Cell PRESS

Novelty signals: a window into hippocampal information processing

Dharshan Kumaran and Eleanor A. Maguire

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, UK

The function of the human hippocampus is a contentious subject among neuroscientists. Theoreticians have long viewed the hippocampus as a computational device, with researchers in humans increasingly adopting this perspective, buoyed by recent reports that its role is not limited to declarative memory. Here, we set out a new strategy for discovering the nature of information processing within the human hippocampus. We argue that novelty responses, measured by functional magnetic resonance imaging, provide a window into the neural representations and computations sustained by the hippocampus. More generally, we suggest that a renewed emphasis on the information processing qualities of the human hippocampus offers the promise of a long awaited union between theoretical and empirical research across species.

Introduction

Ever since the seminal discovery of place cells by O'Keefe and Dostrosky [1] in 1971, neuroscientists have sought to understand the nature of information processing carried out by the rodent hippocampus [2–5]. By contrast, hippocampal research in humans has been dominated by a neuropsychological approach that attempts to characterize the function of the hippocampus according to the types of memory it supports [6,7]. In this account, the hippocampus and surrounding neocortical regions that comprise the medial temporal lobe (MTL) constitute a dedicated declarative memory module concerned specifically with the long term retention and recall of facts and events [6]. As such, non-declarative functions, for example perception and motor skill learning, are viewed to be supported by separate modules instantiated in brain regions outside the MTL, such as posterior neocortex and striatum, respectively [6].

Recently, however, empirical evidence has emerged demonstrating that the human hippocampus has a role outside the traditional domain of declarative memory, in tasks such as short-term memory [8–10], implicit memory [11], imagination [12] and even perception [13,14]. These findings pose considerable challenges for the textbook account of memory [6], which has so long espoused a modular view of brain function grounded in clear-cut distinctions between processes like memory and perception. A highly influential observation has been that patients with MTL damage perform poorly not only on long-term memory tasks but also on short-term memory tasks that involve remembering novel information across brief intervals [15]. For instance, in one recent study, patients with hippocampal amnesia were impaired at remembering the locations of novel objects, but not the objects themselves, even across a delay of a few seconds without distraction [8]. Based on these findings, therefore, it has been suggested that the function of the human hippocampus, like its rodent counterpart, might be best characterized by understanding the nature of neural representations and computations it sustains [16] (Box 1). In this way, it might be possible to explain why the hippocampus is crucial to a range of cognitive functions, from long-term memory to short-term memory, imagination and perception.

Here, we suggest a new strategy for discovering the nature of information processing within the human hippocampus by drawing on its role in novelty detection [17,18]. First, we propose that hippocampal novelty responses, measured using functional magnetic resonance imaging (fMRI) and repetition paradigms, provide evidence concerning the nature of underlying neural representations. Specifically, we suggest that the presence of novelty responses to a given change in sensory inputs (e.g. a sequence of objects) provides evidence that the hippocampus supports the relevant representations (i.e. coding temporal order). Second, we argue that the pattern of novelty responses generated by the hippocampus in response to step-wise changes in sensory inputs can be used to make inferences about the types of computations that are operating. Third, we discuss how cutting-edge high spatial resolution fMRI technology, in combination with the approach we describe, might permit the function of individual subregions within the hippocampus to be defined. More generally, we suggest that a renewed focus on the nature of information processing carried out by the human hippocampus offers the promise of a long awaited convergence between computational modelling and empirical research in rodents and humans.

Hippocampal representations in humans: repetition paradigms and fMRI

Novelty responses distinguish new stimuli or stimulus configurations from those that have been encountered in the past. These signals are generated rapidly and automatically, suggesting they represent a fundamental component of a brain network performing novelty detection [17–20]. Repetition paradigms have been widely used to elicit novelty signals across species, with neural activity measured using methodologies ranging from single-cell recording and immediate-early-gene imaging in animals

Corresponding author: Kumaran, D. (d.kumaran@fil.ion.ucl.ac.uk).

1364-6613/$ – see front matter doi:10.1016/j.tics.2008.11.004 Available online 8 January 2009

Box 1. Neural representations and computations in the hippocampus

The concepts of neural representations and computations are central to neuroscience and to our understanding of the function of the hippocampus [2,3,55]. Representations are not unique to the brain, and are crucial to many aspects of everyday life (e.g. words represent ideas or concepts, names represent specific people). In the brain, neural representations constitute the information that is encoded about a given input (e.g. stimulus) by a given neuron or population of neurons. Different areas of the brain have been shown to represent different kinds of information, ranging from the encoding of simple (V1) and complex unimodal information (inferotemporal cortex) [36], to more abstract constructs derived from highly processed multimodal information such as spatial location in an environment (i.e. hippocampus; Ref. [1]). Recent anatomical evidence indicates that individual hippocampal subregions differ with respect to the inputs they receive and, therefore, the nature of information represented [38]. Parallel processing streams from the medial and lateral entorhinal cortex converge at the level of CA3, but remain separate in CA1. As such, CA3 might be ideally positioned to integrate spatial and non-spatial information into object-place, or perhaps event-context, representations [2,38].

On its own, however, the representation of information in population codes would be of little use. Computations, that is the capacity to transform neural representations, is an essential complementary process [55]. Networks of neurons perform computations by implementing input–output functions, taking the population activity from one set of neurons as the input and producing population activity on another as the output. Detailed knowledge of hippocampal anatomy, in particular the nature of the inputs it receives (i.e. highly processed multimodal information) and unique characteristics of the connectivity and plasticity of individual subregions (e.g. CA3) [38], has inspired researchers to develop computational models of hippocampal function (see Ref. [3] for a recent review). Example computations relevant to the hippocampus are pattern separation and pattern completion. Pattern completion, a process that might occur in the CA3 region of the hippocampus, enables a partial input (e.g. specific face) to be transformed into the recall of the entire stored experience (e.g. birthday party) which is expressed as the output of the network. Pattern separation refers to the process by which interference between similar memories is minimized through the orthogonalization of inputs perhaps mediated by the sparse method of coding adopted by the dentate gyrus [44].

Box 2. fMRA: a tool for investigating the nature of neural representations

fMRA is a tool that has been widely used to infer the nature of object representations within the ventral visual stream, based on the robust finding of repetition-related reductions in neural activity [23]. A key theoretical issue addressed by such studies is the extent to which neural representations in regions such as LOC code information relating to parameters such as object size and viewing angle, or alternatively are 'invariant' to these properties [23,25,26].

The principles of this approach are well illustrated by an early application of fMRA by Grill-Spector and colleagues [26]. Participants were scanned while they viewed a stream of object pictures and performed an incidental (e.g. 1-back) task. Objects presented were either (i) new objects (N), (ii) exact repetitions of previously seen objects (R) or (iii) versions of previously seen objects (i.e. 'lure' objects [L]) varied along a specific dimension, in this case size. New objects elicited relatively greater neural activity than exact repetitions throughout the brain, a phenomenon known as a novelty response. This effect is equivalently referred to as repetition suppression, or neural adaptation, given that neural activity is reduced for exact repeats compared with novel items [23]. The crucial finding in this study, and indeed fMRA experiments, pertains to the response of a given brain region (in this case LOC) to lures. Neural activity in LOC distinguished exact repetitions from lures, providing evidence that neurons in this region code for the size of an object [26]. By contrast, fusiform cortex failed to make this distinction, suggesting that it is 'blind' to object size (Figure I).

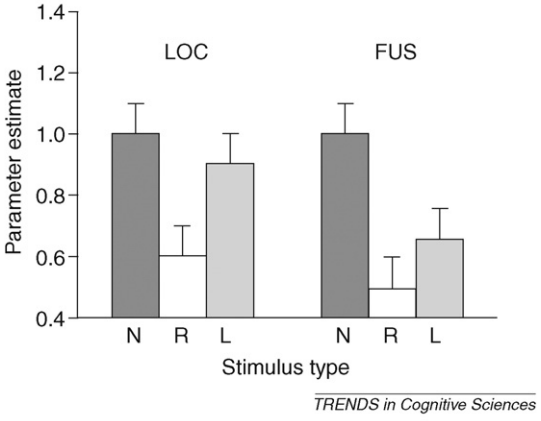

Figure I. Results of a prototypical fMRA experiment designed to investigate where object size is represented in the brain [26]. Condition-specific plots of mean activity in lateral occipital cortex (LOC) and fusiform (FUS) regions-of-interest. Abbreviations: N,new object; R,exact repetition; L,lure object (in this case a previously encountered object presented in a different size). Adapted, with permission, from Ref. [26].

[21,22] to fMRI in humans [23]. One technique that has become particularly popular in recent years is fMR-adaptation (fMRA; see Box 2).

Whereas there is agreement that fMRA is a powerful tool for elucidating the nature of visual representations, the underlying neural mechanisms remain subject to debate [23]. One hypothesis, termed the sharpening or tuning model, indicates that adaptation occurs because poorly selective neurons that fire in response to the original stimulus presentation dropout (i.e. are tuned out) of the neural representation when the stimulus is repeated [23]. Indeed, familiarity-based models of recognition memory propose that the sharpness of neural representations can be used to compute a scalar signal that indexes the familiarity, or novelty, of current sensory input [18,24].

Although numerous studies have employed fMRA to elucidate the nature of visual representations in regions such as the lateral occipital complex (LOC) (e.g. Refs [23,25,26]) and parahippocampal place area (PPA) (e.g. Refs [27]), traditionally the function of the human hippocampus has been investigated using paradigms in which the memory of subjects is explicitly tested [28–32]. However, recent empirical and theoretical work indicate that the hippocampus (and MTL) might have both mnemonic and perceptual functions, suggesting it might be best characterised as part of a continuum of brain regions within an information processing hierarchy, rather than as an isolated memory module [2,8–10,12,13,16,33,34].

In a recent experiment, therefore, we used fMRI and a repetition paradigm to test the hypothesis that the hippocampus supports neural representations of temporal order [35]. Whereas neurons in brain regions such as the LOC are thought to code for object features (e.g. size) [36], anatomical and theoretical considerations indicate that the hippocampus is likely to sustain neural representations coding for more complex parameters, such as associative information (e.g. object–object and object–place associations) [2,37–39]. In our experiment, subjects viewed

sequences of trial-unique objects while performing an incidental target detection task (Figure 1a). Object quartets were re-presented after a short delay in either (i) exactly the same order (S_{rep}: 'repeated'), (ii) where just the last two objects changed order (S_{half}: 'half') or (iii) in an entirely novel order (S_{new}: 'new'). Although our design was conceptually similar to previous experiments employing repetition paradigms (Box 2), we used sequences of objects rather

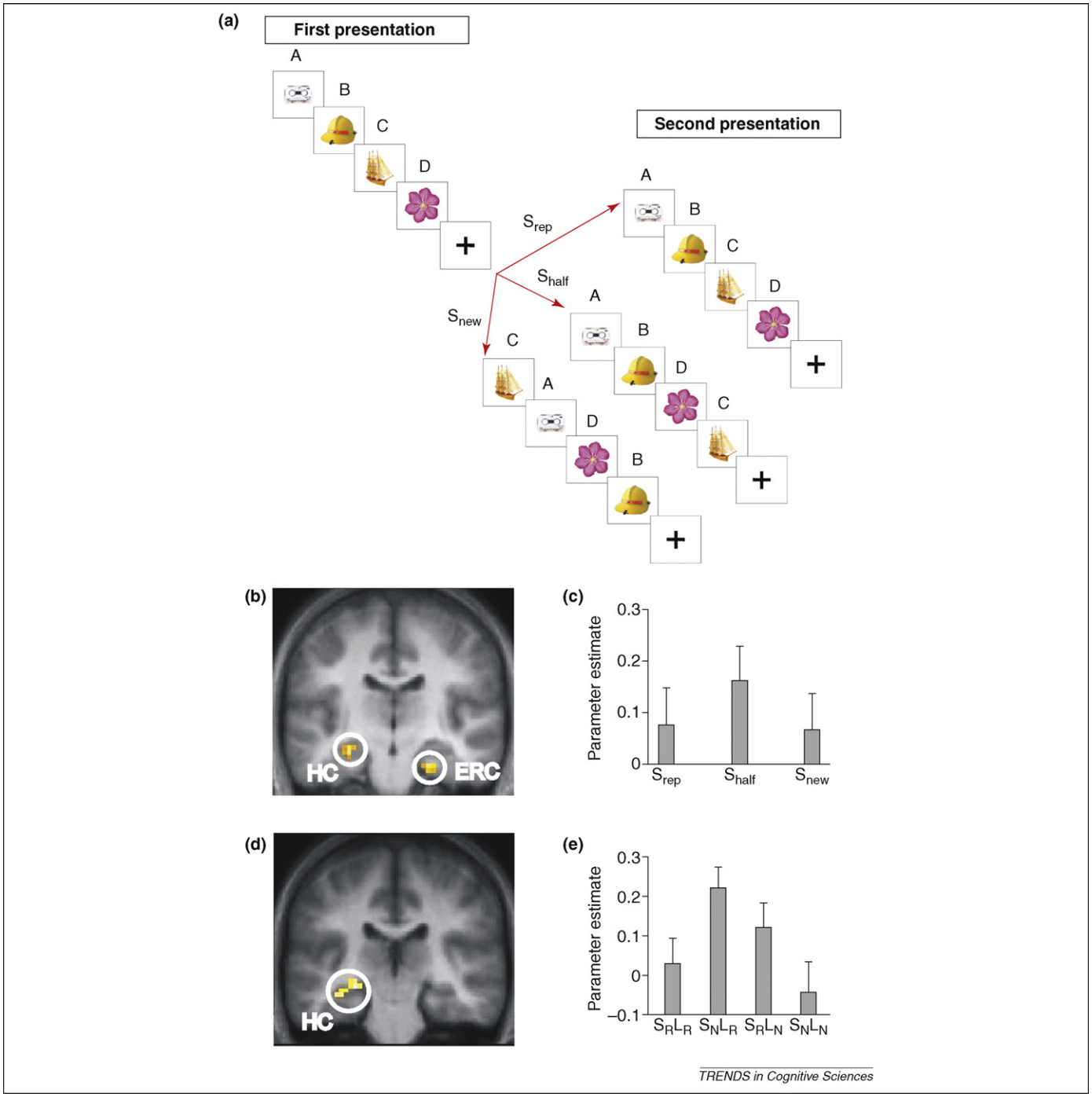

Figure 1. Empirical evidence for sequence representations and match–mismatch computations in the human hippocampus. **(a)** Experimental paradigm used by Kumaran and Maguire [35].Colour pictures of trial-unique objects were presented in groups of four, constituting an object quartet. Objects were presented consecutively for 1 s each, with a 200 msec gap between objects, over a background scene (not shown) which was displayed continuously throughout the presentation of each object quartet (first presentation). After a 2 s period during which a fixation cross was displayed, the same object quartet was re-presented. During the second presentation, object order was either exactly as before (A-B-C-D: S_{rep}), partially rearranged (A-B-D-C: S_{half}) or entirely rearranged (C-A-D-B: S_{new}). Subjects performed an incidental 1-back target detection task throughout the experiment (see Ref. [35] for more details). **(b)** Neural activity in left hippocampus and right entorhinal/perirhinal cortex distinguished partially rearranged object sequences (S_{half}) from exact repetitions (S_{rep}) [35]. Results of comparison of S_{half} and S_{rep} conditions is displayed on average structural image of participants (n=17). Threshold is set at $p<0.005$ uncorrected, for display purposes. **(c)** Condition-specific plots of activity in the peak voxel in the left hippocampus, showing that novelty signals occur selectively in S_{half} condition and not in S_{new} condition [35]. **(d)** Match–mismatch computations outside the temporal domain [41]. The experimental design was similar to Ref. [35], although object quartets were presented consecutively at distinct spatial locations on the screen. During the second presentation, objects were re-presented in either the same (S_R) or different (S_N) order, and in the same (L_R) or different (L_N) spatial locations. Significant hippocampal activity was observed selectively under conditions of match–mismatch (i.e. S_NL_R, S_RL_N). **(e)** Condition-specific plots of activity in the peak voxel in the left hippocampus [41]. Adapted, with permission, from Refs [35,41].

than single objects to probe the nature of associative representations within the hippocampus. Furthermore, we included graded levels of novelty (i.e. half and fully rearranged sequences) to enable us to ask which computations give rise to the observed blood-oxygen-level-dependent (BOLD) responses (see more on this later).

As expected, a distributed network of brain regions showed significantly greater activity during the first presentation of the object sequence relative to the repeated condition, reflecting a stimulus novelty response. Crucially, however, neural activity in the hippocampus distinguished partially rearranged sequences (i.e. half) from exact repetitions (Figure 1b). This finding indicates that neurons within the hippocampus encode the order in which objects are presented during the first presentation, which results in the generation of novelty signals when the same objects are subsequently presented in a partially rearranged order (i.e. half). As such, this finding indicates that the hippocampus is capable of supporting one-shot associative representations coding for the temporal order of events, a capacity proposed to underpin its role in episodic memory [37,40].

To summarize, we suggest that repetition paradigms and fMRI can be used to define which categories of novelty the hippocampus is sensitive to, and therefore elucidate the nature of the neural representations it sustains. Whilst the evidence we have presented supports the contention that the hippocampus supports one-shot neural representations of the temporal order of objects [35], in other work we have shown that conjunctive representations coding for both the spatial and temporal parameters of an episode exist in the hippocampus [41]. In the future, we suggest this approach will be a useful tool in elucidating the function of individual structures within the MTL. For instance, does the hippocampus represent specific aspects of a visual scene, such that damage to this structure leads to impairment on perceptual or short-term memory tasks involving scenes [9,13]? Furthermore, is the perirhinal cortex important for tasks involving feature ambiguity because it sustains neural representations for complex object–feature conjunctions [16]?

Inferring computations from the pattern of fMRI novelty responses

Although the intricate anatomy and connectivity of the hippocampus has inspired a long line of computational modellers beginning with David Marr [3,42,43], it has been technically challenging to subject these hypotheses to empirical testing in humans. Computations such as pattern completion and pattern separation tend to be defined at the neuronal level [3,5,44] (Box 1). To provide direct empirical evidence for the operation of specific computations, therefore, it is necessary to record from individual hippocampal neurons or neuronal ensembles, which although routine in rodents (Box 3), is practically difficult in humans.

Box 3. Computations performed by the rodent hippocampus: empirical data

Different computational functions make different predictions about how graded changes to the external environment (i.e. sensory input) should influence the pattern of hippocampal activity (i.e. output) [3,5,43]. For instance, pattern completion acts to resist environment changes, as far as possible, forcing the reinstatement of stored patterns of activity in the hippocampus. By contrast, pattern separation can be considered an opposing force, acting to amplify even the smallest change in the external environment to create new and orthogonal patterns of neuronal activity.

In rodents, it is possible to directly measure neuronal activity in individual subregions of the hippocampus using single-cell recording techniques or immediate early gene imaging (IEG). Several recent studies have provided empirical evidence for the operation of specific computations in the hippocampus by systematically introducing graded changes in the external environment, and observing how this influences the pattern of hippocampal activity [5,22,52,53].

In one study [22] (Figure I), rats were first exposed to environment A (1st environment), a square enclosure with red balls within it. After a delay period, the rats were then exposed to a 2nd environment which could either be an exact repetition of the previous environment (A), partially novel (A′: square enclosure, red balls in a new configuration) or entirely novel (B: circular enclosure, green triangular shapes – Figure I, *x*-axis). The population of hippocampal neurons active during exposure to first and second environments was assessed using IEG. From this a similarity score was calculated (Figure I, *y*-axis) for both CA3 and CA1, which reflected the extent of overlap between neuronal ensembles activated in response to the first and second environments.

Crucially, CA3 and CA1 differed significantly in terms of their response to environmental change (Figure I). In particular, when the second environment was partially novel (i.e. A′), similarity scores in CA1 were lower than in CA3. These findings indicate that when rats are exposed to an environment that is part novel and part familiar (i.e. A′), CA1 performs pattern separation, thereby creating a new and separate ensemble representation. By contrast, CA3 would seem to respond to intermediate environments (i.e. A′) by performing pattern completion whereby the ensemble code for the original environment (i.e. A) is reinstated. This study, therefore, provides support for the operation of specific computations in individual subregions of the rodent hippocampus, by comparing predicted input–output functions to empirically observed data.

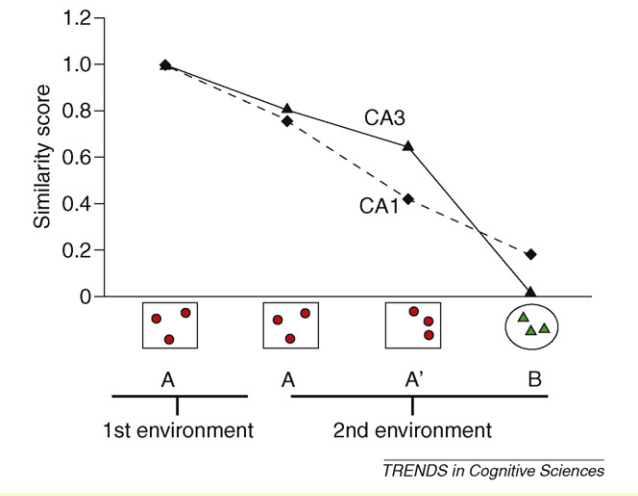

Figure I. Differing responses of CA3, compared to CA1, to graded levels of environmental novelty. Environment type is plotted on the *x*-axis. 1st environment (A). 2nd environment: (i) exact repetition (A), (ii) partially novel (A′) and (iii) entirely novel (B). Similarity score, reflecting the overlap between neuronal ensembles activated during first and second environment is plotted on the *y*-axis. A similarity score of 0 indicates no overlap between ensembles, and a score of 1 implies complete overlap. For sake of clarity, exposure to the first environment (A) is denoted by a similarity score of 1 for use as a reference point. Crucially, similarity scores in CA3 and CA1 differed significantly in response to environmental changes, suggesting a bias towards pattern completion in the former, and pattern separation in the latter [22].

Here, we suggest that novelty responses measured by fMRI can be used as a window into the types of computations performed by the hippocampus. Novelty responses are generated through hippocampal information processing, and therefore can be viewed as a surrogate marker of hippocampal outputs, which parallel the direct recording of hippocampal neuronal activity obtained in rodents. We argue, therefore, that computational models can be empirically tested using fMRI and repetition paradigms, provided these models make clear *a priori* predictions about how the amplitude of observed hippocampal novelty responses should vary in relation to graded levels of novelty present in sensory inputs.

In the previous section, we discussed how the presence of hippocampal novelty responses to rearranged object sequences provides evidence of neural representations of temporal order in this brain region. In fact, the primary aim of our experiment [35] (see also Ref. [41]) was to define which computations underpin the generation of hippocampal novelty signals. Crucially, therefore, our experimental design incorporated a step-wise change in the degree of sensory novelty (i.e. rep<half<new), which allowed us to distinguish between the operation of two competing computational models of hippocampal function. One class of models, termed familiarity-based mechanisms, indicate that novelty signals result from the tendency of familiar stimuli (or stimulus configurations) to trigger neural activity in a reduced population of neurons compared to novel stimuli [18,21,24] (Box 2). As such, familiarity-based models are viewed to respond to novelty *per se*, outputting a linear novelty signal indexing the degree to which current sensory input differs from past experience [18].

By contrast, match–mismatch models account for the genesis of hippocampal novelty signals in an entirely different fashion [18,40,45]. According to this perspective, novelty signals are generated through the operation of a comparator [46,47], which discharges specifically when current sensory input conflicts with what would be expected (based on past experience), rather than more promiscuously in response to novelty *per se*. Under this operating mechanism, novelty or mismatch signals are contingent on an initial match process, which triggers the recall of prior predictions through pattern completion. Because current sensory inputs that are sufficiently different from stored representations will fail to trigger a match process, match–mismatch models predict an inverted U-shaped relationship between the amplitude of hippocampal novelty response (i.e. output) and the degree of sensory novelty (i.e. input) [18].

Our findings [35], therefore, argue strongly against the existence of a familiarity-based mechanism within the hippocampus because novelty signals were not observed under conditions of maximal novelty (i.e. new) where the entire object sequence was rearranged (Figure 1c). Instead, our data provide empirical support for the operation of hippocampal match–mismatch computations, with novelty signals seen selectively in the half condition where initial predictions based on the first two items in the sequence (i.e. match) were subsequently violated (i.e. mismatch). In other work we have demonstrated that the role of hippocampal match–mismatch computations is not restricted to the temporal domain, but also extends to other domains (i.e. spatial; see Figure 1d,e), suggesting a generic role during associative processing [41].

It is important to note that in our experiments [35,41], the novelty or familiarity of associative information was incidental to the task at hand. Interestingly, when subjects are required to use the novelty or familiarity of current sensory inputs to make recognition memory judgements [48] or guide visual search [49], hippocampal activation has been shown to be maximal under conditions of match, rather than mismatch, mirroring the phenomenon of 'match enhancement' observed in monkey inferotemporal or perirhinal cortex [21]. An important avenue for future work will involve exploring whether the amplitude of hippocampal novelty or familiarity signals is modulated by top-down influences that arise from regions such as the prefrontal cortex, and vary as a function of the task being performed.

To summarize, we have shown that, although the presence of novelty responses to a given change in sensory inputs (e.g. sequence) provides evidence of existence of the relevant neural representations (i.e. temporal order), the pattern of novelty responses observed in relation to step-wise levels of sensory novelty can be used to infer the nature of underlying computations.

Measuring novelty responses within human hippocampal subregions

Match–mismatch detection can be considered a multi-component computational function involving pattern completion, a comparison process and mismatch detection [18]. The recent advent of high spatial resolution fMRI offers the hope that it might be possible to define the nature of representations and computations of individual subregions of the human hippocampus (Box 1). A recent experiment by Bakker *et al.* [50] (Figure 2a) employed this methodology in combination with a repetition paradigm. In their study, subjects viewed pictures of everyday objects (e.g. an apple) while performing an incidental task. On each trial the object presented was either new, an exact repetition of a previously viewed object or a slightly different version of a previously presented object (termed a 'lure'). As expected, based on a wealth of previous data, all subregions of the hippocampus exhibited significantly greater BOLD responses in relation to new objects as compared to exact repetition. However, when a lure was presented, novelty responses were observed in CA3/DG, but not in CA1 (or several other areas within the MTL; see Figure 2b). The authors argued that the failure of CA3/DG to show adaptation to lures reflects a bias towards pattern separation [42,44], a computational process which ensures that, as far as possible, non-overlapping neuronal ensembles are activated in response to similar episodes (i.e. lures versus the original object). Furthermore, their contrary finding of adaptation to lures in CA1 is proposed to reflect a bias towards pattern completion.

The observed difference between CA3/DG and CA1 is convincing, and points to a division of function between these two hippocampal subregions. We would suggest, however, that the nature of the difference between

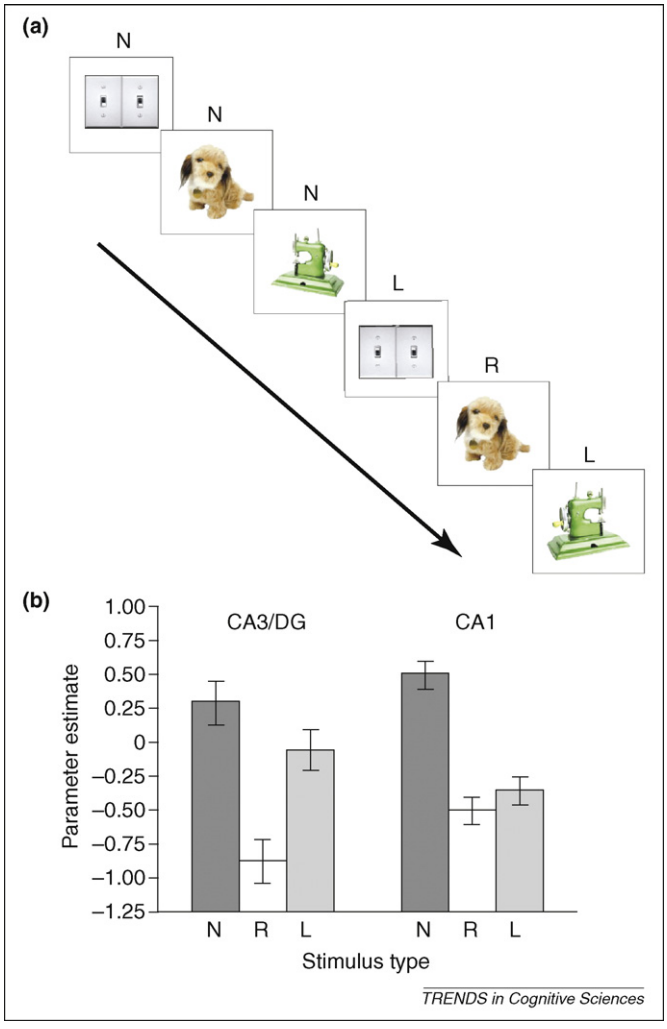

Figure 2. Using high-resolution fMRI to measure novelty responses within human hippocampal subregions. **(a)** Schematic of experimental design used by Bakker *et al.* [50]. Subjects viewed pictures of everyday objects while performing an incidental task (indoor versus outdoor classification). On each trial, a presented object could be either a new object (N), an exact repetition (R) or a slightly different version of a previously encountered object, termed a lure (L). Two examples of lures are shown: the light switches are in a different position in one stimulus, and a sewing machine is shown from a different angle. Note the stimuli shown are illustrative, and are not drawn from the actual stimulus set used in the experiment. **(b)** Condition-specific plots of mean activity in each region-of-interest (CA3/DG and CA1) illustrating the differential response of CA3/DG and CA1 to lure objects. Neural activity in CA3/DG, but not CA1, is significantly greater in response to lure objects compared to exact repetitions. Adapted, with permission, from Ref. [50].

CA3/DG and CA1 remains an open question given the current findings. In particular, Bakker *et al.* [50] used a classical fMRA paradigm highly similar to previous experiments (e.g. Ref. [25]). Notably, analogous findings in such experiments [26], albeit in regions such as the LOC and fusiform cortex, have been viewed to reflect a difference in the nature of information represented, rather than computations performed (Box 2). As such, one possibility is that neural activity in CA3/DG, but not CA1, distinguishes lures from exact repetitions because only this subregion 'sees' and therefore 'represents' the necessary sensory information (e.g. spatial location of light switch; see Figure 2). Alternatively, neural representations of objects in CA1 might be more conceptual (i.e. abstracted) [51]), compared to CA3, leading this region to treat lures and exact repetitions equivalently. We believe that Bakker *et al.* [50] cannot reasonably rule out these alternative explanations for their findings, given little is currently understood about what type of information is coded in individual subregions of the human hippocampus.

Representational explanations aside, we suggest that the experimental design used by Bakker *et al.* [50] makes it difficult to draw inferences concerning the computational functions of CA3/DG and CA1. Whilst the authors included three trial types (exact repetitions, lures and novel objects), novelty was not experimentally manipulated in a step-wise or systematic fashion. Specifically, novel objects differed from previously viewed objects across numerous, often unquantifiable, dimensions (e.g. low-level features, feature conjunctions), explaining their tendency to activate much of the ventral visual stream [23]. By contrast, previous experiments in rodents [5,22,52,53], in addition to our own experiments using fMRI in humans [35,41], have introduced graded changes in sensory inputs to facilitate computational inferences based on the shape of the input–output function observed (Figure 1 and Box 3). Thus, Bakker *et al.*'s [50] finding of increased neural activity in CA3/DG to lures, compared to exact repetitions, might equally reflect pattern separation based on an adaptation type model (Box 2) as the authors conclude, or equally the operation of a match–mismatch detector. Without observing neural activity in CA3/DG in response to graded levels of novelty, they cannot distinguish between these two plausible alternatives.

Bakker *et al.* [50] have taken an interesting step in applying cutting-edge fMRI technology to the important issue of which computations operate within individual hippocampal subregions. Although the divergent response of CA3/DG and CA1 to lures is certainly intriguing, it remains for future studies to define whether this arises because of a difference in function along a representational or computational dimension.

Conclusions and future directions

Recent evidence indicates that the functions of the human hippocampus and adjacent MTL regions might be best understood in terms of the information processing they perform [2,16]. Here, we have outlined a new strategy for tackling this challenging question using novelty responses as a window into the nature of neural representations and computations within the hippocampus. We believe this methodology has the potential to address fundamental issues concerning the function of the hippocampus and related structures. However, it is important to note a limitation of our approach in its current form. We have discussed how competing computational models (familiarity-based mechanisms and match–mismatch computations) make divergent predictions regarding the amplitude of hippocampal novelty responses. Ideally, however, one would test these predictions in a quantitative, rather than qualitative fashion as we and others have done until now [35,41,50], an approach that is routine in the field of reinforcement learning models and decision-making [54].

One obvious, but difficult, solution to this problem is to develop fully functional neural network simulations of the

hippocampus that make quantitative predictions about the amplitude of novelty responses generated in relation to a given sensory input, which can be tested using fMRI. These models should include individual subregions of the hippocampus, incorporate the known characteristics of inputs to these regions and capture a variety of proposed hippocampal computations (e.g. pattern separation, pattern completion and mismatch detection). Though the complexity of the hippocampus makes this a daunting task, computational modelling in combination with cutting-edge fMRI techniques with ever improving spatial resolution, offer the promise of a long-awaited union between theoretical and empirical research across species.

Acknowledgements
The authors thanks Demis Hassabis and Hugo Spiers for useful discussions.

References

1 O'Keefe, J. and Nadel, L. (1978) *The Hippocampus as a Cognitive Map.* OUP
2 Knierim, J.J. *et al.* (2006) Hippocampal place cells: parallel input streams, subregional processing, and implications for episodic memory. *Hippocampus* 16, 755–764
3 Burgess, N. (2006) Computational models of the spatial and mnemonic functions of the hippocampus. In *The Hippocampus Book* (Bliss, T. *et al.*, eds), pp. 715–751, Oxford University Press
4 Morris, R.G. (2006) Theories of hippocampal function. In *The Hippocampus Book* (Bliss, T. *et al.*, eds), pp. 581–715, Oxford University Press
5 Moser, E.I. and Moser, M.B. (2003) One-shot memory in hippocampal CA3 networks. *Neuron* 38, 147–148
6 Squire, L.R. *et al.* (2004) The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306
7 Squire, L.R. (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* 99, 195–231
8 Olson, I.R. *et al.* (2006) Working memory for conjunctions relies on the medial temporal lobe. *J. Neurosci.* 26, 4596–4601
9 Hartley, T. *et al.* (2007) The hippocampus is required for short-term topographical memory in humans. *Hippocampus* 17, 34–48
10 Hannula, D.E. *et al.* (2006) The long and the short of it: relational memory impairments in amnesia, even at short lags. *J. Neurosci.* 26, 8352–8359
11 Greene, A.J. (2007) Human hippocampal-dependent tasks: is awareness necessary or sufficient? *Hippocampus* 17, 429–433
12 Hassabis, D. *et al.* (2007) Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1726–1731
13 Lee, A.C. *et al.* (2005) Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus* 15, 782–797
14 Lee, A.C. *et al.* (2005) Perceptual deficits in amnesia: challenging the medial temporal lobe 'mnemonic' view. *Neuropsychologia* 43, 1–11
15 Jonides, J. *et al.* (2008) The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59, 193–224
16 Bussey, T.J. and Saksida, L.M. (2007) Memory, perception, and the ventral visual-perirhinal-hippocampal stream: thinking outside of the boxes. *Hippocampus* 17, 898–908
17 Ranganath, C. and Rainer, G. (2003) Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202
18 Kumaran, D. and Maguire, E.A. (2007) Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus* 17, 735–748
19 Yamaguchi, S. *et al.* (2004) Rapid prefrontal-hippocampal habituation to novel events. *J. Neurosci.* 24, 5356–5363
20 Strange, B.A. *et al.* (2005) Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw.* 18, 225–230
21 Brown, M.W. and Aggleton, J.P. (2001) Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.* 2, 51–61
22 Vazdarjanova, A. and Guzowski, J.F. (2004) Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *J. Neurosci.* 24, 6489–6496
23 Grill-Spector, K. *et al.* (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23
24 Norman, K.A. and O'Reilly, R.C. (2003) Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev.* 110, 611–646
25 Vuilleumier, P. *et al.* (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat. Neurosci.* 5, 491–499
26 Grill-Spector, K. *et al.* (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203
27 Epstein, R.A. *et al.* (2005) Learning places from views: variation in scene processing as a function of experience and navigational ability. *J. Cogn. Neurosci.* 17, 73–83
28 Kohler, S. *et al.* (2005) Novelty responses to relational and non-relational information in the hippocampus and the parahippocampal region: a comparison based on event-related fMRI. *Hippocampus* 15, 763–774
29 Duzel, E. *et al.* (2003) Human hippocampal and parahippocampal activity during visual associative recognition memory for spatial and nonspatial stimulus configurations. *J. Neurosci.* 23, 9439–9444
30 Johnson, J.D. *et al.* (2008) Multiple repetitions reveal functionally and anatomically distinct patterns of hippocampal activity during continuous recognition memory. *Hippocampus* 18, 975–980
31 Yassa, M.A. and Stark, C.E. (2008) Multiple signals of recognition memory in the medial temporal lobe. *Hippocampus* 18, 945–954
32 Schacter, D.L. and Wagner, A.D. (1999) Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus* 9, 7–24
33 Ranganath, C. and Blumenfeld, R.S. (2005) Doubts about double dissociations between short- and long-term memory. *Trends Cogn. Sci.* 9, 374–380
34 Bird, C.M. and Burgess, N. (2008) The hippocampus and memory: insights from spatial processing. *Nat. Rev. Neurosci.* 9, 182–194
35 Kumaran, D. and Maguire, E.A. (2006) An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biol.* 4, e424
36 Grill-Spector, K. and Malach, R. (2004) The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677
37 Eichenbaum, H. (2004) Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* 44, 109–120
38 Amaral, D.G. and Lavenex, P. (2006) Hippocampal neuroanatomy. In *The Hippocampus Book* (Bliss, T. *et al.*, eds), pp. 37–115, Oxford University Press
39 Cohen, N.J. and Eichenbaum, H. (1993) *Memory, Amnesia and the Hippocampal System.* MIT Press
40 Lisman, J.E. (1999) Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron* 22, 233–242
41 Kumaran, D. and Maguire, E.A. (2007) Match-mismatch processes underlie human hippocampal responses to associative novelty. *J. Neurosci.* 27, 8517–8524
42 McClelland, J.L. *et al.* (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457
43 Marr, D. (1971) Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 262, 23–81
44 McNaughton, B.L. and Morris, R.G. (1987) Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* 10, 408–415
45 Hasselmo, M.E. and Schnell, E. (1994) Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: computational modeling and brain slice physiology. *J. Neurosci.* 14, 3898–3914
46 Gray, J.A. (1982) *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System.* Clarendon Press

47 Vinogradova, O.S. (2001) Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus* 11, 578–598
48 Hannula, D.E. and Ranganath, C. (2008) Medial temporal lobe activity predicts successful relational memory binding. *J. Neurosci.* 28, 116–124
49 Preston, A.R. and Gabrieli, J.D. (2008) Dissociation between explicit memory and configural memory in the human medial temporal lobe. *Cereb. Cortex* 18, 2192–2207
50 Bakker, A. *et al.* (2008) Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* 319, 1640–1642
51 Quiroga, R.Q. *et al.* (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107
52 Leutgeb, S. *et al.* (2004) Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* 305, 1295–1298
53 Leutgeb, J.K. *et al.* (2007) Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* 315, 961–966
54 Corrado, G. and Doya, K. (2007) Understanding neural coding through the model-based analysis of decision making. *J. Neurosci.* 27, 8178–8180
55 deCharms, R.C. and Zador, A. (2000) Neural representation and the cortical code. *Annu. Rev. Neurosci.* 23, 613–647