

Visual masking and RSVP reveal neural competition

Christian Keyzers and David I. Perrett

A test visual stimulus is harder to recognize when another stimulus is presented in close temporal vicinity; presenting stimuli in close spatial vicinity of a test stimulus reduces its visibility; presenting a stimulus to one eye can render invisible another stimulus presented to the other eye; and perceiving one interpretation of an ambiguous image prevents the simultaneous perception of other visual interpretations. A single, neurophysiological theory, which may be called 'neural competition' might explain all these phenomena: when two alternative neural visual representations co-exist in the brain, they compete against each other.

'Visual pattern masking' refers to the fact that a 'target' or 'test' stimulus will be reported with less accuracy if presented in the temporal vicinity of other stimuli called 'masks' (Fig. 1). Visual masking has been extensively studied by psychologists [1–3]. More recently, several single-neuron studies have started to unravel its physiological bases [4–11]. These physiological results suggest an explanation of masking in terms of competition between neural representations.

In brief, two sequentially presented images activate neural representations in the brain that will overlap in time because of neural persistence. Yet in the natural world at a given time, two objects cannot occupy the same space. We propose that the nervous system acknowledges this fact by preventing the simultaneous perception of two objects at the same location. When the sensory input to the brain activates two populations of neurons that represent different objects at the same place and time, the competition between those two incompatible representations ensures that, at any given time, a single representation quickly wins. This article aims to support the claim that competition underlies masking and to show that the same mechanism can explain visual crowding, binocular rivalry, flash suppression, monocular rivalry and multi-stable images.

Neural persistence and masking

When a brief stimulus is presented, it remains subjectively visible for some time after it is no longer physically present [12,13]. This 'visual persistence' predicts that neurons should respond to a stimulus for longer than the physical duration of the stimulus (Fig. 2a). Neurophysiological investigations confirm the existence of this neural persistence, with neurons

higher in the visual hierarchy showing longer durations of persistence. In the retina, cells respond for ~60 ms longer than stimulus duration [6], independent of the presence of a mask. However, in the lateral genicular nucleus (LGN), inferior temporal cortex (IT) and the anterior superior temporal sulcus (STSa), the duration of persistence depends on the existence and timing of a mask that follows the test stimulus. Without a mask, persistence lasts up to about 300 ms [4,7–11]. A mask immediately after the stimulus shortens the duration of that persistence to ~60 ms in STSa (Fig. 3). This neural persistence is automatic: it occurs despite the subjects not memorizing the stimuli.

Psychological theories of visual masking

In the past, psychologists have explained visual masking [14] using the 'interruption' and the 'integration' theory. The 'competition theory' we propose incorporates and extends these two theories.

The integration theory postulates that masking occurs because the visual system resembles a long exposure photograph: two temporally neighbouring stimulus events will be integrated into a single blended image. Figure 2b illustrates the neurophysiological predictions of this theory. The interruption theory postulates that the processing of the first stimulus is abandoned in favour of the new stimulus when the latter is presented. This interruption leaves the processing of the first stimulus unfinished, thereby impairing its perception (Fig. 2c).

'...competition differs from interruption because it makes no assumptions about the temporal order of stimuli...'

The competition account we propose (Fig. 2d) postulates that when a target stimulus is followed by a masking stimulus, two contradicting forces determine the neuronal response in higher visual cortex. First, persistence stretches the duration of the response to the first stimulus beyond its physical duration (Figs 2a, 3a). This results in temporal overlap in the neural representations of both stimuli, even if the stimuli themselves never overlapped in time. If the two representations cannot be interpreted as a single plausible percept (e.g. transparency), the activity in the two sets of neurons will compete against each other. During a transition phase of ~60 ms, the representation of the old stimulus decreases and that of the new stimulus, increases (Fig. 2d). The new stimulus wins the competition because it is represented by the stronger initial phase of neuronal responses, whereas the

Christian Keyzers*
Istituto di Fisiologia
Umana, Università di
Parma, Via Volturmo 39,
I-43100 Parma, Italy.
*e-mail: keyzers@unipr.it

David I. Perrett
School of Psychology,
University of St Andrews,
St Andrews, Fife,
UK KY16 9JU.

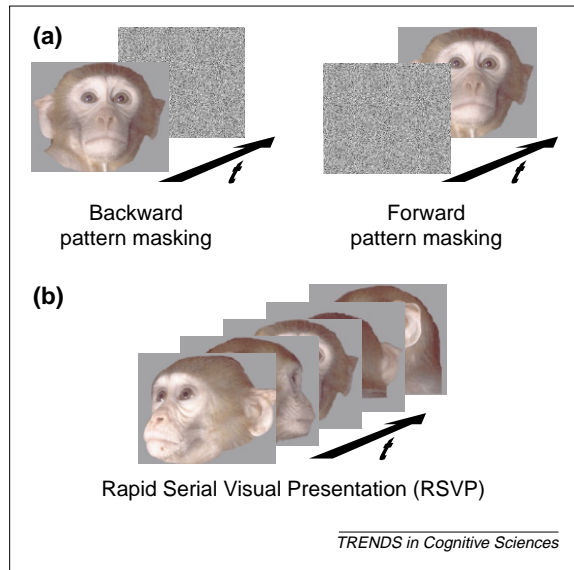


Fig. 1. Paradigms used in the study of visual masking. (a) In backward pattern masking, the target image (face) is followed by a masking stimulus (random pattern) that spatially overlaps the target stimulus. The task of the subject is typically to indicate perception of the target stimulus. The stimulus onset asynchrony (SOA), i.e. the time between the onset of the stimulus and that of the mask, is crucial to performance. If SOA is less than ~ 300 ms, the accuracy with which the face is reported will decrease. In forward masking, the target is presented after the masking stimulus. If SOA is less than ~ 100 ms the face will be reported with less accuracy. (b) In rapid serial visual presentation (RSVP), more than two stimuli are shown one after another: each image acting as the forward mask for the next image and a backward mask for the preceding image. The task of the subject, e.g. 'respond when the face is looking straight at you', makes a particular image a target and the others distracters or masks. If images are presented with SOA of less than ~ 200 ms, the accuracy with which a particular image is reported is reduced.

older stimulus is represented by an already weakened response (Fig. 2d).

There is a fundamental difference between interruption and competition. In the interruption theory, the mechanism responsible for masking is essentially temporal: the newer stimulus interrupts the older stimulus. Competition, however, goes both ways: the new stimulus competes against the old (backward masking) and the old stimulus competes against the new stimulus (forward masking). It is a different phenomenon that favours the new pattern in competition: responses to stationary patterns decrease over time (Fig. 3a,b). Hence, the new stimulus generally wins but is nevertheless affected by the earlier stimulus. The fact that competition is not restricted to the time domain means that it can account for greater number of perceptual phenomena.

Neurophysiological findings

Physiological studies have investigated the effect of masking on single neurons. Studies performed in the 1960s and early 1970s used masking stimuli that produced neural responses very similar to those of the target stimulus. This renders interpretation difficult, as responses cannot be unambiguously attributed to one or the other

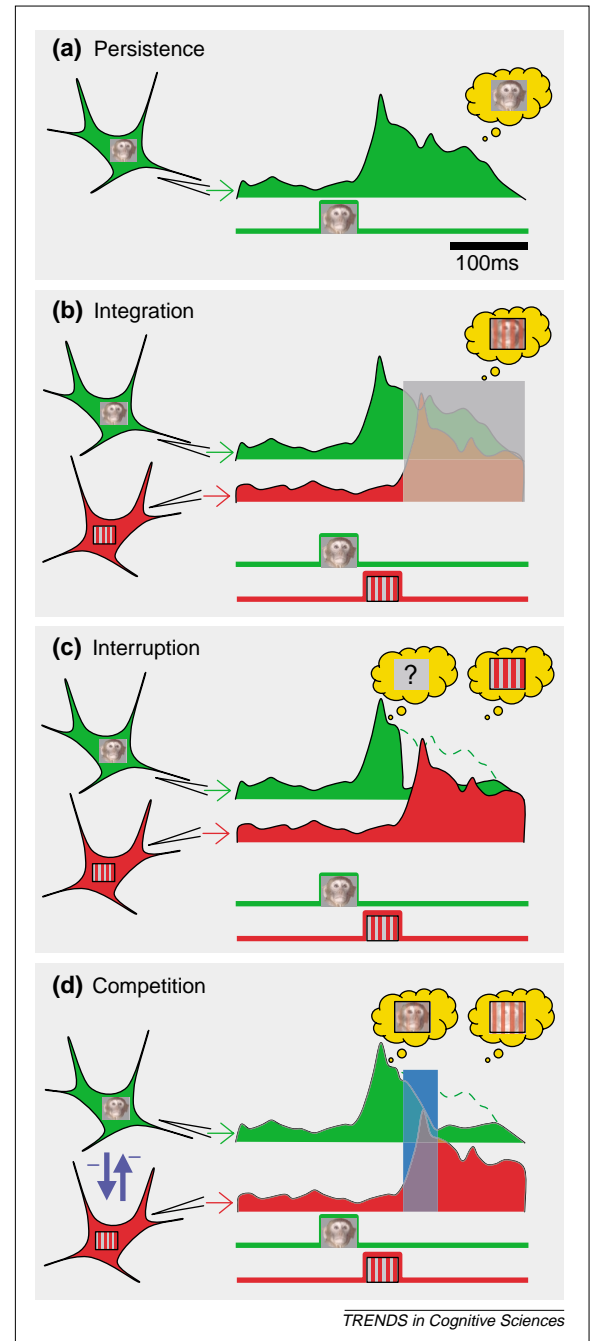


Fig. 2. Neurophysiological interpretations of the theories of visual masking. Two hypothetical cells are recorded simultaneously: one (green) responding to a face and the other (red) to a grating. (a) Persistence predicts that the brief presentation of a face results in activity in the green cell lasting longer than the time the face was physically present (see Fig. 3a) and allows continued perception of the face (thought bubble). (b) Integration predicts that presentation of a face followed by a grating results in persisting responses to both stimuli, with the overlap between the activity in the red and green neuron (grey zone) triggering the perception of an integrated stimulus containing both the grating and the face. (c) Interruption predicts that the same stimulus sequence results in the response of the green neuron discontinuing abruptly (broken green line) when the red neuron is activated by the presentation of the grating. (d) Competition predicts that the two neurons compete against each other (blue arrows). The activity in response to the face continues after it has disappeared from the screen, because of persistence. When the bars are presented, they activate the red cell. The two cells are now active at the same time (blue zone). As the red cell is in the strong transient phase of response, it wins the competition. After about 60 ms, only the red neuron remains active.

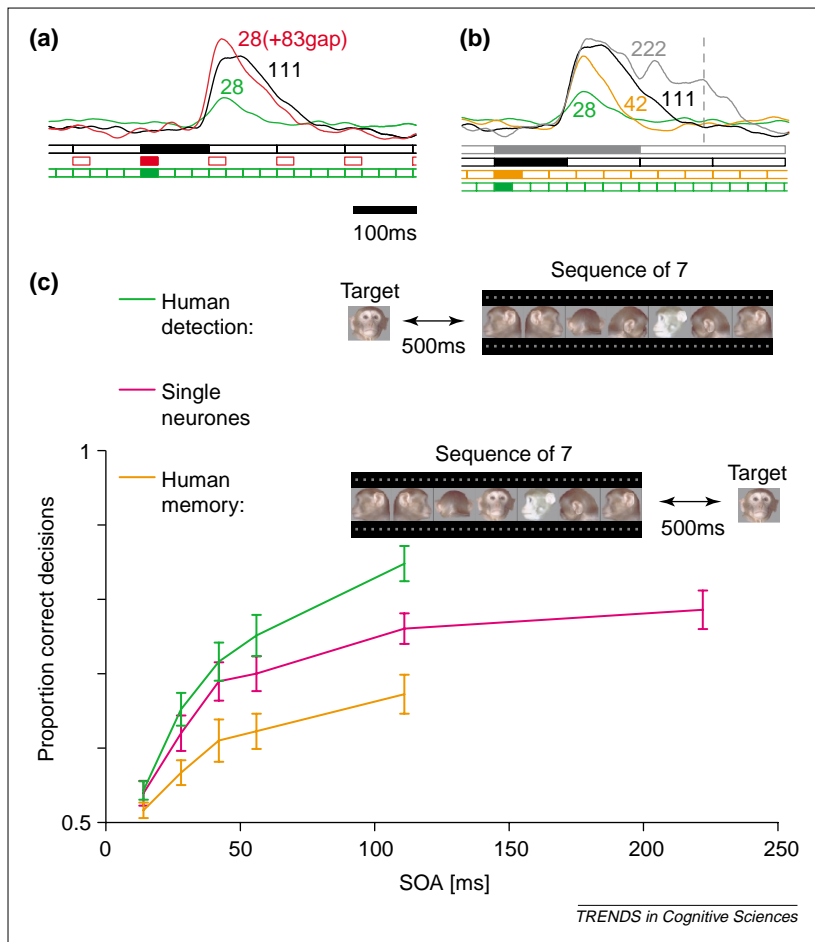


Fig. 3. (a) Neural persistence during rapid serial visual presentation (RSVP) sequences. The boxes under the curves represent the stimulus timing for three conditions in which temporal cortex (STSa) cells were recorded while a monkey fixated an RSVP image sequence [10,11]. The cells responded to one of the stimuli used in the RSVP sequences. Filled boxes denote the occurrence of this effective stimulus; open boxes denote other stimuli in the sequence. The curves represent the latency aligned population response of 34 STSa neurons and can be read as the response of an average neuron. The average neuron started responding after ~110 ms. In the 'black' condition, each stimulus was presented for 111 ms and was directly followed by the next stimulus. Responses lasted for ~170 ms. In the 'red' condition, a shorter 28 ms stimulus was presented, and the next stimulus appeared after a 83 ms gap with a blank screen. The red response nevertheless lasted as long and was as strong as the black one, demonstrating strong neural persistence. Stimuli presented every 28 ms without gap produced a much smaller response lasting ~90 ms [10]. This demonstrates that it is the representation of the next stimulus that terminates responses, but that this termination takes ~60 ms. (b) Conventions as above with the stimulus duration [and stimulus onset asynchrony (SOA)] indicated next to each curve. Responses outlasted stimulus duration for ~60 ms in all conditions [11]. This is illustrated for the grey curve (SOA = 222 ms) with the grey line below marking the time 222 ms after response onset. This persistence creates a temporal overlap in the brain between the representations of consecutive stimuli. Responses show an initial peak followed by a gradual decline that biases neural competition in favour of new stimuli. (c) In the detection task, human subjects were shown a single target image followed by an RSVP sequence either containing the target or not. The proportion of correct responses is shown as a function of SOA (green curve, chance = 0.5). In the memory task, the RSVP sequence was shown before the target, rendering the task more difficult (orange curve). The performance of single STPa neurons in detecting the same RSVP targets (estimated with ROC analysis [11], pink curve) closely followed that of human observers.

stimulus. Levick and Sacks [6] presented two identical flashes in sequence to the retina. When flashes were separated by less than 60 ms, a single fused neural response was produced, identical to that produced in response to a single flash lasting from the onset of the first to the end of the second flash. Is this response the result of the second flash replacing the response to the first (interruption), or did the second flash fuse with the first (integration)?

No-one knows. Competition theory predicts that a second flash identical to the first should not induce competition, because both activate the same neurons. The same problem of interpretation holds for other studies [4,5].

More-recent studies have used more-complex stimuli, such as line-drawn geometric shapes [9], and photographs of faces and natural objects [7,8,10,11]. Masking stimuli were chosen to elicit little response. In temporal cortex (area IT [7–9] and STSa [10,11]), presenting a masking stimulus after the target stimulus reduces the duration of the neural representation of the target (Fig. 3a,b). Responses to the stimulus depend on the stimulus onset asynchrony (SOA; Fig. 1) and last for SOA + ~60 ms [11] (Fig. 3a,b). Shortening the SOA not only shortens the neural responses, but also renders the neural representation of the stimulus less accurate [8,10] (Fig. 3c). Indeed, if human subjects are asked to detect a particular face in a rapid serial visual presentation (RSVP) sequence, the accuracy of their performance decreases with decreasing SOA in a way that mimics the performance of single neurons in STSa (Fig. 3c). The equivalence of psychophysical and cellular performance suggests that STSa cell activity is the neural correlate of perceptual masking.

Why are the responses shortened by the occurrence of the mask? Is it because of integration? According to integration theory, both the old and the new stimulus are represented jointly, and there is no reason why the old stimulus should remain represented for 60 ms longer than SOA and then disappear from the neural representation. Is it because of interruption? No: interruption predicts that one stimulus is processed at a time, yet during ~60 ms, both the old and the new stimulus are processed in the brain (Fig. 3b). However, competition predicts that when neurons start responding to the mask stimulus (e.g. broken grey line for 222 ms stimuli in Fig 3b), a conflict occurs between the persisting response to the old stimulus and the new response to the mask. This competition lasts ~60 ms, during which time the neural representation of the new stimulus becomes increasingly prominent and that of the old stimulus diminishes.

The proposal that masking arises from neural competition is not merely another way of saying 'the new stimulus interrupts the old, but this takes ~60 ms': competition differs from interruption because it makes no assumptions about the temporal order of stimuli. Even simultaneously presented stimuli compete. Interruption does not apply to such situations.

Neural competition in the spatial domain

In visual crowding (Fig. 4a), a target stimulus becomes harder to detect when surrounded by irrelevant stimuli [15]. Neurophysiological studies [16–18] show that the amplitude of the response to an effective stimulus for a neuron in V1, V2, V4 or IT is

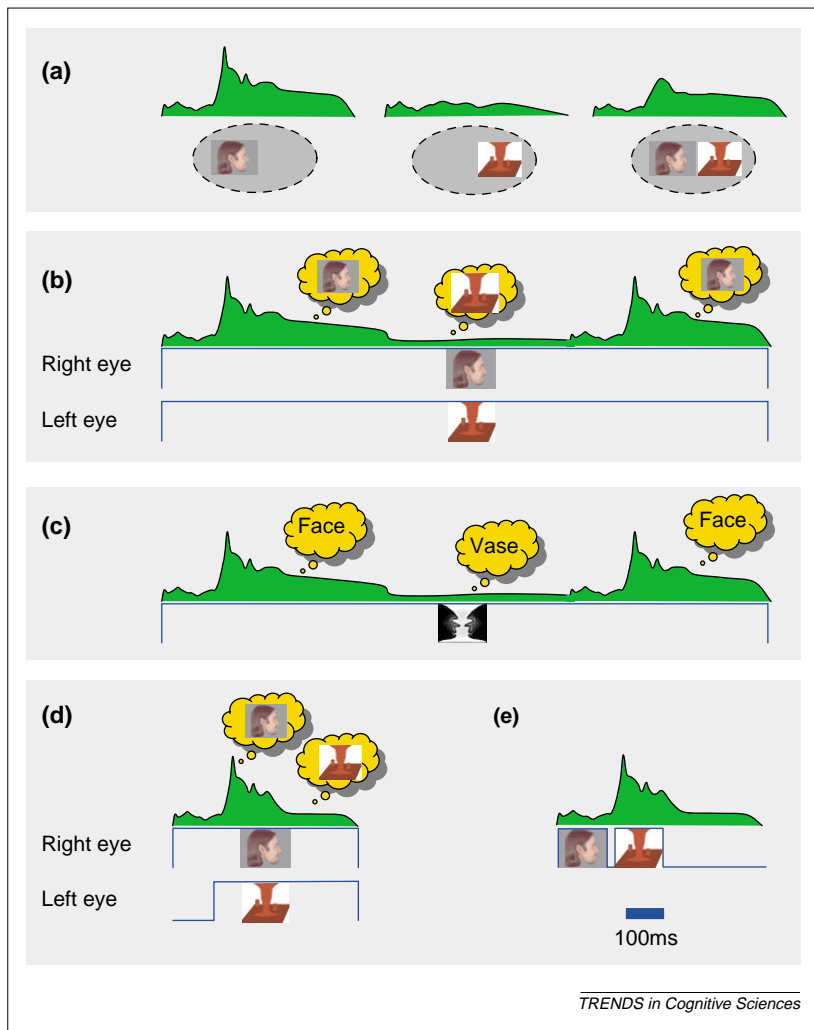


Fig. 4. Overview of experimental paradigms and findings supporting the concept of neural competition. All responses are from a hypothetical neuron. The responses of the neuron are shown in green, the timing of the stimuli in blue. In (a), (c) and (e), the same stimulus is shown to both eyes, whereas in (b) and (d), two different stimuli are shown one to each eye. (a) Left: response when the effective stimulus (in profile) is presented alone in the receptive field (surrounded by a broken line). Centre: absent response to an ineffective stimulus: a vase on a table. Right: reduced response when both effective and ineffective stimuli are presented simultaneously in the receptive field [17]. (b) Binocular rivalry: the effective stimulus is presented to the right eye, while the ineffective stimulus is presented simultaneously to the left eye. Perception alternates between the two objects. The neural response in temporal cortex alternates between the representations of the two stimuli in register with perception [24]. (c) Monocular rivalry: the ambiguous vase-face display is presented throughout, yet perception alternates between either the vase or the face. The neural response switches accordingly (suggested in Ref. [31]). (d) Flash suppression: the face is shown to the right eye and the vase is presented additionally after a short time to the left eye. There is therefore a temporal overlap between the two stimuli. Perception then switches to the new stimulus, and the response is shortened compared with the presentation of the effective stimulus alone [24] (see Fig. 3a). (e) Backward masking: the face is first presented, causing a response in the cell and when the vase replaces the face, the response decreases. Here, unlike in (d), the stimuli do not physically overlap in time. Response duration is about 60 ms longer than the stimulus onset asynchrony [7–11].

reduced when another less effective stimulus is placed in the receptive field of the neuron. These results can be simulated using feed-forward competitive neural networks [19], where the recorded neuron combines inhibitory inputs from neurons representing the ineffective stimulus together with excitatory inputs from neurons that represent the effective stimulus.

Neural responses to moving stimuli often decrease less over time than do responses to static

stimuli [20]. Neural competition would therefore predict that surrounding a static target with moving distracters should bias competition in favour of the moving distracters. This is indeed the case: static stimuli can be rendered invisible by moving distracters [21].

Competition also occurs in binocular rivalry, where two different stimuli are presented, one to each eye (Fig. 4d). In this situation, the visual system does not allow the perception of both stimuli simultaneously. Instead, the two stimuli engage in a winner take all competition, and perception alternates between the two stimuli [22]. Single cell investigations show that in V1 both ocular inputs are represented simultaneously [23]. The two rival inputs then appear to compete against each other actively in V4 [23], and in IT/STSa [24] neurons, they seem to represent exclusively the pattern that is perceived (Fig. 4d). Functional magnetic resonance imaging studies initially suggested that rival patterns competed in V1 [25,26], but it now appears likely that these findings reflect top-down influences from higher areas where competition is resolved [27–29].

Binocular rivalry is a good illustration of mutually exclusive stimuli. Under normal conditions, the two ocular inputs are complementary and the brain extracts a single 3-D interpretation from interocular disparities. When different stimuli are presented to the two eyes, it is impossible to fuse them into a single 3-D interpretation: they are mutually exclusive and competition occurs. Just as in masking, this competition takes time: if the stimuli are presented for less than ~150 ms, both are perceived [30].

Similar mechanisms appear to apply to perceptual phenomena called 'monocular rivalry' (Fig. 4c). The multistable Rubin's vase-face display, for example, is perceived as two profile faces or a single vase – but never both. Brain activity during the observation of this display resembles that during binocular rivalry [31]. When a stimulus contains both monocular and binocular rivaling regions, the perceptual switches in the two regions synchronize [32], suggesting that the two forms of rivalry share a common mechanism: competition. Competition might therefore be a general principle of visual processing whenever the brain has to decide between alternative percepts or is faced with multiple active neural representations.

Neural competition in the temporal domain

Flash suppression [24,33] bridges the apparent gap between masking and binocular rivalry (Fig. 4d). One pattern is presented continuously to one eye. After 150 ms, a 'new' pattern is additionally presented to the second eye. Perception always switches abruptly to the 'new' pattern. This case is similar to both binocular rivalry (Fig. 3b), because stimuli are presented dichoptically, and masking (Fig. 3e), because stimuli are offset in time, and the new stimulus wins. Dichoptic presentation is not

Questions for future research

- What are the differences between different cases of competition (visual crowding, binocular rivalry, monocular rivalry and masking)?
- Can attention bias the competition between sequentially presented stimuli as it biases competition in visual crowding [18]?
The phenomenon called 'attentional blink' shows that this seems to be the case [39], moreover brain activity during attentional blink and crowding is similar [40].
- What determines whether two representations are mutually exclusive or whether they can co-exist as with transparency?

necessary for competition to occur [31,32]. The only difference between masking and flash suppression is that the competing stimuli physically overlap in time in suppression, but not in masking. Yet neural persistence can make the neural representation of the temporally separate stimuli overlap in time for ~60 ms (Fig. 3a,b). During these ~60 ms, neural activity in masking is thus comparable with flash suppression, and competition can account for the dynamic aspects of perception in both situations. Indeed, in both cases, the new stimulus wins over the old stimulus, because the response to the old stimulus is already weakened and disadvantaged by the time the new stimulus occurs.

Competition can take three forms: lateral, feed-forward and feedback competition. In lateral competition, the neuronal populations that represent the competing stimuli inhibit each other through connections within the same level of visual processing. In V1, there is ample evidence for such lateral competition [34,35]. In feed-forward competition, incompatible representations at one level can compete through connections to a higher level [19]. Finally, feedback connections that descend from higher areas could determine which neural representations are maintained at lower processing stages [27–29,36]. The three mechanisms are not mutually exclusive.

Competition need not always be resolved in similar ways for all aspects of a stimulus. In both masking and binocular rivalry, one aspect (e.g. shape) of two competing stimuli can sometimes compete strongly to the point of rendering one stimulus shape invisible, while other aspects (e.g. movement) compete less. The latter aspect of the otherwise invisible stimulus can then be falsely attributed to the visible stimulus [37,38].

Conclusions and perspectives

Single-cell recordings therefore suggest that competition between conflicting neural representations can account for visual masking, visual crowding, binocular rivalry, flash suppression and monocular rivalry. Indeed, competition is probably an even more general solution that the brain employs to deal with the ambiguities of everyday vision. The visual system always needs to solve a seemingly simple task: what real 3-D objects are responsible for a given retinal input? Unfortunately, multiple interpretations often exist for any given retinal input. Our brain makes 'us' aware of only one of these interpretations at a time by letting the alternatives compete neuronally. Usually, the best alternative receives the strongest synaptic input because more of its features are present in the image. Ambiguous figures (e.g. the vase–face figure) arise from the rare cases when two or more alternatives fit the image equally well, and hence receive comparable synaptic input. Perception then switches back and forth between the different possibilities [29]. The concept of competition can explain how this is achieved: early visual areas process multiple alternatives equally, but they compete against each other before reaching later processing stages and awareness. For two successively presented images, a conflict occurs between the persisting activity to the old stimulus and the activity to the new stimulus. The new stimulus has an advantage, because the peak of a neural activity occurs at the beginning of the responses. As a result, competition ensures that a succession of images yields a succession of unique precepts, rather than a complex and blurry awareness of all images at the same time. The cost of this mechanism is masking: if images are presented faster than one image every 200 ms, competition reduces both the quality of neural representations and our ability to recognize any particular image (Fig. 3c). It is interesting that during normal vision, saccades occur at a rate of one fixation every 300 ms – a rate sufficiently slow to prevent the potential cost of masking.

In conclusion, neural competition provides a parsimonious neural mechanism with which to explain a wealth of visual phenomena. We hope that by unifying such a diversity of perceptual phenomena, the idea of competition will further our conceptual understanding of vision and encourage a fruitful exchange of knowledge between researchers working on topics that had formerly been considered profoundly different.

Acknowledgements

C.K. was supported by Boehringer Ingelheim Foundation and EU Marie-Curie fellowships. We thank E. Kohler, M.A. Umiltá, B. Kast and A. Perrett for their useful comments. Without D-K. Xiao's invaluable expertise, the data leading to these ideas would never have been collected.

References

- 1 Felsten, G. and Wasserman, G.S. (1980) Visual masking: mechanisms and theories. *Psychol. Bull.* 88, 329–353
- 2 Breitmeyer, B.G. (1984) *Visual Masking: An Integrative Approach*, Oxford University Press
- 3 Enns, J.T. and Di Lollo, V. (2000) What's new in visual masking. *Trends Cogn. Sci.* 4, 345–352
- 4 Schiller, P.H. (1968) Single unit analysis of backward visual masking and metacontrast in the cat lateral geniculate nucleus. *Vis. Res.* 8, 855–866
- 5 Fehmi, L.G. *et al.* (1969) Electrophysiological correlates of visual perceptual masking in monkeys. *Exp. Brain Res.* 7, 299–316
- 6 Levick, W.R. and Sacks, J.L. (1970) Responses of the cat retinal ganglion cells to brief flashes of light. *J. Physiol.* 206, 677–700
- 7 Rolls, E.T. and Tovee, M.J. (1994) Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc. R. Soc. London B Biol. Sci.* 257, 9–15
- 8 Rolls, E.T. *et al.* (1999) The neurophysiology of backward visual masking: information analysis. *J. Cogn. Neurosci.* 11, 300–311
- 9 Kovacs, G. *et al.* (1995) Cortical correlate of pattern backward masking. *Proc. Natl. Acad. Sci. U. S. A.* 92, 5587–5591
- 10 Keysers, C. *et al.* (2000) Minimal response duration for single cells in the macaque temporal cortex. *Soc. Neurosci. Abstr.* 26, 1842
- 11 Keysers, C. *et al.* (2001) The speed of sight. *J. Cogn. Neurosci.* 13, 90–101

- 12 Coltheart, M. (1980) Iconic memory and visible persistence. *Percept. Psychophys.* 27, 183–228
- 13 Coltheart, M. (1980) The persistences of vision. *Philos. Trans. R. Soc. London B Biol. Sci.* 290, 57–69
- 14 Kahneman, D. (1968) Method, findings, and theory in studies of visual masking. *Psychol. Bull.* 70, 404–425
- 15 Loomis, J.M. (1978) Lateral masking in foveal and eccentric vision. *Vision Res.* 18, 335–338
- 16 Luck, S.J. *et al.* (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42
- 17 Chelazzi, L. *et al.* (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.* 80, 2918–2940.
- 18 Reynolds, J.H. *et al.* (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753
- 19 Desimone, R. (1998) Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. London B Biol. Sci.* 353, 1245–1255.
- 20 Oram, M.W. and Perrett, D.I. (1996) Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J. Neurophysiol.* 76, 109–129
- 21 Bonneh, Y.S. *et al.* (2001) Motion-induced blindness in normal observers. *Nature* 411, 798–801
- 22 Logothetis, N.K. (1998) Single units and conscious vision. *Philos. Trans. R. Soc. London B Biol. Sci.* 353, 1801–1818
- 23 Leopold, D.A. and Logothetis, N.K. (1996) Activity changes in early visual-cortex reflect monkeys percepts during binocular-rivalry. *Nature* 379, 549–553
- 24 Sheinberg, D.L. and Logothetis, N.K. (1997) The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3408–3413
- 25 Polonsky, A. *et al.* (2000) Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nat. Neurosci.* 3, 1153–1159
- 26 Tong, F. and Engel, S.A. (2001) Interocular rivalry revealed in the human cortical blind-spot representation. *Nature* 411, 195–199
- 27 Logothetis, N.K. *et al.* (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157
- 28 Blake, R. and Logothetis, N.K. (2002) Visual competition. *Nat. Rev. Neurosci.* 3, 13–21
- 29 Leopold, D.A. and Logothetis, N.K. (1996) Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* 3, 254–264
- 30 Wolfe, J.M. (1983) Influence of spatial frequency, luminance, and duration on binocular rivalry and abnormal fusion of briefly presented dichoptic stimuli. *Perception* 12, 447–456
- 31 Andrews, T.J. *et al.* (2000) Ambiguous figures reveal neural correlates of perceptual awareness in human visual cortex. *Soc. Neurosci. Abstr.* 26, 1843
- 32 Andrews, T.J. and Purves, D. (1997) Similarities in normal and binocularly rivalrous viewing. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9905–9908
- 33 Wolfe, J.M. (1984) Reversing ocular dominance and suppression in a single flash. *Vision Res.* 24, 471–478
- 34 Bringuier, V. *et al.* (1999) Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* 283, 695–699
- 35 Crook, J.M. *et al.* (1998) Evidence for a contribution of lateral inhibition to orientation tuning and direction selectivity in cat visual cortex: reversible inactivation of functionally characterized sites combined with neuroanatomical tracing techniques. *Eur. J. Neurosci.* 10, 2056–2075
- 36 Lamme, V.A. (2001) Blindsight: the role of feedforward and feedback corticocortical connections. *Acta Psychol.* 107, 209–228
- 37 Herzog, M.H. and Koch, C. (2001) Seeing properties of an invisible object: feature inheritance and shine-through. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4271–4275
- 38 Andrews, T.J. and Blakemore, C. (1999) Form and motion have independent access to consciousness. *Nat. Neurosci.* 2, 405–406
- 39 Chun, M.M. and Potter, M.C. (1995) A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 109–127
- 40 Marois, R. *et al.* (2000) Neural correlates of the attentional blink. *Neuron* 28, 299–308

Constraining the neural representation of the visual world

Shimon Edelman

Understanding the perception of all but the most impoverished and artificial scenes presents a different and probably far greater challenge from understanding face recognition, reading, or identification (or even categorization) of single objects. Central issues in the interpretation of structured objects and scenes are reviewed, starting with fundamentals such as the meaning of seeing. A theoretical approach to this formidable task is outlined, motivated by some recent developments in neuroscience and neurophilosophy.

What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is. [David Marr, *Vision*]

A common notion of vision, consistent with this excerpt from the first paragraph of David Marr's seminal book [1], is gained by considering the

predicament of a person with a searchlight placed in a pitch-dark room full of unfamiliar furniture. One would hope that, by swinging the beam around, the observer will be able to recognize the objects present in the room (a cat here, an aquarium there, etc.) – a task that no longer appears as daunting as it used to because its computational nature is now better understood [2,3]. There is, however, more to high-level vision than recognizing and mentally labeling one object after another, just as there is more to our visual world than a list of objects in the field of view that can be ticked off. Unless viewed in the unusual conditions of darkness with the aid of a searchlight, objects present themselves to us embedded in scenes, combined and recombined in a highly variable, yet structured, manner.

Vision as scene description

It is tempting to draw a parallel between the structure of composite objects and scenes and that of natural languages. However, this analogy, which motivates 'structural description' theories of object representation [4], leads the quest for a comprehensive theory of visual representation to a dilemma. On the one hand, the need to deal explicitly with structure does not arise in recognition tasks [5,6]; furthermore, a scene that affords a satisfactory description by a noun-phrase observational sentence ('*lo, a tabby cat*') fails to give the human language system a run for its money.

Shimon Edelman
Dept of Psychology,
232 Uris Hall, Cornell
University Ithaca,
NY 14853-7601, USA.
e-mail: se37@cornell.edu
http://kybele.psych.cornell.edu/~edelman