VISUAL OBJECTS IN CONTEXT

Moshe Bar

We see the world in scenes, where visual objects occur in rich surroundings, often embedded in a typical context with other related objects. How does the human brain analyse and use these common associations? This article reviews the knowledge that is available, proposes specific mechanisms for the contextual facilitation of object recognition, and highlights important open questions. Although much has already been revealed about the cognitive and cortical mechanisms that subserve recognition of individual objects, surprisingly little is known about the neural underpinnings of contextual analysis and scene perception. Building on previous findings, we now have the means to address the question of how the brain integrates individual elements to construct the visual experience.

Think of a giraffe, a basketball or a microscope. It is hard to imagine seeing any of them without a background and other objects. Our experience with the visual world dictates our predictions about what other objects to expect in a scene, and their spatial configuration. For example, seeing a steering wheel inside a car sets expectations about where the radio, ashtray and mirrors might be. These predictable properties of our environment can facilitate perception, and in particular object recognition (BOX 1). Recognizing someone's hand, for instance, significantly limits the possible interpretations of the object on that person's wrist to either a watch or a bracelet; it is not likely to be a chair or an elephant. This *a priori* knowledge allows the visual system to sensitize the corresponding visual representations (of a watch and a bracelet) so that it is easier to recognize the surrounding objects when we attend to them. In fact, these context-driven predictions can allow us to choose not to attend to this object at all if none of the possible identities 'suggested' by the context are of immediate interest.

Representing and processing objects in groups that tend to be found together might explain why recognition of an object that is highly associated with a certain context facilitates the recognition of other objects that share the same context[1,2]. Is this clustering reflected in the cortical analysis of contextual associations? How does contextual knowledge facilitate recognition of individual objects in a scene? What cortical areas are involved and how does information flow in the brain when contextual representations are activated? The primary goal of this article is to review the current knowledge about this field, and to propose what is further needed to answer these questions. Cognitive studies from the past 30 years are reviewed with recent physiological and neuroimaging data, and a theoretical proposal regarding how contextual knowledge facilitates object recognition is described.

There are many studies on the broad subject of context, considerably more than can be covered here. I will concentrate on the visual context of objects; the 'glue' that binds objects in coherent scenes. Within the underlying definition, each context (for example, an airport or a zoo) is a prototype that has infinite possible exemplars (specific scenes). In these prototypical contexts, certain elements are present with certain likelihoods, and the spatial relations among these elements adhere to typical configurations. Visual objects are contextually related if they tend to co-occur in our environment, and a scene is contextually coherent if it contains items that tend to appear together in similar configurations.

Understanding how contextual associations and object recognition are accomplished is essential for any complete theory of the brain. Studies of the organization of cortical representations have focused on groups of objects with the same basic-level name (such as faces, chairs or flowers)[3–6], generally ignoring the effects of context and the typical co-appearance of related objects. This necessary research has allowed us to progress to

Martinos Center at Massachusetts General Hospital, Harvard Medical School, 149 Thirteenth Street, Charlestown, Massachusetts 02129, USA.
e-mail: bar@nmr.mgh.harvard.edu

doi:10.1038/nrn1476

Box 1 | The powerful effects of context

The idea promoted throughout this review is that context-based predictions make object recognition more efficient. Furthermore, in cases where recognition cannot be accomplished quickly based only on the physical attributes of the target, contextual information can provide more relevant input for the recognition of that object than can its intrinsic properties (see REFS 131,151 for computational demonstrations). The hairdryer in the left panel and the drill in the right panel are identical objects: contextual information uniquely resolves ambiguity in each case.

The benefits of biasing recognition processes to cohere with an activated context, however, are accompanied by occasional inaccuracies. For example, you can probably read: IVAUTRE REIVEWS IVEOURCSEICNE as Nature Reviews Neuroscience reasonably quickly, without wasting time on details, and therefore without noticing at least seven syntactic errors. (A counter-example where context is not strong enough to make the observer ignore details is: NaTuRe ReViEwS NeUrOsCiEnCe, where attention is drawn to the irregularity, and the observer therefore misses the advantage from context.) Similarly, objects that follow the presentation of a contextual scene are misrecognized if they look like an object that belongs in the context[17]. False memory[26], boundary extension[32] and change blindness[152,153] are additional demonstrations of how contextually driven expectations can 'taint' subjective perception. Such effects might be considered as manifestations of the efficiency and flexibility of our perceptual and semantic mechanisms, which possibly, but rarely, fail. Similar contextual biases have been demonstrated in studies with linguistic stimuli (for example, see REFS 154–156).

considering objects within scenes and context. After all, our visual environment consists of contextually bound scenes, and research in this direction is ecologically most valid.

Representation of visual context

Objects can be related in various dimensions. For example, a hairdryer and a hairbrush are contextually related; two different hairdryers are different exemplars of the same BASIC-LEVEL CONCEPT; and a hairdryer and a drill look physically similar and so are perceptually related. Our brains can distinguish even subtle differences along these dimensions so that, for instance, physically similar objects are still labelled as different objects (for example, the cell phone and the calculator in FIG. 1), and two visually different objects can still be given the same basic-level name (for example, the two phones in FIG. 1). These relationships are usually explored using PRIMING, where improvement in performance after exposure to a stimulus can reveal the characteristics of underlying representations and their relations. What representation framework would make these intricate distinctions possible? Each object does not need to be represented in rich detail in multiple cortical regions — instead, different regions could represent features of objects that are relevant for the dimensions along which objects are grouped in these regions. For example, objects might be grouped by physical appearance in the occipital visual cortex[7–9], by basic-level categories in the anterior temporal cortex[4,10,11], by contextual relations in the parahippocampal cortex (PHC)[12], and by semantic relations in the prefrontal cortex (PFC)[13]. In addition, the grouping of objects might be represented by stored relationships. In the framework promoted here, different brain regions represent different possible grouping relations, and there is one centralized, detailed object representation component that serves all of these relations 'on demand'.

How are contextual representations of associated objects stored so that cortical processing can take advantage of predictable aspects of our environment? A recurring proposal is that prototypical contexts might be represented in structures that integrate information about the identity of the objects that are most likely to appear in a specific scene with information about their relationships. These contextual structures are referred to here as 'context frames'[2] but have also been described as schemata[14–18], scripts[19] and frames[20–22]. In general, these structures can be viewed as sets of expectations that can facilitate perception.

We know very little about how the brain arranges and retains such contextually associated information, although cognitive studies have provided important insights. For example, Biederman defined five types of relations that characterize a scene[1]: support (most objects are physically supported rather than float), interposition (for example, occlusion), probability (the likelihood that certain objects will be present in a scene), position (the typical positions of some objects in some scenes) and size (the familiar relative size of objects) (see also REF. 23). Objects that violate these relations in a scene are generally processed more slowly and with more errors[1]. These findings not only indicate what information is represented about scenes, but also that the semantic context of a scene might be extracted early enough to affect our perception of individual objects in it, possibly in a 'top–down' manner.

Context frames are assumed to be derived from exposure to real-world scenes. The extent to which the information in these frames is abstract, perceptually concrete or exists on multiple levels of abstraction[21–23] is unclear. Some evidence for abstract representation of scenes comes from the phenomenon of boundary extension[24,25] — a type of memory distortion in which observers report having seen not only information that was physically present in a picture, but also information that they have extrapolated outside the scene's boundaries. Similarly, in visual false memory experiments, participants report that they 'remember' having seen, in a previously presented picture, objects that are contextually related to that scene but that were not in the picture[26]. Such memory distortions might be byproducts of an efficient mechanism for extracting and encoding the gist of a scene.

Context frames can be viewed as prototypical representations of unique contexts (for example, a library), which guide the formation of specific instantiations in episodic scenes (for example, our library). It might be possible to generalize knowledge stored in context frames

BASIC-LEVEL CONCEPTS
The level of abstraction that carries the most information, and at which objects are typically named most readily. For example, subjects would recognize an Australian Shepherd as a dog (that is, basic-level) more easily than as an animal (that is, superordinate-level) or as an Australian Shepherd (that is, subordinate-level).

PRIMING
An experience-based facilitation in perceiving a physical stimulus. In a typical object priming experiment, subjects are presented with stimuli (the primes) and their performance in object naming is recorded. Subsequently, subjects are presented with either the same stimuli or stimuli that have some defined relationship to the primes. Any stimulus-specific difference in performance is taken as a measure of priming.

Figure 1 | **Some of the intricate object relations that are accommodated in the brain.** Objects that look very similar can be represented and recognized as different objects, whereas objects that look very different can be recognized as the same basic-level objects.

to instances of scenes where relations are novel but plausible[27], although information about relational plausibility (a person holding a dog is plausible, whereas a dog holding a person is not) might be represented independently as general world knowledge outside specific context frames. A central prediction that stems from this proposed co-representation of contextually related objects is that processing of 'typical' items and relations will be faster than processing of novel items and relations, and this has been supported by many studies. Context frames provide sets of expectations that can guide perception and action, and they can influence our exploration of a scene using eye movements and attention. Context frames can also modulate memory encoding and retrieval (memory can be improved when the encoding context is reinstated at retrieval[28]). As will be explained later, context frames can be activated by coarse, global scene information. It is proposed (and has been shown computationally[29]) that it is possible to construct a coarse representation of a scene that bypasses object identities, where the scene is represented as a single entity. This rudimentary information can provide a shortcut for automatic activation of high-level semantic information by relatively low-level perceptual information.

In summary, typical arrangements in our environment are represented in context frames, which provide expectations that facilitate the perception of other scenes that can be represented by the same context. Objects and relations that are sufficiently characteristic of the context are extracted and recognized readily, on the basis of global information and expectation-based shortcuts provided by defaults in the frame. The recognition of atypical objects and relations requires further scrutiny mediated by fine detail and elaborated analysis of local features. During recognition, an object can activate a context frame (or a set of frames), and a frame can activate an object (or a set of objects)[2]. Our understanding of these representations is limited, and the concept of context frames is helpful in guiding our search. Identifying the exact structure of these representations will subsequently help us to understand how context frames are activated to facilitate our perception of the visual world.

Context and object recognition

We seem to be able to take advantage of visual regularities in our environment, as contextual knowledge facilitates perception and cognition in many domains. Context-based facilitation of visual perception has been reviewed previously[30–33], and only the most relevant studies are highlighted here.

A typical scene structure that follows physical and contextual semantic rules facilitates recognition[1], at least compared with situations in which these rules are violated. When subjects are presented with a scene of a familiar context, such as a kitchen, objects that are consistent with that context (such as a loaf of bread) are recognized more easily than objects that would not be expected in that context (for example, a drum)[17]. These findings support the idea that context facilitates object recognition by activating context frames.

Context also facilitates the recognition of related objects even if these objects are ambiguous when seen in isolation[2] (FIG. 2); an ambiguous object becomes recognizable if another object that shares the same context is placed in an appropriate spatial relation to it.

In recognition, the goal is to determine the identity of viewed objects, despite possible variations in appearance[34,35]. Expectations derived from the context frame about the identity of other objects, as well as their position, orientation, size and so on, could therefore greatly facilitate the recognition of related objects.

If contextual information facilitates object recognition, one might expect that it would be easier to recognize a fixated object in a contextually coherent scene than in isolation. Two studies that tested this prediction found that an individually presented object was actually recognized more easily than the same object when it was embedded in a coherent scene[36,37]. However, at least two possible confounds make it hard to isolate the contribution of context *per se*. First, segmenting an individual object from its background is a consuming process, and is likely to make recognition more difficult, even in contextually coherent scenes. Second, attentional distraction from the scene might have affected the execution of response in those studies[36,37], rather than the perception of the target. In addition, the results of one study that addressed this issue indirectly[38] indicated that a contextually consistent background facilitates object recognition compared with the effect of a meaningless background that was equated for visual appearance. This adds support to contextual facilitation of recognition.

At what level of processing might context facilitate the recognition of individual objects? One possibility is that context is extracted from the scene so rapidly that it can facilitate the perceptual analysis of individual objects and therefore directly promote their recognition[1,17,38,39]. A slight variation of this idea is that when a context frame is activated it might sensitize the representation of all the objects associated with it, so that when the input image

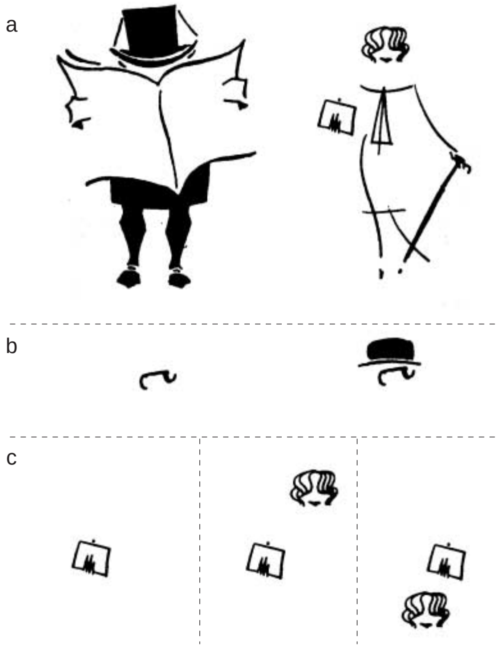


Figure 2 | **Resolution of ambiguous objects by context.** **a** | Two examples of the incomplete but recognizable figures from which the stimuli were derived (created by the artist Haro; reproduced, with permission, from REF. 156 © (1960) Elsevier Science). These figures are missing many features, but are nevertheless recognizable. **b** | The glasses on the left are ambiguous, but when a related object (hat) is placed next to them, in an appropriate spatial relation, they become recognizable. **c** | The object on the left is ambiguous in isolation. Placing a related object (head) next to it disambiguates its identity, as a purse, if placed in a typical, relative spatial position (middle), but does not affect its recognition when placed inappropriately (right).

has been sufficiently analysed to be compared with memory, this contextual activation facilitates convergence into the most likely interpretation[2,21,40,41]. The third possibility is that object recognition and contextual scene analysis are functionally separate and interact only at a later, semantic stage[30,42]. The first two alternatives have been criticized on the grounds of a possible response bias inherent to the experimental design of early studies, where cuing observers to a specific object identity and/or position in the scene could be selectively helpful for guessing when the scene is contextually consistent with the object[30]. In other words, if subjects identified the scene's context (for example, a street) they could infer that a mailbox was present at the cued location even without perceiving the mailbox. If the cued object was contextually incongruent (for example, a blender at the same position in the street), however, subjects could not respond correctly without first recognizing the object.

In a study addressing the bias effect, Henderson and Hollingworth reported results that support their view of functional separation[30] (but see REFS 37,44). Nevertheless, there is continuing evidence that context exerts its effect relatively early during object recognition. First, the response bias criticism focused on the biasing position of probable target objects. However, there are at least five relations that are present in coherent typical scenes[1]. So, whereas the criticism is legitimate with regard to position, it does not account for findings about the contribution of context when compared with violations of other relations (such as support and size).

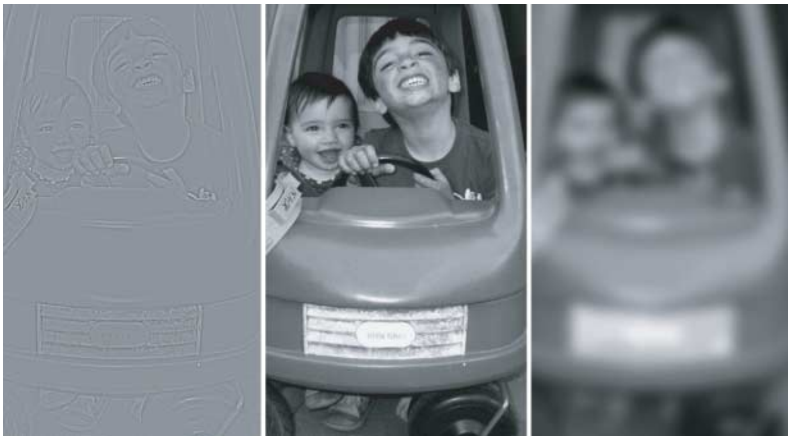
Second, in recent experiments where the possibility of response bias was specifically controlled[37,44], context contributed significantly to object recognition, although response bias probably contributed to part of the improvement that was previously attributed exclusively to context[44]. In addition, object recognition facilitates the recognition of a scene's background[37], which constitutes its context, indicating a bidirectional exchange between the two processes. Therefore, scene recognition does not seem to proceed in parallel, separated 'channels', but instead is a more interactive process that integrates contextual information to facilitate object recognition, and uses object identities to promote the understanding of a scene.

Another reason why opinions about the role of contextual information in object recognition have been mixed might be that object recognition is very efficient. As a result, a clear, prototypical, isolated object (usually on a computer screen in the laboratory) is recognized in less than 150 ms (REFS 45,46). In real life, however, clutter, occlusion, shading, viewing angles and other factors make recognition harder, and in these realistic situations recognition can benefit from other sources. Context, as well as familiarity, non-contextual expectations, top–down facilitation[40,47] and movement[48], might all facilitate object recognition. In other words, object (and scene) recognition vary in difficulty, and added sources might facilitate recognition at increasing levels of difficulty. When studying the possible effects of these auxiliary factors, the difficulty of the task must be manipulated. Studying these influences when recognition is atypically easy might give the impression that other sources do not contribute to the process.

Our discussion of contextual facilitation in object recognition emphasizes recognition of a fixated, target object. However, given the proposed structure of context frames and the sets of expectations they elicit, it would be predicted that the recognition of expected non-target objects would also be facilitated, perhaps even more than the recognition of the target object. Indeed, contextual associations promote the deployment of attention towards associated objects[49], thereby facilitating their recognition compared with non-associated objects. Similarly, contextual understanding helps us to determine, consciously or not, where to look next[30].

What is the benefit of the representation of common relations, and how do they influence the recognition of individual objects? From a computational standpoint, it is clear that representing such regularities allows an efficient generalization in new situations[50] as well as analysis shortcuts produced by the expectations that the context frames provide, possibly mediated by long-range cortical connections[51]. More generally, contextual facilitation might be mediated by lowering the response thresholds (increasing the sensitivity) of the cortical representations of anticipated objects.

Box 2 | Spatial frequencies and the information they convey



Different spatial frequencies convey different information about the appearance of a stimulus. High spatial frequencies represent abrupt spatial changes in the image (such as edges), and generally correspond to configural information and fine detail (left panel). Low spatial frequencies, on the other hand, represent global information about the shape (such as general orientation and proportions) (right panel). The centre panel shows the original image containing the entire spectrum. The initial perception of global scene information, carried by low spatial frequencies, might mediate the rapid extraction of gist information from scenes.

MAGNETOENCEPHALOGRAPHY (MEG). A non-invasive technology for functional brain mapping, which provides superior millisecond temporal resolution. It measures magnetic fields generated by electric currents from active neurons in the brain. By localizing the sources of these currents, MEG is used to reveal cortical function.

Rapid extraction of context

For contextual information to assist the recognition process, it has to be extracted rapidly and subsequently to generate guiding expectations. How quickly context is extracted from a scene or an object has been the subject of extensive research.

We can recognize visual scenes in a 'glance'[14]. A study in which pictures of objects were briefly presented[52] provided evidence that semantic meaning about context is extracted from the input at an early stage, possibly even before perceptual processing is complete (see also REF. 53). Although some reports indicate an absence of semantic priming in subliminal presentations (for example, REF. 54), the inconsistency might be due to differences in experimental conditions and varying degrees of subjects' non-awareness. Subjects can understand a visual scene with exposure durations of around 100 ms (REFS 14,55,56), and might be able to extract semantic information about context from presentations as brief as 80 ms (REF. 37). Another study indicates that contextual information is extracted before observers can saccade towards the portions of the picture that were rated as contributing most to the context of the scene, and possibly even before the recognition of individual objects[39]. Furthermore, observers process the most informative portions of an image earliest[57].

How is contextual meaning extracted so rapidly? I propose that this swift extraction is mediated by global cues that are conveyed by low spatial frequencies in the image[29,58,59], and that details conveyed by the high spatial frequencies are analysed later (BOX 2). The global shape information that is conveyed by the low spatial frequencies can activate 'scene schema'[1,17] or context frames, although frames can also be activated by individual objects[2,12].

Given that contextual information is extracted rapidly, presumably on the basis of coarse representations, does contextual processing require awareness? Context can be processed implicitly[60,61], and it can be learned incidentally (without explicit intentions)[31,62]. Furthermore, subjects can categorize visual objects in contextual scenes in the 'near absence' of attention[63]. Finally, contextual information can be automatically activated by a scene and can subsequently interfere with task performance[64]. Implicit access to semantic information about context does not need to be direct. Instead, contextual processing might use a shortcut whereby high-level semantic information is activated by coarse input, even before this input has been identified (see the proposed model below and also REF. 65 for a similar treatment of subliminal semantic activation in words).

Event-related potential (ERP) studies have provided some insight into the neural dynamics that underlie the rapid extraction of context. For example, ERPs can distinguish between visual categories for individual objects in as little as 75–80 ms (REF. 45). In other studies[66,67], ERP signals distinguished between new and old contexts around 100–200 ms after stimulus onset, depending on the task. Finally, preliminary data from a study that combined functional magnetic resonance imaging (fMRI) and MAGNETOENCEPHALOGRAPHY (MEG)[68] indicate that activity that is directly related to contextual processing develops first in the PHC, which has previously been implicated in contextual processing[12], and in the fusiform gyrus around 130 ms after stimulus onset. A second wave of differential activation develops there around 230 ms after stimulus onset. The functional significance of these two waves has yet to be determined, but we propose that the first is a manifestation of a quick and coarse activation of the scene's representation and the second reflects a richer representation, incorporating the full spatial bandwidth.

Cortical processing of context

In contrast to the number of behavioural studies that have addressed contextual processing, little has been revealed about the underlying neural mechanisms. Most of the related research has focused on associative processing, which can be considered as the building blocks of context.

Structures in the medial temporal lobe, including the hippocampus, PHC, and perirhinal and entorhinal cortices, are thought to be important in associative processing[69–72]. Unfortunately, there is not enough evidence to make a clear functional distinction between the subdivisions of the medial temporal lobe. For example, the hippocampus receives input from many sources, and there is some evidence that it emphasizes associative rather than single-item representations under some conditions[73] but not others[74]. More generally, there are active debates about which sub-region within the medial temporal cortex mediates associative versus non-associative representations, episodic versus semantic memory, spatial versus contextual analysis, and familiarity versus recognition judgements.

Human neuroimaging studies have started to address the cortical basis of context and scene processing. These studies revealed a region in the PHC that responds preferentially to topographical information and spatial landmarks[75–77] — the parahippocampal place area (PPA). The invariance of the representations and processes in the PPA to viewing position are currently being characterized[78] in an analogous manner to what has been characterized behaviourally[79,80]. This region might have an important role in large-scale integration[81], and there is an increasing assumption that it is a module for analysing scenes[82–84] (but see REF. 85). However, although the PHC seems to be part of a cortical contextual network[12], treating it as a special module for scene representation and analysis might be overly simplistic.

We see the world in scenes. This implies that the cortical mechanism that mediates scene perception integrates the output of many processes, which analyse different aspects of the input into one smooth and coherent scene. Consider individual objects. The various components of representing and recognizing individual objects (shape, identity and so on) have been attributed to a network[86] that includes the lateral occipital cortex[8] and the fusiform gyrus[87], each of which is larger than the PPA. Could this large network merely mediate pre-processing before the information is synthesized into a comprehensive visual scene in a relatively small region in the PHC? This is plausible, but because most attention has been allocated so far to the cortical mechanisms of object recognition, and considerably less to scene and visual contextual analysis, our view of the cortical mechanisms allocated for each faculty might be distorted. In any event, it would seem that the suggestion that 'the PPA is for scenes' is only part of the story. The framework for integrating cues into scenes might be more complex, possibly including regions that have previously been implicated in processing individual objects, and where the PPA might be responsible only for some aspects of scene processing.

An even larger-scale network is predicted to be involved in the representation and processing of contextual scenes. Consider that scenes are not always bound by semantic context. Scenes can be: coherent in their visual properties exclusively, without adhering to semantic context and to our knowledge of the physical world (a zebra reading a book on a cloud); physically but not contextually coherent (a zebra reading a book on the street); or coherent also in semantic context (a zebra grazing on tall grass in the wilderness). Evaluating these dimensions requires more information and incorporates an increasingly complicated set of constraints and experience-based rules. Therefore, to analyse real-world contextual scenes, the brain would rely not only on the circuitry that subserves visual scene perception, but also on pre-existing representations of common associations and typical relations.

With regard to contextual processing *per se*, the field of ERP measurements has been particularly active. The main catalyst has been the discovery and characterization of THE N400 SIGNAL[88]. Using the N400 phenomenon, Ganis and Kutas[89] studied the dynamics of contextual effects in scene perception, mainly by comparing the responses to conditions in which a target object and its background were congruent and incongruent. On the basis of the spatial distribution of the response across the scalp, they suggested that schemata (context frames) might be stored and activated in the anterior temporal lobes. The spatial resolution provided by ERP does not allow further localization of the source of this signal. However, intracranial recordings in humans[90,91] have shown that activity in regions of the medial temporal lobe, including the PHC, is modulated by contextual information in words. These findings are consistent with our proposal that the PHC stores information about contextual associations[12], which will be elaborated below.

It is not clear whether the same semantic system subserves the representation and processing of contextual information conveyed by words and pictures. The 'dual-code' view[92,93] posits multiple semantic systems, and the 'single-code' view[94,95] posits a unitary system. Experimentally distinguishing between these alternatives has proven difficult. Contextual information for words and pictures affects the ERP measurement similarly, but the picture-related N400 seems to be more frontal, and the word-related N400 to be more occipital[96–98]. These studies, and a related study using positron emission tomography (PET)[99], indicate that the contextual systems that mediate word and picture processing might be similar but not completely overlapping. It is possible that the semantic representations of pictures and words use a shared system, but that their processing uses different circuits[98]. That the semantic information conveyed by words and pictures is processed differently, beyond differences in perceptual analysis, is supported by other consistent differences. For example, words are read quicker than the corresponding pictures can be named, but pictures of objects are categorized faster than the corresponding object names[100,101]. Interestingly, when comparing the cortical processing of semantic context conveyed by written and spoken words, responses to both seem to be initiated by modality-specific circuitry, but then elaborated primarily in amodal regions[102].

These findings indicate only a partial overlap between the mechanisms that mediate semantic analysis conveyed by different formats, and they demonstrate a role for both the medial temporal lobe and the PFC in contextual analysis. The involvement of the PFC in contextual processing has been demonstrated in the past, for example, in mediating face–name associations, using fMRI[103]. Prefrontal contextual processing has further been reported in studies of the N400 effect that used intracranial depth recordings[104], and in studies that combined fMRI with MEG[105] and with ERP[106]. This frontal activity occurred along with medial temporal activity[107], possibly reflecting the interaction between these two regions[108].

How is the clustering of objects into typical contexts reflected in the cortical processing of contextual associations? To address this question, we compared the fMRI signal elicited during the recognition of visual objects that are highly associated with a certain context (for example, a bowling pin) with that elicited by objects

THE N400 SIGNAL
Originally described as a negative deflection in the event-related potential waveform occurring approximately 400 ms following the onset of contextually incongruent words in a sentence. It has consistently been linked to semantic processing. Although it is probably one of the best neural signatures of contextual processing, its exact functional significance has yet to be elucidated.

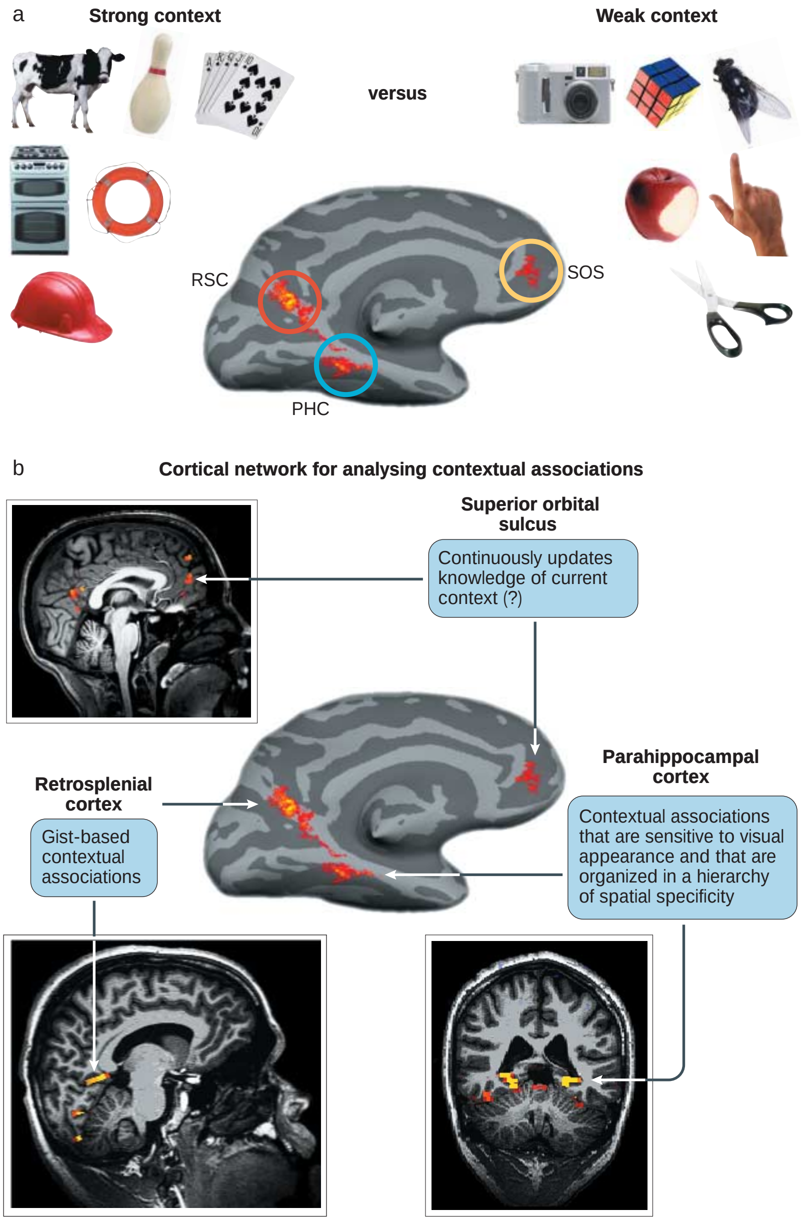


Figure 3 | **Cortical areas involved in processing context. a** | A functional magnetic resonance imaging (fMRI) statistical activation map representing the difference between perceiving objects that are strongly associated with a specific context and perceiving objects that are not associated with a unique context. This is a medial view of the left hemisphere, shown using a precise computer reconstruction where the sulci have been exposed by 'inflation'. The parahippocampal cortex (PHC) is circled in blue; the retrosplenial cortex (RSC) is circled in red; the superior orbital sulcus (SOS) is circled in yellow. Note that in all experimental conditions, subjects viewed similar looking colour photographs of meaningful, everyday common objects that were equally recognizable. Consequently, activation due to low-level processes was presumably subtracted out, and the differential activation map shown here represents only processes that are related to the level of contextual association. **b** | The cortical network for contextual associations among visual objects, suggested on the basis of existing evidence. Other types of context might involve additional regions (for example, hippocampus for navigation[125] and Broca's area for language-related context). Modified, with permission, from REF. 12 © (2003) Elsevier Science.

that are not associated with any unique context (for example, a camera)[12] (FIG. 3).

The first and largest focus of differential activity was in the posterior PHC. This site encompasses the PPA, which has been reported to respond selectively to houses and other environmental landmarks[75–77]. A second focus of activation was found in the retrosplenial cortex (RSC), which has also been implicated in the analysis of spatial information[109–111]. In addition, the processes in the PHC site seemed to be sensitive to visual appearance, whereas the RSC was more insensitive to specific stimulus appearance. Consequently, we proposed that both the PHC and the RSC represent familiar associations, but with a different level of abstraction. Finally, a third focus, revealed only in our follow-up event-related fMRI and MEG experiments[68], was found in the superior orbital sulcus (SOS). We propose that this region integrates information from several sources to create a continuously updated representation of the current context, and that it uses this knowledge for top–down facilitation of scene and object recognition.

The association of the PHC and RSC with the perception of places leads to two possible interpretations of these results. First, perceiving the contextual objects (for example, a roulette wheel) might have indirectly activated the corresponding places (a casino) and, consequently, elicited an fMRI signal in regions that have been associated with the perception of places. Alternatively, the PHC and RSC might mediate the representation and processing of familiar contextual associations in general, rather than places *per se*. In many cases, sets of associations correspond to landmarks, which generally associate objects with places[41], but the PHC and RSC processes might also involve non-spatial sets of associations. To distinguish between these alternatives, we compared the fMRI activation elicited by spatial, place-specific contexts (such as 'street') with the signal elicited by non-spatial contexts (such as 'romance').

Both spatial and non-spatial contexts elicited significant differential activation in the PHC and the RSC, supporting our hypothesis that the PHC and RSC sites mediate the general analysis of contextual associations, rather than of place-related associations exclusively. Notably, the spatial and non-spatial contexts activated slightly different, non-overlapping subregions of the PHC: the spatial contexts elicited a stronger signal in a relatively posterior part of the PHC focus, possibly encompassing the PPA, whereas the signal for the non-spatial contexts peaked more anteriorly. Our generalization of the role of the PHC to non-spatial as well as spatial associations is supported by recent studies[112–116] (but see REF. 117), and by reports that only 8% of the input to the PHC consists of visuospatial information[118,119]. Its multimodal inputs might further indicate that the PHC binds together more than just visual components.

Interestingly, the PHC[72,120] and the RSC[121] have also been associated with aspects of episodic memory. The proposal that emerged from our findings is that these two regions process familiar associations between individual constituents, which provide a basis both for episodic memories and for navigation. Consequently, the proposal that these regions mediate contextual associations provides a framework that bridges these seemingly unrelated functions.

This proposal might also shed light on disputes concerning the function of other medial temporal regions. A related account is proposed every few years about various medial temporal structures, predominantly the hippocampus[71,122–124]. The main rationale for having attributed a role in navigation to hippocampal cells[125,126], as well as to parahippocampal regions in humans[75–77], is the use of place-related paradigms. But these tasks can generally be viewed, instead, as mediated by contextual associations. We do not argue that these regions are not involved in spatial analysis, but the terminology might have to be modified to accommodate the non-spatial responses of these medial temporal regions.

We have consistently found activity related to visual context in the PHC[12]. In this study, we did not obtain consistent context-related activity in the hippocampus and the perirhinal cortex. This might be because the perirhinal cortex represents relatively simple, paired associations[118,127], which serve as building blocks for the more global contextual associations that are represented in the PHC.

It is unclear what exactly is being represented in the PHC regarding associations. I propose that the PHC serves as a switchboard-like 'multiplexer' of associations between items that are represented in detail elsewhere, allowing flexible use of a representation system. This proposal is reminiscent of Kosslyn's definition of associative memory and tokens[41]. In the framework proposed here, each object's representation 'stands alone' in the inferior temporal cortex (ITC) and can be connected to one of its possible association sets depending on guidance signals from the PHC.

Visual objects are represented in detail in the ITC, and there is no reason to believe that detailed representations are replicated elsewhere. Instead, a multiplexing system that maps a particular situation to its corresponding set of connective associations would allow the brain to use one multi-purpose representation system for visual objects. Furthermore, if associations between objects were implemented by direct connections that are co-activated automatically whenever one of the objects is presented, then seeing a television set would immediately make us think about a living room, an appliance store, news, talk shows and so on, regardless of the specific context. In a typical situation, however, we need to activate only the relevant subset of existing associations, and therefore it does not seem efficient to connect all of the associations of an object to its representation. In this framework, the PHC represents information about recurring regularities in our environment so that, for example, a television set in the living room would co-activate sofa and coffee table, but a television set in an electronic store would activate other appliances, a cashier and so on (FIG. 4). Associations are not always symmetrically bi-directional; scissors might activate the representation of paper, but paper might not activate the representation of scissors. This reinforces the need for indirect associative connections rather than simple 'lines'. Most sets of associations can be considered as context frames, but not all sets are about context *per se* (for example, association of different exemplars of the same object, conjunction of properties such as smell, taste and sound). This flexible representation system is not limited to visual objects and can be generalized to other modalities.

Support for this proposal comes from neurophysiological studies of memory and perception in the temporal lobe in monkeys. The response to an object's perceptual features occurs in area TE (positioned laterally to the medial temporal lobe) before it occurs in the medial temporal lobe, but information about associations between such objects elicits an earlier response in the medial temporal lobe than in area TE[128]. Moreover, lesioning regions in the medial temporal lobe eliminates the representation of paired visual associations in the intact region TE[118,129]. These findings support the proposal that knowledge about associations is stored in the medial temporal lobe, but perceptual representations are stored in the visual inferior temporal cortex. Such paired association responses appear in the perirhinal cortex rapidly, and so are believed to be mediated by a bottom–up mechanism that connects two TE neurons representing two individual items to a single perirhinal neuron representing the association between them[127]. Additional support for indirect associative activation comes from recent fMRI studies in humans[84,130]. For example, a face-sensitive area in the fusiform gyrus was shown also to be active for images of a person with a blurred face, presumably owing to contextual expectations[130].

The multiplexer can be considered as a type of distributed representation[50], in which the components of a scene do not need to be spatially adjacent in the cortex, but rather are activated in their centralized representation system as dictated by the stored set of associations in the PHC. The associations might be activated using BAYESIAN inference methods[131,132], and are reinforced through mechanisms such as HEBBIAN-BASED LEARNING[133] and long-term potentiation (LTP)[134].

The PHC receives polysensory input through the RSC, and the cingulate gyrus, visuospatial information from the posterior parietal cortex in the dorsal visual stream, auditory input from the superior temporal gyrus, somatosensory information through the insula, and visual shape input through areas TE/TEO and the perirhinal cortex[70]. Therefore, it seems that the PHC receives the input that is required for mediating global contextual associations. Furthermore, the PHC shows an ERP N400 effect for semantically incongruent stimuli, which might indicate that it signals the violation of a familiar set of associations and alerts the observer to an aspect of the environment that requires attention.

In summary, contextual processing involves regions in the medial temporal lobe, the PFC and the RSC. It is necessary to understand the functional division of labour between these sites. Furthermore, it is important to differentiate which aspects of the observed dynamics can be attributed solely to contextual processing, and which are a manifestation of direct contextual facilitation of object recognition.

BAYESIAN METHODS
Use *a priori* probability distributions derived from experience to infer optimal expectations. They are based on Bayes' theorem, which can be seen as a rule for taking into account history information to produce a number representing the probability that a certain hypothesis is true.

HEBBIAN LEARNING
Builds on Hebb's learning rule that the connections between two neurons will strengthen if the neurons fire simultaneously. The original Hebbian rule has serious limitations, but it is used as the basis for more powerful learning rules. From a neurophysiological perspective, Hebbian learning can be described as a mechanism that increases synaptic efficacy as a function of synchrony between pre- and postsynaptic activity.

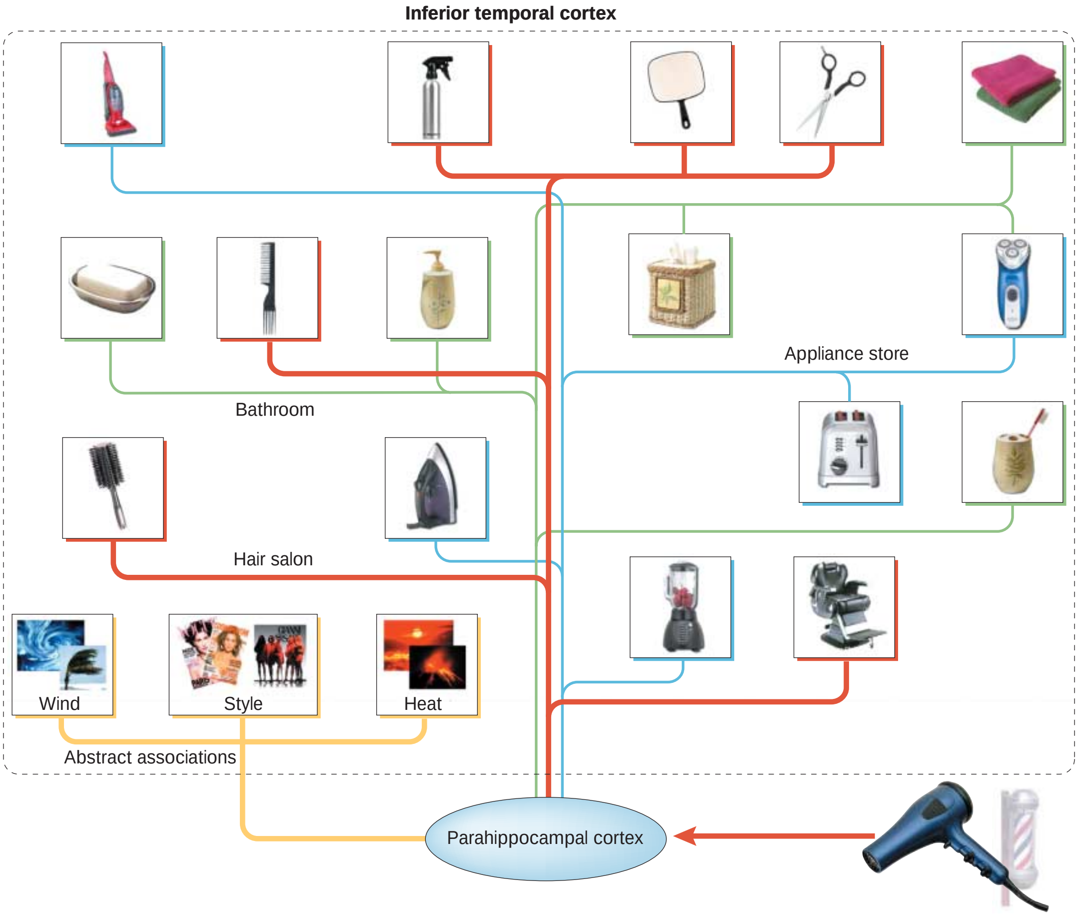


Figure 4 | **The proposed model of how information in the parahippocampal cortex might activate visual representations in the inferior temporal cortex in a flexible manner.** A particular object (for example, a hairdryer) can be associated with several context frames (for example, a hair salon, a bathroom, an appliance store), as well as with abstract concepts (heat, style and wind). The specific context (for example, hair salon), which is established by global scene information and/or by contextual cues that are provided by the recognition of other objects in the scene (for example, a barber pole), dictates which of the experience-based association sets should be activated to facilitate recognition in the particular situation. Coloured lines link different association sets.

A model for contextual facilitation

A context frame represents prototypical information about a unique context, and contains information about the identities and typical spatial arrangements of objects that tend to co-appear within that context. This information can be considered as a set of expectations about the environment, which, once activated, are tested against incoming information. Unfilled 'slots' in these frames[20] are filled by default values that are based on stereotypical expectations. The rapid activation of these frames, as discussed above, can be triggered by either global scene information (low spatial frequencies; BOX 2) or by key objects in the scene[2,12]. In this section, I will describe a specific model for how contextual activation facilitates object recognition.

At the heart of this model is the observation that a coarse, low-spatial-frequency representation of an input image is usually sufficient for rapid object recognition. Specifically, the low-spatial-frequency image of a scene is typically sufficient for deriving a reliable guess about the context frame that needs to be activated, and a low-spatial-frequency image of a single target object is sufficient for limiting its possible interpretations. The intersection of these two sources of information would result in a unique identification.

The model is illustrated in FIG. 5. A blurred, low-spatial-frequency representation is projected early and rapidly from the visual cortex to the PFC and PHC. In the PHC, this image activates an experience-based guess about the context frame that needs to be activated. This contextual information is projected to the ITC, where a set of associations that corresponds to the relevant context is activated (FIG. 4). In parallel, the same blurred image, but with the target object selected by foveal vision and attention, activates information in the PFC that subsequently sensitizes the most likely candidate interpretations of that individual object[40]. In the ITC, the intersection between the representations of the objects associated with the specific context and the candidate interpretations of the target object results in the reliable selection of a single identity. This representation (for example, a car) is then refined and further instantiated (for example, as an old convertible Mustang), with specific detail gradually arriving in higher spatial-frequency information.

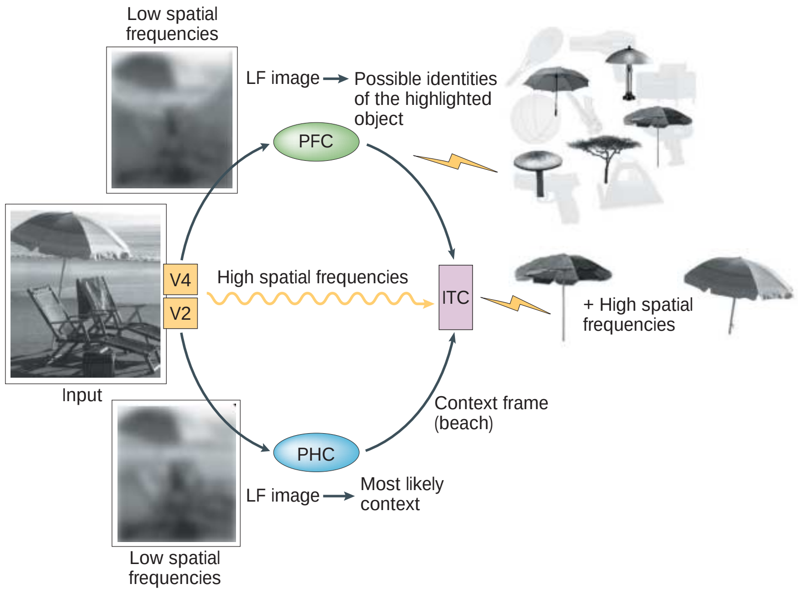


Figure 5 | **The proposed model for the contextual facilitation of object recognition.** The early intersection of the association set in the context frame with the candidate interpretations of the individual target object results in rapid recognition of that object as a generic beach umbrella. The exact representation of the specific exemplar is subsequently derived from the later arrival of higher spatial frequencies. Several of the specific cortical mechanisms have yet to be characterized, and the assignment of functions to specific cortical regions in the proposed model might be refined as more data become available. In particular, current reports make it plausible that other medial temporal structures, in addition to the parahippocampal cortex (PHC), might contribute to the analysis of various aspects of associations. For simplicity, only the relevant connections and flow directions are illustrated here. ITC, inferior temporal cortex; LF, low frequency; PFC, prefrontal cortex; V2 and V4, early visual areas. 'Lightening strike' symbol represents activation of representations.

This model focuses on contextual facilitation of visual object recognition, and two particular exclusions should be noted. First, the model is not intended to explain contextual influences in letter, word or sentence recognition. Recognition in the language domain, for which many accounts have been proposed[135], presumably benefits from context through mechanisms other than those that mediate object recognition, in spite of several similarities. Second, contextual information involving human faces, which can be considered as uniquely important visual objects, might also be analysed separately from other objects. On the one hand, recognizing a face outside of its typical context (your dentist on the beach) will be harder than when it benefits from a consistent context. In the present model, such conflicts will be expressed by the wrong expectations being elicited by the context frame. On the other hand, it is not clear how helpful low spatial frequencies are with regard to limiting the possible identities of faces, and it is therefore not clear how faces would be incorporated into this proposal. Face identification is in many respects analogous to subordinate-level individuation, and even for non-face objects, such knowledge is available only after the arrival of details in higher spatial frequencies.

In support of this model, anatomical studies have shown both ascending and descending connections between visual areas[136,137], and these connections might mediate both bottom–up and top–down cortical processing[41,138–140]. I previously proposed a detailed mechanism for how such top–down processing would be triggered to facilitate object recognition[40]. In the proposed framework, low spatial frequencies in the image are extracted quickly and projected from early visual areas to the PFC. This projection is considerably faster than the thorough bottom–up analysis, and therefore is predicted to use fast anatomical connections — possibly the magnocellular pathway, which propagates low-spatial-frequency information early and rapidly[141,142]. The global information that is conveyed by low spatial frequencies is typically sufficient to activate a small set of probable candidate interpretations of the input ('initial guesses'). When the input representation is associated with one of the candidates, recognition is accomplished and the other initial guesses are no longer active. Preliminary and unpublished data from my laboratory support this model by showing that: differential activity that is diagnostic of recognition success develops in the orbital PFC significantly earlier than it does in the temporal cortex, as shown by MEG physiological recordings (M.B. *et al*, submitted); the fMRI signal in the orbital PFC is significantly stronger for low spatial frequencies than for high spatial frequencies[143]; and orbital PFC activity increases as a direct function of the number of alternative interpretations that can be produced about the object image on the basis of its low spatial frequencies[144].

This model is expanded in this article from the recognition of individual, isolated objects to entire scenes. The expansion makes the proposal more ecologically valid and accounts for the processes that are involved in the extraction of context and its use for recognizing objects in scenes. It has been proposed[40] that the orbital PFC might contain a 'look-up table' that maps the low-spatial-frequency appearances of objects to their most probable interpretations. Along the same lines, the PHC could map low-spatial-frequency representations to the most likely context frames, which contain information about possible objects (FIG. 4) and their spatial arrangement. Unlike the PFC, the PHC has not been shown explicitly to receive direct magnocellular connections from early visual cortex. However, it does receive massive visual input (in addition to input from other modalities) and, on the basis of its anatomical architecture and output connections, it is sometimes considered part of the visual system[145].

In addition to psychophysical and computational demonstrations[29,39,59], single-unit recordings in monkeys also indicate that low spatial frequencies are extracted from scenes earlier than high spatial frequencies, and that this global information can be sufficient for scene categorization[146]. Activity in the ITC is initially broadly tuned and represents only the global features (the low spatial frequencies) of a stimulus[147,148]. Later, 51 ms after the onset of the global response[148], neurons in the ITC also represent the fine attributes of the image, presumably propagated by the high spatial frequencies. So, the ITC responds to low-spatial-frequency information before it receives high-spatial-frequency information.

According to this model, early in scene analysis, the ITC has access to a sensitized set of associated objects from the PHC (where the level of sensitization of each object's representation depends on the strength of its association with the specific context), and an 'initial guess' from the PFC containing the most likely interpretations of the target object. To select the correct identity from these initial guesses, an intersection operation, which can be considered as the neuronal equivalent of an 'AND' function followed by selective inhibition and excitation, is performed. For example, if the PFC 'suggests' that the object might be a television set, a microwave or a fireplace, and the PHC suggests that the context is a kitchen, the microwave alternative is selected, and all other candidates can be suppressed. Note, however, that if the context in this example were a living room, the output of the intersection would still remain ambiguous, because a television and a fireplace are equally likely. In such cases, final recognition is accomplished only after the arrival of more detail, which is conveyed by the higher spatial frequencies.

Shifting gaze and/or attention would make the PFC shift its focus to different objects of interest, which in turn would result in a different set of initial guesses being transmitted to the ITC. However, assuming that the scene's context has not changed between shifts, there will be little or no change in the context frame projected from the PHC. Furthermore, the top–down projection from the PFC might not be as crucial for the development of expectations in the ITC as the projection that stemmed from focusing on the first object. In that sense, the identification of the context and of the first object bootstraps the recognition, at some level, of the complete scene. That observers fixate first on the most informative aspects of a scene[30] indicates that our system might operate to maximize the extraction of contextual information from the first fixation.

Although context frames have been suggested to represent both object identities and their spatial relations, this model emphasizes identities more than relations. This is partly because not much is known about the cortical analysis of spatial arrangements (or about analysis of the other relations that define a scene), and partly because the availability of contextual and identity information is typically sufficient for object recognition. Indeed, when scenes are presented briefly, information about objects' identities seems to be acquired earlier than information about their locations[149]. The representation of typical spatial relations can nevertheless be a powerful principle for guiding expectations, attention and eye movements, and so is certain to have a central role.

What is the role of the magnocellular projections from early visual cortex to the ITC[150]? These projections might provide the ITC with a spatial template in which to 'anchor' the interpretations derived from the initial guesses. Note that the proposed projections from the PFC and PHC to the ITC lack information about the spatial arrangement of the scene elements, which can be supplemented by the direct magnocellular projection of the blurred scene from early visual cortex to the ITC.

Importantly, all aspects of the model can be addressed experimentally. Furthermore, it can account for many relevant findings. For example, individual objects in isolation seem to be recognized more easily than if they are embedded in contextually congruent scenes[36,37]. By definition, isolated objects occur without background. Consequently, their low-frequency appearance will be less ambiguous than when they are surrounded by other objects, resulting in a more precise initial guess from the PFC to the ITC, and therefore more efficient recognition compared with individual objects in a scene background. Second, the identity of an individual object can help subjects to recognize the background scene[37]. In the proposed model, the initial recognition of a semantically informative object in the scene, mediated by the pathway from the PFC to the ITC, can result in a projection from the ITC to the PHC that will elicit the context frame that corresponds to the recognized object. This pathway is expected to be more helpful when the recognition of the background is not straightforward, for example, in atypical instances of familiar contexts. Third, the identity of an otherwise ambiguous object can be disambiguated by the presence of another object that is contextually related to the target object[2]. In such cases, the recognition of one object would activate a context frame that would improve the imprecise initial guess that was produced by the ambiguous object. In addition, when both objects are ambiguous, their visual and spatial properties can activate the correct context frame, which then activates a set of expectations that facilitates their relatively late but successful recognition. This mechanism might also explain why a particular object can be interpreted differently in different contexts (BOX 1), and how an object that does not belong in the context would elicit an N400 effect in the PHC after all the alternatives from the PFC have been compared with those activated by the context frame and no intersection has been found.

Conclusions

Our brain takes advantage of common associations among objects in the environment to facilitate visual perception and cognition. The cortical network that mediates the processing of such contextual associations and their interface with object recognition involves regions in the PHC, PFC and RSC. Important open questions include: how are context frames represented in the cortex, and what triggers their activation? How is contextual information translated into expectations? How does context facilitate object recognition? How is gist information from a contextual scene represented in the brain? And how do motivation and attention modulate contextual processing? I have proposed a testable model for the rapid use of contextual associations in recognition, in which an early projection of coarse information can activate expectations about context and identity that, when combined, result in successful object recognition. With progress in spatiotemporal neuroimaging and theoretical formulation, we are certainly on the verge of exciting discoveries about the behavioural and cortical mechanisms that combine visual elements into rich, coherent scenes.

1. Biederman, I., Mezzanotte, R. J. & Rabinowitz, J. C. Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* **14**, 143–177 (1982).
A seminal study that characterizes the rules that govern a scene's structure and their influence on perception.
2. Bar, M. & Ullman, S. Spatial context in recognition. *Perception* **25**, 343–352 (1996).
3. 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).
4. Puce, A., Allison, T., Asgari, M., Gore, J. C. & McCarthy, G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* **16**, 5205–5215 (1996).
5. Martin, A. *et al.* Neural correlates of category-specific knowledge. *Nature* **379**, 649–652 (1996).
6. Ishai, A. *et al.* Distributed representation of objects in the human ventral visual pathway. *Proc. Natl Acad. Sci. USA* **96**, 9379–9384 (1999).
7. Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. The lateral occipital complex and its role in object recognition. *Vision Res.* **41**, 1409–1422 (2001).
8. Malach, R., Levy, I. & Hasson, U. The topography of high-order human object areas. *Trends Cogn. Sci.* **6**, 176–184 (2002).
9. Tanaka, K. Neuronal mechanisms of object recognition. *Science* **262**, 685–688 (1993).
10. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
11. Downing, P. E. *et al.* A cortical area selective for visual processing of the human body. *Science* **293**, 2470–2473 (2001).
12. Bar, M. & Aminoff, E. Cortical analysis of visual context. *Neuron* **38**, 347–358 (2003).
Defines the cortical regions that are directly involved in the contextual analysis of visual objects.
13. Gabrieli, J. D., Poldrack, R. A. & Desmond, J. E. The role of left prefrontal cortex in language and memory. *Proc. Natl Acad. Sci. USA* **95**, 906–913 (1998).
14. Biederman, I. *et al.* On the information extracted from a glance at a scene. *J. Exp. Psychol.* **103**, 597–600 (1974).
15. Bartlett, F. C. *Remembering: A Study in Experimental and Social Psychology* (Cambridge Univ. Press, Cambridge, UK, 1932).
16. Mandler, J. M. in *Memory Organization and Structure* (ed. Puff, C. R.) 259–299 (Academic, New York, 1979).
17. Palmer, S. E. The effects of contextual scenes on the identification of objects. *Mem. Cogn.* **3**, 519–526 (1975).
One of the earliest and most compelling reports of contextual influences on object recognition.
18. Piaget, J. *The Child's Construction of Reality* (Routledge & Kegan Paul, London, 1955).
19. Schank, R. C. in *Theoretical Issues in Natural Language Processing* (eds Schank, R. C. & Nash-Weber, B.) 117–121 (Tinlap, Arlington, Virginia, 1975).
20. Minsky, M. in *The Psychology of Computer Vision* (ed. Winston, P. H) 211–277 (McGraw-Hill, New York, 1975).
21. Friedman, A. Framing pictures: the role of knowledge in automatized encoding and memory for gist. *J. Exp. Psychol. Gen.* **108**, 316–355 (1979).
A thorough study of the concept of frames in contextual representations
22. Barsalou, L. W. in *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization* (eds Kittay, E. & Lehrer, A.) 21–74 (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1992).
23. Mandler, J. M. & Johnson, N. S. Some of the thousand words a picture is worth. *J. Exp. Psychol. Hum. Learn. Mem.* **2**, 529–540 (1976).
24. Intraub, H. *et al.* Boundary extension for briefly glimpsed photographs: do common perceptual processes result in unexpected memory distortions? *J. Mem. Lang.* **35**, 118–134 (1996).
25. Gottesman, C. V. & Intraub, H. Wide-angle memories of close-up scenes: a demonstration of boundary extension. *Behav. Res. Methods Instrum. Comput.* **31**, 86–93 (1999).
26. Miller, M. B. & Gazzaniga, M. S. Creating false memories for visual scenes. *Neuropsychologia* **36**, 513–520 (1998).
27. Hock, H. S. *et al.* Real-world schemata and scene recognition in adults and children. *Mem. Cogn.* **6**, 423–431 (1978).
28. Cutler, B. L. & Penrod, S. D. in *Memory in Context: Context in Memory* (eds Davies, G. M. & Thomson, D. M.) 231–244 (John Wiley & Sons Ltd, New York, 1988).
29. Oliva, A. & Torralba, A. Modeling the shape of a scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vision* **42**, 145–175 (2001).
Provides computational demonstrations that low spatial frequencies are generally sufficient for scene categorization.
30. Henderson, J. M. & Hollingworth, A. High-level scene perception. *Annu. Rev. Psychol.* **50**, 243–271 (1999).
A systematic review that elaborates on the opposition to the notion that context can facilitate object recognition.
31. Chun, M. M. Contextual cueing of visual attention. *Trends Cogn. Sci.* **4**, 170–178 (2000).
32. Intraub, H. The representation of visual scenes. *Trends Cogn. Sci.* **1**, 217–222 (1997).
33. Palmer, S. E. *Vision Science. Photons to Phenomenology* (MIT Press, Cambridge, Massachusetts, 1999).
34. Lowe, D. G. *Perceptual Organization and Visual Recognition* (Kluwer, Boston, 1985).
35. Ullman, S. Aligning pictorial descriptions: an approach to object recognition. *Cognition* **32**, 193–254 (1989).
36. Murphy, G. L. & Wisniewski, E. J. Categorizing objects in isolation and in scenes: what a superordinate is good for. *J. Exp. Psychol. Learn. Mem. Cogn.* **15**, 572–586 (1989).
37. Davenport, J. L. & Potter, M. C. Scene consistency in object and background perception. *Psychol. Sci.* **15**, 559–564 (2004).
38. Boyce, S. J., Pollatsek, A. & Rayner, K. Effect of background information on object identification. *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 556–566 (1989).
39. Metzger, R. L. & Antes, J. R. The nature of processing early in picture perception. *Psychol. Res.* **45**, 267–274 (1983).
40. Bar, M. A cortical mechanism for triggering top–down facilitation in visual object recognition. *J. Cogn. Neurosci.* **15**, 600–609 (2003).
Describes some of the conceptual bases for the model of contextual facilitation that is proposed in this review.
41. Kosslyn, S. M. *Image and Brain* (MIT Press, Cambridge, Massachusetts, 1994).
42. de Graef, P., de Troy, A. & d'Ydewalle, G. Local and global contextual constraints on the identification of objects in scenes. *Can. J. Psychol.* **46**, 489–508 (1992).
43. Hollingworth, A. & Henderson, J. M. Does consistent scene context facilitate object perception? *J. Exp. Psychol. Gen.* **127**, 398–415 (1998).
44. Auckland, M., Cave, K. R. & Donnelly, N. Perceptual errors in object recognition are reduced by the presence of context objects. *Abstr. Psychon. Soc*. **8**, 109 (2003).
45. VanRullen, R. & Thorpe, S. J. The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* **13**, 454–461 (2001).
46. Potter, M. C. & Faulconer, B. A. Time to understand pictures and words. *Nature* **253**, 437–438 (1975).
This paper reports evidence for the speed with which a scene can be comprehended.
47. Ullman, S. *High-Level Vision* (MIT Press, Cambridge, Massachusetts, 1996).
48. Gibson, J. J. *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston, 1979).
49. Moores, E., Laiti, L. & Chelazzi, L. Associative knowledge controls deployment of visual selective attention. *Nature Neurosci.* **6**, 182–189 (2003).
50. Rumelhart, D. E., McClelland, J. E. & The PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* Vol. 1 (MIT Press, Cambridge, Massachusetts, 1986).
51. Sigman, M. *et al.* On a common circle: natural scenes and Gestalt rules. *Proc. Natl Acad. Sci. USA* **98**, 1935–1940 (2001).
52. McCauley, C. *et al.* Early extraction of meaning from pictures and its relation to conscious identification. *J. Exp. Psychol. Hum. Percept. Perform.* **6**, 265–276 (1980).
53. Carr, T. H. *et al.* Words, pictures, and priming: on semantic activation, conscious identification, and the automaticity of information processing. *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 757–777 (1982).
54. Bar, M. & Biederman, I. Subliminal visual priming. *Psychol. Sci.* **9**, 464–469 (1998).
55. Potter, M. C. Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem.* **2**, 509–522 (1976).
56. Intraub, H. Rapid conceptual identification of sequentially presented pictures *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 115–125 (1981).
57. Loftus, G. R. in *Eye Movements and Psychological Processes* (eds Senders, J. & Monty, R.) 499–513 (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1976).
58. Schyns, P. G. & Oliva, A. Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception* **26**, 1027–1038 (1997).
59. Schyns, P. G. & Oliva, A. From blobs to boundary edges: evidence for time- and spatial- dependent scene recognition. *Psychol. Sci.* **5**, 195–200 (1994).
An elegant study showing that observers can categorize a scene briefly on the basis of the low-spatial-frequency content in the image.
60. Chun, M. M. & Jiang, Y. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* **36**, 28–71 (1998).
A convincing demonstration that contextual information can be learned without awareness.
61. Chun, M. M. & Phelps, E. A. Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neurosci.* **2**, 844–847 (1999).
62. Good, M., de Hoz, L. & Morris, R. G. Contingent versus incidental context processing during conditioning: dissociation after excitotoxic hippocampal plus dentate gyrus lesions. *Hippocampus* **8**, 147–159 (1998).
63. Li, F. F. *et al.* Rapid natural scene categorization in the near absence of attention. *Proc. Natl Acad. Sci. USA* **99**, 9596–9601 (2002).
64. Mathis, K. M. Semantic interference from objects both in and out of a scene context. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 171–182 (2002).
65. Kouider, S. & Dupoux, E. Partial awareness creates the 'illusion' of subliminal semantic priming. *Psychol. Sci.* **15**, 75–81 (2004).
66. Tsivilis, D., Otten, L. J. & Rugg, M. D. Context effects on the neural correlates of recognition memory: an electrophysiological study. *Neuron* **31**, 497–505 (2001).
67. Olson, I. R., Chun, M. M. & Allison, T. Contextual guidance of attention: human intracranial event-related potential evidence for feedback modulation in anatomically early temporally late stages of visual processing. *Brain* **124**, 1417–1425 (2001).
68. Kassam, K. S., Aminoff, E. & Bar, M. Spatial-temporal cortical processing of contextual associations. *Soc. Neurosci. Abstr*. 128.8 (2003).
69. Squire, L. R., Stark, C. E. L. & Clark, R. E. The medial temporal lobe. *Annu. Rev. Neurosci.* **27**, 279–306 (2004).
A clear and thorough review of the controversy surrounding the functional distinction of the various sub-regions within the medial temporal lobe.
70. Brown, M. W. & Aggleton, J. P. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nature Rev. Neurosci.* **2**, 51–61 (2001).
71. Eichenbaum, H. The hippocampus and declarative memory: cognitive mechanisms and neural codes. *Behav. Brain Res.* **127**, 199–207 (2001).
72. Schacter, D. L. & Wagner, A. D. Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus* **9**, 7–24 (1999).
73. Giovanello, K. S., Verfaellie, M. & Keane, M. M. Disproportionate deficit in associative recognition relative to item recognition in global amnesia. *Cogn. Affect. Behav. Neurosci.* **3**, 186–194 (2003).
74. Stark, C. E. & Squire, L. R. Simple and associative recognition memory in the hippocampal region. *Learn. Mem.* **8**, 190–197 (2001).
75. Aguirre, G. K. *et al.* The parahippocampus subserves topographical learning in man. *Cereb. Cortex* **6**, 823–829 (1996).
76. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
This paper coined the term 'parahippocampal place area' (PPA).
77. Maguire, E. A. *et al.* Knowing where things are: parahippocampal involvement in encoding object locations in virtual large-scale space. *J. Cogn. Neurosci.* **10**, 61–76 (1998).
78. Epstein, R., Graham, K. S. & Downing, P. E. Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* **37**, 865–876 (2003).
79. Sanocki, T. & Epstein, W. Priming spatial layout of scenes. *Psychol. Sci.* **8**, 374–378 (1997).
80. Christou, C. G. & Bülthoff, H. H. View dependence in scene recognition after active learning. *Mem. Cogn.* **27**, 996–1007 (1999).
81. Levy, I. *et al.* Center-periphery organization of human object areas. *Nature Neurosci.* **4**, 533–539 (2001).
Provides a systematic alternative view of the organization of the visual cortex.
82. Epstein, R. A. The cortical basis of visual scene processing. *Visual Cogn.* (in the press).
83. Nakamura, K. *et al.* Functional delineation of the human occipito-temporal areas related to face and scene processing. A PET study. *Brain* **123**, 1903–1912 (2000).
84. Stern, C. E. *et al.* The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **93**, 8660–8665 (1996).
85. Gaffan, D. Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection. *J. Cogn. Neurosci.* **6**, 305–320 (1994).
86. Bartels, A. & Zeki, S. Functional brain mapping during free viewing of natural scenes. *Hum. Brain Mapp.* **21**, 75–85 (2004).

87. Bar, M. *et al.* Cortical mechanisms of explicit visual object recognition. *Neuron* **29**, 529–535 (2001).
88. Kutas, M. & Hillyard, S. A. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* **207**, 203–205 (1980).
89. Ganis, G. & Kutas, M. An electrophysiological study of scene effects on object identification. *Brain Res. Cogn. Brain Res.* **16**, 123–144 (2003).
Reports interesting observations about the temporal dynamics of contextual analysis in scene recognition.
90. Smith, M. E., Stapleton, J. M. & Halgren, E. Human medial temporal lobe potentials evoked in memory and language tasks. *Electroencephalogr. Clin. Neurophysiol.* **63**, 145–159 (1986).
91. McCarthy, G. *et al.* Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *J. Neurosci.* **15**, 1080–1089 (1995).
92. Paivio, A. *Imagery and Verbal Processes* (Holt, Reinhart, & Winston, New York, 1971).
93. Paivio, A. Dual coding theory: retrospect and current status. *Can. J. Psychol.* **45**, 255–287 (1991).
94. Glaser, W. R. Picture naming. *Cognition* **42**, 61–105 (1992).
95. Riddoch, M. J. *et al.* Semantic systems or system? Neuropsychological evidence re-examined. *Cogn. Neuropsychol.* **5**, 3–25 (1988).
96. Holcomb, P. J. & McPherson, W. B. Event-related brain potentials reflect semantic priming in an object decision task. *Brain Cogn*. **24**, 259–276 (1994).
97. West, W. C. & Holcomb, P. J. Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* **12**, 1024–1037 (2000).
98. Federmeier, K. D. & Kutas, M. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *J. Exp. Psychol. Learn. Mem. Cogn.* **27**, 202–224 (2001).
99. Vandenberghe, R. *et al.* Functional anatomy of a common semantic system for words and pictures. *Nature* **383**, 254–256 (1996).
100. Smith, M. C. & Magee, L. E. Tracing the time course of picture — word processing. *J. Exp. Psychol. Gen.* **109**, 373–392 (1980).
101. Glaser, W. R. & Dungelhoff, F. J. The time course of picture-word interference. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 640–654 (1984).
102. Marinkovic, K. *et al.* Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* **38**, 487–497 (2003).
103. Sperling, R. *et al.* Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. *Neuroimage* **20**, 1400–1410 (2003).
104. Halgren, E. *et al.* Spatio-temporal stages in face and word processing. 2. Depth-recorded potentials in the human frontal and Rolandic cortices. *J. Physiol. (Paris)* **88**, 51–80 (1994).
105. Dale, A. M. *et al.* Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* **26**, 55–67 (2000).
One of the best demonstrations of high-resolution spatiotemporal imaging, with a clear description of the theoretical background.
106. Kuperberg, G. R. *et al.* Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J. Cogn. Neurosci.* **15**, 272–293 (2003).
107. Burgess, N. *et al.* A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage* **14**, 439–453 (2001).
108. Simons, J. S. & Spiers, H. J. Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Rev. Neurosci.* **4**, 637–648 (2003).
109. Maguire, E. A. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand. J. Psychol.* **42**, 225–238 (2001).
110. Cooper, B. G. & Mizumori, S. J. Temporary inactivation of the retrosplenial cortex causes a transient reorganization of spatial coding in the hippocampus. *J. Neurosci.* **21**, 3986–4001 (2001).
111. Vann, S. D. & Aggleton, J. P. Extensive cytotoxic lesions of the rat retrosplenial cortex reveal consistent deficits on tasks that tax allocentric spatial memory. *Behav. Neurosci.* **116**, 85–94 (2002).
112. Düzel, E. *et al.* Human hippocampal and parahippocampal activity during visual associative recognition memory for spatial and nonspatial stimulus configurations. *J. Neurosci.* **23**, 9439–9444 (2003).
113. Burwell, R. D. *et al.* Corticohippocampal contributions to spatial and contextual learning. *J. Neurosci.* **24**, 3826–3836 (2004).
114. Mendez, M. F. & Cherrier, M. M. Agnosia for scenes in topographagnosia. *Neuropsychologia* **41**, 1387–1395 (2003).
115. Henke, K. *et al.* Human hippocampus associates information in memory. *Proc. Natl Acad. Sci. USA* **96**, 5884–5889 (1999).
116. Jackson, O. & Schacter, D. L. Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. *Neuroimage* **21**, 456–462 (2004).
117. Hayes, S. M. *et al.* An fMRI study of episodic memory: retrieval of object, spatial, and temporal order information. *Behav. Neurosci.* (in the press).
118. Buckley, M. J. & Gaffan, D. Perirhinal cortex ablation impairs configural learning and paired-associate learning equally. *Neuropsychologia* **36**, 535–546 (1998).
119. Insausti, R., Amaral, D. G. & Cowan, W. M. The entorhinal cortex of the monkey: II. Cortical afferents. *J. Comp. Neurol.* **264**, 356–395 (1987).
120. Ranganath, C. & D'Esposito, M. Medial temporal lobe activity associated with active maintenance of novel information. *Neuron* **31**, 865–873 (2001).
121. Valenstein, E. *et al.* Retrosplenial amnesia. *Brain* **110**, 1631–1646 (1987).
122. Hirsh, R. The hippocampus and contextual retrieval of information from memory: a theory. *Behav. Psychol.* **12**, 421–444 (1974).
123. Redish, A. D. The hippocampal debate: are we asking the right questions? *Behav. Brain Res.* **127**, 81–98 (2001).
124. Miller, R. *Cortico-Hippocampal Interplay and the Representation of Contexts in the Brain*. *Studies of Brain Function.* Vol. 17 (Springer, Berlin, 1991).
125. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
126. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res*. **34**, 171–175 (1971).
127. Naya, Y., Yoshida, M. & Miyashita, Y. Forward processing of long-term associative memory in monkey inferotemporal cortex. *J. Neurosci.* **23**, 2861–2871 (2003).
128. Naya, Y., Yoshida, M. & Miyashita, Y. Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science* **291**, 661–664 (2001).
129. Higuchi, S. & Miyashita, Y. Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proc. Natl Acad. Sci. USA* **93**, 739–743 (1996).
130. Cox, D., Meyers, E. & Sinha, P. Contextually evoked object-specific responses in human visual cortex. *Science* **304**, 115–117 (2004).
131. Torralba, A. Contextual priming for object detection. *Int. J. Comput. Vision* **53**, 153–167 (2003).
132. Kersten, D., Mamassian, P. & Yuille, A. Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**, 271–304 (2004).
133. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
134. Dudai, Y. *The Neurobiology of Memory* (Oxford Univ. Press, Oxford, 1989).
135. McClelland, J. L. & Rumelhart, D. E. An interactive activation model of context effects in letter perception: part 1. An account of basic findings. *Psychol. Rev.* **88**, 375–407 (1981).
136. Felleman, D. J. & Van Essen, V. C. Distributed hierarchical processing in primate visual cortex. *Cereb. Cortex* **1**, 1–47 (1991).
137. Rempel-Clower, N. L. & Barbas, H. The laminar pattern of connections between prefrontal and anterior temporal cortices in the rhesus monkey is related to cortical structure and function. *Cereb. Cortex* **10**, 851–865 (2000).
138. Ullman, S. Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex* **1**, 1–11 (1995).
Provides a theory and compelling demonstrations for the existence and role of bidirectional processes in the cortex.
139. Grossberg, S. How does a brain build a cognitive code? *Psychol. Rev.* **87**, 1–51 (1980).
140. Graboi, D. & Lisman, J. Recognition by top–down and bottom–up processing in cortex: the control of selective attention. *J. Neurophysiol.* **90**, 798–810 (2003).
141. Merigan, W. H. & Maunsell, J. H. How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* **16**, 369–402 (1993).
142. Bullier, J. & Nowak, L. G. Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr. Opin. Neurobiol.* **5**, 497–503 (1995).
143. Schmid, A. M. & Bar, M. Selective involvement of prefrontal cortex in visual object recognition. *Soc. Neurosci. Abstr*. 161.8 (2002).
144. Schmid, A. M. & Bar, M. Activation of multiple candidate object representations during top–down facilitation of visual recognition. *Soc. Neurosci. Abstr*. 128.5 (2003).
145. Pandya, D. N., Seltzer, B. & Barbas, H. in *Comparative Primate Biology, Vol. IV: Neurosciences* (eds Staklis, H. D. & Erwin, J.) 39–80 (Alan R. Liss, New York, 1988).
146. Mannan, S. K., Ruddock, K. H. & Wooding, D. S. Fixation patterns made during brief examination of two-dimensional images. *Perception* **26**, 1059–1072 (1997).
147. Tamura, H. & Tanaka, K. Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb. Cortex* **11**, 384–399 (2001).
148. Sugase, Y. *et al.* Global and fine information coded by single neurons in the temporal visual cortex. *Nature* **400**, 869–873 (1999).
149. Antes, J. R. Recognizing and localizing features in brief picture presentations. *Mem. Cogn.* **5**, 155–161 (1977).
150. Nowak, L. G. & Bullier, J. in *Cerebral Cortex: Extrastriate Cortex in Primate* (eds Rockland, K., Kaas, J. & Peters, A.) 205–241 (Plenum, New York, 1997).
151. Torralba, A. & Oliva, A. Statistics of natural image categories. *Network* **14**, 391–412 (2003).
152. Rensink, R., O'Regan, J. & Clark, J. To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* **8**, 368–373 (1997).
153. Simons, D. J. & Levin, D. T. Change blindness. *Trends Cogn. Sci.* **1**, 261–267 (1997).
154. Haber, R. N. & Schindler, R. M. Errors in proofreading: evidence of syntactic control of letter processing. *J. Exp. Psychol. Hum. Percept. Perf.* **7**, 573–579 (1981).
155. Morris, A. L. & Harris, C. L. Sentence context, word recognition, and repetition blindness. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 962–982 (2002).
156. Kanwisher, N. G. Repetition blindness: type recognition without token individuation. *Cognition* **27**, 117–143 (1987).
156. Green, R. T. & Courtis, M. C. Information theory and figure perception: the metaphor that failed. *Acta Psychol. (Amst.)* **25**, 12–35 (1966).

Acknowledgements

I would like to thank members of my lab, E. Aminoff, H. Boshyan, M. Fenske, A. Ghuman, N. Gronau and K. Kassam, as well as A. Torralba, N. Donnelly, M. Chun, B. Rosen and A. Oliva for help with this article. Supported by the National Institute of Neurological Disorders and Stroke, the James S. McDonnell Foundation (21st Century Science Research Award in Bridging Brain, Mind and Behavior) and the MIND Institute.

Competing interests statement

The author declares no competing financial interests.

Online links

FURTHER INFORMATION

Bar laboratory: http://www.nmr.mgh.harvard.edu/~bar/Lab

Access to this interactive links box is free online.