VISUAL OBJECT UNDERSTANDING

Thomas J. Palmeri and Isabel Gauthier*‡*

Visual object understanding includes processes at the nexus of visual perception and visual cognition. A traditional approach separates questions that are more associated with perception — how are objects represented by high-level vision — from questions that are more associated with cognition — how are objects identified, categorized and remembered. However, to understand the bridge between perception and cognition, it is fruitful to abandon any sharp distinction between perceptual and cognitive aspects of visual object understanding. We provide a selective review of research from both the Object Recognition and Perceptual Categorization literatures, highlighting relevant behavioural, neuropsychological, neurophysiological and theoretical research into the representations and processes that underlie visual object understanding in humans and primates.

IDENTIFICATION
A decision about an object's unique identity. Identification requires subjects to discriminate between similar objects and involves generalization across some shape changes as well as physical translation, rotation and so on.

CATEGORIZATION
A decision about an object's kind. Categorization requires generalization across members of a class of objects with different shapes.

**Department of Psychology, Center for Integrative and Cognitive Neuroscience, and ‡Vanderbilt Vision Research Center, Vanderbilt University, 301 Wilson Hall, Nashville, Tennessee 37203, USA*
Correspondence to T.J.P or I.G. e-mails: thomas.j.palmeri@vanderbilt.edu; isabel.gauthier@vanderbilt.edu
doi:10.1038/nrn1364

Our ability to recognize an object as one that we have seen before, even under very different viewing conditions, effortlessly disguises the tremendous computational challenges that are presented to our visual system by variations in location, viewpoint and lighting. Such variations conspire to present a markedly different stimulus to the eyes even though the same physical object is being seen. We can also uniquely identify individual objects in a class — not just faces, but also other objects, such as particular cars, houses and animals. This requires us to ignore changes in viewpoint, position, size and lighting but to tolerate some changes in shape, although other differences in shape can imply different identities. And we can also tell what kind of object something is. For IDENTIFICATION, we must discriminate between physically similar objects, but for CATEGORIZATION, we must generalize across physically different objects, and the amount of generalization varies with levels of category abstraction.

Issues in object understanding

To recognize, identify or categorize an object involves comparing its visual representation with some representation of stored knowledge. This raises a number of fundamental questions.

How are objects represented by the visual system? Are object representations abstract three-dimensional descriptions, or are they tied more closely to the two-dimensional image of an object? Are different representations used to identify unique objects and to generalize across categories? Are there specialized representational systems for certain categories?

How is object knowledge represented? Are there specific representations for object identity but abstract representations of object categories? Or can the same object representations be used adaptively to recognize, identify and categorize objects?

What mechanisms underlie visual object understanding? Research on visual object understanding often formalizes mechanisms using computational models. Formal models are powerful tools for understanding how combinations of representations and processes can lead to adaptive behaviour, often in ways that counter intuition[1].

How does the visual object understanding system change with experience? Are qualitatively different forms of knowledge used at different levels of experience, or are the same representations transformed in a more quantitative manner from those used by novices to those used by experts?

Two relatively independent areas of research into visual cognition have examined important aspects of visual object understanding: Object RECOGNITION and Perceptual Categorization. Despite addressing many of the same issues, the two areas have focused on different aspects of those problems, with surprisingly little overlap[2,3]. Nevertheless, they have ultimately arrived at many complementary conclusions regarding the computational bases of visual object understanding.

One source of the rift between them is the classic demarcation between perception and cognition[4,5]. Early research in Object Recognition arose from work on visual perception and focused on how the visual system creates a perceptual description of an object[6]. Theories were grounded in models of perception, and empirical studies were based on psychophysics. Early research in Perceptual Categorization, by contrast, was rooted in cognitive science and focused on the structure of conceptual knowledge[7]. Theories were grounded in models of semantic memory, and empirical studies were based on cognitive research methods. This early demarcation is reflected in the limited number of recent research papers with citations that cross the two areas. Also, Object Recognition researchers and Perceptual Categorization researchers present their scientific work in journals and conference sessions that too rarely reach individuals in the other area.

This traditional rift is consistent with a modularized view of visual cognition[8], where the creation of visual representations is uninfluenced by knowledge or goals. But, if categorization does not start where perception ends[2], then issues of perceptual representations, knowledge representations and how they are used must be studied concurrently[4,9].

In this review, we provide a selective review of behavioural, neuropsychological, neurophysiological and theoretical research from the Perceptual Categorization and Object Recognition literatures, using the term 'visual object understanding' to subsume these two literatures. We show how the two literatures have begun to converge on computational issues regarding representations and processes that underlie how we recognize, identify and categorize objects. We also review research issues that have been central to recent developments in both areas, such as those concerning modular models of visual object understanding, evidence for interactions between perception and conception, and the importance of studying visual understanding in a dynamic perspective that takes into account learning and expertise.

Computational models

Computational models have had central theoretical roles in both Object Recognition and Perceptual Categorization. Verbal descriptive theories often suffer from hidden, unstated or overlooked assumptions. By contrast, computational models specify hypothesized representations and processes in sufficient detail that they can be instantiated in mathematical equations or computer simulations, allowing theories to be more explicitly tested[1,10].

Traditionally, computational models of Object Recognition and Perceptual Categorization have focused on different stages of visual processing. Object Recognition models typically provide a detailed description of the format of object representations, with less emphasis on how evidence for a particular identification or categorization is generated and used to make a decision. By contrast, Perceptual Categorization models often make simplifying assumptions about object representations but provide detailed descriptions of how representations are used to make decisions[11]. Despite such differences, models from the two fields show striking parallels in theoretical development. Most specifically, early models assuming that representations were abstract have been challenged by proposals that representations are closely tied to specific experiences.

Models from Object Recognition. Early models from the Object Recognition literature assumed that the fundamental goal of vision is to create a faithful description of objects in the world, reconstructing the three-dimensional structure of objects and their spatial relations[6]. One intuitive proposal for constructing a three-dimensional object description, recognition-by-components[12], and its neural network instantiation[13], represent each object by a small number of three-dimensional primitives called GEONS, combined with their spatial relationships in an object-centred reference frame. Objects can be recognized independently of viewpoint (under specific conditions[14]) because geons are defined by a combination of non-accidental properties (such as parallelism or curvilinearity) that are invariant over viewing position.

These early STRUCTURAL DESCRIPTION models were challenged on empirical, computational and theoretical grounds[15]. Reliable detection of geons is based on reliable detection of edges and vertices, a notoriously challenging problem in computer vision[15] that the model largely bypasses. In these early models, structural descriptions lacked metric information about size and shape, making many cases of within-category discrimination[16] nearly impossible; more recent structural description models preserve some metric information[17]. Moreover, recognition of both novel and familiar objects depends on viewpoint and other specific aspects of object experience.

Of the alternatives to structural description theories[18,19], perhaps the most successful are IMAGE-BASED models. Rather than creating a viewpoint-invariant structural description, these models represent an object in terms of its similarity to two-dimensional views that are stored in memory. Accordingly, behavioural performance and neural responses should depend on viewpoint. Simple template-matching models are often presented as foils for sophisticated feature-analysis and structural description models[20]. However, the power of sophisticated image-based models in computer vision applications[21] and in accounting for psychophysical[22] and neurophysiological data[23,24] belie any intuitive implausibility of these models.

A key computational challenge faced by image-based models is that whereas relatively simple transformations of size, translation and picture-plane rotation can bring an object view into correspondence with a stored view[18], no simple transformations can compensate for differences in illumination, depth-plane rotation or shape; for example, the frontal view of a face does not contain sufficient information to recreate the side view of a face.

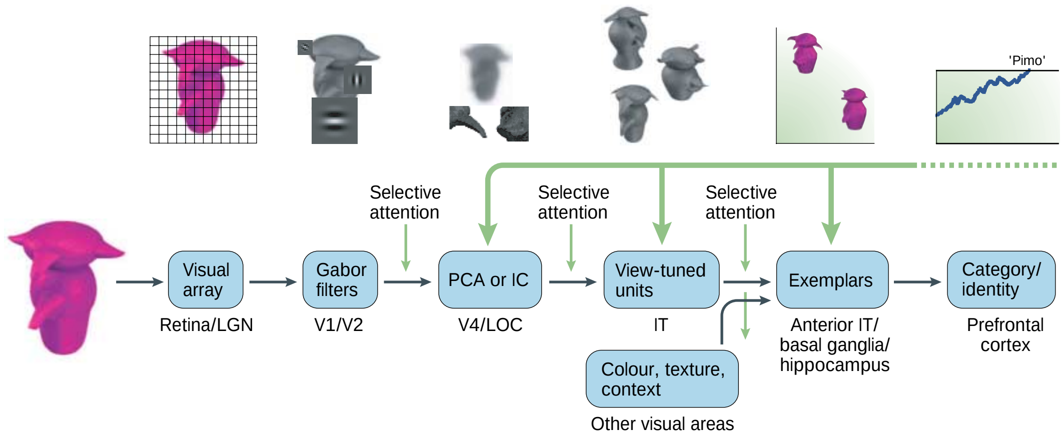
RECOGNITION
A decision about whether an object has been seen before. We can recognize an object seen just moments before — as in many experiments from Object Recognition — or we can recognize an object seen on an earlier occasion — as in many experiments from Perceptual Categorization and the memory literature. Recognition involves generalization across size, location, viewpoint and illumination.

GEONS
(Geometric ions). Simple viewpoint-independent volumetric primitives that are the building blocks of object representation for recognition-by-components theory.

STRUCTURAL DESCRIPTION
A qualitative representation of an object in terms of its three-dimensional primitives (for example, 'geons') and their relative positions. Many structural descriptions are devoid of metric information regarding quantitative aspects of the primitives (specific shapes and sizes) and their positions (specific spatial locations).

IMAGE-BASED
A representation of an object that preserves much of the richness of the perceived two-dimensional image. It is viewpoint-specific, or represented in an egocentric frame of reference, and might contain information about illumination, colour and material (but is often proposed to be largely scale- and translation-invariant).

Box 1 | Summary of some visual object understanding models



The figure summarizes a broad class of computational models of visual object understanding, illustrating stages of visual processing and their possible neural loci; both the physiology and computational theories are simplified for purposes of illustration. A three-dimensional object is first represented as a two-dimensional spatial array along the retina and lateral geniculate nucleus (LGN); in many models from the Object Recognition field, the input is simply an array of luminance values. The two-dimensional array is processed in early visual areas (V1 and V2) according to orientation and spatial frequency; in some models from Object Recognition, the input array is processed by a bank of frequency- and orientation-tuned Gabor filters[27,30,155]. Additionally, some models have proposed implicit[27] or explicit[50] processing in early visual areas to create a scale- and translation-invariant image-based representation, although the need for such explicit preprocessing has been questioned on the basis of neurophysiological evidence. Various kinds of intermediate object representation have been proposed for intermediate stages of visual processing, such as areas V4 and TEO/lateral occipital cortex (LOC)[70,121]. Some models from Object Recognition propose local viewpoint-dependent features of intermediate complexity (IC)[30,50,107], others propose holistic components that are obtained through techniques such as principal components analysis (PCA)[155], and some ignore intermediate representations entirely[27]. Although not shown, some models have proposed direct mappings from such intermediate representations to category knowledge[106,107]. These mappings could allow early representations, such as those in V4 or TEO, to support basic-level judgements. Object shape is represented according to the activity of a population of view-tuned units in inferotemporal (IT) neurons[24,74]; novel views of an object are represented by similarity to stored views, and novel objects are represented by similarity to stored objects (FIG. 1). Essentially, this distributed view-tuned representation provides a low-dimensional shape description[27]. According to some theories[27], view-tuned units are directly connected to category or identity knowledge. However, in other theories an intermediate exemplar (or exemplar-like) representation is needed[156]. Models from the Perceptual Categorization field (accounting for categorization, identification and recognition) typically begin with a low-dimensional representation of an object in terms of its shape, colour and other dimensions; these dimensions are either taken from known psychophysical mappings or are derived using techniques such as multidimensional scaling. One crucial component of some models from Perceptual Categorization is that parts, properties or dimensions of an object can be selectively attended according to their diagnosticity[35,46]. An exemplar (or exemplar-like) representation is a local conjunction of shape, colour, texture and so forth in multidimensional space. Efficient creation of such configural exemplar representations could depend on the hippocampus[115,119,120] but might not[42]. The exemplars might be stored in anterior IT[78] or in the basal ganglia[68]. The activity of an exemplar is proportional to its similarity (distance in psychology space) to the presented object. Category (or identity) might be encoded by units in the prefrontal cortex[78]; some theories formalize the dynamics of the categorization decision process as an accumulation of relative evidence for various responses, such as identifying that particular GREEBLE as 'Pimo'[49,102].

VIEWPOINT-INDEPENDENT (OR DEPENDENT) PERFORMANCE
Behavioural performance that is invariant of viewing position and independent of experience with particular views is said to be viewpoint-independent. By contrast, viewpoint-dependent performance depends systematically on experience with specific views of an object.

GREEBLES
Novel objects that, like faces, all share a common spatial configuration. Their features can be varied systematically to test aspects of object recognition and feature perception.

The solution is to store multiple views of a single object, compensating for changes in viewing conditions[25,26], and to store multiple views of multiple objects, compensating for changes in shape[27]. VIEWPOINT-INDEPENDENT recognition is possible when a sufficient number of views are stored[28]. Novel objects can be represented in a distributed fashion by their similarity to a relatively small number of stored views of known objects[27].

Image-based models can be instantiated in neural networks with a representational layer of view-tuned radial basis function (RBF) units[29] (BOX 1). The response of each RBF depends on the degree of correspondence between the input and the RBF's stored view, being maximal when the two match[27,28,30]. RBF units can be characterized as receptive fields in a multidimensional representational space (FIG. 1), with similar views activating the same units and dissimilar views activating different units. Early models assumed that implicit preprocessing corrected for scale and translation[28], and a more recent model[30] has added a hierarchy of processing layers[31] to create scale- and translation-invariant representations (but see REF. 32).

There has been a recent trend towards hybrid models that combine viewpoint-specific and viewpoint-invariant mechanisms. Some of these models postulate parallel systems that represent objects using image-based and structural description representations[33]. Others integrate metric information into a structural description model[17].

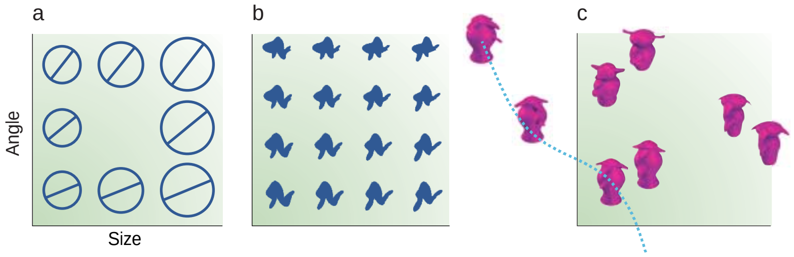


Figure 1 | **Multidimensional representations of objects.** Many models from Object Recognition and Perceptual Categorization assume that objects are represented in a multidimensional psychological space. Similarity between objects (or between views of an object) is inversely proportional to distance in the space. Dimensional representations are sometimes derived from known psychophysical mappings (**a**), with each psychological dimension representing a basic stimulus-processing channel, such as orientation or size[11,35]. Alternatively, dimensional representations can be derived using various psychological scaling techniques, such as multidimensional scaling (**b**)[36,37]. Some Object Recognition theorists[27] have described how a low-dimensional shape space (**c**) for complex objects can emerge from a distributed pattern of similarity to particular stored views. Change in shape (represented along the plane of the figure) and changes in viewpoint (represented by the dotted blue trajectory that is roughly orthogonal to the plane of the figure) are captured within such low-dimensional representations. Novel views of objects are recognized by interpolating between stored views of objects. Extrapolation beyond the range of stored views is rather limited. Categorization of a novel object is a function of the similarity to stored objects. Category typicality effects emerge because prototypical objects are similar to many stored objects, whereas atypical objects are similar to few stored objects and might be similar to objects in other categories.

A recent model predicts the linear costs that are observed in object recognition by additively integrating contributions from object structure and image-based views[34].

Models from Perceptual Categorization. Many models from Perceptual Categorization begin with simplified assumptions about object representations, most commonly assuming that objects are represented in a multidimensional psychological space[11] (FIG. 1) with similar objects close together in that space and dissimilar objects far apart. Such multidimensional representations are often derived from known psychophysical mappings[11,35] or using various psychological scaling techniques[36,37]. Whereas models from Perceptual Categorization share assumptions about object representations, they differ markedly with respect to how object knowledge is represented.

Prototype models assume that object categories are represented abstractly, as the central tendency of experienced members of that category[38]. Objects are categorized according to their relative similarity to stored prototypes; prototypicality effects — whereby certain objects are deemed better members of a category — emerge because certain items are more similar to the abstracted prototype. But whereas categorization depends on similarity to a prototype, identification and recognition of specific objects relies on other independent representations from specific experiences[39,40].

Decision-boundary models[11] are multidimensional generalizations of signal detection theory, with psychological space carved into response regions by linear and non-linear boundaries. Decisions are based on what region of psychological space an object representation occupies. Categorization versus identification decisions are made by carving psychological space with decision boundaries into coarse versus fine response regions. A combination of perceptual noise in the location of an object in psychological space and decisional noise in the location of the decision boundary leads to errors[11] and response-time variability[41]. Object knowledge is abstract in the sense that a decision boundary for categorization or identification is abstracted from specific experience. Representations of specific experiences are not used to categorize or identify, so recognizing that an object was seen on an earlier occasion relies on an independent episodic memory system[42].

By contrast, exemplar models assume that recognition, categorization and identification depend on stored instances of experienced objects[43,44]. Categorization (or identification) judgements are based on the relative similarity of an object to exemplars of categories (or unique objects). Recognition is based on the overall familiarity of an object, irrespective of its category or identity[45], and explicit recognition memory might also depend on contextual representations that are bound to object representations[44]. A successful class of exemplar models[43,46,47] assumes that attention is allocated to psychological dimensions in a task-specific manner, effectively 'stretching' and 'shrinking' psychological space along relevant and irrelevant dimensions, respectively (FIG. 2). Such task-specific, dimension-selective attention is necessary to account for recognition, categorization and identification using the same exemplar representations[46,48,49].

Commonalities between models. View-based models from Object Recognition and exemplar-based models from Perceptual Categorization are complementary. Both image-based theories[50,51] and exemplar-based theories[45] have articulated how the same representations can be used to recognize, identify and categorize. However, they differ in focus. Image-based theories describe detailed mechanisms for how an image-based representation is created from visual features and demonstrate the sufficiency of image-based representations for many important aspects of visual object understanding. But details of how an object representation is used to recognize, identify or categorize the object sometimes involve little more than directly mapping an object representation to a response label[27,50]. By contrast, exemplar-based models — as well as other models from Perceptual Categorization — begin with simplified assumptions about object representations, but describe in detail the process of how object representations are used to recognize, identify or categorize objects, accounting for quantitative patterns of errors and response times across different kinds of task over the course of learning. Marrying image-based models and exemplar-based models is a natural theoretical linkage, but the details remain to be elucidated.

Despite their theoretical successes, both view-based and exemplar-based models have been criticized for apparently requiring that each experienced view of every object for all categories is explicitly represented[44,52]. But it has been shown that storing just a modest number of experiences — a few score to several score, depending on stimulus complexity[53] and category complexity[54] — can produce nearly the same performance as storing all

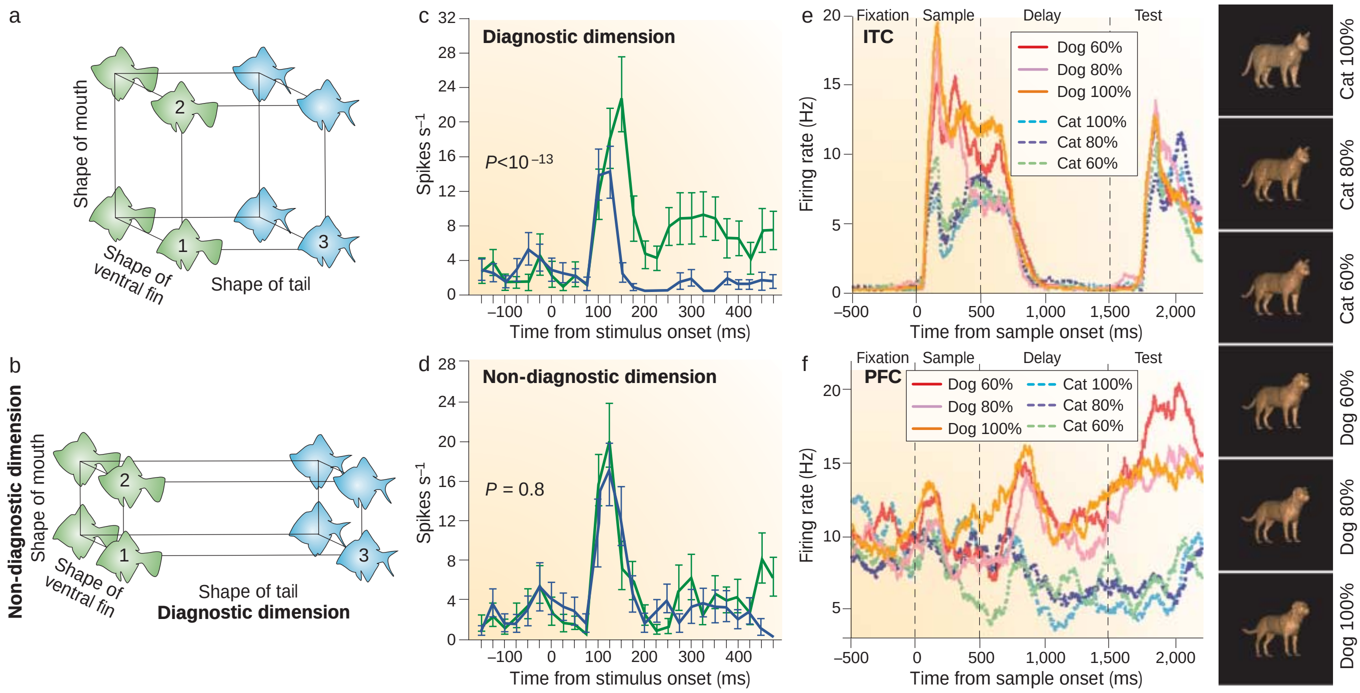


Figure 2 | **Neurophysiology of Perceptual Categorization. a** | A simplified version of stimuli in multidimensional psychological space[76]. In this depiction, the stimuli in category A are shaded green and the stimuli in category B are shaded blue (stimuli were monochromatic in the experiments). **b** | According to exemplar-based models of perceptual categorization[35,43,46] selective attention stretches the space along diagnostic dimensions and shrinks it along non-diagnostic dimensions. **c**,**d** | Sigala *et al.*[76] found cells in inferotemporal (IT) neurons that were selective for at least one of the stimulus dimensions. Spike rate is shown as a function of time from stimulus onset for two representative cells. Cells that responded to a diagnostic dimension (for example, tail shape) showed discriminative responses for particular feature values along that dimension (for example, a pointed tail) (**c**). The figure shows the cell response for the best feature value (green) versus the worst feature value (blue) along the dimension. Cells that responded to non-diagnostic dimensions (for example, the shape of the ventral fin) showed little discriminative response between the best and worst feature value (**d**). **e**,**f** | Freedman *et al.*[78] had monkeys learn to categorize a continuous space of 'cats' and 'dogs' with test stimuli generated by systematically morphing across a continuum from 100% 'cat' to 100% 'dog'. Cells in IT and prefrontal cortex (PFC) were recorded while monkeys performed a category-matching task in which a sample was shown, after a delay a test item was presented, and the monkey had to release a bar if the sample and test were from the same category. Part **e** shows spike rate as a function of time for a category-sensitive IT cell during presentation of the sample, during delay and during presentation of the test item. IT cells showed more significant responses during stimulus processing than during the delay. IT cells also showed significant within-category variability in spike rate. Part **f** shows spike rate as a function of time for a category-sensitive PFC cell. PFC cells show significant activity during the delay. PFC cells also showed more between-category variability than within-category variability in spike rate. Panels **c** and **d** adapted, with permission, from REF. 76 © (2002) Macmillan Magazines Ltd. Panels **e** and **f** adapted, with permission, from REF. 78 © (2003) Society for Neuroscience.

experienced views of all experienced objects. Image-based and exemplar-based models are computationally viable and representationally tractable. And, as described below, they are also consistent with behavioural and neurophysiological evidence.

The role of abstraction

A hallmark of visual object understanding is the ability to generalize — for example, to recognize objects from new viewpoints or to categorize new objects. Does this imply that mental representations are abstract[55]?

Abstraction from Object Recognition. Objects can usually be recognized effortlessly from any viewpoint, indicating that a viewpoint-independent representation might be abstracted. However, naming, matching and priming experiments have revealed costs of viewpoint changes for both novel and familiar objects (for review, see REF. 56), indicating instead that recognition might be supported by viewpoint-specific representations. Studies that use familiar objects and find no effect of viewpoint (for example, REFS 14,57) can be difficult to interpret (apart from the challenge of making any strong argument from null results). Because most objects are experienced from many viewpoints, apparent viewpoint-independent behaviour could instead depend on a collection of viewpoint-specific representations. So, our extensive experience with common objects[58] makes them unsuitable to compare the predictions from multiple-view[59] and structural description models[14,60].

Even after substantial experience of viewing novel objects from a limited set of viewpoints, recognition from unfamiliar viewpoints depends on the angular distance from the nearest familiar viewpoint[26,61]. Even recognition of single geons shows orientation-dependent performance[62], rendering them unlikely building blocks for a theory of viewpoint-invariant recognition. Further evidence against abstract object representations includes findings that recognition also depends on changes in size, position, colour and illumination[63].

Abstraction from Perceptual Categorization. Everyday experience also indicates that knowledge about object categories is abstract[55]. Indeed, early theories assumed

Box 2 | Rules and exemplars

Accompanying the emerging consensus for exemplar-based or exemplar-like representations that underly important aspects of knowledge representation has been a re-emerging interest in whether such specific representations might be supplemented by more abstract representations. Some contemporary theories of perceptual categorization have hypothesized a role for both abstract rules (or prototypes) and specific exemplars[42,134,157,158]. Part of the impetus for such hybrid accounts comes from the recognition that people are sometimes instructed with explicit rules for categorizing objects[146], and that when faced with the task of learning novel categories without any explicit instruction people might attempt to form explicit categorization rules[64,109,157,159]. The expression of such explicit rules could eventually be supplemented by — or at least influenced by — stored exemplar information[52,102,134,145].

We can draw an intuitive distinction between modes of cognition that are explicit and deliberate and those that seem more implicit and automatic[160]. But a real challenge is to relate such an intuitive distinction to an actual distinction in the kinds of mental representation that underlie various aspects of human cognition[134]. There is real controversy over what patterns of observed behaviour provide an appropriate signature for distinguishing between explicit (rule-based) and implicit (exemplar-based) categorization[135,159,161,162]. Converging evidence from behavioural studies, computational modelling, neuropsychology and functional brain imaging will be needed to resolve such basic issues regarding the experiential and abstract bases of human knowledge.

that abstract rules defined category membership[64] (BOX 2). But membership in natural categories is often probabilistic, not rule-defined[7]. When subjects learn novel, ill-defined categories, membership is a graded function of prototypicality[7,38]. Moreover, category prototypes that are never explicitly learned are often classified more accurately than other category members[65,66]. These results indicated that category learning involves abstracting a category prototype[39,65], but later work showed that such prototype effects are consistent with exemplar theories that assume no prototype abstraction[44,67]. As with image-based models from Object Recognition, exemplar generalization accounts for interpolation of the unseen prototype from the surrounding stored exemplars (FIG. 1). Just as recognizing an object is influenced by particular stored views, categorizing an object is influenced by particular stored exemplars. There is an emerging consensus that exemplar-based[35,46] or at least exemplar-like[68] representations underlie important aspects of category representations (BOX 2).

BASIC LEVEL
The level at which object descriptions (both functional and perceptual attributes) maximize a combination of informativeness and distinctiveness. Typically, the basic level is the entry level of recognition. Exceptions include atypical category members (such as penguin, palm tree).

ENTRY LEVEL
The first level of abstraction at which a perceived object triggers its representation in memory. Empirically it is the fastest level at which observers can verify that an object can be given a particular label at some level of the hierarchy (for example, canary, bird or animal).

Neural evidence. Image-based and exemplar-based models are also supported by behavioural and neurophysiological results from monkeys. Responses of inferotemporal (IT) neurons to objects largely depend on stimulus size[69] and orientation[69,70]; even accepted notions of retinal position invariance in IT[71] have been challenged in recent work[72,73]. Relatively few neural responses in IT are invariant to position, size or viewpoint[74]. When trained on particular object views, monkeys recognize novel views according to their similarity to experienced views, and neurons respond in a similar graded fashion[24]. When monkeys are trained to categorize objects, their behaviour is consistent with exemplar generalization and not prototype abstraction[75], and IT neurons respond selectively to particular exemplars[37,76–78]. Interestingly, neurons often show more within-category discrimination than between-category discrimination, indicating a distributed representation that emphasizes exemplar-specific information rather than highlighting category-specific information[27,37]. Neural responses are also modulated by the diagnosticity of stimulus dimensions[76], consistent with learned selective attention, the cornerstone of many exemplar models[43,46,79,80] (FIG. 2). Category-specific responses are observed in the prefrontal cortex[78], which could be the category output layer that is assumed in all models from Perceptual Categorization and Object Recognition[50] (FIG. 2). Alternatively, a recent theory of category learning implicates the striatum in mediating the mapping from object representations in IT to category representations in the prefrontal cortex[42].

Although neurophysiology mainly reveals image-based responses in IT cortex, human functional magnetic resonance imaging (fMRI) studies provide evidence for both image-based and more abstract processing in homologous regions of human cortex[81,82]. There also seems to be a plurality of image-based mechanisms: one study compared mental rotation with recognition of the same novel objects and found that, despite the same behavioural effects of viewpoint, mental rotation engaged the parietal lobe but recognition engaged the fusiform gyrus[83]. So, the transformations that are used in object recognition might not be the same as those used in mental rotation, as had been suggested[61].

Levels of categorization

Objects can be categorized at several levels of abstraction (for example, animal, mammal, cat, Abyssinian, Max). By some accounts, discriminating similar objects (discerning Fido from Fifi) and generalizing across a category of objects (calling both of them dogs) are competing goals that require different representational systems[6,12,33,74]. One system is thought to be optimal for categorization at a BASIC[84] or ENTRY LEVEL[85], and the other for identification at the exemplar or subordinate level of categorization[16]. However, categorization can occur at more than two levels of a conceptual hierarchy, and no one has proposed a unique representational system for every level. Clearly, a flexible system is needed to account for categorization. This has led to the investigation of whether a single representational system can support categorization at multiple levels of abstraction[51,86].

Category levels from Object Recognition. This identification–categorization distinction is an undercurrent throughout debates between structural-description and image-based theorists. Abstract structural descriptions of objects at the same basic level[84] — for example, different cats — have the same geon description. Such descriptions are devoid of the metric information that is necessary to discriminate between members of the same class (but see REF. 87). Some argue that basic-level categorization is the fundamental goal of vision, with identification relying on features other than object shape[12]. By contrast, early tests of image-based theories emphasized discrimination at the subordinate level[23,61]. It is possible that structural descriptions support basic-level categorization and image-based representations support subordinate-level categorization[74]. Subordinate-level judgements have been suggested to depend more on viewpoint than on basic-level decisions[88]. However, presentation of novel objects among distractors of varying discriminability revealed that judgements of highly discriminable object sets (akin to a basic-level discrimination) were just as viewpoint-dependent as judgements of less discriminable object sets (akin to a subordinate-level discrimination)[89].

Recently, image-based theorists have argued that categorization at all levels can be accomplished using image-based representations[27,50,63]. Although theoretical work supports the sufficiency of a single, flexible system, some empirical evidence indicates that there might be multiple systems. Evidence primarily from lateralized-presentation studies indicates that objects can be represented simultaneously in an image-based and in an abstract fashion by different systems. For instance, the left hemisphere is more efficient at categorization, whereas the right hemisphere is more efficient at encoding exemplars[33,90,91]. Event-related potentials (ERPs)[92] reveal that the additional perceptual processing that is necessary for subordinate-level categorization (for example, terrier) is associated with early activity in the occipito-temporal cortex, whereas the semantic processing that is necessary for superordinate judgements (for example, animal) triggers later activity in frontal areas[92] (and see also REF. 93).

Category levels from Perceptual Categorization. Early work in Perceptual Categorization also suggested that identification and categorization used distinct representations and processes. Object similarities can be derived from stimulus–response confusions by subjects in an identification task. However, these identification-derived similarities could not account for categorization performance[94], indicating that qualitatively different representations supported identification and categorization. However, although object similarities vary systematically across tasks, the same underlying multidimensional psychological space can support both categorization and identification[43]. Selective attention to dimensions[35,46,48] can change stimulus similarities in a task-dependent manner depending on the relative diagnostic values of dimensions for identifying or categorizing objects. From a computational standpoint, |a common representational substrate can be used adaptively according to task demands[48,50,63].

Expertise and levels of categorization. One virtue of considering visual object understanding in terms of levels of categorization is in providing a framework for contrasting the performance of experts and novices. Novices make basic-level categorizations (for example, bird) more quickly than either more subordinate (robin) or superordinate (animal) categorizations. But when categorizing objects within their domain of expertise, experts (for example, dog experts) make subordinate-level judgements (terrier) as fast as basic-level judgements (dog) and are more likely to use subordinate-level labels in speeded naming[95,96]. In addition, Tanaka[97] showed that typical adults can be considered face experts in that they can identify faces (for example, Bill Clinton) as quickly as they categorize them at the basic level (human). Our expertise in recognizing faces can be contrasted with our expertise in recognizing letters. Although the goal of face recognition (like expert bird or dog recognition) is subordinate-level identification, the goal of expert letter identification is to distinguish letters at the basic level while ignoring within-class variations in font and writing style. In ongoing research in Gauthier's laboratory, comparisons of novices and experts with Chinese characters and experts with Roman characters indicate that expertise leads to a relative shift in categorization, but instead of subordinate-level judgements becoming as fast as basic-level judgements, judgements at the basic level become faster with expertise[98]. The levels of categorization framework can lead to a taxonomy of different kinds of expertise that might recruit different neural substrates.

One issue concerns the computational inferences that can be drawn by contrasting response times at different levels of categorization. The entry level can be interpreted as the first categorization stage that needs to be completed before processing at more subordinate or superordinate levels can begin. Indeed, subordinate-level processing of certain specific categories has been proposed to proceed only after an initial 'detection' stage, which categorizes an object as a member of that basic-level class — for instance, a face is only identified after it is categorized as a face[99,100]. Alternatively, basic-level judgements could be completed earlier because these categorizations can be made using coarse perceptual information that is available earlier than the detailed perceptual information required for more specific discriminations (BOX 1). Decisions about category membership can be made on the basis of partial perceptual information, with perceptual processing cascading into decision processes[47].

In understanding the flow of information processing, one challenge is to account for how experts achieve subordinate-level recognition as fast as basic-level recognition[95,96]. Some theories from Object Recognition have suggested that the development of perceptual expertise involves a switch from one representational system to another. For instance, experts might represent objects using a holistic system, similar to the one that is postulated for faces[101]. Alternatively, some theories from Perceptual Categorization suggest that rapid expert subordinate-level processing, as well as emergent holistic effects, can be explained by changes in memory representations[49,102] or decision stages[103] (see section on learning and perceptual expertise). In addition, expertise could entail the creation of new features to discriminate members of different classes[3] (see section on interactions between perception and conception).

Neural evidence. Neurophysiological evidence is consistent with cascaded processing at different levels of abstraction. Moving rostrally along IT, the response properties of neurons tend to increase in complexity[70,74]. However, efferents from IT are not limited to cells that encode entire objects. Rather, the entire span of cortex from area V4 through areas TEO and TE in the macaque projects to frontal and striatal areas (BOX 1). This includes the frontal eye fields, where neurons that respond to the categorization of visual objects[104] begin to respond before neurons in anterior IT[105]. Therefore, object representations might be distributed across many areas. Representations that are available after relatively early visual processing, for example, in area V4 or TEO, might be sufficient for basic-level judgements, but more detailed representations, formed later in area TE, might be needed for more specific decisions[106]. Image-based 'features' of intermediate complexity, possibly akin to responses of neurons in TEO, could be sufficient to categorize types of objects (such as faces or cars) but are probably insufficient to discriminate objects within a class[107].

In humans, fMRI studies[93] reveal that subordinate-level judgements (as compared to basic-level decisions) recruit more anterior regions of the occipito-temporal stream, including the fusiform face area. A CASCADE MODEL that emphasizes the importance of perceptual overlap between exemplars has also been proposed to account for category-selective deficits in brain-damaged individuals[108]. Because evidence from any single technique might be compatible with both cascade and serial implementations, converging evidence at many scales (temporal and spatial) will probably be required to test specific models of the flow of processing that is responsible for categorization at different levels of abstraction.

Modularity

Whether categorization and identification are subserved by independent systems is just one of the MODULARITY debates that permeate the Perceptual Categorization and Object Recognition literatures. Are there domain-specific, informationally encapsulated subsystems[8] for recognizing certain objects or performing certain object-understanding tasks?

Modular explanations abound for dissociations and double dissociations within normal individuals on different tasks[33,109], between brain-damaged individuals on the same tasks[101,110], and in functional brain imaging studies[111,112]. According to this logic, finding two patients with opposite patterns of deficits and spared function indicates the existence of modules. Unfortunately, this logic is not flawless. As Shallice[113] stated, "If modules exist, then double dissociations can reveal them. However, finding double dissociations is no guarantee that modules exist." Logical arguments[113], simulated deficits with computational models[114–116], and alternative analyses of fMRI data[117] all reveal that non-modular systems can yield double dissociations. Unquestionably, parts of the brain are specialized for different ways of representing, integrating or processing information. However, what distinguishes most modular approaches is the claim that some areas are exclusively devoted to one domain or task. By contrast, many non-modular accounts acknowledge that subsystems might be relatively more important for a domain or task, but emphasize interactions between systems in the service of adaptive behaviour[118–120]. There are some interesting contrasts in the way in which modularity issues have been manifest in the Object Recognition and Perceptual Categorization fields.

Researchers from Object Recognition have traditionally discussed modularity of content: are there specific modules that are devoted to particular kinds of object? According to one framework, object recognition is supported by a general-purpose mechanism in the lateral occipital cortex (LOC)[121]; fMRI reveals greater activation of the LOC by objects than by random patterns, whether those objects are defined by luminance, motion, texture or otherwise. This mechanism might be supplemented by a small number of modules for processing special categories such as faces[122], places[123] or body parts[111] (but see BOX 3). Interestingly, a similar pattern of specialization for faces and perhaps for bodies, but not for hands, fruits or man-made objects, has been reported in fMRI studies of macaques[124].

The Perceptual Categorization literature has been embroiled in debates regarding the modularity of memory systems[125,126] — are there specific modules that are devoted to particular tasks, irrespective of object category? Theories that propose multiple memory systems[125] suggest that categorization and recognition are subserved by independent systems. Evidence for this comes from a double dissociation in which amnesics can categorize new dot patterns as members of a previously seen category but cannot recognize dot patterns that they have seen before[40,127], whereas patients with Parkinson's disease cannot learn probabilistically defined categories but perform normally on tests of explicit recognition[42,128] (BOX 2).

In both fields, claims of modularity have been disputed. Neurons are selective for particular objects, including faces and other complex objects[24], but selectivity alone is not sufficient to support claims of a truly modular organization. As a rule, modularity tends to be invoked when neurons that are selective for a category cluster into spatially contiguous patches of cortex, something more easily seen with functional imaging than with neurophysiological recordings. Such localized hot spots of activity are found for a small number of object categories, including faces, places and body parts[111,122,123]. Furthermore, brain lesions sometimes seem to affect only one of these clusters, as in PROSOPAGNOSIA[129]. However, even brain lesions that lead to selective deficits are often large, diffuse or impossible to localize, and the distribution of category-selective areas seen with neuroimaging is far from a perfect match to the distribution of brain damage in patients.

Spatial clustering of selective neurons might be tied to claims of modularity because it seems reasonable to attribute a special function to a dense network of interconnected neurons with similar selectivity. However, our understanding of how these local cortical networks function is limited (but see REFS 76,130). In truth, the link between spatial clustering of selectivity and modularity has often been more assumed than explicitly deliberated. Spatially clustered selective neurons might be revealed to be parts of larger interactive, non-modular networks[117], or more spatially distributed systems might meet many of the characteristics that are otherwise expected of modules. Even if there are cognitive modules, we do not know what they should correspond to in neural terms (BOX 3).

CASCADE MODELS
Cascade models are those in which the later stages of information processing can begin before the completion of earlier stages, unlike discrete models in which computations at any given stage are completed before the subsequent step is engaged.

MODULARITY
A thesis concerning the structure of the mind that is based on special-purpose computational mechanisms termed 'modules'. Fodor[8] proposed that modules are innate, that they perform their operations on a specific input or domain (for example, faces or speech) and that their operations are informationally encapsulated (not accessible to any other module).

PROSOPAGNOSIA
Originally defined as the inability to gain a sense of familiarity from known faces, prosopagnosia also now includes a deficit in the perception of faces. It typically occurs in the context of visual agnosia — a visual deficit in object recognition — and only a few cases have been suggested to present with a face-specific deficit.

Box 3 | Are faces special?

Two related debates centre around whether faces are special. First, how can we explain the pattern of activity across extrastriate cortex in response to different categories? One view focuses on the maximal response of a cluster of neurons[163]. This leads to the conclusion that most classes of objects are recognized using a general-purpose system that is distributed across a large portion of cortex but that there are a few 'hot spots' for 'special' classes of objects, such as faces, places and body parts[35,123,163,164]. Alternatively, there might be a distributed and continuous representation of all categories of objects across the cortex, in which neurons that respond non-maximally nonetheless participate crucially in object representations[117]. According to the first perspective, faces are represented only in the fusiform face area (FFA); but from the other perspective, faces are coded in a distributed manner across the cortex, with the FFA simply being an area of maximal activity.

Another debate concerns the origins of the category specialization that is observed in extrastriate cortex, such as in the FFA[122]. Specialization itself is not challenged, but the question is how best to characterize it, as well as its causes. The replicability of the pattern of specialization for a few categories across individuals[117] indicates that innate or maturational constraints might govern the development of cortical representation. One view is that the 'face module' is innately programmed for the unique geometry of faces[165]. This was supported by findings that newborns prefer upright to inverted faces[166]. However, this preference might not be unique to faces *per se*: newborns prefer the version of any pattern that has more elements at its top[167]. An alternative is that more general constraints, such as gradients of eccentricity[168] or a continuum from local parts to holistic representations[169], govern organization across the visual cortex. Category selectivity correlates with these cortical biases because of the processing biases that are associated with certain objects through experience[118,170]. For instance, faces are recognized individually more than other objects[93] and we develop expertise with this task. Behavioural[144,149], functional magnetic resonance imaging[148,171] and event-related potential (ERP)[172,173] evidence reveals that experts can process non-face objects such as cars, dogs, birds and novel objects in a manner that is similar to face processing, using the same brain areas, and with neural responses with the same latency (see also REF. 174). In addition, face recognition by car experts — but not by car novices — shows interference, both behaviourally and in ERPs, from concurrent processing of cars[150], indicating that in experts, face and car processing share important neural resources.

In Perceptual Categorization, arguments against modular accounts have relied on demonstrations that non-modular models can account for dissociations between categorization and recognition in neuropsychological populations[115,116,120,131–133] and those between different kinds of categorization strategy in normal individuals[134,135]. Mirroring arguments from the Object Recognition field, it has been argued that categorization and recognition tasks often differ in important ways beyond judging category membership versus judging familiarity. For example, for the widely used dot pattern tasks[39,40,127] a single-system exemplar model predicts that categorization performance will be far less affected by memory impairment than recognition performance because of the similarities between stimuli used in those tasks[115,131]. A single-system exemplar model that assumes degraded memory in amnesia provides a parsimonious account of the behaviour of amnesics and normal individuals; recent arguments to the contrary[136] have been challenged[120]. In addition, in some experimental paradigms, memory for category members might not even be required. When subjects are deceived into believing that they have received subliminal exposure to training items — no items having been presented — their performance on categorization tests is the same as that observed in patients and controls who were exposed to training items[120,131,133].

Interactions between perception and conception

In addition to horizontal modularity for different kinds of object or kinds of memory, the traditional distinction between Object Recognition and Perceptual Categorization can be seen as a form of vertical modularity from the perceptual to the conceptual[8]. Traditionally, visual perception was thought to create the representational input (the domain of Object Recognition) to a conceptual system that identified or categorized objects (the domain of Perceptual Categorization). Recently, more 'interactive' solutions have been proposed[3–5].

As discussed earlier, in one traditional model, an object is first categorized at the basic level[84]. Categorization at more subordinate levels requires further perceptual processing[85,92], and categorization at superordinate levels requires further semantic processing[84,85]. This framework is supported by an ERP study that dissociates subordinate and superordinate processing both spatially and temporally in the brain[92]. However, such clear distinctions can break down in cases of perceptual expertise, where equivalent performance is seen at the basic and subordinate categorization[95,96].

But experts also acquire a rich knowledge base of information. For example, bird experts know not only what an Indigo Bunting looks like, but also that it eats seeds, berries and herbs, and can navigate using the stars. One study of a patient with category-specific visual agnosia indicates that perceptual and conceptual knowledge might not be independent[137]. Patient ELM was an expert on brass instruments before his brain injury. Surprisingly, he can learn arbitrary pairings between names of brass instruments and novel abstract shapes better than when the names are those of string instruments. The use of expert concepts led to facilitation in a perceptual task with arbitrary visual stimuli, indicating an interaction between perception and conceptual knowledge. Such interactions might occur even in the absence of expertise. Gauthier *et al.*[138] asked subjects to associate novel objects with non-visual semantic attributes. The semantic attributes overlapped either significantly, in that different objects shared many of the same attributes, or not at all. After learning, subjects performed matching judgements across viewpoint changes on these objects. They made fewer errors after learning non-overlapping semantic features during training, even though the matching task made no reference to the semantic features. So, conceptual knowledge might be invoked involuntarily to influence perceptual judgements. A similar study using fMRI found that conceptual associations with novel objects recruited the frontal lobe during a simple perceptual matching task[139]. The frontal areas involved are active during semantic encoding, retrieval and generation[140].

Along similar lines, computational arguments[3] and empirical evidence[9] have shown that category learning can influence perceptual representations. After category learning, perceptual discriminations can be enhanced along dimensions that are relevant for the learned categorization[9]. Such perceptual learning could be due to enhanced receptive field properties or an increased number of units coding a diagnostic dimension. Although IT neurons seem to encode stimuli in a way that emphasizes diagnostic dimensions[37,76], the neural mechanisms involved are still unclear[80]. Perhaps there is a more complex form of perceptual learning in which the underlying dimensional descriptions adapt to the categorization task that is being learned[3]. The mechanisms of such feature creation remain to be elucidated, but image-based parts of intermediate complexity[107] might be a starting point.

Finally, perhaps the greatest swing away from a modular view of visual object understanding counters the classic view that conceptual knowledge is amodal and separate from episodic memory, and instead suggests that abstract semantic knowledge is grounded in specific episodes of perception, action and emotion[4]. According to this view, your knowledge of a kiwi fruit is the conjunction of the visual, somatosensory, motor, olfactory and gustative states that are stored each time you interact with one. Such conjunctions could be represented by neurons in 'convergence zones'[141], and the process of reactivating such states in modal systems has been described as 'running simulations'. These simulations can combine stored information, making it possible to experience concepts in novel ways that diverge from actual experience. The implementational details of this framework remain speculative, but it is consistent with behavioural results showing that sensorimotor variables can affect performance in feature listing or property verification — for example, subjects are more likely to produce 'roots' as a property for rolled-up lawn than for lawn[142] — and is consistent with neuroimaging results showing that the use of concepts activates processing in modality-specific areas — for example, motion-related features (such as hops, jumps or walks) recruit cortical areas that overlap with those that are engaged by biological motion perception[143].

Learning and perceptual expertise

Historically, both the Perceptual Categorization and Object Recognition literatures were largely polarized around binary questions, for example: are object representations viewpoint-dependent or viewpoint-independent[12,61]? Is perceptual categorization based on exemplars, prototypes or rules[35]? Are faces and objects processed in the same way[144]? Recently, a more dynamic approach has been favoured. Novice subjects can demonstrate visual object understanding in qualitatively different ways than experts. Fully understanding the development of learning and perceptual expertise will demand a synergy of several fields of research, as perceptual expertise might reflect changes in perceptual representations, perceptual knowledge and conceptual knowledge.

A novice searching for Morel mushrooms might use well-memorized rules, categorizing mushrooms as edible or poisonous explicitly. With experience, a mushroom gatherer shifts from this slow, attention-demanding mode to a more rapid and automatic mode of expert categorization. Instance theory[52] posits shifts from strategic rule-based processes to exemplar-based memory retrieval as the basis for the development of automaticity in a range of domains (BOX 2), including perceptual categorization[49,102,134]. People might initially categorize using rules[102,145–147], but with experience they start to retrieve exemplars from memory[49,102,134]; expert performance is faster, less deliberate and less attention-demanding than categorization by novices, because memory retrieval is faster, less deliberate and less attention-demanding than most rule use[52].

The automatic strategies that are acquired by perceptual experts can sometimes confound study results. For instance, comparisons between object and face processing often indicate that faces are processed in a 'special' way (BOX 3). However, subjects are more experienced at recognizing faces individually. This might influence performance on identification tasks, but face experts — typical adults — might also automatically recognize faces individually when they are not explicitly instructed to do so. In an fMRI study in which car or bird experts selectively attended to the identity or the location of faces, birds or cars[148], novices (birders looking at cars or car experts looking at birds) showed more activity in the fusiform face area (FFA) during identity than location judgements. By contrast, there was little difference in FFA activity between identity and location judgements for experts with birds and cars, just as for faces. So, expertise might promote automatic processing of identity, even when identity is irrelevant.

Although perceptual expertise confers benefits to the expert, some behavioural effects of expertise are best described as interference. For example, subjects might be asked to attend to part of an object, but experts show a larger influence of the parts they were told to ignore, a 'holistic processing' effect[149,150]. The same effects are obtained with faces[101,151]. Expertise carries a cost: the loss of flexibility of processing for objects in a highly trained domain.

Comparisons between novices processing objects and experts processing faces have provided evidence for two representational systems, one representing objects in terms of their parts and the other representing objects as relatively undifferentiated wholes[101]. During the acquisition of perceptual expertise, there could be a switch to using holistic representations, rendering experts unable to attend selectively to parts. However, when subjects are tested at various timepoints during the acquisition of expertise, holistic processing seems to occur at different times for different parts of an object[149]: there were intermediate steps between apparent part-based and holistic processing when only a subset of an object's parts seemed to be bound together. Therefore, rather than a representational switch, the acquisition of expertise could rely on fine tuning and quantitative changes in the same representational system.

Before we can fully understand how processes and representations become 'holistic' with expertise, clarifications about the mechanisms that underlie holistic processing, or what it means to form a holistic representation, are needed[152]. Until recently, accounts of holistic processing were more verbal descriptions than specified models[149,153]. Image-based representations[27,28] provide one natural substrate for holistic or unitized representations[154]. It might also be useful to decompose holistic processing into specific aspects of information processing: whether components of an object are processed independently, whether those components are processed in parallel, whether there are capacity limitations, and so forth[103,152]. Although some holistic processing effects could emerge as a result of underlying holistic representations, others could instead emerge because of decision processes that integrate information about the dimensions of an object[103].

Conclusion

We are just beginning to understand how people learn to recognize, identify and categorize novel objects, in some cases developing expertise in a domain. These challenges highlight the importance of investigating changes in perceptual representations, knowledge representations and processes that act on those representations. The line between perception and cognition has been blurred. Despite their historic differences, current theories from Object Recognition and from Perceptual Categorization have begun to consider complementary problems and have converged on similar solutions. Ultimately, a complete understanding of visual object understanding will demand an integration of the best theoretical constructs from the Object Recognition and Perceptual Categorization fields.

1. Jennings, C. & Aamodt, S. Computational approaches to brain function. *Nature Neurosci.* **3**, Suppl. 1160 (2000).
2. Schyns, P. G. Diagnostic recognition: task constraints, object information, and their interactions. *Cognition* **67**, 147–179 (1998).
3. Schyns, P. G., Goldstone, R. L. & Thibaut, J.-P. The development of features in object concepts. *Behav. Brain Sci.* **21**, 1–54 (1998).
This paper brings together several lines of evidence in favour of a theory that category learning involves the flexible creation of new perceptual features.
4. Barsalou, L.-W. Perceptual symbol systems. *Behav. Brain Sci.* **22**, 577–660 (1999).
Argues against amodal theories of conceptual knowledge, but instead proposes a theory that abstract conceptual knowledge is grounded in perceptual experiences.
5. Goldstone, R. L. & Barsalou, L. W. Reuniting perception and conception. *Cognition* **65**, 231–262 (1998).
6. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, 1982).
7. Rosch, E. Cognitive representations of semantic categories. *J. Exp. Psychol. Gen.* **104**, 192–233 (1975).
8. Fodor, J. A. *Modularity of Mind* (MIT Press, Cambridge, Massachusetts, 1983).
9. Goldstone, R. Influences of categorization on perceptual discrimination. *J. Exp. Psychol. Gen.* **123**, 178–200 (1994).
10. Hintzman, D. L. Human learning and memory: connections and dissociations. *Ann. Rev. Psychol.* **41**, 109–139 (1990).
11. Ashby, F. G. *Multidimensional Models of Perception and Cognition* (Lawrence Erlbaum, Hillsdale, New Jersey, 1992).
A classic tutorial volume that brings together theories from mathematical psychology on categorization and identification, theories from psychometrics on similarity, choice and preference, and theories from visual perception. The unifying theoretical theme is that all assume probabilistic multidimensional representations of perceptual and cognitive information.
12. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
13. Hummel, J. E. & Biederman, I. Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* **99**, 480–517 (1992).
14. Biederman, I. & Gerhardstein, P. C. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J. Exp. Psychol. Hum. Percept. Perform.* **19**, 1162–1182 (1993).
Defines the conditions under which a specific structural description theory (RBC) predicts depth invariance for object recognition.
15. Edelman, S. Computational theories of object recognition. *Trends Cogn. Sci.* **1**, 296–304 (1997).
16. Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P. & Fiser, J. Subordinate-level object classification reexamined. *Psychol. Res.* **62**, 131–153 (1999).
17. Hummel, J. E. & Stankiewicz, B. J. Two roles for attention in shape perception: a structural desription model of visual scrutiny. *Vis. Cogn.* **5**, 49–79 (1998).
Presents a structural description model of object recognition that integrates both categorical and metric information to account for recognition at both the basic and subordinate levels.
18. Ullman, S. Aligning pictorial descriptions: an approach to object recognition. *Cognition* **32**, 193–254 (1989).
19. Mel, B. SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput.* **9**, 977–804 (1997).
20. Pinker, S. Visual cognition: an introduction. *Cognition* **18**, 1–63 (1984).
21. Poggio, T., Bülthoff, H. & Lee, S. W. in *BMCV 2000* (Springer, Seoul, 2000).
22. Bülthoff, H. H. & Edelman, S. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl Acad. Sci. USA* **89**, 60–64 (1992).
23. Bülthoff, H. H., Edelman, S. Y. & Tarr, M. J. How are three-dimensional objects represented in the brain? *Cereb. Cortex* **5**, 247–260 (1995).
24. Logothetis, N. K. & Pauls, J. Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb. Cortex* **5**, 270–288 (1995).
25. Ullman, S. & Basri, R. Recognition by linear combinations of models. *IEEE PAMI* **13**, 992–1006 (1991).
26. Tarr, M. J. Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychon. Bull. Rev.* **2**, 55–82 (1995).
Proposes and provides empirical support for a multiple-views-plus-rotation model in which view-specific representations are matched to percepts through a normalization procedure akin to mental rotation.
27. Edelman, S. *Representation and Recognition in Vision* (MIT Press, Cambridge, Massachusetts, 1999).
28. Poggio, T. & Edelman, S. A network that learns to recognize three-dimensional objects. *Nature* **343**, 263–266 (1990).
29. Poggio, T. & Girosi, F. Regularization algorithms for learning that are equivalent to multilayer networks. *Science* **247**, 978–982 (1990).
30. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**, 1019–1025 (1999).
An image-based model from Object Recognition that combines and extends preceding models: early layers of the hierarchical model create representations that are size- and translation-invariant. These representations are matched to view-tuned image-based units, which in turn activate category and identity nodes.
31. Fukushima, K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**, 193–202 (1980).
32. Schneider, R. & Riesenhuber, M. *A Detailed Look at Scale and Translation Invariance in a Hierarchical Neural Model of Visual Object Recognition*. CBCL Paper 218/AI Memo 2002-011 (Massachusetts Institute of Technology, Cambridge, Massachusetts, 2002).
33. Marsolek, C. Dissociable neural subsystems underlie abstract and specific object recognition. *Psychol. Sci.* **107**, 111–118 (1999).
Provides empirical support for a dual systems theory of object recognition that includes an abstract subsystem for categorization in the left hemisphere and an exemplar-specific subsystem for individuation in the right hemisphere.
34. Foster, H. G. & Gilson, S. J. Recognizing novel three-dimensional objects by summing signals from parts and views. *Proc. R. Soc. Lond. B* **269**, 1939–1947 (2002).
35. Nosofsky, R. M. Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* **115**, 39–61 (1986).
Presents a unified exemplar-based model of categorization and identification, the generalized context model, by integrating the context model of categorization with the similarity-choice model of identification. Because of selective attention to dimensions, similarities can vary systematically between categorization and identification.
36. Op de Beeck, H., Wagemans, J. & Vogels, R. The effect of category learning on the representation of shape: dimensions can be biased, but not differentiated. *J. Exp. Psychol. Gen.* **132**, 491–511 (2003).
Provides new results that might temper theories of novel feature creation during category learning. Extensive category learning can bias the processing of already separable shape dimensions but cannot differentiate already integral shape dimensions.
37. Op de Beeck, H., Wagemans, J. & Vogels, R. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neurosci.* **4**, 1244–1252 (2001).
Responses of neurons in IT cortex are characterized as low-dimensional representations of complex shape; interestingly, these neurons often show more within-category discrimination than between-category discrimination, indicating a distributed representation that emphasizes exemplar-specific information rather than highlighting category-specific information.
38. Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. *J. Exp. Psychol.* **77**, 353–363 (1968).
39. Smith, J. D. & Minda, J. P. Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 800–811 (2002).
40. Knowlton, B. & Squire, L. R. The learning of categories: parallel brain systems for item memory and category knowledge. *Science* **262**, 1747–1749 (1993).
41. Ashby, F. G. A stochastic version of general recognition theory. *J. Math. Psychol.* **44**, 310–329 (2000).
42. Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. & Waldron, E. M. A formal neuropsychological theory of multiple systems in category learning. *Psychol. Rev.* **105**, 442–481 (1998).
Presents a neuropsychological theory of category learning that assumes competition between separate implicit procedural learning (mediated by the caudate nucleus and other structures) and verbal rule (mediated by the anterior cingulate, prefrontal cortex and other structures) systems.
43. Nosofsky, R. M. Choice, similarity, and the context theory of classification. *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 104–114 (1984).
44. Hintzman, D. L. 'Schema abstraction' in a mutiple-trace memory model. *Psych. Rev.* **93**, 411–428 (1986).
45. Nosofsky, R. M. Exemplar-based accounts of relations between classification, recognition, and typicality. *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 700–708 (1988).
46. Kruschke, J. K. ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* **99**, 22–44 (1992).
A connectionist, exemplar-based model of category learning that combines the generalized context model of categorization with an error-driven learning rule to allow learning of selective attention to dimensions and of associations between exemplars and categories.

47. Lamberts, K. Information-accumulation theory of speeded categorization. *Psychol. Rev.* **107**, 227–260 (2000).
Presents a process model of categorization that accounts for both response probabilities and response times. The model assumes that perceptual features are sampled probabilistically over time according to their salience until sufficient information has been acquired to generate a response. Under time pressure, responses are driven by salient features. Under deliberate decisions, responses are driven by diagnostic features.
48. Nosofsky, R. M. in *Rational Models of Cognition* (eds Oaksford, M. & Chater, N.) (Oxford Univ. Press, Oxford, 1998).
49. Nosofsky, R. M. & Palmeri, T. J. An exemplar-based random walk model of speeded classification. *Psychol. Rev.* **104**, 266–300 (1997).
50. Riesenhuber, M. & Poggio, T. Models of object recognition. *Nature Neurosci.* **3**, Suppl., 1199–1204 (2000).
51. Tarr, M. J. & Vuong, Q. C. in *Steven's Handbook of Experimental Psychology* 3rd edn Vol. 1 (eds Pashler, H. & Yantis, S.) 287–314 (John Wiley & Sons, New York, 2002).
52. Logan, G. D. Toward an instance theory of automatization. *Psychol. Rev.* **95**, 492–527 (1988).
53. Tjan, B. S. & Legge, G. E. The viewpoint complexity of an object recognition task. *Vision Res.* **38**, 2335–2350 (1998).
54. Rosseel, Y. Mixture models of categorization. *J. Math. Psychol.* **46**, 178–210 (2002).
55. Barsalou, L. W. in *Building Object Categories* (ed. Rakison, D.) (Erlbaum, Mahwah, New Jersey, in the press).
56. Tarr, M. J. & Bülthoff, H. H. Image-based object recognition in man, monkey and machine. *Cognition* **67**, 1–20 (1998).
57. Booth, M. C. A. & Rolls, E. T. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex* **8**, 510–523 (1998).
58. Palmer, S., Rosch, E. & Chase, P. in *Attention and Performance IX* (eds Long, J. & Baddeley, A.) 135–151 (Lawrence Erlbaum, Hillsdale, New Jersey, 1981).
59. Tarr, M. J. & Bülthoff, H. H. Is human object recognition better described by geon-structural-descriptions or by multiple-views? *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 1494–1505 (1995).
60. Wallis, G. & Bülthoff, H. Learning to recognize objects. *Trends Cogn. Sci.* **3**, 22–31 (1999).
61. Tarr, M. J. & Pinker, S. Mental rotation and orientation-dependence in shape recognition. *Cogn. Psychol.* **21**, 233–282 (1989).
62. Tarr, M. J., Williams, P., Hayward, W. G. & Gauthier, I. Three-dimensional object recognition is viewpoint-dependent. *Nature Neurosci.* **1**, 275–277 (1998).
Tests a fundamental prediction from a specific structural description model (RBC), according to which simple volumes called geons are the building blocks of a system that leads to depth-invariant recognition. Several experiments using various tasks and stimulus conditions find that geon recognition is viewpoint-dependent.
63. Tarr, M. J. in *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes* (eds Peterson, M. A. & Rhodes, G.) 177–211 (Oxford Univ. Press, Oxford, 2003).
64. Nosofsky, R. & Palmeri, T. J. A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychon. Bull. Rev.* **5**, 345–369 (1998).
65. Homa, D. Abstraction of ill-defined form. *J. Exp. Psychol. Hum. Learn. Mem.* **4**, 407–416 (1978).
66. Reed, S. K. Pattern recognition and categorization. *Cogn. Psychol.* **3**, 382–407 (1972).
67. Medin, D. L. & Schaffer, M. M. Context theory of classification learning. *Psychol. Rev.* **85**, 207–238 (1978).
68. Ashby, F. G. & Waldron, E. M. On the nature of implicit categorization. *Psychon. Bull. Rev.* **6**, 363–378 (1999).
69. Perrett, D. I., Oram, M. W. & Ashbridge, E. Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* **67**, 111–145 (1998).
70. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
71. Tovee, M. J., Rolls, E. T. & Azzopardi, P. Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *J. Neurophysiol.* **72**, 1049–1060 (1994).
72. Op de Beeck, H. & Vogels, R. Spatial sensitivity of macaque inferior temporal neurons. *Comp. Neurol.* **426**, 505–518 (2000).
73. DiCarlo, J. J. & Maunsell, J.-H. R. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J. Neurophysiol.* **89**, 3264–3278 (2003).
74. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
75. Sigala, N., Gabbiani, F. & Logothetis, N. K. Visual categorization and object representation in monkeys and humans. *J. Cogn. Neurosci*. **14**, 187–198 (2002).
Monkeys and humans learned novel object categories with feedback. Comparative fits of computational models to observed response probabilities indicated that neither monkeys nor humans learned categories by abstracting a prototype, but instead based categorization decisions either on similarity to exemplars or with respect to a linear decision boundary.
76. Sigala, N. & Logothetis, N. K. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415**, 318–320 (2002).
Monkeys learned novel categories of objects with four varying feature dimensions, only two of which were diagnostic for categorization. Single-unit recordings of cells in IT cortex revealed neural responses sensitive to dimensional diagnosticity, consistent with the construct of dimensional selective attention found in models such as the generalized context model.
77. Vogels, R., Biederman, I., Bar, M. & Lorincz, A. Inferior temporal neurons show greater sensitivity to nonaccidental metric shape differences. *J. Cogn. Neurosci.* **15**, 444–453 (2001).
78. Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. A comparison of primate prefrontal and temporal cortices during categorization. *J. Neurosci.* **23**, 5235–5246 (2003).
Monkeys learned to categorize 'dogs' and 'cats' that parametrically varied in shape. Neurons in IT cortex seem more involved in encoding object shape whereas neurons in PFC seem more involved in encoding object category, object memory and meaning.
79. Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C. & Glauthier, P. Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Mem. Cogn.* **22**, 352–369 (1994).
80. Gauthier, I. & Palmeri, T. J. Visual neurons: categorization-based selectivity. *Curr. Biol.* **12**, R282–284 (2002).
81. Kourtzi, Z., Erb, M., Grodd, W. & Bülthoff, H. H. Representation of the perceived 3D object shape in the human lateral occipital complex. *Cereb. Cortex* **13**, 911–920 (2003).
82. James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S. & Goodale, M. A. Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron* **35**, 793–801 (2002).
83. Gauthier, I. *et al.* BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron* **34**, 161–171 (2002).
84. Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. Basic objects in natural categories. *Cogn. Psychol.* **8**, 382–439 (1976).
85. Jolicoeur, P., Gluck, M. & Kosslyn, S. M. Pictures and names: making the connection. *Cogn. Psychol.* **16**, 243–275 (1984).
86. McClelland, J. L. & Rumelhart, D. E. Distributed memory and the representation of general and specific information. *J. Exp. Psychol. Gen.* **114**, 159–188 (1985).
87. Stankiewicz, B. J. Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 913–932 (2002).
88. Tarr, M. J. & Bülthoff, H. H. Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 1494–1505 (1995).
89. Hayward, W. G. & Williams, P. Viewpoint dependence and object discriminability. *Psychol. Sci.* **11**, 7–12 (2000).
Refutes a common prediction that object recognition should be viewpoint-invariant when an object set is highly discriminable. The results support the notion that differences in object geometry are more important than the difficulty of the categorization in terms of accounting for viewpoint effects.
90. Laeng, B., Zarrinpar, A. & Kosslyn, S. M. Do separate processes identify objects as exemplars versus members of basic-level categories? Evidence from hemispheric specialization. *Brain Cogn.* **53**, 15–27 (2003).
91. Vuilleumier, P., Henson, R. N., Driver, J. & Dolan, R. J. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neurosci.* **5**, 491–499 (2002).
92. Tanaka, J., Luu, P., Weisbrod, M. & Kiefer, M. Tracking the time course of object categorization using event-related potentials. *Neuroreport* **10**, 829–835 (1999).
93. Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P. & Gore, J. C. Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Curr. Biol.* **7**, 645–651 (1997).
94. Shepard, R. N., Hovland, C. I. & Jenkins, H. M. Learning and memorization of classifications. *Psychol. Monogr.* **75** (1961).
95. Johnson, K. E. & Mervis, C. B. Effects of varying levels of expertise on the basic level of categorization. *J. Exp. Psychol. Gen.* **126**, 248–277 (1997).
96. Tanaka, J. W. & Taylor, M. Object categories and expertise: is the basic level in the eye of the beholder? *Cogn. Psychol.* **23**, 457–482 (1991).
97. Tanaka, J. W. The entry point of face recognition: evidence for face expertise. *J. Exp. Psychol. Gen.* **130**, 534–543 (2001).
Supports the notion that people are experts with faces by showing that faces are more frequently identified, are equally quick to identity at the subordinate level (an individual) as the basic level (human), and that face recognition is not impaired by brief presentations in the way the recognition of other objects is.
98. Wong, A. C.-N. & Gauthier (2003). The basic level as the entry point of expert letter recognition. *Annu. Meet. Cogn. Neurosci. Soc. Abstr*. C295 (2003)
99. Liu, J. & Kanwisher, N. Stages of processing in face perception: an MEG study. *Nature Neurosci.* **5**, 910–916 (2002).
100. Bentin, S., Deouell, L. Y. & Soroker, N. Selective visual streaming in face recognition: evidence from developmental prosopagnosia. *Neuroreport* **10**, 823–827 (1999).
101. Farah, M. J., Wilson, K. D., Drain, M. & Tanaka, J. W. What is 'special' about face perception? *Psychol. Rev.* **105**, 482–498 (1998).
An empirical and review paper on behavioural evidence supporting the hypothesis that faces are recognized in a more holistic fashion than other objects.
102. Palmeri, T. J. Exemplar similarity and the development of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* **23**, 324–354 (1997).
103. Wenger, M. J. & Ingvalson, E. M. A decisional component of holistic encoding. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 872–892 (2002).
Describes the construct of holistic representation in terms of informational independence, informational separability and decisional separability adopted from general recognition theory; only informational independence and informational separability can be related to traditional notions of holistic representations. But only evidence for decisional separability was observed.
104. Schall, J. D. On building a bridge between brain and behavior. *Annu. Rev. Psychol.* (in the press).
105. Schmolesky, M. T. *et al.* Signal timing across the macaque visual system. *J. Neurophysiol.* **79**, 3272–3278 (1998).
106. Tjan, B. Adaptive object representation with hierarchically-distributed memory sites. *Adv. Neural Inf. Process. Syst.* (in the press).
107. Ullman, S., Vidal-Naquet, M. & Sali, E. Visual features of intermediate complexity and their use in classification. *Nature Neurosci.* **5**, 682–687 (2002).
Provides computational evidence that image-based 'features' of intermediate complexity can be sufficient to categorize objects at the basic level.
108. Humphreys, G. W., Lamote, C. & Lloyd-Jones, T.-J. An interactive activation approach to object processing: effects of structural similarity, name frequency, and task in normality and pathology. *Memory* **3**, 535–586 (1995).
109. Waldron, E. M. & Ashby, F. G. The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychon. Bull. Rev.* **8**, 168–176 (2001).
110. Moscovitch, M., Winocur, G. & Behrmann, M. What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *J. Cogn. Neurosci.* **9**, 555–604 (1997).
111. Kanwisher, N., Downing, P., Epstein, R. & Kourtzi, Z. in *The Handbook on Functional Neuroimaging* (eds Cabeza, R. & Kingstone, A.) 109–152 (MIT Press, Cambridge, Massachusetts, 2001).
112. Patalano, A. L., Smith, E. E., Jonides, J. & Koeppe, R. A. PET evidence for multiple strategies of categorization. *Cogn. Affect. Behav. Neurosci.* **1**, 360–370 (2001).
113. Shallice, T. *From Neuropsychology to Mental Structure* (Cambridge Univ. Press, Cambridge, 1988).
114. Plaut, D. C. Double dissociation without modularity: evidence from connectionist neuropsychology. *J. Clin. Exp. Neuropsychol.* **17**, 291–321 (1995).
115. Nosofsky, R. & Zaki, S. Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. *Psychol. Sci.* **9**, 247–255 (1998).
116. Kinder, A. & Shanks, D. R. Amnesia and the declarative/nondeclarative distinction: a recurrent network model of classification, recognition, and repetition priming. *J. Cogn. Neurosci.* **13**, 648–669 (2001).

117. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
118. Gauthier, I. What constrains the organization of the ventral temporal cortex? *Trends Cogn. Sci.* **4**, 1–2 (2000).
119. Poldrack, R.-A. *et al.* Interactive memory systems in the human brain. *Nature* **29**, 546–550 (2001).
Provides evidence that both the medial temporal lobes (emphasized early in learning) and the basal ganglia (emphasized later in learning) are involved in category learning.
120. Palmeri, T. J. & Flanery, M. A. in *The Psychology of Learning and Motivation* vol. 41 (ed. Ross, B. H.) (Academic, San Diego, California, 2002).
Critically evaluates recent evidence for single memory system versus multiple memory system models of categorization and old–new recognition memory.
121. Malach, R. *et al.* Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl Acad. Sci. USA* **92**, 8135–8139 (1995).
122. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
123. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
124. Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B. & Tootell, R. B. Faces and objects in macaque cerebral cortex. *Nature Neurosci.* **6**, 989–995 (2003).
125. Squire, L. R. & Zola, S. M. Episodic memory, semantic memory, and amnesia. *Hippocampus* **8**, 205–211 (1998).
126. Roediger III, H. L., Buckner, R. L. & McDermot, K. B. in *Memory: Systems, Process, or Function?* (eds Foster, J. K. & Jelicic, M.) 31–65 (Oxford Univ. Press, Oxford, 1999).
127. Squire, L. R. & Knowlton, B. Learning about categories in the absence of memory. *Proc. Natl Acad. Sci. USA* **92**, 12470–12474 (1995).
128. Knowlton, B. J., Mangels, J. A. & Squire, L. R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).
129. Farah, M. J., Levinson, K. L. & Klein, K. L. Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia* **33**, 661–674 (1995).
130. Rolls, E. Memory systems in the brain. *Annu. Rev. Psychol.* **51**, 599–630 (2000).
131. Palmeri, T. J. & Flanery, M. A. Learning about categories in the absence of training: profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychol. Sci.* **10**, 526–530 (1999).
132. Zaki, S. R., Nosofsky, R. M., Jessup, N. M. & Ungerzagt, F. W. Categorization and recognition performance of a memory-impaired group: evidence for single-system models. *J. Int. Neuropsychol. Soc.* **9**, 394–406 (2003).
133. Zaki, S. R. & Nosofsky, R. M. A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cogn. Affect. Behav. Neurosci.* **1**, 344–359 (2001).
134. Johansen, M. K. & Palmeri, T. J. Are there representational shifts during category learning? *Cogn. Psychol.* **45**, 482–553 (2002).
135. Nosofsky, R. M. & Johansen, M. K. Exemplar-based accounts of 'multiple-system' phenomena in perceptual categorization. *Psychon. Bull. Rev.* **7**, 375–402 (2000).
136. Smith, J. D. & Minda, J. P. Thirty categorization results in search of a model. *J. Exp. Psychol. Learn. Mem. Cogn.* **26**, 3–27 (2000).
137. Dixon, M. J., Desmarais, G., Gojmerac, C., Schweizer, T. A. & Bub, D. N. The role of premorbid expertise on object identification in a patient with category-specific visual agnosia. *Cogn. Neuropsychol.* **19**, 401–419 (2002).
An elegant case study of a patient with category-specific agnosia. This paper indicates that a better knowledge of attributes for a category of expertise (in this case brass instruments) can facilitate learning of arbitrary associations between concepts and novel shapes.
138. Gauthier, I., James, T. W., Curby, K. M. & Tarr, M. J. The influence of conceptual knowledge on visual discrimination. *Cogn. Neuropsychol.* **20**, 507–523 (2002).
139. James, T. W. & Gauthier, I. Brain areas engaged by involuntary access to novel conceptual information during visual judgments. *Vision Res.* (in the press).
140. Murtha, S., Chertkow, H., Beauregard, M. & Evans, A. The neural substrate of picture naming. *J. Cogn. Neurosci.* **11**, 399–423 (1999).
141. Damasio, A.-R. *Descartes' Error: Emotion, Reason, and the Human Brain* (G. P. Putnam's Sons, New York, 1994).
142. Barsalou, L. W., Solomon, K. O. & Wu, L. L. in *Cultural, Typological, and Psychological Perspectives in Cognitive Linguistics: The Proceedings of the 4th Conference of the International Cognitive Linguistics Association* Vol. 3. (eds Hiraga, M. K., Sinha, C. & Wilcox, S.) (John Benjamins, Amsterdam, 1999).
143. James, T. W. & Gauthier, I. Auditory and action semantic feature types activate sensory-specific perceptual brain regions. *Curr. Biol.* (in the press).
144. Diamond, R. & Carey, S. Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. Gen.* **115**, 107–117 (1986).
145. Allen, S. W. & Brooks, L. R. Specializing the operation of an explicit rule. *J. Exp. Psychol. Gen.* **120**, 3–19 (1991).
146. Palmeri, T. J. & Nosofsky, R. M. Recognition memory for exceptions to the category rule. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 548–568 (1995).
147. Nosofsky, R. M., Palmeri, T. J. & McKinley, S. C. Rule-plus-exception model of classification learning. *Psychol. Rev.* **101**, 53–79 (1994).
148. Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neurosci.* **3**, 191–197 (2000).
149. Gauthier, I. & Tarr, M. J. Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 431–446 (2002).
150. Gauthier, I., Curran, T., Curby, K. M. & Collins, D. Perceptual interference supports a non-modular account of face processing. *Nature Neurosci.* **6**, 428–432 (2003).
Provides evidence that expertise with cars can interfere with face processing, specifically at the level of holistic processing taking place as early as 170 ms after the stimulus.
151. Young, A. W., Hellawell, D. & Hay, D. Configural information in face perception. *Perception* **10**, 747–759 (1987).
152. Wenger, M. J. & Townsend, J. T. in *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges.* (eds Wenger, M. J. & Townsend, J. T.) (Erlbaum, Hillsdale, New Jersey, 2001).
153. Maurer, D., LeGrand, R. & Mondloch, C. J. The many faces of configural processing. *Trends Cogn. Sci.* **6**, 255–260 (2002).
154. Goldstone, R. L. Unitization during category learning. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 86–112 (2000).
Systematically explores the question of whether new perceptual units can be created during category learning.
155. Dailey, M. N. & Cottrell, G. W. Organization of face and object recognition in modular neural network models. *Neural Netw.* **12**, 1053–1073 (1999).
156. Riesenhuber, M. & Poggio, T. in *Biologically Motivated Computer Vision* (eds Lee, S.-W., Bulthoff, H. H. & Poggio, T.) 1–9 (Springer, Berlin Heidelberg, 2000).
157. Erikson, M. A. & Kruschke, J. K. Rules and exemplars in category learning. *J. Exp. Psychol. Gen.* **127**, 107–140 (1998).
158. Smith, E.-E., Patalano, A. L. & Jonides, J. Alternative strategies of categorization. *Cognition* **65**, 167–196 (1998).
159. Erickson, M. A. & Kruschke, J. K. Rule-based extrapolation in perceptual categorization. *Psychon. Bull. Rev.* **9**, 160–168 (2002).
An example of the recent theoretical and empirical research from Perceptual Categorization that proposes both exemplar-based and rule-based components to category learning.
160. Sloman, S. A. The empirical case for two systems of reasoning. *Psychol. Bull.* **119**, 3–22 (1996).
161. Nosofsky, R. & Kruschke, J. K. Single-system models and interference in category learning: commentary on Waldron and Ashby (2001). *Psychon. Bull. Rev.* **9**, 175–180 (2001).
162. Ashby, F. G., Waldron, E. M., Lee, W. W. & Berkman, A. Suboptimality in human categorization and identification. *J. Exp. Psychol. Gen.* **130**, 77–96 (2001).
163. Kanwisher, N. & Spiridon M. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* **35**, 1157–1165 (2002).
164. Downing, P. E., Jiang, Y., Shuman, M. & Kanwisher, N. A cortical area selective for visual processing of the human body. *Science* **293**, 2470–2473 (2001).
165. Farah, M. J., Rabinowitz, C., Quinn, G. E. & Liu, G. T. Early commitment of neural substrates for face recognition. *Cogn. Neuropsychol.* **17**, 117–124 (2000).
166. Morton, J. & Johnson, M. H. CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychol. Rev.* **98**, 164–181 (1991).
167. Turati, C., Simion, F., Milani, I. & Umilta, C. Newborns' preference for faces: what is crucial? *Dev. Psychol.* **38**, 875–882 (2002).
Elegant set of studies with newborns which suggests that the bias to prefer an upright to an inverted face-like configuration can be explained by a preference for any pattern with more elements in its upper part.
168. Hasson, U., Levy, I., Behrmann, M., Hendler, T. & Malach, R. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* **34**, 479–490 (2002).
Provides support from fMRI for a centre–periphery model of occipito-temporal cortex, according to which objects requiring foveation (for example, faces) are associated with a centre-biased representation and objects requiring the integration of large-scale features (for example, buildings) are represented in part of the cortex that is periphery-biased.
169. Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M. & Malach, R. A hierarchical axis of object processing stages in the human visual cortex. *Cereb. Cortex* **11**, 287–297 (2001).
170. Malach, R., Levy, I. & Hasson, U. The topography of high-order human object areas. *Trends Cogn. Sci.* **6**, 176–184 (2002).
171. Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P. & Gore, J. C. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neurosci.* **2**, 568–573 (1999).
172. Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J. & Crommelinck, M. Expertise training with novel objects leads to left lateralized face-like electrophysiological responses. *Psychol. Sci.* **13**, 250–257 (2002).
173. Tanaka, J. W. & Curran, T. A neural basis for expert object recognition. *Psychol. Sci.* **12**, 43–47 (2001).
The first ERP study to provide evidence for expertise effects (with dog and bird images) on the N170 face-selective potential.
174. Rousselet, G. A., Mace, M. J. & Fabre-Thorpe, M. Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J. Vis.* **3**, 440–455 (2003).
Measures processing speed and the inversion effect for faces and non-face objects in the context of natural scenes. The evidence argues against the involvement of a special face module or a mental rotation mechanism, and supports the use of features of intermediate complexity.

Acknowledgements

Supported by grants from the NSF, NIMH, NEI and James S. McDonnell Foundation. The authors wish to thank M. J. Tarr and the members of the Perceptual Expertise Network (funded by the James S. McDonnell Foundation) for helpful discussions. We also thank M. Graf for detailed comments on an earlier version of this paper.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

FURTHER INFORMATION

Isabel Gauthier's laboratory:
http://www.psy.vanderbilt.edu/faculty/gauthier/

Thomas Palmeri's homepage:
http://www.psy.vanderbilt.edu/faculty/palmeri/home.html

Access to this interactive links box is free online.

Online links

Isabel Gauthier's laboratory: http://www.psy.vanderbilt.edu/faculty/gauthier/

Thomas Palmeri's homepage: http://www.psy.vanderbilt.edu/faculty/palmeri/home.html

At-a-glance

- Two relatively independent areas of visual cognition research examine important aspects of visual object understanding: Object Recognition and Perceptual Categorization. These areas have focused on different aspects of the same problems, with surprisingly little overlap. Nevertheless, they have ultimately arrived at complementary conclusions regarding the computational bases of visual object understanding.
- Traditionally, computational models in Object Recognition provide a detailed description of the format of object representations, whereas Perceptual Categorization models emphasize how representations are used to make decisions. Both image-based theories and exemplar-based theories have articulated how the same representations can be used to recognize, identify and categorize objects.
- Although intuition suggests that object recognition is effortless regardless of changes in viewpoint, and that knowledge about object categories is abstract, there is much evidence to the contrary. Just as recognizing an object is influenced by particular stored views, categorizing an object is influenced by particular stored exemplars. Image-based and exemplar-based models are supported by behavioural, neurophysiological and functional imaging results. There is also some renewed support for abstraction, and new hybrid models attempt to integrate structural descriptions with image-based representations and to integrate abstract category representations with exemplar-based representations.
- Objects can be categorized at several levels of abstraction (for example, animal, mammal, cat, Abyssinian, Max). Some argue that basic-level categorization is the fundamental goal of vision, with identification relying on features other than object shape, whereas early tests of image-based theories emphasized discrimination at the subordinate level. Recently, image-based theorists have argued that categorization at all levels can be accomplished using image-based representations. Early work in Perceptual Categorization suggested that identification and categorization used distinct representations and processes, but recent evidence indicates that a common representational substrate can be used adaptively according to task demands.
- Researchers in Object Recognition have traditionally discussed modularity of content: are there specific modules devoted to particular kinds of objects? The Perceptual Categorization literature focused on debates regarding the modularity of memory systems: are there specific modules devoted to particular tasks, irrespective of object category? In both fields, claims of modularity have been disputed, relying primarily on demonstrations that non-modular models can account for dissociations.
- Traditionally, visual perception was thought to create the representational input to a conceptual system that identified or categorized objects in a linear fashion. Recently, more 'interactive' solutions have been proposed. The evidence indicates that there is an interaction between perception and conceptual knowledge, and that category learning can influence perceptual representations.
- A new dynamic approach emphasizes the role of learning in most questions of interest in visual object understanding. Novices can demonstrate visual object understanding in qualitatively different ways than experts: for instance, people might initially categorize using rules but with experience start to retrieve exemplars from memory. Experience with certain categories leads to specialization in the visual system: for example, experts can process non-face objects such as cars, dogs, birds and novel objects in a manner similar to faces, using the same brain areas and with neural responses with the same latency.
- Despite their historic differences, current theories of Object Recognition and Perceptual Categorization have begun to consider complementary problems and have converged on similar solutions. Ultimately, a complete understanding of visual object understanding will demand an integration of the best theoretical constructs from Object Recognition and Perceptual Categorization.

Biographies

Thomas Palmeri received his B.S. in cognitive science from Carnegie Mellon University, and in 1999 he receved his Ph.D. in cognitive psychology from Indiana University. He is currently an associate professor in the Department of Psychology at Vanderbilt University. His research combines human behavioural research, computational modelling and cognitive neuroscience techniques to study perceptual categorization, memory and the development of expertise.

Isabel Gauthier received her B.S. in psychology from Université du Québec à Montréal, Canada, and her Ph.D. in cognitive psychology from Yale University in 1998. She is currently an assistant professor in the Department of Psychology and the Vanderbilt Vision Research Center at Vanderbilt University. Her work uses a cognitive neuroscience approach to the study of object, face and letter recognition, with an emphasis on the acquisition of perceptual expertise.