimuli [14,15].

Applying MAX to the real world

But before we can conclude that the MAX function is a major coding mechanism in object vision, several points must be clarified. In their study, Gawne and Martin used high-contrast stereotyped black and white patterns presented on a uniform grey background (Walsh patterns). In order to generalize their discovery to real world situations it will therefore be necessary to test neurons in the ventral pathway with more realistic stimuli, possibly photographs of real world objects embedded in natural scenes, which can sometimes have very low contrasts. Can neurons in the ventral pathway encode several objects in parallel or are serial routines required each time we face a complex environment? This ‘ultimate’ and crucial test, still to be performed, is illustrated schematically in Fig. 1b.

Clear predictions can be made about neuronal responses according to the three coding schemes discussed here (the weighted-average model, weighted-sum model, and the MAX model). Recent evidence in favour of the parallel hypothesis has been provided by a study showing that IT neurons can still fire selectively and rapidly to objects in the context of real world scenes [16]. It is possible that whereas subjects might have problems processing multiple stimuli, neurons in IT could potentially behave differently, at least at the beginning of their response.

Implementation of the MAX function

A final point concerns the mechanism by which the MAX function might be implemented. Gawne and Martin noted

that when two stimuli are presented together, the time course of the response to the combined stimulus often matched the profile produced by the stimulus producing the shortest latency response when presented in isolation. It is as if the earliest response masks the effectiveness of the later one. Given that more effective stimuli tend to produce responses with shorter latencies, this sort of temporal mechanism could potentially explain why the stronger stimulus determines the response. Interestingly, this kind of effect can be related to the sort of temporal winner-take-all mechanism proposed by Thorpe [17], in which the earliest firing inputs generate inhibition that reduces or even blocks the effect of those afferent inputs that fire later.

A similar sort of desensitisation mechanism based on activation of populations of fast-spiking interneurons has also been used more recently to implement a rank-order coding scheme in which neurons can be made selective to the order of firing of their inputs [18,19]. However, Gawne and Martin reported a MAX effect even for neurons presented with stimuli eliciting similar response latencies, so it could be that the temporal mechanism is not the only way to implement the MAX function. To solve this controversial issue, it might be necessary to perform intracellular recordings while presenting two stimuli that elicit responses with identical latencies when presented alone.

Whatever the final outcome, it is clear that recent interest in non-standard computational algorithms signals the emergence of a new kind of research strategy – one in which computational principles are directly tested using neurophysiological methods.

References

- 1 Logothetis, N. and Sheinberg, D. (1996) Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621
- 2 Tanaka, K. (1996) Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139
- 3 Wallis, G. and Rolls, E.T. (1997) Invariant face and object recognition in the visual system. *Prog. Neurobiol.* 51, 167–194
- 4 Thorpe, S. *et al.* (1996) Speed of processing in the human visual system. *Nature* 381, 520–522
- 5 Rousseelet, G.A. *et al.* (2002) Parallel processing in high-level categorization of natural images. *Nat. Neurosci.* 5, 629–630
- 6 Riesenhuber, M. and Poggio, T. (1999) Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025
- 7 Gawne, T.J. and Martin, J.M. (2002) Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J. Neurophysiol.* 88, 1128–1135
- 8 Reynolds, J.H. *et al.* (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753
- 9 Chelazzi, L. *et al.* (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.* 80, 2918–2940
- 10 Chelazzi, L. *et al.* (2001) Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11, 761–772
- 11 Tovee, M.J. *et al.* (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. *J. Neurophysiol.* 70, 640–654
- 12 Rolls, E.T. *et al.* (1999) The neurophysiology of backward visual masking: information analysis. *J. Cogn. Neurosci.* 11, 300–311
- 13 Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222
- 14 Missal, M. *et al.* (1997) Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb. Cortex* 7, 758–767

- 15 Missal, M. *et al.* (1999) Shape interactions in macaque inferior temporal neurons. *J. Neurophysiol.* 82, 131–142
- 16 Sheinberg, D.L. and Logothetis, N.K. (2001) Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.* 21, 1340–1350
- 17 Thorpe, S.J. (1990) Spike arrival times: a highly efficient coding scheme for neural networks. In *Parallel Processing in Neural Systems* (Eckmiller, R., Hartman, G. and Hauske, G., eds) pp. 91–94, Elsevier
- 18 VanRullen, R. and Thorpe, S.J. (2001) Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* 13, 1255–1283
- 19 Thorpe, S. *et al.* (2001) Spike-based strategies for rapid processing. *Neural Netw.* 14, 715–725

1364-6613/03/\$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S1364-6613(03)00023-8

The nature of cognitive development

Scott P. Johnson

Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14850, USA

Theories of cognitive development have led to enduring and fierce arguments that have been long on rhetoric but short on evidence. Constructivist theory has roots in Piagetian notions of cognitive development as proceeding from self-directed action during infancy. Nativist theories subsequently became popular by producing claims of cognitive precocity, but left open many central questions concerning mechanisms of development. Now, a new view of constructivism is experiencing a renaissance, having achieved greater psychophysical, computational and neural plausibility.

Debates concerning developmental origins of human concepts are enduring and relentless, rooted in the ancient and more recent philosophizing of such eminent thinkers as Plato, Aristotle, Descartes and Hume. Inaccessibility to any empirical contributions to the argument hampered emergence of a viable theory of concept acquisition until the last century, when Jean Piaget devised techniques to elicit behaviors in infants that were suggestive of a developing system of knowledge, centering on the great themes discussed originally by Kant: object, space, time and causality [1]. The essence of Piaget's theory is 'constructivism': the building of concepts from simpler perceptual and cognitive precursors. Consider, for example, development of the 'object concept' – the notion that objects maintain their existence and properties, such as location or trajectory of motion, in the absence of direct perceptual input. On Piaget's theory, newborn infants have no object concept; they have to learn, with the aid of a set of basic reflexes (see Box 1). This constructivist view of development was challenged by the nativist view, according to which object concepts arise earlier than co-ordinated manual behavior, so they must be innate (Box 1).

Constructivism strikes back

A recent article describing an information-processing approach to cognitive development in infancy holds promise for the goal of accounting for infant acquisition of new knowledge. Cohen and colleagues [2] propose a set of principles that make explicit the progression towards new

knowledge in a bottom-up fashion. The Cohen *et al.* account is diametrically opposed to the competent-infant view held by nativism: instead of a group of high-level cognitive capacities that remains constant across development (such as deductive reasoning and object representation), Cohen *et al.* posit a set of general-purpose sensory, perceptual, and lower-level cognitive processes that are operational at birth and serve to guide knowledge acquisition across domains. These include sensory processing of auditory and visual input, short-term memory, allocation of attention, and primitive categorization. Principal developmental changes concern *expansion* of these rudimentary skills, the 'content' of information, and what constitutes a 'unit' of information.

Central to the theory is the idea that units of information are elaborated and enhanced with improvements in information-processing skills. Initially, the system has access only to information that is relatively simple, but with time, infants integrate the lower-level units of information into more complex, higher-level units, these higher-level units serve as the components for even more complex units, and so forth. There is a bias to attend to the most complex units that the system can handle. If it fails to process the highest units (owing, say, to immature attentional capacity or increased cognitive load), then lower units are utilized. This hierarchical approach to cognitive development is repeated across domains, underlying skill acquisition and proficiency over a range of tasks, throughout our lives in fact. Concepts are thus formed incrementally and progressively, in a manner broadly consistent with Piagetian theory but not tied to a single, determinate developmental mechanism, such as self-directed manual activity. Information-processing theory provides an excellent account of the development of object perception in early infancy, which proceeds in part-to-whole fashion: infants respond initially only to the components of occluded-object displays (i.e. those surfaces that are directly visible), failing to take account of those object constituents that are hidden from view.

Cohen and colleagues' extensive and elegant work on development of understanding of physical causality (see [3] for a review) provides another cardinal example of these principles. Leslie and Keeble [4] had proposed a

Corresponding author: Scott P. Johnson (sj75@cornell.edu).