

# How position dependent is visual object recognition?

Dwight J. Kravitz, Latrice D. Vinson and Chris I. Baker

Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA

**Visual object recognition is often assumed to be insensitive to changes in retinal position, leading to theories and formal models incorporating position-independent object representations. However, recent behavioral and physiological evidence has questioned the extent to which object recognition is position independent. Here, we take a computational and physiological perspective to review the current behavioral literature. Although numerous studies report reduced object recognition performance with translation, even for distances as small as 0.5 degrees of visual angle, confounds in many of these studies make the results difficult to interpret. We conclude that there is little evidence to support position-independent object recognition and the precise role of position in object recognition remains unknown.**

## Introduction

One of the biggest challenges faced by the visual object recognition system is to enable rapid and accurate recognition despite vast differences in the retinal projection of an object produced by changes in, for example, viewing angle, size, position in the visual field or illumination [1–4]. Such ‘invariance’ is often considered one of the key characteristics of object recognition [4–7]. Changes in position (translations) are among the simplest of these transformations, because only the retinal position of the projection of an object is affected, and not the projection itself [2]. Although it is often assumed that objects can be recognized independently of retinal position [8], the behavioral evidence is limited. In this review, we critically evaluate the behavioral studies of position dependence in visual object recognition from a computational and physiological perspective. We find that the behavioral data on position independence are inconclusive. Furthermore, these studies do not test several key predictions from neurophysiology, including the effect of translations between eccentricities and hemifields, making it difficult to understand the relationship between behavior and the proposed neural substrate. We argue that whereas the balance of the available evidence argues against complete position independence [9–14], the role of position in visual object recognition remains essentially unknown.

## What is visual object recognition?

We use a definition of visual object recognition similar to that of previous authors [1,3]. For our purposes ‘visual object’ refers to a conjunction of a complex set of visual

features. Successful recognition of such an object requires that the response to a current percept be in some way consistent enough with the internal representation of a previous percept to at least partially invoke it [1,7,15]. This formulation of visual object recognition defines it, fundamentally, as a process of comparison between the current percept and preexisting visual object representations.

The ability to name an object is often taken as the strongest evidence for successful recognition, and indeed naming an object requires that the current percept be in some way matched to a visual representation that is associated with the semantic label. However, naming is clearly not a necessary component of visual recognition because nonverbal animals are capable of recognition [3], and we can recognize a previously viewed object even if we cannot name it. In fact, to investigate the effects of position on *visual* recognition it is highly desirable to minimize the influence of semantics, especially verbal labels, which would not be expected to be affected by changes of position.

In this review, we first discuss computational and physiological issues highlighting the importance of position in the comparisons underlying object recognition. These issues provide a framework for a critical evaluation of the behavioral evidence. Such evidence has often been discussed in terms of invariance [10,12,14,16–18] (whether performance is completely insensitive to translation), constancy [10,19] (whether performance reflects the stable properties of the object rather than the changing retinal image) or tolerance [11,20] (the extent to which performance is maintained despite translations). Here we will adopt the term position dependence, which we use to describe the degree to which translation affects object recognition behaviorally.

## The importance of position in object recognition

Given the comparison model of object recognition described above, there are two types of preexisting object representations that might underlie object recognition in the context of position changes. Both make specific behavioral predictions about the degree to which experience with an object at one position will affect recognition during later presentations of that object at different positions (transfer).

The first possibility is that the preexisting representations are specific to the object but independent of its precise retinal projection [4,5,7,21]. Incoming perceptual events would then need to be transformed or rerepresented for comparison with this position-independent representation [4,9,15,22–24] (see Box 1 for details of specific computational models). The extreme prediction of this framework

Corresponding author: Baker, C.I. (bakerchris@mail.nih.gov).

### Box 1. Computational models of object recognition

#### Position-independent models

##### HMAX model [5,23,24]

This model uses a hierarchical design to transform a current retinal projection into a position-independent reference frame. Each level of the hierarchy integrates over progressively larger areas of the visual field. Integration uses the MAX operator, causing any layer to reflect only the strongest response from its input. Units in the region atop this hierarchy, putatively anterior inferior temporal cortex, respond to their preferred object independent of its position in the visual field. Note that the HMAX model was originally presented as a model of roughly only the central 4° of vision.

##### Dynamic routing circuit [22,89]

In this model an explicit 'dynamic routing circuit' is used to remap the current input into a position-independent reference frame. The transformation is achieved by 'control neurons' (putatively residing in the pulvinar), which modify the synaptic strengths of intracortical connections.

#### Position-specific models

##### (Coarse coding of shape fragments) + retinotopy [25,27]

This approach proposes that object representations are selective for both positions and objects. Because each object has multiple representations each tuned to a different position, translated retinal projections need not be transformed for comparison.

##### Fragment-based hierarchy [18,25,26,28]

This model also maintains multiple position-specific representations of objects and object features, but goes further and proposes a hierarchy that assembles complex features from simpler ones. Although its hierarchy is similar to the HMAX model, this model learns its featural representations rather than having them built into the model. Thus, the model will not produce position independence, because recognition will be better at positions where an object commonly occurs. This sort of coding also maintains position information, which can greatly aid recognition [25].

is that there should be complete position independence at a behavioral level (Figure 1a) such that the effect of previous exposure is equivalent across the visual field. Because the initial exposure either evoked (in the case of familiar objects) or created (in the case of novel objects) a position-independent visual representation, subsequent presentations will evoke this same representation regardless of their position.

The second possibility is that instead of a single position-independent representation, there are multiple experience-based representations tuned to particular objects at particular locations [25,26] (Box 1). With this model no transformation into a position-independent form is required. Instead, each retinal projection is compared either with a representation specific to that projection or with an interpolation of representations tuned to similar projections [27,28]. The extreme prediction of the multiple representation framework is position-specific behavior (Figure 1c) – the ability to recognize an object in one retinal location is unrelated to recognition at other locations because there is no shared representation between positions.

Alternatively, with overlapping position-specific representations the multiple representation framework could predict that object recognition would be graded (Figure 1b) such that there is a monotonic decrease in the effect of previous experience with increasing translation distance. Small translations would have a limited

effect, because the overall response would remain similar. Larger translations would engage an increasingly different set of representations, increasing the effect of translation. In this case, the interesting question becomes over what spatial range does behaviorally relevant transfer occur?

Given this framework provided by computational considerations, we next turn to physiology and evaluate which aspects of the behavior will be most informative in integrating the neural data and theory.

#### Position dependence in the ventral visual pathway

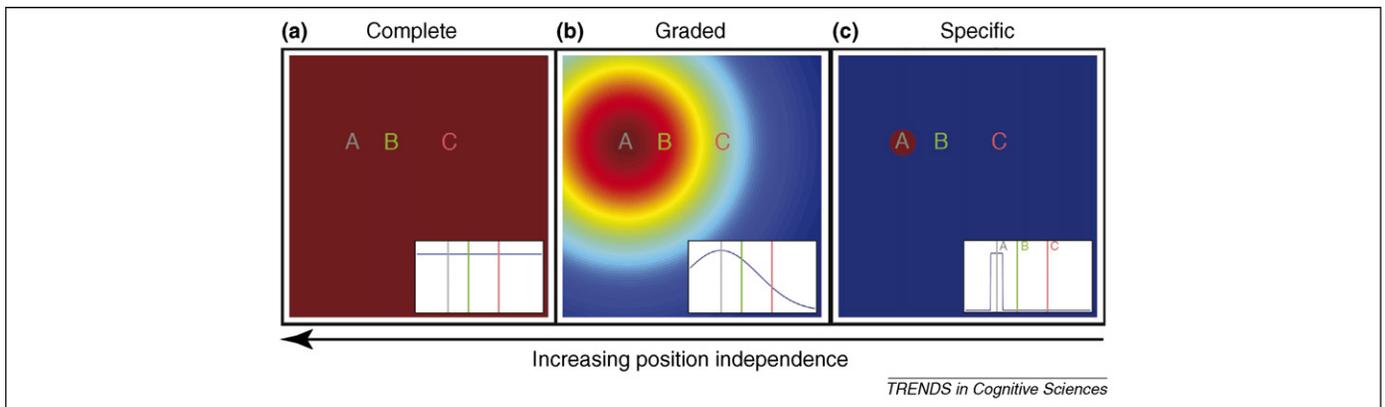
The cortical system supporting object recognition is often described as a ventral visual pathway extending from primary visual cortex (V1) through a series of hierarchical processing stages (V2–V4) to the anterior parts of the inferior temporal (IT) cortex [29], a region crucial for visual object recognition [30,31]. Here we focus primarily on the response properties of neurons in monkey IT, which respond selectively to visual objects.

#### Size of receptive fields

In terms of evaluating position dependence, one of the key properties of neurons is the size of their receptive fields (RFs – the range of retinal positions over which stimuli elicit responses). In V1, RFs are typically small [ $\sim 1$  degree ( $^\circ$ ) of visual angle], consistent with a position-specific representation, but RF size increases as you move along the ventral visual pathway [32–34]. Early studies of anterior IT emphasized the presence of large receptive fields ( $>20^\circ$ ) [35–37] and a largely preserved object preference (or selectivity) within RFs [38–40], suggesting that, despite the position specificity of early visual areas, the problem of transforming current percepts into position-independent reference frames was somehow being solved [15]. However, more recent studies of anterior IT [20,41,42] have emphasized the presence of small receptive fields ( $<5^\circ$ ), consistent with position-dependent representations, and recent human imaging studies have reported retinotopic maps beyond V4 [43,44]. The most quantitative and systematic study of IT RFs to date [41] reported a range of RF sizes from 2.8 to 26°, with a mean size of 10°, and large variability in response within RFs (Figure 2). Thus, although there are some neurons with large RFs there is a wide distribution of RF size.

Such heterogeneity in RF size makes it difficult to predict the degree of behavioral position dependence from the responses of single neurons. Whereas IT neurons with small RFs could give rise to position specificity (consistent with a multiple representations framework), cells with large RFs could support position-independent performance (consistent with a single representation framework). Furthermore, RF size has been reported to vary with changes in task demands [20,45] and the presence of other objects in a visual scene [46,47], making position dependence even more difficult to predict from RF size measured under passive viewing with isolated objects.

However, object representations are unlikely to arise directly from the responses of individual cells but rather from a population-level response across an ensemble of IT neurons [4]. Thus, predicting the degree of position dependency in object recognition depends on understanding how



**Figure 1.** Position dependency in object recognition. The colorplots depict the impact of experience with an object at location A on subsequent recognition of that object throughout a visual hemifield (transfer). The colors range from dark red (best performance) to dark blue (worst performance). The inset graphs show cross-sections through the colorplots taken along a horizontal line through locations A, B and C. At one end of the continuum, in the complete model (a), predicted by the most extreme single representation models, transfer is equivalent at all locations. At the other extreme, in the specific model (c), there is no transfer to novel locations. The graded model (b) (instantiated here as a Gaussian) predicts decreasing transfer with increasing distance from the location of the initial experience with the object.

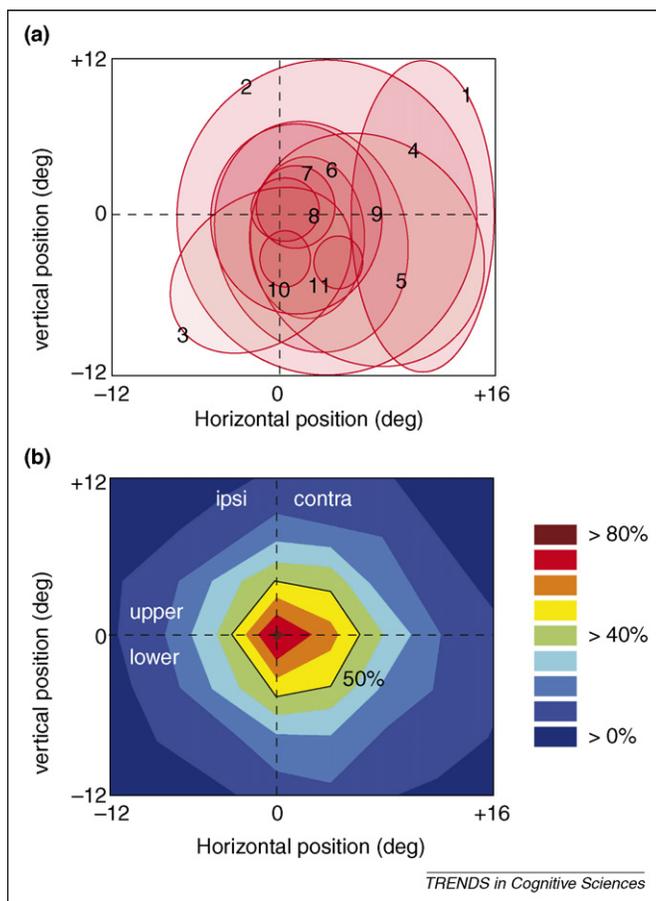
responses are aggregated across neurons. This aggregation might be simple and largely linear, enabling us to use averaging or classifiers on the single-neuron data to model the population response. If so, then we might predict largely independent performance, as was found by a linear

classifier that could accurately predict object category after translations of  $4^\circ$  [48]. However, the aggregation across neurons might also be far more complex, as in the ‘untangling framework’ [4], which proposes that IT neurons interact to create object representations whose properties cannot be trivially predicted from the response of single neurons.

#### Spatial distribution of receptive fields

**Contralateral bias.** Visual information from the right and left visual fields is initially projected to the contralateral hemisphere only (even within the fovea) [49]. Therefore, position independence across the two hemifields would require interhemispheric transfer of information. If there is complete and efficient transfer early in the ventral visual pathway, position independence could be achieved in later areas such as anterior IT. However, IT has several characteristics that suggest minimal interhemispheric transfer in earlier areas. First, lesion studies suggest IT is necessary for interhemispheric transfer to occur [50,51] and transfer might be limited before IT [52]. Second, IT RFs are generally centered within the contralateral field and extend on average only  $3^\circ$  into the ipsilateral field [35,36,41] (Figure 2). When stimuli are present in both the contralateral and ipsilateral hemifields, the response is dominated by the contralateral stimulus [53–55]. A contralateral bias has also been reported in regions of human cortex thought to be crucial for object recognition [56–58]. Furthermore, unilateral lesions of IT in monkeys can produce contralateral deficits in object recognition with little or no impairment on the ipsilateral side [59]. Although a study [52] of humans with unilateral anterior IT lobectomies found no evidence for any contralateral (or indeed ipsilateral) deficits in object recognition tasks, these lesions were much more anterior than regions thought to be crucial for object recognition in humans (such as the lateral occipital complex [60]).

Collectively, these findings suggest that even within anterior regions of the ventral visual pathway there are two largely independent groups of neurons, each responsive predominantly to stimuli in one hemifield. Therefore, a translation within a hemifield should show less position dependence than one between hemifields (i.e. across the vertical midline), which engages largely different sets of



**Figure 2.** Receptive field characteristics of neurons in monkey IT cortex. Examples of 11 individual receptive fields (top) and the mean population ( $n = 77$ ) receptive field (bottom) redrawn from published data [41]. (a) Individual receptive fields vary widely in size and location in the visual field, although the centers of the RFs tend to be in the contralateral hemifield and most cover the fovea. (b) The mean RF plot was produced by normalizing the response of each neuron to its maximum firing rate and then averaging across neurons. This mean RF shows a clear contralateral bias in the population of neurons and an average width of  $10^\circ$ . For each neuron, the point at which the response dropped to 50% of the maximum was taken as the boundary of the RF. Ipsi, ipsilateral hemifield; contra, contralateral hemifield.

neurons. The requirement for interhemispheric transfer of visual information could be a major constraint on the potential for position independence across visual fields.

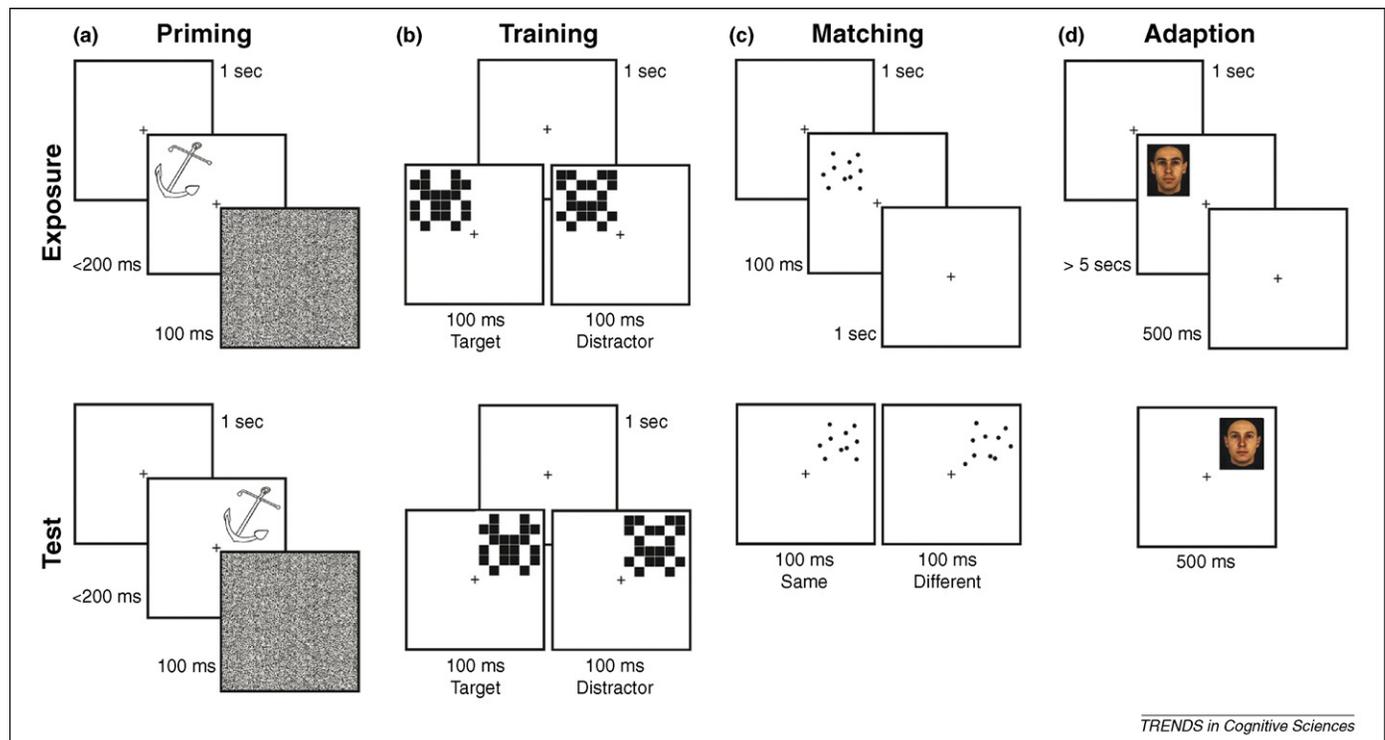
### Foveal bias and eccentricity

Computationally, it has been proposed that the effect of translation (or indeed any affine transformation) can be estimated from a single view of an object at a given location [2], implying that transforming a translated object for comparison with a preexisting representation is relatively simple. This assumes a uniform and consistent sampling of the visual field at all locations. However, retinal photoreceptors are not distributed evenly, but are more concentrated in the fovea than in the periphery, leading to differences in image sampling with eccentricity (distance from fovea) [61]. Thus, for objects presented at different eccentricities, the pattern of response even in the earliest levels of visual processing will be dramatically different, making transformation into a position-independent reference frame difficult. The heterogeneity in retinal sampling is reflected in a foveal bias in the cortical visual hierarchy, with decreasing cortical magnification with eccentricity

[62,63]. In IT, the RFs of most neurons (98%) include the fovea [35] and 50% have a peak response at or near the fovea [41]. Eccentricity biases have also been reported throughout the ventral visual pathway in humans [64,65]. Behavioral studies of position dependence must control for eccentricity, either by ensuring that all stimuli are presented at equal eccentricities, or by explicitly testing the effect of translations between eccentricities while controlling for basic acuity differences. The physiology suggests that less transfer should be observed for translations between eccentricities than within an eccentricity.

### Behavioral studies of position dependence

Physiological considerations (RFs, retinal sampling) suggest there should be some effect of translations on object recognition, especially those between hemifields and eccentricities. However, without the behavioral output of the system it is impossible to know whether these characteristics have a role in determining the degree of position dependence. Most of the formal models of object recognition (Box 1) attempt to implement some aspect of the physiology into their architecture. Thus, the behavior



**Figure 3.** Behavioral investigations of position dependence. Schematic representations of the four paradigms used to investigate position independence. Each paradigm has an exposure phase (top row) and a test phase (bottom row). In the test phase, stimuli are presented at either the same position as in the exposure phase (not shown) or in a different position (shown here). Comparison of performance between the same and different test positions provides a measure of position dependence. The typical durations of each presentation are given and the example stimuli follow specific studies. **(a)** Priming [66]. These studies employ a simple task, like naming or categorization, which remains the same throughout the experiment. The relative amount of priming (improved performance at test for the exposure location over other locations) between same and different location test trials establishes the amount of transfer between locations. **(b)** Training [11]. The exposure phase of these paradigms consists of a long training period in which participants are taught to discriminate between a target and a set of distractors in a particular location. During the test phase participants continue to make the same discrimination either at the same location or in a different location. The relative ability to make the discrimination at different locations serves as the measure of transfer. **(c)** Matching [12]. During exposure a single stimulus is briefly presented followed by a short delay. Immediately afterward another stimulus is presented at either the same or a different location and participants must determine whether it was identical to or different from the first stimulus. In these paradigms a unique stimulus must be used in every trial otherwise priming will occur between trials at all of the locations and lead to an overestimation of position independence [10]. **(d)** Adaption [77]. This paradigm has mostly been used with faces (but see Ref. [88]), which can be defined as a direction and a distance from the average face within a face space [75]. Face adaption paradigms make use of the behavioral observation that exposure to a face (adaption) alters the perception of a subsequent test face such that it appears more dissimilar to the adapted face and more like the face on the other side of the average (the anti-face). The variation in the strength of this effect with the relative location of the adapted and test face indicates the position dependency of the adapted representation. The faces shown in panel (d) are from the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany face database as used by [77].

is crucial not only for verifying the accuracy of a model in replicating human performance, but also for helping to establish which aspects of the physiology should be included in any model.

Despite the importance of behavior, there have only been a limited number of investigations into position dependence. Further, these studies focus only on the simple case of recognition of a single object presented without the clutter or context that would be present in natural scenes. Although this is an impoverished situation, establishing the degree of position dependence under these conditions is an important first step. Four main paradigms have been used – priming, training, matching and adaptation (Figure 3) – all of which first provide experience with an object in one position of the visual field (exposure) to evoke or create (for novel objects) an internal visual representation. They then measure the relative effects of that exposure on subsequent presentations of the same object at the exposure location and other locations in the visual field (test). Studies have tested a range of translations (0.5–10°) and have primarily focused on whether or not there is any effect of translation (i.e. whether or not recognition is completely position independent) rather than systematically establishing the degree of position independence. In reviewing this literature, we focus on each paradigm, highlight its findings and limitations, and look for a consistent pattern of results across different testing conditions.

### Priming

Two prominent priming studies show some evidence for complete position independence with translations of 4.8° [16] and 10° [17]. Both of these studies use supraliminal priming, in which participants were familiarized with the names of the stimuli beforehand and were consciously aware of the stimuli as they were presented. These tasks require explicit use of semantic verbal labels, which would probably be insensitive to position, and might lead to an overestimation of the position independence of *visual* object recognition. One study [16] explicitly tested for the effect of semantics, reporting slightly reduced priming for different exemplars with the same name as the primed object. However, because this study did not include unprimed objects during test, the baseline improvement in the task over time could not be established. Without this control condition to compare against, the authors could only conclude that some amount of the priming was visual. Because much of the observed improvement between exposure and test could have been semantic, this study probably overestimated the degree of position independence.

Furthermore, the same group [66,67] later reported evidence against complete position independence using a similar priming task with shorter presentation times. Under these conditions, objects were subliminally presented, reducing the chances that a semantic representation was engaged. One of these studies [66] included a control condition, which allowed the authors to conclude that there was some, but not complete, transfer to novel locations during test (Figure 4a). Consistent with this report, two other priming studies [68,69] report reduced priming with translations. Overall, the priming studies

suggest that there is some position dependence, but all suffer from a potential semantic confound that could lead to underestimation of the degree of position dependence.

### Training and matching

Three matching studies reported some evidence for complete independence with equivalent performance across translations. However, the first study [19] used translations so small relative to the size of the objects that there was a large overlap between the two positions (and thus large overlap of visual features). The second [70] found complete independence only when symmetric but not asymmetric stimuli near fovea (1°) were reflected across a meridian. The third study [10] found complete independence with animal-like stimuli but because of the small number of exemplars (six in total), cross-trial priming might have contaminated the results. When unique exemplars were used on every trial, complete position independence disappeared. In fact, most of the training and matching studies [10–14,71,72] found a significant decrement in discrimination performance with translations varying from 0.5° to 2°.

The training and matching studies, however, have generally used simple and abstract stimuli (e.g. random dot clouds, but see Refs [10,19]), making it unclear what they suggest about object-level processing [73] (Figure 4). Participants could easily have adopted strategies dependent on the low-level properties of the stimuli that could have relied on neural mechanisms in early position-dependent areas of the ventral visual pathway (such as matching the independent elements closest to the fovea or the principal axis of orientation).

Further, all of the matching studies except one [12] suffer from an attentional confound in that the initial presentation of the object effectively cues the location of the second presentation in same-position trials. Because attention is known to decrease monotonically with distance from a cue [74], the reduction in transfer at the different position could be an entirely attentional phenomenon.

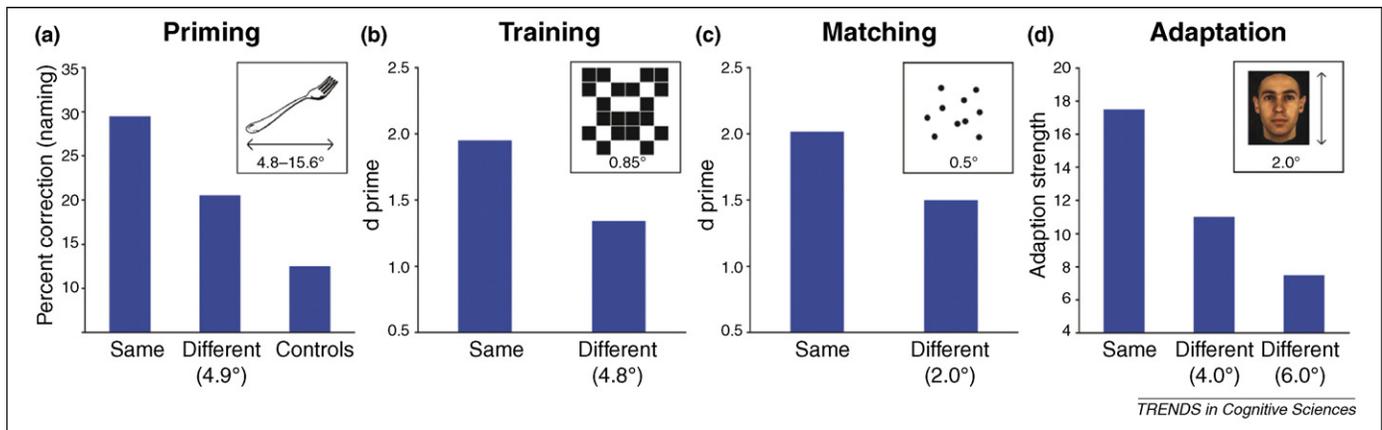
Overall, the training and matching studies suggest that there is some position dependence but, given the confounds described, these studies could be overestimating its degree.

### Adaptation

As in the other paradigms, adaptation studies have found evidence both for and against complete independence. For example, a face adaptation study [75] reported equivalent adaptation despite retinal translations of up to 6°. However, the stimuli were larger (11.25°) than the translations, producing overlapping presentations that could lead to an overestimation of position independence.

Other studies [76,77] have reported evidence against complete independence. In particular, a recent study [77] found a systematic reduction in the strength of adaptation with increasing distance from the adaptor (Figure 4), which provides some evidence against position specificity (Figure 1c) by showing a graded decrease in transfer with increasing translation distance.

However, adaptation studies suffer from all the potential confounds and difficulties of the previous three paradigms: generalizability, semantic effects and attentional



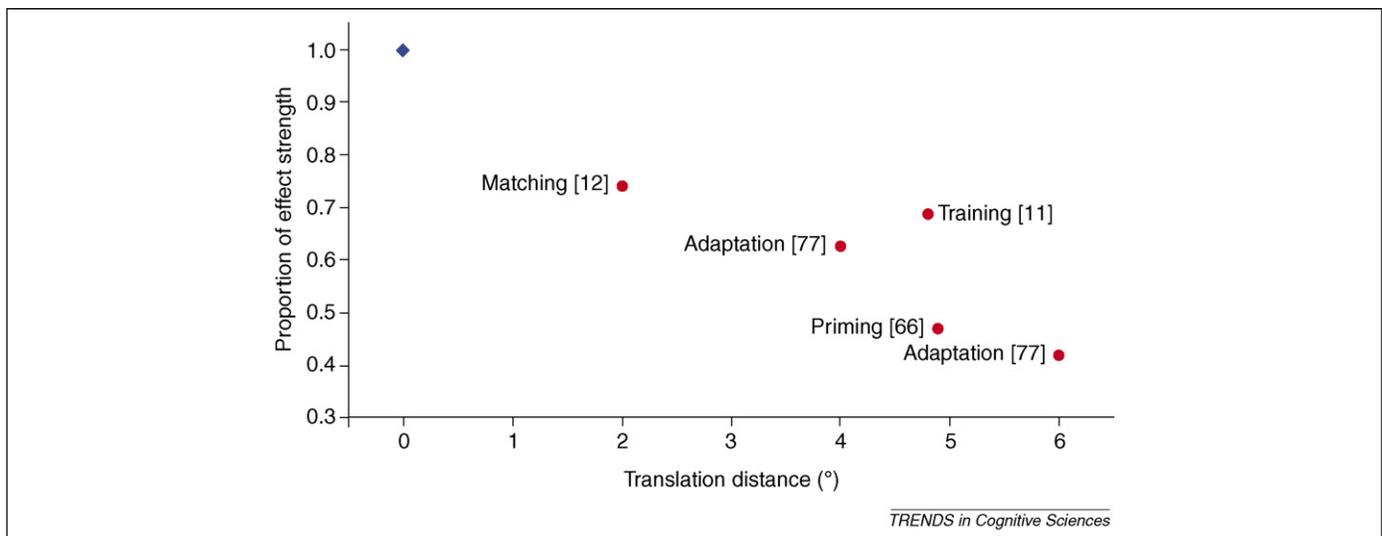
**Figure 4.** Behavioral results across the four paradigms. Each panel depicts a key result from each of the four primary behavioral paradigms (see Figure 3 for details of these paradigms). ‘Same’ indicates performance when there was no change in position between exposure and test. The distance of the translation in the ‘Different’ position trials is indicated on the axis. Insets depict the stimulus used in each experiment. (a) Priming [66]. These data were adapted from a priming study with the best controls for the effects of semantic priming. The control trials are objects presented during test that were not seen during exposure, which provides a measure of unprimed performance. The intermediate performance observed in the ‘Different’ position trials suggests that there is some but not complete priming, which goes against both the complete and specific models (Figure 1a,c). (b) Training [11]. The reduction in performance in the ‘Different’ position argues against position independence, but without a baseline measure of performance, position specificity cannot be ruled out. (c) Matching [12]. These data were adapted from an experiment that included a cue between exposure and test to control for the attentional confound (see ‘Training and matching’ in main text). (d) Adaptation [77]. These data were adapted from a study on face adaptation. The decreasing effect strength argues strongly against both the complete and specific models (Figure 1a,c).

cueing. Most studies have thus far focused on faces, which arguably comprise a special stimulus class that might not generalize well to other objects. These studies also rely on category judgments, which are inherently semantic (e.g. gender, identity). Finally the test stimulus must be presented immediately following the adaptor, meaning that the adaptor might be functioning as a spatial cue.

*What do the behavioral studies tell us?*

The behavioral data suggest there is some impact of even small (0.5°) translations on object recognition, arguing against complete independence (Figure 1a). Further, there is some evidence for at least partial transfer across positions [66,77], arguing against position specificity

(Figure 1c). However, all the behavioral studies suffer from potential confounds, making it difficult to interpret the results. Interestingly, in spite of these confounds, an aggregation of the results of the most controlled studies from each paradigm shows a largely monotonic decrease in the amount of transfer with increasing distance (Figure 5), implying graded position dependency (Figure 1b) for translations within an eccentricity. These findings support the multiple representation framework. Further, combined with the ability of even anterior IT to support position-specific learning [20,78], the behavioral data call into question an increasingly prevalent logic that attempts to localize the cortical regions underlying a behavioral phenomenon through its position dependence [79,80].



**Figure 5.** Comparison of behavioral data. This figure plots behavioral data from the four studies illustrated in Figure 4. To place the data on the same graph, data from each study were normalized relative to performance when there was no shift in position between exposure and test (same position). This calculation fixes the proportion of transfer at 1 for translations of 0° (blue diamond). The amount of transfer after a change in position is then expressed as a proportion of the transfer observed in the same position (red circles). The largely monotonic nature of the decrease in transfer with distance best agrees with the graded model (Figure 1b), suggesting that recognition performance with translations greater than 6° will be severely impaired.

### Box 2. Questions for future research

- How can we establish a strong link between the physiology and behavior? Relating the two requires that a behavioral change be systematically reflected in some aspect of the physiology. Experience could produce this change and confirm which aspects of the physiology are key in producing behavioral position dependence.
- What effect does long-term experience have on the position dependence of the representation of objects? Objects that only occur in constrained portions of the visual field or are common and must be distinguished from one another might encourage the formation of markedly position-dependent representations. Experience differences such as these might relate to the heterogeneity of RF sizes observed in IT.
- How is position dependence affected by the presence of unrelated objects in the scene (clutter) and context? RF properties of IT neurons are known to change in these circumstances, and an explicit measure of the behavioral effect would help in establishing the relationship between the physiology and behavior.
- If visual representations are position dependent, how are they associated with position-independent semantic representations?

### Concluding remarks

A complete understanding of object recognition requires the integration of physiological, computational and behavioral evidence. Although the current behavioral data argue against complete position independence, future research will need to address what factors (such as task or long-term experience) affect the degree of position dependence and which properties of IT neurons are reflected in behavior (Box 2).

### The importance of experience

Although the behavioral evidence shows that there is probably no automatic transfer across all spatial positions after a single exposure, long-term experience could modulate the degree of position dependence. The representation of an object might be affected both by the statistics of its appearance on the retina and on the sorts of tasks performed on it. For example, when monkeys were trained extensively to make fine-grained discriminations on small visual stimuli ( $<1^\circ$ ) in a limited number of positions, the responses of IT neurons were greatly position dependent [20] with small RFs. Longer-term experience might affect large-scale cortical organization producing eccentricity biases in ventral visual cortex [65,81].

Alternatively, experience with stimuli across multiple spatial positions could encourage the formation of more position-independent representations. One recent model suggested that extensive experience with object primitives (e.g. angles, parts) at a variety of positions could give rise to multiple position-specific representations that nonetheless produce responses similar enough to support completely independent recognition (see Box 1, 'Fragment-based hierarchy'). It is also possible that complete independence could arise from temporally adjacent exposure to an object. In the natural environment, objects move across the retina smoothly, which might enable the visual system to associate different presentations of the same object together despite the distinct responses they evoke (e.g. due to changes in lighting, position, size or orientation) [82]. This idea is supported by a recent study reporting that temporal

association could cause the foveal view of one object to be associated with the peripheral view of a different object [8] – 'breaking position invariance'.

### Physiological implications

The behavioral evidence against complete position independence is consistent with the current physiological evidence. However, the behavioral data speak little to the specific neural predictions. One priming study [67] included translations across the vertical and horizontal meridians, but the data from these two types of translations were collapsed, providing no insight into the relative amount of transfer (although the authors did report greater transfer within a quadrant than between quadrants). Another study [10] included a comparison of within- and between-hemifield translations, and revealed no difference in transfer, although this was a matching study subject to the attentional confound. Further, no study systematically tested the effect of eccentricity changes on position dependence. In most studies, the eccentricity of the locations tested was equated. In others, the locations varied in eccentricity, but the data were either collapsed across these positions [66] or the change in eccentricity was ignored [12,13]. Further research is needed to address these issues.

### Computational implications

The behavioral evidence against complete independence frees computational models of visual object processing from a significant constraint. Position dependence is consistent with learning in a biologically plausible system, whose inputs must be largely position specific to be in keeping with the small RFs in early visual cortex. The effects of previous exposure will therefore necessarily be limited to a subset of positions in the input. Achieving complete independence (Figure 1a) in this context is difficult, generally requiring either specialized subsystems [11,22], or a heavy reliance on large receptive fields in object-selective units [83]. An approach in which some position dependence is maintained [25,84] is likely to be a better and more parsimonious implementation of cortical visual object processing.

Having position-specific object representations could be useful for object recognition in complex scenes [85]. Objects have a tendency to occur in particular positions and in particular spatial relationships with other objects. If the visual object recognition system maintains position frequency information, it can be used as a constraint to aid in the recognition of ambiguous or occluded figures. It can also be used to resolve the general content of scene, providing a cue that can resolve ambiguities and provide an initial spatial map of information [86,87]. An analogous process has been proposed to aid in the ability to assemble object fragments into coherent objects [25].

### Acknowledgements

We would like to thank Hans Op de Beeck, Marlene Behrmann, Lalitha Chandrasekher, Jim DiCarlo, Daniel Dilks, Stephanie Manchin, Alex Martin, Mortimor Mishkin, Julianne Rollenhagen and Rebecca Schwarzlose for their insightful comments on an earlier version of this manuscript. We would also like to thank Hans Op de Beeck for contributing to Figure 2.

## References

- 1 Edelman, S. (1999) *Representations and Recognition in Vision*, MIT Press
- 2 Riesenhuber, M. and Poggio, T. (2000) Models of object recognition. *Nat. Neurosci.* 3 (Suppl), 1199–1204
- 3 Ullman, S. (1997) *High-Level Vision: Object Recognition and Visual Cognition*, MIT Press
- 4 DiCarlo, J.J. and Cox, D.D. (2007) Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341
- 5 Riesenhuber, M. and Poggio, T. (1999) Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025
- 6 Marr, D. and Nishihara, H.K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B. Biol. Sci.* 200, 269–294
- 7 Marr, D. (1982) *Vision*, W.H. Freeman
- 8 Cox, D.D. et al. (2005) 'Breaking' position-invariant object recognition. *Nat. Neurosci.* 8, 1145–1147
- 9 Graf, M. (2006) Coordinate transformations in object recognition. *Psychol. Bull.* 132, 920–945
- 10 Dill, M. and Edelman, S. (2001) Imperfect invariance to object translation in the discrimination of complex shapes. *Perception* 30, 707–724
- 11 Dill, M. and Fahle, M. (1997) The role of visual field position in pattern-discrimination learning. *Proc. Biol. Sci.* 264, 1031–1036
- 12 Dill, M. and Fahle, M. (1998) Limited translation invariance of human visual pattern recognition. *Percept. Psychophys.* 60, 65–81
- 13 Foster, D.H. and Kahn, J.I. (1985) Internal representations and operations in the visual comparison of transformed patterns: effects of pattern point-inversion, position symmetry, and separation. *Biol. Cybern.* 51, 305–312
- 14 Nazir, T.A. and O'Regan, J.K. (1990) Some results on translation invariance in the human visual system. *Spat. Vis.* 5, 81–100
- 15 Cavanagh, P. (1978) Size and position invariance in visual-system. *Perception* 7, 167–177
- 16 Biederman, I. and Cooper, E.E. (1991) Evidence for complete translational and reflectional invariance in visual object priming. *Perception* 20, 585–593
- 17 Fiser, J. and Biederman, I. (2001) Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Res.* 41, 221–234
- 18 Ullman, S. and Soloviev, S. (1999) Computation of pattern invariance in brain-like structures. *Neural Netw.* 12, 1021–1036
- 19 Ellis, R. et al. (1989) Varieties of object constancy. *Q. J. Exp. Psychol.* 41, 775–796
- 20 DiCarlo, J.J. and Maunsell, J.H. (2003) Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J. Neurophysiol.* 89, 3264–3278
- 21 Hummel, J.E. and Biederman, I. (1992) Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* 99, 480–517
- 22 Olshausen, B.A. et al. (1995) A multiscale dynamic routing circuit for forming size-invariant and position-invariant object representations. *J. Comput. Neurosci.* 2, 45–62
- 23 Serre, T. et al. (2007) A quantitative theory of immediate visual recognition. *Prog. Brain Res.* 165, 33–56
- 24 Serre, T. et al. (2007) A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6424–6429
- 25 Edelman, S. and Intrator, N. (2000) (Coarse coding of shape fragments) plus (retinotopy) approximate to representation of structure. *Spat. Vis.* 13, 255–264
- 26 Ullman, S. (2007) Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn. Sci.* 11, 58–64
- 27 Edelman, S. and Intrator, N. (2003) Towards structural systematicity in distributed, statically bound visual representations. *Cogn. Sci.* 27, 73–109
- 28 Ullman, S. and Evgeniy, B. (2004) Recognition invariance obtained by extended and invariant features. *Neural Netw.* 17, 833–848
- 29 Ungerleider, L.G. and Mishkin, M. (1982) *Two Cortical Visual Systems*, MIT Press
- 30 Logothetis, N.K. and Sheinberg, D.L. (1996) Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621
- 31 Tanaka, K. (1996) Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139
- 32 Boussaoud, D. et al. (1991) Visual topography of area TEO in the macaque. *J. Comp. Neurol.* 306, 554–575
- 33 Kobatake, E. and Tanaka, K. (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* 71, 856–867
- 34 Tanaka, K. et al. (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* 66, 170–189
- 35 Desimone, R. and Gross, C.G. (1979) Visual areas in the temporal cortex of the macaque. *Brain Res.* 178, 363–380
- 36 Gross, C.G. et al. (1972) Visual properties of neurons in inferotemporal cortex of the macaque. *J. Neurophysiol.* 35, 96–111
- 37 Richmond, B.J. et al. (1983) Visual responses of inferior temporal neurons in awake rhesus monkey. *J. Neurophysiol.* 50, 1415–1432
- 38 Ito, M. et al. (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* 73, 218–226
- 39 Lueschow, A. et al. (1994) Inferior temporal mechanisms for invariant object recognition. *Cereb. Cortex* 4, 523–531
- 40 Schwartz, E.L. et al. (1983) Shape-recognition and inferior temporal neurons. *Proc. Natl. Acad. Sci. U. S. A.* 80, 5776–5778
- 41 Op De Beeck, H. and Vogels, R. (2000) Spatial sensitivity of macaque inferior temporal neurons. *J. Comp. Neurol.* 426, 505–518
- 42 Logothetis, N.K. et al. (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563
- 43 Larsson, J. and Heeger, D.J. (2006) Two retinotopic visual areas in human lateral occipital cortex. *J. Neurosci.* 26, 13128–13142
- 44 Brewer, A.A. et al. (2005) Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat. Neurosci.* 8, 1102–1109
- 45 Merriam, E.P. et al. (2006) Remapping in human visual cortex. *J. Neurophysiol.* 97, 1738–1755
- 46 Aggelopoulos, N.C. and Rolls, E.T. (2005) Scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. *Eur. J. Neurosci.* 22, 2903–2916
- 47 Rolls, E.T. et al. (2003) The receptive fields of inferior temporal cortex neurons in natural scenes. *J. Neurosci.* 23, 339–348
- 48 Hung, C.P. et al. (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866
- 49 Lavidor, M. and Walsh, V. (2004) The nature of foveal representation. *Nat. Rev. Neurosci.* 5, 729–735
- 50 Seacord, L. et al. (1979) Role of inferior temporal cortex in interhemispheric transfer. *Brain Res.* 167, 259–272
- 51 Gross, C.G. and Mishkin, M. (1977) The neural basis of stimulus equivalence across retinal translation. In *Lateralization in the Nervous System* (Haxnad, R.W. et al., eds), pp. 109–122, Academic Press
- 52 Biederman, I. et al. (1997) High level object recognition without an anterior inferior temporal lobe. *Neuropsychologia* 35, 271–287
- 53 Chelazzi, L. et al. (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.* 80, 2918–2940
- 54 Sato, T. (1988) Effects of attention and stimulus interaction on visual responses of inferior temporal neurons in macaque. *J. Neurophysiol.* 60, 344–364
- 55 Sato, T. (1989) Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp. Brain Res.* 77, 23–30
- 56 Hemon, C.C. et al. (2007) A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS ONE* 2, e574
- 57 Niemeier, M. et al. (2005) A contralateral preference in the lateral occipital area: sensory and attentional mechanisms. *Cereb. Cortex* 15, 325–331
- 58 McKyton, A. and Zohary, E. (2007) Beyond retinotopic mapping: the spatial representation of objects in the human lateral occipital complex. *Cereb. Cortex* 17, 1164–1172
- 59 Merigan, W.H. and Saunders, R.C. (2004) Unilateral deficits in visual perception and learning after unilateral inferotemporal cortex lesions in macaques. *Cereb. Cortex* 14, 863–871
- 60 Malach, R. et al. (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. U. S. A.* 92, 8135–8139
- 61 Curcio, C.A. et al. (1987) Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* 236, 579–582
- 62 Sereno, M.I. et al. (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893

- 63 Van Essen, D.C. *et al.* (1984) The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Res.* 24, 429–448
- 64 Hasson, U. *et al.* (2002) Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34, 479–490
- 65 Levy, I. *et al.* (2001) Center-periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539
- 66 Bar, M. and Biederman, I. (1998) Subliminal visual priming. *Psychol. Sci.* 9, 464–469
- 67 Bar, M. and Biederman, I. (1999) Localizing the cortical region mediating visual awareness of object identity. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1790–1793
- 68 McAuliffe, S.P. and Knowlton, B.J. (2000) Long-term retinotopic priming in object identification. *Percept. Psychophys.* 62, 953–959
- 69 Newell, F.N. *et al.* (2005) The interaction of shape- and location-based priming in object categorisation: Evidence for a hybrid “what plus where” representation stage. *Vision Res.* 45, 2065–2080
- 70 Dill, M. and Fahle, M. (1999) Display symmetry affects positional specificity in same-different judgment of pairs of novel visual patterns. *Vision Res.* 39, 3752–3760
- 71 Cave, K.R. *et al.* (1994) The representation of location in visual images. *Cognit. Psychol.* 26, 1–32
- 72 Kahn, J.I. and Foster, D.H. (1981) Visual comparison of rotated and reflected random-dot patterns as a function of their positional symmetry and separation in the field. *Q. J. Exp. Psychol.* 33, 155–166
- 73 Bricolo, E. and Bulthoff, H.H. (1992) Translation-invariant features for object recognition. *Perception* 21, 59
- 74 Henderson, J.M. and Macquistan, A.D. (1993) The spatial distribution of attention following an exogenous cue. *Percept. Psychophys.* 53, 221–230
- 75 Leopold, D.A. *et al.* (2001) Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94
- 76 Kovacs, G. *et al.* (2005) Position-specificity of facial adaptation. *Neuroreport* 16, 1945–1949
- 77 Afraz, S-R. and Cavanagh, P. Retinotopy of the face aftereffect. *Vision Res.* (in press)
- 78 Yang, T. and Maunsell, J.H. (2004) The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.* 24, 1617–1626
- 79 Ahissar, M. and Hochstein, S. (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464
- 80 Fahle, M. (2004) Perceptual learning: A case for early selection. *J. Vis.* 4, 879–890
- 81 Kanwisher, N. (2001) Face and places: of central (and peripheral) interest. *Nat. Neurosci.* 4, 455–456
- 82 Wallis, G. and Rolls, E.T. (1997) Invariant face and object recognition in the visual system. *Prog. Neurobiol.* 51, 167–194
- 83 Riesenhuber, M. and Poggio, T. (2002) Neural mechanisms of object recognition. *Curr. Opin. Neurobiol.* 12, 162–168
- 84 Edelman, S. (2002) Constraining the neural representation of the visual world. *Trends Cogn. Sci.* 6, 125–131
- 85 Rousselet, G.A. *et al.* (2004) How parallel is visual processing in the ventral pathway? *Trends Cogn. Sci.* 8, 363–370
- 86 Oliva, A. and Torralba, A. (2007) The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527
- 87 Biederman, I. *et al.* (1982) Scene perception: detection and judging object undergoing relational violations. *Cognit. Psychol.* 14, 143–177
- 88 Suzuki, S. and Cavanagh, P. (1998) A shape-contrast effect for briefly presented stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1315–1341
- 89 Olshausen, B.A. *et al.* (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* 13, 4700–4719

## Elsevier joins major health information initiative

Elsevier has joined with scientific publishers and leading voluntary health organizations to create patientINFORM, a groundbreaking initiative to help patients and caregivers close a crucial information gap. patientINFORM is a free online service dedicated to disseminating medical research.

Elsevier provides voluntary health organizations with increased online access to our peer-reviewed biomedical journals immediately upon publication, together with content from back issues. The voluntary health organizations integrate the information into materials for patients and link to the full text of selected research articles on their websites.

patientINFORM has been created to enable patients seeking the latest information about treatment options online access to the most up-to-date, reliable research available for specific diseases.

**For more information, visit [www.patientinform.org](http://www.patientinform.org)**