PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2011) **366**, 596–610
doi:10.1098/rstb.2010.0188

Review

Vision and the representation of the surroundings in spatial memory

Benjamin W. Tatler[1,*] **and Michael F. Land**[2]

[1] *School of Psychology, University of Dundee, Dundee DD1 4HN, UK*
[2] *School of Life Sciences, University of Sussex, Sussex BN1 9QG, UK*

One of the paradoxes of vision is that the world as it appears to us and the image on the retina at any moment are not much like each other. The visual world seems to be extensive and continuous across time. However, the manner in which we sample the visual environment is neither extensive nor continuous. How does the brain reconcile these differences? Here, we consider existing evidence from both static and dynamic viewing paradigms together with the logical requirements of any representational scheme that would be able to support active behaviour. While static scene viewing paradigms favour extensive, but perhaps abstracted, memory representations, dynamic settings suggest sparser and task-selective representation. We suggest that in dynamic settings where movement within extended environments is required to complete a task, the combination of visual input, egocentric and allocentric representations work together to allow efficient behaviour. The egocentric model serves as a coding scheme in which actions can be planned, but also offers a potential means of providing the perceptual stability that we experience.

Keywords: egocentric; allocentric; object memory; natural tasks; position memory; saccade

1. INTRODUCTION

We use our eyes to get the information we need to perform the tasks of everyday life. Some of this information may be in the image on the retina at the time, but usually it is not. If we need to find a mug to make a cup of coffee, that mug is unlikely to be in the central retinal region where we can easily recognize it. It is more likely to be either in peripheral vision where resolution is too poor to identify it, or else it is not in the field of view at all, in which case a reorientation is needed to locate it. This simple example shows that we require, in memory, a representation of the surroundings that is adequate to act as a basis for directing foveal vision to places where it is needed.

In principle, if the brain were able to join up and remember the entire series of images provided by the retinae each time we move our eyes, then we would have a complete panoramic memory that could be used to guide future actions. However, not only is the storage capacity implied by such a proposal unrealistically immense, but there is also a great deal of empirical evidence against it. How, then, is the spatial information required by an active visual system obtained, stored, updated and made available?

In this article, we will first briefly review current thinking about what is and what is not retained each time we move our eyes. This leads directly to the nature of the representations of space that are built up while viewing a scene. Inevitably, much of what has been learned has come from studies of static images where no complex actions are involved, and where the requirements of the representations involved are less exacting than those involving action in three-dimensional space. We then consider tasks that involve the manipulation of objects in proximate space but do not involve bodily relocation. Following this, we discuss the kinds of representation needed to operate in extended spaces that require movement around the environment, for example, within a kitchen while preparing food. Here, the problem is to retain a panoramic memory whose spatial contents are updated as we rotate and translate within the environment—the 'egocentric model'.

Finally, we return to the question of why the continuous visual world we experience is so different from the temporally and spatially disjointed series of images provided by the retinae. Could it be that our subjective gaze direction is anchored not to the retinal image, but to the continuous representation we use to guide our actions?

2. WHAT IS (NOT) RETAINED ACROSS EYE MOVEMENTS?

Our subjective experience of continuous visual perception of our surroundings in spite of the temporally discontinuous and spatially restricted information supplied by the eyes inevitably poses the question: how does the brain achieve perceptual stability despite the nature of the input supplied by the eyes? This question has been asked by researchers since the saccade-and-fixate strategy of the oculomotor system was first observed [1]. Initially, it was assumed that perceptual continuity arose from continued visual sampling during saccades (e.g. [2,3]). However, this notion was

[*] Author for correspondence (b.w.tatler@dundee.ac.uk).

One contribution of 11 to a Theme Issue 'Visual stability'.

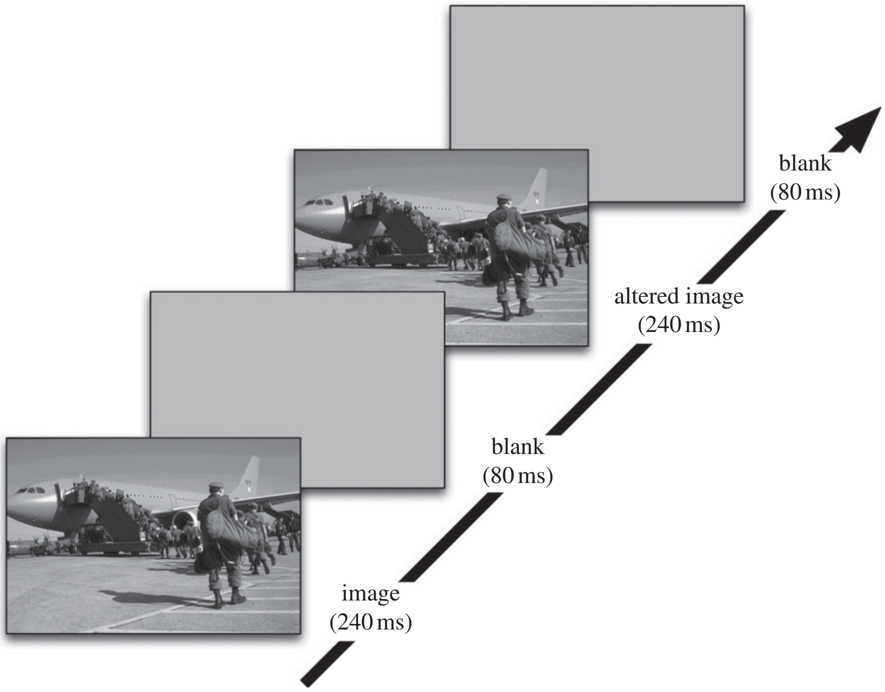

Figure 1. The 'flicker paradigm', which has been used extensively to study change detection. Here, a change is introduced to the scene during a brief interruption to viewing. The whole cycle shown above repeats until the participant detects the change. The scene images are supplied courtesy of Ron Rensink.

quickly and effectively dismissed when Erdmann & Dodge [4] happened to notice that while using a mirror to observe the eye movements of participants reading text, they were never able to see their *own* eyes moving. Thus, perception is suspended during saccades. Emphasis thereafter shifted to trying to understand how the brain might use the information sampled during fixations to construct an internal representation capable of giving rise to our complete and detailed subjective experience. For some time, it was thought that the pictorial contents of each fixation were fused to construct a point-by-point complete picture of the scene within the brain (e.g. [5]). However, an increasing body of research showed that there was little evidence to suggest that the pictorial content of fixations might be integrated [6–8]. In spite of these arguments against internal veridical representations, the notion of a complete picture in the brain survived until the advent of change-detection studies in the 1990s [9]. A considerable volume of research has now demonstrated that observers can fail to notice large changes to scenes if they are timed to coincide with saccades [10], blinks [11] or brief artificial interruptions to viewing ([12], figure 1). Change blindness was largely interpreted as providing a strong case against the notion of point-by-point pictorial representations [9]. If the pictorial content of each fixation were retained, then, it has been argued, it should be trivial to detect changes to the colour, location or even presence of an object. Following this logic, change-detection studies have been used to suggest that the pictorially veridical information from each fixation is lost every time we move our eyes.

Of course, one potential criticism of change-detection studies is that the changes that occur are typically very unlike those we may encounter in a natural context: clothes change colour, buildings move and foliage disappears instantaneously, during some brief interruption to viewing. A number of studies have therefore considered change detection in situations that are more ecologically valid. Extensions of change-detection research into dynamic scenes have suggested that detection is at least as difficult as was found for static scene viewing. Levin & Simons [13] found that changes to colour, presence, position and identity of objects during cuts between camera viewpoints were rarely detected. Even the main actor in a sequence can be changed during an editorial cut without always being noticed by the observer. Wallis & Bülthoff [14] extended this work by considering whether there were differences in detection ability depending upon the type of change made in a dynamic setting. These authors created movies of virtual environments simulating observer motion through the environment and compared this situation to static viewing of the same scenes. Wallis and Bülthoff found that the ability to detect the appearance/disappearance of an object was the same in the static and dynamic situations. However, for object orientation, position and, to a lesser degree, colour, detection performance was worse during the simulated observer motion than during static scene viewing. Further support for difficulties in detecting colour changes across cuts in movies has been provided by Angelone *et al.* [15]. Hirose *et al.* [16] have explored object memory across changes in viewpoints in movies and suggested that position information is represented differently to other object properties in such dynamic scenes. This may reflect greater difficulties associated with spatial representations across changes in viewpoint when viewing dynamic scenes, compared with representing other object properties.

In their now-classic study, Simons & Levin [17] demonstrated that if two actors changed places

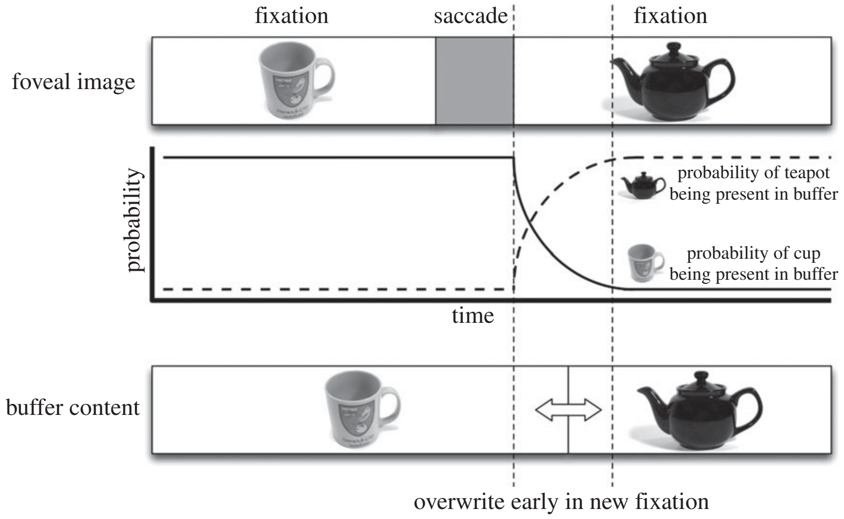
fixation
saccade
fixation
foveal image
probability
probability of teapot being present in buffer
probability of cup being present in buffer
time
buffer content
overwrite early in new fixation

Figure 2. Schematic of Tatler's proposed transient retention of visually rich information across saccades, until overwritten by the content of the new fixation. Redrawn for Land & Tatler [19].

during a brief interruption to an ongoing conversation (provided by a passing door!), people often failed to notice that they were holding a conversation with a different person from the one who they began talking to. This and the studies above have all been used to argue that much of the visual information must be lost whenever viewing is interrupted.

Further evidence for failure to retain visually rich information in natural settings has been provided by Tatler [18]. This study did not employ a change-detection paradigm, but instead tested what visual information participants could access while engaged in a natural, everyday activity. Participants were interrupted by turning out the lights in a blacked-out room while they were making a cup of tea. When this occurred, they were able to give pictorially rich descriptions of the information that was the target of their foveal vision as the lights went out. However, they were unable to describe what they had been looking at prior to this. The stark contrast in reportability between the final target of fixation and prior targets argues for transient or no retention of visually rich information. However, a common error when attempting to report the final target of fixation at the point of interruption was for participants to mistakenly report the content of their penultimate fixation; this mistaken report was given with the same degree of detail and confidence as the correct reports of final fixation contents. Moreover, the probability that this type of error occurred was related to the time between the start of the final fixation and the time that it was interrupted by the lights going out. The existence of this class of error can only be explained if rich visual information is retained across saccades and for a short time into the new fixation, until it is overwritten by the content of the new fixation (figure 2).

3. MEMORY FOR STATIC SCENES

The phenomenon of change blindness has renewed interest in the nature of scene representation and a variety of explanations of how we encode and retain information from the visual environment have been argued.

(a) *Current views of scene memory*

First, change blindness has been used to argue that we do not construct an internal representation of our visual environment at all [20]. According to this view, the most reliable source of information about our surroundings is the world itself. Because we have highly mobile eyes, we can redirect our foveae to the regions of the world we wish to scrutinize with relatively little cost and so there is no need to interrogate an internal representation rather than the world itself [21]. The absence of any internal representation raises a number of questions about how our perceptions of the world arise: for example, if we only ever have access to the current retinal image, how do we form three-dimensional perceptions of objects? Here, O'Regan & Noë [20] appeal to Gibsonian notions of sensorimotor contingencies, whereby perceptions arise from the changes that occur on the retina across eye movements: the changes depend crucially on the three-dimensional structure of objects and so can be used to reveal this structure.

Second, Rensink [22] proposed that some visual detail can survive saccades, as long as it is the subject of focal attention. Thus, a limited number of *proto-objects* can be maintained as an object representation, but all unattended visual detail is lost. The attended information can be retained as a coherent object representation only for as long as it receives focal attention. Rensink [22] further proposed that this limited attentional coherence of visual detail was integrated with more abstract and higher level information about the gist and layout of scenes (figure 3). In this view, therefore, information survives beyond the end of a fixation, but only if and while attended. The number of items that can be preserved beyond the end of a single fixation is also very limited.

Third, a number of authors have suggested that visual representations may be less sparse than suggested by Rensink. Indeed, one could argue that change blindness need not imply that representation must be sparse or absent, and that failure to detect change may be due to a number of possible reasons [23]. For example, representations may be formed

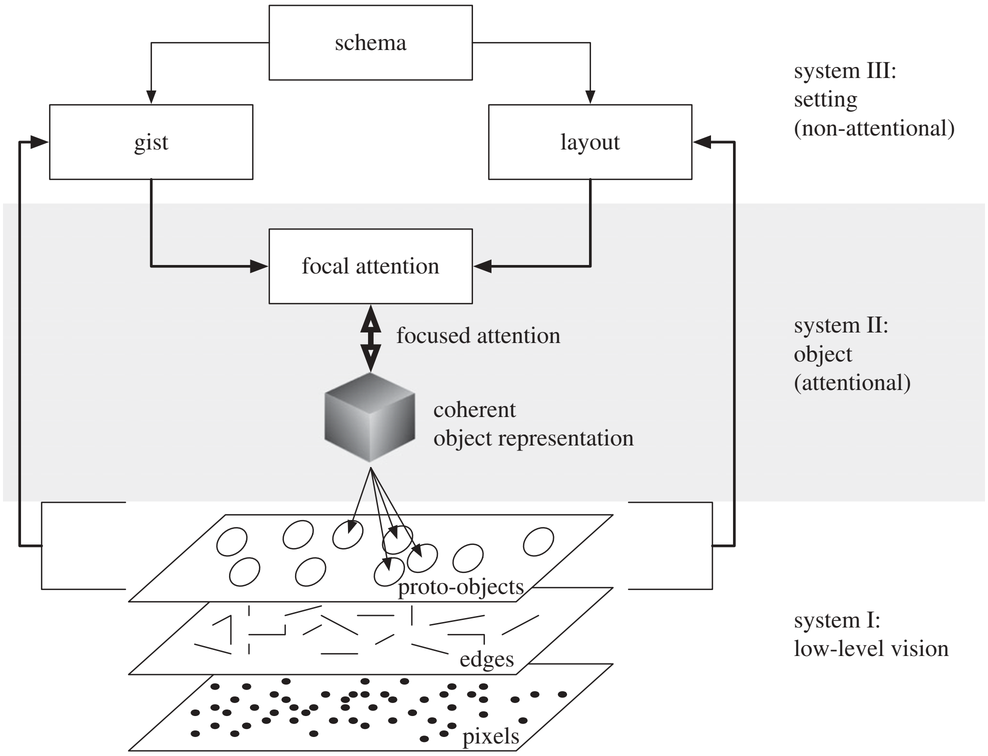

Figure 3. Rensink's proposed architecture for scene representation. Redrawn from Rensink [22] for Land & Tatler [19].

that are not accessible to conscious scrutiny and therefore cannot be used to report the change. Support for this possibility comes from studies that have shown better than chance localization of a change in a stimulus array, even when the participants report that they are unaware of any changes (e.g. [24]). Another possibility is that while point-to-point visual detail may be lost from each fixation, other, more abstract information may be retained, but may be insufficient for supporting change detection. Both of these possibilities suggest that representations are formed that survive fixations, a notion supported by a growing number of studies demonstrating that object property information appears to be extracted and retained from the scenes viewed (e.g. [25–28]). However, while there is general agreement that information survives the fixation, the nature of the retained information remains the topic of continued research and debate. One possibility is that object information may be encoded into a limited number of *object files* [29–31]. These object files are temporary representations of objects maintained across several saccades. However, this representation scheme remains quite sparse, with an upper limit of three to five object files being able to be maintained at any time. Once the upper limit is reached, new object files can only be encoded and retained at the expense of existing files.

In contrast to the sparse scheme suggested by the object file account, some authors have suggested that richly detailed visual representations are formed when viewing scenes (e.g. [32–34]) that contain a large amount of detailed visual information and can survive for extended periods of time. Hollingworth argues that because observers can detect changes to objects that are as small as a change in orientation, the representations that underlie this detection must be visually rich in order to support such subtle distinctions. A similar case for high-capacity, visually rich memory for objects has been proposed by Brady *et al.* [35]. After viewing 2500 objects for just 3 s each, observers were able to discriminate previously seen objects from paired distractors with impressive ability. This was even the case when distractors were different exemplars from the same object category or were the same object at a different orientation. The authors argue that such fine discriminations for objects drawn from such a large memorized set implies that object memory must be both rich in visually precise detail and immense in capacity.

While Hollingworth and Brady *et al.* argue for visually rich representations, other authors have interpreted essentially rather similar findings in different ways. Melcher [27,36] proposed the involvement of a higher level medium-term memory, with representations being less strictly visual and more abstracted. Tatler has also favoured a more abstract account of representation [28,37,38], but which can still include a large amount of information describing the objects in the scene. Recently, Pertzov *et al.* [39] have argued for a similar scheme of representation to that discussed by Tatler. It is hard to find empirical evidence that really favours any particular one of the interpretations suggested by Hollingworth, Melcher and Tatler and the nature of information retained from fixations remains an open question.

Common to all of these accounts is the finding that information survives beyond a single fixation and typically accumulates over prolonged viewing. While Hollingworth, Melcher and Pertzov have argued for a general increase in accumulated information over time, Tatler has suggested that different object properties are integrated into representations over different time scales. In general, during viewing, the overall scene gist and spatial layout seem to be extracted earlier than more detailed information about the properties of individual objects [28,40]. When multiple object properties were tested at the same time, Tatler *et al.* [38] found that object identity and colour did not accumulate over multiple fixations of an object, with maximal performance in response to questions testing these properties being reached after only a single fixation of the object. Conversely, position information continued to accumulate over multiple refixations of the object.

(b) *The interplay between vision and memory for static scene viewing*

If representations of the visual scenes we observe are formed, it is reasonable to assume that they might influence ongoing inspection behaviour. Certainly, there is evidence that saccade programming can involve not only immediate visual input but also remembered information. When viewing a series of isolated fixation targets, saccades can be launched to remembered locations of previously presented targets [41]. For more complex scenes, a brief preview of the scene has been shown to facilitate subsequent search for a target object [42]. Using a gaze-contingent moving window paradigm, Castelhano and Henderson showed that search time was faster following a whole-scene preview than following no preview or a preview of a different scene. This result suggests that scene information encoded during the preview period played a role in programming the saccades launched during the search epoch. Information encoded from complex scenes can also influence inspection behaviour over much longer time scales. Repeated viewings of scenes decrease search times, even when repetitions of the scene are separated by several intervening trials [43].

Oliva *et al.* [44] considered the interplay between vision and memory by presenting scenes that extended beyond the bounds of what was visible on the monitor at any one time. Panoramic virtual scenes were presented by panning a virtual camera across an extended scene such that the observer was presented with a moving image on the monitor. Scenes were shown twice: once to learn a set of objects present in the panorama, and subsequently to decide whether each of the learned objects was present or absent. The nature of the responses in the test phase was varied such that visual information at test and memory from training were differentially informative. Participants forced to rely on either visual or remembered information alone were able to complete the search task. However, when both sources of information were present, search behaviour was dominated by the immediate visual information. Taken together, these results argue that remembered information can influence ongoing gaze behaviour, but that for viewing static scenes gaze relocations are primarily under the control of immediate visual input.

(c) *Frames of reference for programming saccades*

However rich or sparse the information accumulated across saccades, the question arises as to the form in which they are stored, and in particular whether the representations are compensated for the changes in eye direction that result from each saccade. Following a saccade, an object that was in one location on the retina, or on any retinotopic representation in the brain, will now be somewhere else. This means that, if a number of saccades intervene between seeing an object and returning to it, a straightforward representation of the object's original location in retinotopic coordinates will not provide the right vector to allow a return saccade to be made.

There are three ways round this [41]. The memory representation might be kept in retinotopic coordinates, but with each intervening eye movement monitored, perhaps by efference copy signals, and summed vectorially so that when a return saccade is made, it is compensated for the intervening path of the eye. Alternatively, the representation could be stored in head-based coordinates, with object location stored as the sum of retinal location and proprioceptively monitored eye-in-head position. Retrieval in such a scheme simply involves making a saccade based on the difference between current gaze direction and object location, both in head-based coordinates. Thirdly, objects can be located with reference to exocentric cues; that is, to the positions of other objects in a scene. This requires an indexing of the identities and locations of scene landmarks in a quasi-pictorial representation that is not necessarily tied to any one physical frame of reference. Coding of remembered visual information in exocentric coordinate frames, providing a scaffold for immediate visual input, has been suggested on several other occasions [27,28,45]. Karn *et al.* [41] favour a combination of head-centred and exocentric reference frames. Others favour a spatial updating scheme based on retinal coordinates, but with mechanisms in the parietal cortex for translating this representation into other, head or body-centred, coordinate frames [46,47]. This possibility of transforming between multiple coordinate frames has been the subject of much research and is particularly important in the context of visuomotor tasks in extended environment; we will return to this issue in §5. As we shall see in §5, during active tasks, it will also be necessary to assume that there are representations of object locations that include parts of the surroundings that are outside the current visual field.

The theoretical perspectives developed by the various authors in the sections above have, in general, been derived from experimental paradigms either mostly or wholly within the realm of static scene viewing. Is it reasonable to consider whether the same representational structures and processes that have been described for static scene viewing would be found in more natural, dynamic settings. Understanding representation in the context of natural settings is important because the role of internal representation must surely be to assist us in reaching our behavioural goals (see also [19]). We are certainly not the first to raise concerns about the use of static scene viewing paradigms for eye movement research, and the need to consider more dynamic settings. For example, Henderson and colleagues have raised this concern on a number of occasions [48–50]. Hayhoe has also argued the need to study representation in the context of natural tasks and has suggested that the representational processes under such circumstances may be very different from those under static scene viewing conditions [51].

In the sections that follow, we will consider the questions of visual representation and memory in the context of tasks carried out wholly in proximal space and those that require movement through a larger environment. These two situations place potentially differing requirements on any representational system.

4. MEMORY DURING MANIPULATIONS IN PROXIMATE SPACE

An important aspect of natural visual environments is that we tend to interact with the scene rather than simply observe it. Therefore, any representational scheme that supports natural behaviour must be flexible enough to deal with how our actions influence the world. Under these circumstances, enduring memories such as those that have been suggested from the static viewing paradigms discussed above may not be useful and indeed may even interfere with efficient interaction with the world. For example, we do not want an enduring memory of the previous location of an object once we have moved it. It may therefore be that being involved in an active manipulation of the environment places different demands upon the representational processes and structures that underlie vision.

(a) *Evidence for limited moment-to-moment memory*

Ballard, Hayhoe and colleagues have used virtual reality tasks in which participants interact with objects in proximal space, in order to study eye movements and representation in the context of an active task. In a task in which participants used coloured blocks to reconstruct a visible model, the eye movement strategies employed revealed a tight coupling between vision and action [52]. In this task, each goal completion requires knowledge only of the colour of each block and the position in the model at which it should be placed. Despite such limited demands on memory, participants typically looked twice at the model that they had to copy during each cycle of selecting and placing a block: once before selecting the next block, then again before placing it in the construction area (figure 4). This result was interpreted by the authors as suggesting that the two fixations served very different purposes: the first to encode the colour of the next required block, the second to extract the information about where to place the block. Such limited information in the context of this simple naturalistic task is far more consistent with the views of scene representation expressed by O'Regan and Rensink than it is with the more extensive representational schemes discussed by Hollingworth, Melcher and Tatler. Later work from the same authors, however, has shown that memory during the block-copying task may not be as limited as they initially suggested [53]. At the start of each model-building trial, people tended to fixate the relevant block in the model area twice, as described above. However, over the course of the trial, there was a shift towards using a strategy in which the model was not returned to. Instead, after adding a block to the model, the eyes moved straight to the resource area to select the next block, and from there to the construction area to guide the placement of this new block. This latter strategy is only possible if details have been remembered from previous fixations of the model. The gradual shift towards this memory-based strategy later in the trial implies some degree of information accumulation over time.

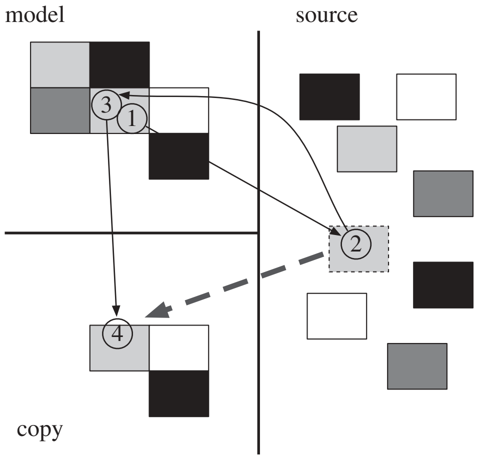

Figure 4. Ballard's block-copying task, illustrating the most common visual strategy by participants. Typically, participants fixate a block in the model (1) before fixating a block of the corresponding colour in the source area (2). Once the block is picked up and in transit towards the copy area (dashed grey arrow), a refixation of the block in the model is made (3), presumably to gather information about where to place the selected block. Finally, the location at which the block will be placed is fixated (4).

While Ballard's block-copying paradigm may not be an ideal surrogate for understanding the nature of representations that might underlie natural behaviour, it does point to the possibility that information is only encoded when it is required for the immediate task goals. The notion of only gathering and retaining information at the times when that information is required for the current behavioural goal has been explored and extended by Hayhoe and colleagues, using more semantically distinguishable objects and environments. Triesch *et al.* [54] used a virtual block-sorting task to consider the influence of introducing changes at critical times during the execution of the behaviour. In this task, blocks of two different heights were sorted by placing them on one of two conveyor belts (figure 5). Three conditions were used to vary the relevance of height information at various points in the task. In the first condition, participants were asked to pick up bricks from front to back in the virtual space and place all bricks on the nearest conveyor belt. Thus, brick height was relevant to

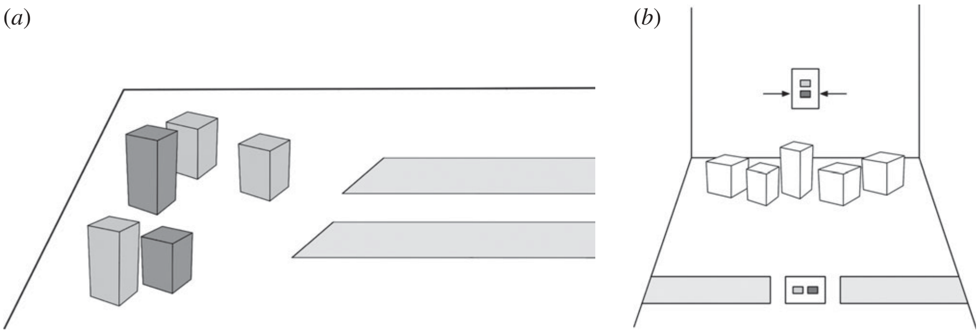

Figure 5. The block-sorting tasks employed by (*a*) Triesch *et al.* and (*b*) Droll and Hayhoe.

neither the pick-up or put-down decisions. In the second condition, participants were asked to pick up all the taller bricks first and place each on the front conveyor, and then to pick up the shorter bricks and also place them on the nearest conveyor belt. Thus, in condition 2, brick height was relevant to the pick-up but not to the put-down sections of the task. In the third condition, participants were asked to place all the taller bricks on the closer conveyor belt and then to place all the shorter bricks on the far conveyor belt. Thus, in condition 3, brick height was relevant to both the pick-up and put-down decisions. In 10 per cent of all trials, the height of the brick was changed between pick-up and put-down (i.e. while in the participant's hand). Detection of these changes increased as the height of the brick became more relevant throughout the task: 2 per cent of changes were detected in condition 1, 20 per cent in condition 2 and 45 per cent in condition 3. This result argues elegantly that whether information about the height of the brick was retained stably throughout the task depended critically on whether and when the height was relevant to the task.

The importance of task relevance through time in representations for visuomotor tasks was explored further by Droll & Hayhoe [55] in which the predictability that a cue would be relevant later in the block-sorting task was varied. In this paradigm, blocks were defined by four properties: height, width, colour and texture (figure 5). Visual cues were presented both for the pick-up and put-down decisions. These cues indicated not only which feature to use for the two decisions, but also, in the case of the put-down cue, how to use this feature to sort the blocks between two virtual conveyor belts in front of the participant. In the most predictable condition, the same single feature was used for both the pick-up and put-down decisions. Under these circumstances, refixations of the block once it had been picked up were rare, indicating a reliance on the remembered state of the feature when making the decision about which conveyor belt to place it on. These authors also used an unpredictable condition in which a single cue was used to pick up blocks, but any one of the four features could be selected at random for the put-down cue. In this condition, refixations of the block after it had been picked up, but before it was placed on a conveyor belt were common. This was true even when the put-down cue was the same as the pick-up cue (which occurred in 25% of trials). This result implies that if it can be predicted that information will be required at a later stage of the task, it can be retained stably until needed. However, if it is not predictable that the information will be needed again, that property is not retained and the eyes are used to gather information as and when it is needed for the task.

At this stage, we should compare this seemingly limited and selective scheme of representation derived from visuomotor tasks performed in proximate space with the more comprehensive and detailed representational schemes described in the context of static scene viewing. Certainly, when viewing static scenes (even static real environments in the case of [38]), there is a large volume of evidence to suggest that much object information can be retained stably throughout viewing and recalled when tested after the trial [27,28,56]. Not only can apparently detailed representations be found for static scene viewing paradigms, but also there is compelling evidence that information is encoded incidentally, and does not require that objects be the target of active memorization [57]. Castelhano & Henderson compared visual memory for a memorization task and a visual search task. These authors found that memory performance was still good for objects in the search task, where there had been no expectation that the object information would be required later as the memory test was unexpected. All of these results from static scene viewing paradigms are very different from those discussed above for visuomotor tasks. Why should such a selective and task-dependent representation be found in active settings, when it is possible to encode and remember much more? A number of possible explanations can be suggested here. First, it may simply be that when engaged in a visuomotor task we employ a principle of efficiency, expending resources only upon maintaining representations of information that are necessary and only for the times at which they are required. Second, it may be that the dynamic nature of the scene places constraints upon representations that are not present when viewing a static scene. For example, memory for object positions in a dynamic setting presents a very different set of problems from memory for position in a static scene. In the latter, a single index of the target with respect to the scene will suffice. However, for a dynamic situation, position information must encompass movements by the observer, movements of elements in the environment and changes in object locations as a result of active manipulations by the observer. These additional demands of encoding information in a dynamic setting may require additional or even alternative representational schemes to those employed when encoding static pictures.

One possible limitation of the dynamic tasks that have been discussed so far is that the apparent sparsity of representation may be confounded by the semantic (and visual) similarities of the objects used in the paradigms. Coloured or textured blocks used on repeated trials may result not only in difficulties for maintaining distinct representations of the component objects in a task, but may also result in interference between trials. Such an interference effect for semantically similar displays has been found in the context of static scene viewing [36]. It will therefore be important in the sections that follow to draw widely from active task settings to include more natural tasks with semantically distinguishable objects.

(b) *Evidence for temporally extended spatial memory*

The tasks discussed in §4*a* suggest a scheme of representation, where much of the available information is only encoded when needed and only retained if required later in the task. However, evidence from similar tasks carried out in proximate space has provided evidence for a slightly different form of temporal extension to representations, involving preemptive information-gathering. When washing one's hands [58] or making a sandwich [59], fixations are occasionally made to objects that are not involved in the current part of the task but will be the focus of an upcoming act. These 'look-ahead' fixations may occur several seconds before that object is used but have a measurable benefit for the efficiency with which the target is later re-acquired [60].

In order for these look-ahead fixations to have a measurable behavioural consequence, some information gathered during these fixations must persist long enough for it to aid the later location of the object. Further evidence for the functional significance of looking ahead comes from the observations that objects that were the focus of completed portions of the task are never looked back to [58]. Thus, the look-ahead fixations are not incidental looks, but are likely to be functional. For natural task settings, it is hard to estimate exactly what information about an object might be extracted during these look-ahead fixations. However, this must minimally be some spatial information about where the object is, in order to produce the observed differences in how quickly and how accurately the object is fixated later in the task when it is the target of the current act. These studies show that information-gathering can be proactive, seeking out information that will be needed in the near future and retaining this information until it is required.

(c) *The balance between vision and memory in peripheral vision*

In §3*b*, we discussed studies which demonstrated that saccades can be programmed on the basis of immediate visual input or memory depending upon the relative availability of these two sources of information. Here, we consider the relative roles of vision and memory when targets of visuomotor tasks are present within peripheral vision. It is usually assumed that we use vision to locate objects that are within our peripheral field of view, but there is evidence that this is not always the case, and that memory may be equally important even for objects that are plainly visible. The resolution of peripheral vision is poor, and the angle that an object must subtend to be identified increases dramatically and approximately linearly with its angular distance from the fovea [61]. Very approximately, if a letter 0.1° high can be identified in the fovea, it needs to be 1° high at an eccentricity of 10°, and 6° at 60°. Aivar *et al.* [62] used a variant of Ballard's block-copying paradigm to consider whether saccades to blocks in peripheral vision are guided primarily by vision or memory. In this task, the layout of the blocks in the resource area (figure 4) was changed when the participants looked away from this area. Provided the participants had sufficient time to familiarize themselves with the layout of the environment before the first change was made, saccades to the resource area were launched to the remembered locations of blocks rather than to the actual post-change locations. This is in spite of the fact that the resource area was within peripheral vision at the time that these saccades were launched. Aivar's result clearly implicates an important role for memory in planning saccades to targets in peripheral vision.

Brouwer & Knill [63] used a virtual visually guided reaching task to consider the relative use of vision and memory for guiding action. In this task, two virtual objects had to be picked up and placed in a trash bin. In some trials, the position of the second target was moved by a small amount while the first was being moved to the trash. While participants never noticed this, it did have a noticeable influence on their behaviour. Essentially, this perturbation means that vision and memory were in conflict when the arm movement towards the second target was executed. Brouwer and Knill found that both vision and memory played a role in the targeting decision, but the relative weighting of vision and memory in planning the reach to the second target depended critically on the visibility of the second target. Targeting high-contrast objects involved a greater reliance on visual information than did targeting low-contrast objects, where remembered position was relied upon more. From this, it seems that the targeting system uses a blend of what is in immediate vision, and what is available from the current representation of the surroundings.

5. SPATIAL MEMORY DURING ACTIVE TASKS REQUIRING MOVEMENT

In the above section, we have argued that when engaged in an active task performed within proximate space, the representational scheme that underlies such behaviour appears to be rather limited. However, the demands placed upon memory and representation in such settings are in some ways rather reduced compared with the potential requirements for representation in situations where task completion involves movement through an extended environment. For example, when preparing food in a kitchen we need to know not only about the work surface we are facing, but also about those on either side or behind us. One central problem that any representational scheme must solve in a real-world setting, which is not present in static scenes, is the spatial reference frame in which we must represent our surroundings.

(a) *Spatial organization in natural settings*

When comparing spatial organization in pictures and the real world, two issues are immediately apparent. First, the scales at which spatial information occurs are very different. Second, pictures cannot readily be used to distinguish between the range of frames of reference in which space may be coded in natural settings.

Montello [64] has suggested that we can classify space into four different categories. Figural spaces are smaller than the body and include objects and pictures. Vista spaces are larger than this but only encompass what can be seen by an observer from a single viewpoint. Environmental spaces go beyond what a single observer can see from a single vantage point, but are bounded by what a human can reasonable explore on foot. Geographical space is that beyond the exploration capabilities of a single individual. In natural tasks, at least figural and vista space will be important, but in many cases, understanding our surroundings in environmental space is also important (such as understanding where different rooms are in our house, or shops in other parts of town). Pictures are therefore problematic in two ways. First, they cannot encompass the larger scales of spatial information that we need to understand in natural settings. Second, they compress vista space into figural space: a whole vista is presented within the bound of a picture, which itself occupies figural space.

Given the different levels at which space can be represented, one question that arises is whether we have entirely separate representations for each of these levels, or whether there is cross-talk between them. That is, does the representation of our current vista (e.g. a room) interact with our representation of the environmental space outside the room? Hirtle & Jonides [65] used a variety of methods to test participants' recall for a real extended environment, and concluded that levels of representation were nested hierarchically. In contrast, Brockmole & Wang [66,67] found no evidence for cross-talk between vista representation and the representation of environmental space.

Insights into the frames of reference in which we might encode and retain information about scenes can be made by considering the effect of changes in viewpoint when viewing static or dynamic scenes. Simons & Wang [68] used an array of objects on a tabletop to explore the influences of changing viewing position (by walking between two viewing positions) and retinal projection (by rotating the tabletop) upon change detection. Changing the retinal projection between views by rotating the table resulted in poorer change detection. However, changing the retinal projection by the same amount by asking the observer to walk to a new viewing location did not have any detrimental effect on change detection. The importance of generating the movement between viewpoints was demonstrated by sitting participants in a wheeled chair and wheeling them (with eyes closed) to the new location. This manipulation resulted in poorer change-detection performance. Wang & Simons [69] conducted a series of follow-up experiments to reinforce the suggestion that viewers can update representations across active changes in viewpoint, but not across passive changes in viewpoint. For dynamic movie sequences, the ability to encode information across viewpoints is unclear. Garsoffky *et al.* [70] found recognition accuracy to be higher when scene memory was tested using the same viewpoint as experienced by the viewer when watching a movie sequence than when the viewpoint at test did not match that at encoding. This result is consistent with a viewpoint-dependent representation. However, Garsoffky *et al.* [71] showed no such cost of viewpoint change when recognizing computer-animated basketball scenes, consistent with a viewpoint-independent representation. While the evidence suggests that active exploration of the world is essential for being able to integrate information across viewpoints, the coding scheme in which the information is represented remains unclear from these studies. However, the importance of active exploration and the ability to use changes in viewpoint both imply that spatial coding of objects can occur in coordinate frames that are neither retinocentric nor wholly exocentric.

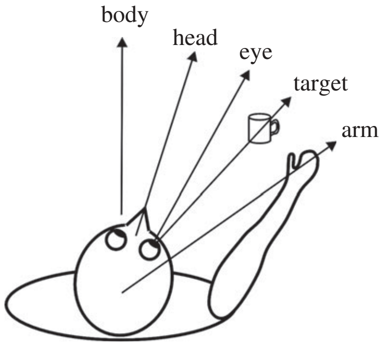

Figure 6. Frames of reference for visuomotor tasks. The required movement to grasp the mug is the angle from arm to target. This is the angle from body-to-arm minus the sum of the angles from target-to-fovea, eye-in-head and head-on-body. In practice, eye, head and body are often aligned before such a grasp movement, but such alignment is not essential.

(b) *Transformations between frames of reference*

The coordinate frame in which space may be represented in the brain has been the topic of much research [46,72–74]. It is clear that muscular movement plans must ultimately be coded in limb-centred coordinates. Similarly, visual information must initially be coded in retinotopic space. Indeed, it is clear that in the context of natural behaviour, a range of different spatial coding schemes must be involved and must act in parallel (figure 6). However, it seems likely

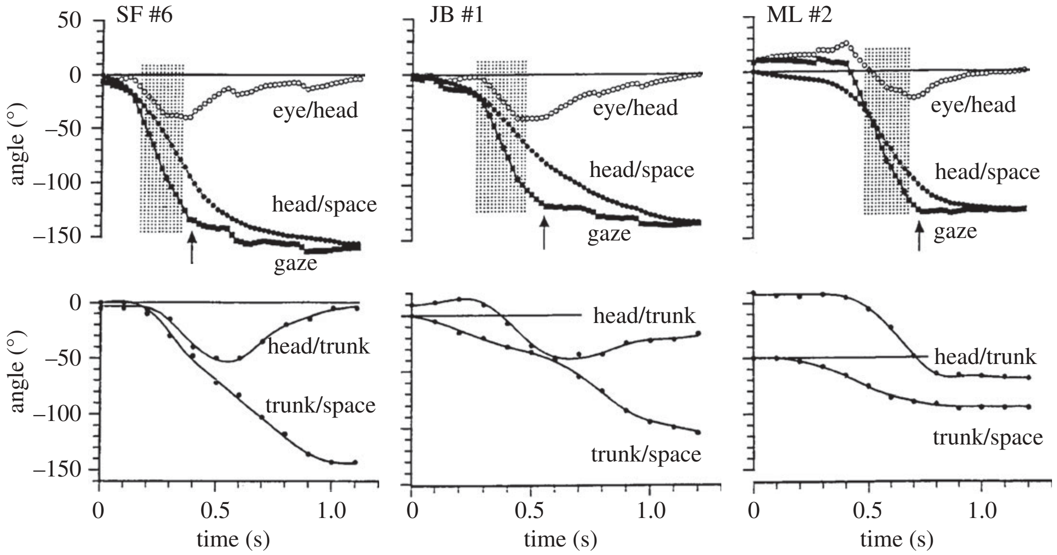

Figure 7. Details of three large gaze saccades (100–150°) made during turns from one work surface to another while making tea in a kitchen. Upper plots show the rotations of the eyes in the head, the head in space and the sum of the two, i.e. the gaze rotation. Lower plots show the somewhat variable contributions of the trunk and the neck (head/trunk). Arrows indicate the onset of the vestibulo-ocular reflex, and the end of the initial gaze saccade. The hatched areas indicate blinks, when the pupil was not visible and thus vision was not occurring.

that efficient coordination of multi-sensory input and motor output must involve transformation between the various parallel frames of reference for spatial coding. Converging evidence suggests that such transformations are possible and that the parietal cortex is crucially implicated in multimodal spatial organization.

One question is whether the parietal cortex simply handles the transformations between multiple frames of reference or combines across representations to form a master representation of space. Chang *et al.* [75] found evidence for parietal transformation between eye- and hand-centred representations, consistent with a single representation of eye–hand distance. Whether the parietal cortex forms a master map or simply handles the transformation between representational frames of reference, it is clear that efficient behaviour relies upon the integration of information coded in a range of different frames of reference. It should also be noted that frames of reference for sensory processing and motor responses are all in some way centred on the individual rather than in exocentric coordinates.

(c) *Spatial memory in natural tasks*

In natural tasks, we often find gaze relocations to objects that are currently outside our field of view. In a study of tea making, Land and colleagues found that gaze changes of up to 180° were often made to objects on other surfaces in the kitchen [76]. Sometimes these were made with a series of saccades, but frequently they were completed with a single saccade that involved combined rotations of the eyes, head and trunk. Importantly, the movement of the gaze was continuous until the target was reached (figure 7).

Most of these gaze shifts were accompanied by a long blink, so that vision would have been impossible for most of the movement. This means that the complete gaze movement must have been pre-planned. Typically, these gaze saccades were off-target by about 10°, and were followed 200–300 ms later by a second small saccade that brought the fovea onto the target (figure 8). All this suggests that the system that allows gaze to target unseen objects has access to the same transient egocentric representation of the surroundings that makes it possible to locate, and point to, objects in the world around us when they are not currently visible. The resolution of this representation is not good enough to allow exact targeting, but it seems that it is sufficient to bring the foveae close enough to the target to allow a second saccade to be made under visual control.

(d) *Allocentric maps and egocentric models*

Recent accounts of the way we encode information about objects, places and routes in the world around us propose that we have two kinds of spatial representation: allocentric and egocentric (e.g. [73,77]). While we highlighted the established notion that there are a variety of egocentric coding schemes in §5*b*, our primary concern here is to argue for the utility of egocentric coding for spatial representation in natural tasks rather than to determine which of the possible ego centres is at the heart of this coding scheme.

The allocentric representation is map-like (figure 9*a*). It is indexed to a world-based coordinate system, is independent of our current location and heading and survives over extended periods of time. This representation must of course be built up from vision over time, but does not rely on immediate visual input. Longer term memories of the present or similar environments are integrated into this representation.

When walking in a natural environment, there is evidence for the storage of information about objects encountered on the route, which is consistent with

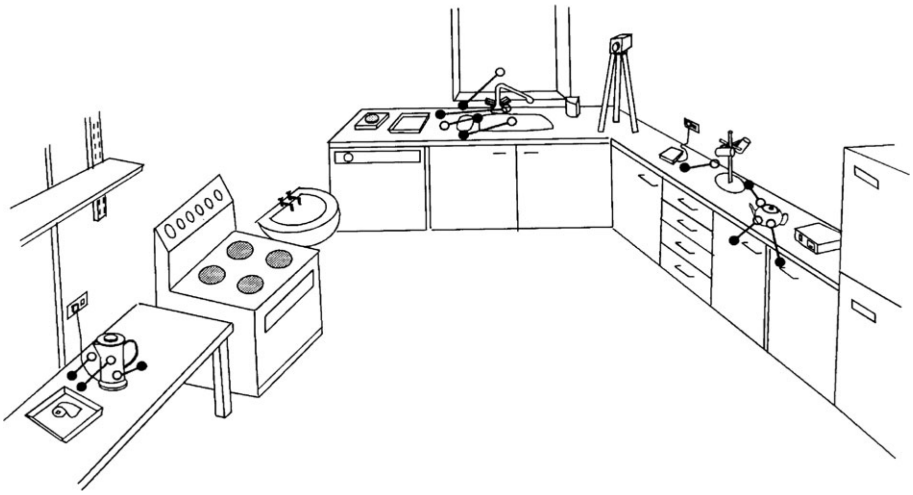

Figure 8. Landing positions of large single saccades (greater than 90°) made by one subject (J.B.) to kitchen objects while making a cup of tea. Black dots show the initial landing positions and open dots the final positions after a secondary correction saccade. The average size of the secondary saccades was 8°.

an updating of a world-based allocentric representation. Droll & Eckstein [78] asked participants to walk a course around a building eight times. A variety of objects were arranged close to the path that the participants walked. While the participants walked this course, changes were made to nine of the objects located near the path. These changes were made to objects while they were out of sight for the participants. When simply instructed to walk the route, participants were very unlikely to detect these object changes (5% detection). However, when asked to prepare for an object memory test that would follow the experiment, participants were far more likely to detect the change (32% detected) and also spent longer looking at individual objects. This result is consistent with the notion of encoding information in world-centred coordinates and also suggests that, like the studies discussed in §4, such representations are selective and task-based.

The other kind of spatial representation, the egocentric representation, is temporary, and based on the directions of objects relative to our current body position and heading in the space around us (figure 9*b*). It is this second representational frame that allows us to act upon our environment, for the purposes of locating, reaching for and manipulating objects. We can think of the egocentric model as containing low-resolution information about the identities and locations of objects throughout the 360° space around us. It is available for making targeted movements of gaze or arm irrespective of whether or not it is supplemented by direct visual information. Of course, the view we see is not the same thing as the egocentric model. The seen world has detail, colour and movement, none of which is an obvious property of the model. We are conscious of what is in the field of view, particularly the central region around the fovea, to a much greater degree than we are of objects outside the visible part of our surroundings. Nevertheless, in familiar places, we are aware of what is outside the current field of view and are able point to or make saccades to unseen objects with reasonable accuracy (figure 8). The egocentric model can be updated from the allocentric map by a process akin to map-reading: finding one's location on the map and matching one's current heading to it. It can also be refreshed by direct visual input, adding or correcting the locations of particular features.

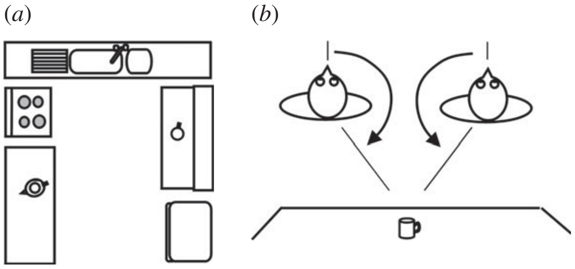

Figure 9. (*a*) Allocentric representation of a kitchen. This is independent of location and viewpoint. (*b*) Egocentric representation showing that the action required to reach the mug depends on the relation of the mug to the actor in egocentric space.

Although authors differ in the emphasis placed on each kind of representation in this dual scheme [73,74], the idea of a combination of an enduring and a temporary store generally accords well with people's intuition of how they operate in space. There is now a great deal of evidence for the existence of both kinds of representations in the brain, with the allocentric map located in the hippocampus and the medial temporal lobe, the egocentric model in the parietal lobe and translations from one to the other occurring in the retrosplenial cortex [73]. A number of lines of evidence favour the precuneus on the medial face of the parietal lobe as a likely location of the egocentric model (e.g. [79]).

To be of continuing use, the egocentric model must always be oriented so that it is aligned with the current field of view. Thus, as we move through our environment, the model must be constantly updated and rotated to match our body rotations. Rotations of the body must be accompanied by corresponding rotations of the egocentric model in order for this model to serve the planning and execution of the *next* motor command.

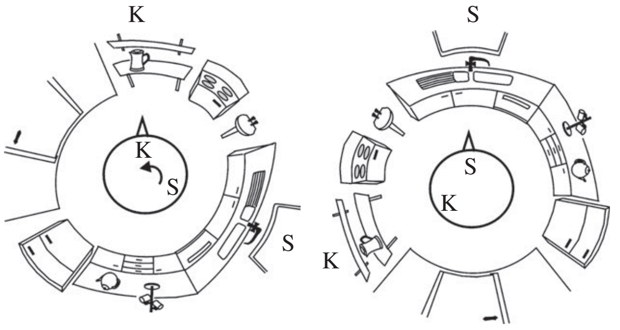

Figure 10. Panoramic view of the same kitchen as in figure 8, showing the locations of the kettle (K) and the sink (S) with respect to the viewer's centre of gaze, before and after a turn from one to the other. The viewer's egocentric model must rotate by an angle equal and opposite to the rotation of gaze, if the objects in the room are to remain in the same remembered relationships to the trunk.

If gaze is rotated 110° clockwise, the egocentric model must rotate 110° anti-clockwise (figure 10).

Although the prospect of a model of the world rotating in the brain seems alarming, there is a precedent. Duhamel *et al.* [80] found that the cells in the lateral intraparietal (LIP) area 'remap' the locations of stimuli when the eyes move. The receptive fields of the whole array of LIP neurons shift in such a way that the new target becomes the centre of the array about 80 ms before the saccade begins. We are proposing a similar 'software' transformation here for the egocentric model.

In order to consider the consequences of head and body rotation for updating the egocentric model, it is worth returning to the issue of what should be the centre of the egocentric model we describe. This model could be centred around gaze, the head or the body, and the consequences of rotations of each of these components would be different when updating the egocentric map depending upon which forms the ego-centre. For example, if the head were rotated without body rotation, the consequences for the egocentric model would be very different if it is coded in head- or body-centred coordinates. For the purposes of the present discussion, we wish to remain somewhat agnostic about the frame of reference for the centre of the egocentric model. This is in part because, for much of the time that we are involved in natural tasks, there is a co-alignment of at least the head and the body, and often of gaze too: we tend to move such that we bring eyes, head and body in line with the target of the current manipulation and it is only in the transitions between each manipulation that these three components become unaligned. The consequence of this is that when the plan is made to move on to the next object, it is usually planned and initiated at a time when gaze, head and body are all in close alignment. Even if this were not so, the ability of the parietal cortex to translate between different egocentric reference frames, as emphasized by Colby & Goldberg [46], makes it difficult to design a test that would distinguish between them. The fact that the egocentric model must itself rotate during movement (figure 10), and that the head has its own rotation sensor in the vestibular system, perhaps argues in favour of a primary role for a head-centred representation.

To illustrate the way vision and the egocentric map interact, let us consider the example of intending to pick up a mug, situated somewhere within or outside the range of peripheral vision (target T in figure 11). Its location can be obtained, at least roughly, from the egocentric model. But to grasp the mug, more details are required. Before picking up the mug, accurate coordinates of its location relative to the body are required, together with the direction in which the handle is pointing, so that the hand can be preshaped accordingly. For this level of detail, foveal vision is required, and so a gaze shift is needed to bring the fovea to bear on the mug. The gaze-directing system then consults the egocentric model about the likely whereabouts of the mug, and directs the foveae of the eyes to it. (This may involve body and head movements as well as eye movements.) After a gaze movement based on coordinates from the model, and in many cases a further eye movement based on its seen retinotopic location, gaze is brought as close to the target as it needs to be. Having acquired the target, the visual system is now in a position to supply the motor system of the limbs with the information needed to formulate the required action. Further actions may ensue. Filling a kettle, finding coffee or a tea-bag, pouring hot water into the mug, then milk and so on. Each action requires one or more foveal fixations to provide new information, but the gaze movement system also needs to refer back to the egocentric model to find the locations of new objects as they are required.

This dual scheme of representation, in which the egocentric map is updated on the basis of both sensory input and reading from the allocentric map, offers an efficient coding scheme in which our action plans can be executed within a space constructed from sensory and remembered information. Such a dual scheme allows for the potential of varying our reliance on the two types of information depending upon the relative reliability and availability of these types of information. Moreover, the relative reliance upon vision and memory can impact upon our moment-to-moment behaviour in two ways. First, and as described here, we may rely more upon remembered information from our allocentric map when constructing the egocentric map. Second, we may vary the reliance upon immediate retinocentric visual information and the egocentric map depending upon the reliability and availability of visual information [60]. Within the visual field itself, the balance between information from vision and memory is likely to vary with eccentricity, with vision dominating in the region around the fovea and being increasingly supplemented by egocentric memory towards the periphery. The topography of this balance has yet to be explored.

6. STABILITY OF THE VISUAL WORLD REVISITED

Unlike the temporally and spatially disjointed series of images provided by the retinae, our phenomenal visual world is seamless and stable. Many possible explanations have been put forward for this stability,

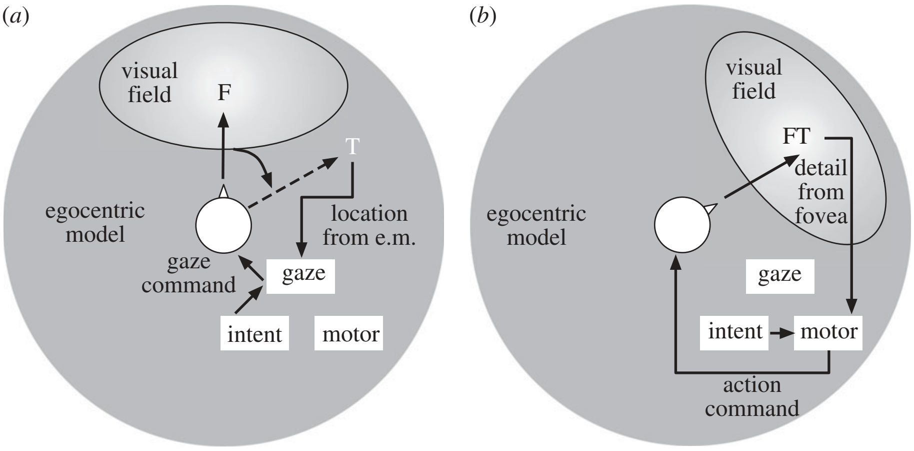

Figure 11. Planning to locate and reach for a target (T). (*a*) The interplay between vision (oval centred around the fovea, F) and egocentric representation (grey background centred around the head). In this example, we consider the situation where the observer intends to reach for a target (T) that is outside the field of view and not currently foveated. First, a gaze shift is planned to bring the fovea to bear upon the target. This gaze shift is planned using information from the egocentric model, which itself is furnished by information from ambient vision in the past and from the allocentric representation. (*b*) The situation after the gaze shift to the new target (T). As gaze shifts clockwise from F to T, so vision is re-centred around T and the egocentric map in the head is rotated anti-clockwise to re-centre around T. The manual reach can now be executed using motor commands planned using information provided by the fovea.

including the spatial updating of receptive fields by efference copy that occurs in parietal area LIP and other regions of the cortex [47,80], the fact that much of the pictorial content of the image is discarded with each saccade [9] lessening the need to integrate successive fixations or the possibility that we simply ignore the discrepancies because we know the world to be stable [81]. Here, we introduce a further suggestion, namely that the current direction of gaze is anchored not only to the instantaneous visual scene, but also to the current egocentric model.

The attraction of this idea is that the egocentric model provides a continuous panoramic layout, so that moving the direction of regard around it need not involve the kinds of visual dislocation presented by the behaviour of the retinal image itself. Since the egocentric model rotates as gaze rotates (figure 10), objects within the model retain their spatial relationships with the external world as we look around it. Thus, if our conscious readout of the layout of the world derives from the egocentric model, this layout will stay still as gaze rotates, and this is indeed the way the world seems to us. We are not suggesting that the model is a substitute for the detailed pictorial information contained in the visual input itself, but rather that it acts as the reference frame to which gaze changes are indexed. It is true that the resolution of the egocentric model is low, but then so is our ability to detect displacements during a saccade [6]. The indexing only has to be good enough to paper over the cracks.

This suggestion may also help to explain why, when looking around a familiar scene, we feel no discontinuity when making saccades into the regions beyond our current field of view. There are no surprises because we already have an outline of what is to be found there. Indeed, as the observations of Brouwer & Knill [63] make clear, the egocentric model overlaps and interacts with the visual input, and as figure 8 demonstrates, it can provide location information on its own when objects are out of sight. Thus, the egocentric model can provide a geometrical base within which the locations of objects can be stored temporarily, before being passed on to more permanent allocentric memory.

We thank the editor and two anonymous reviewers for their helpful suggestions on the previous version of this manuscript. This work is supported by a research grant awarded to B.W.T. by The Leverhulme Trust (F/00143/O) and a grant awarded to M.F.L. from the Gatsby Foundation.

REFERENCES

1 Wade, N. J. & Tatler, B. W. 2005 *The moving tablet of the eye: the origins of modern eye movement research.* Oxford, UK: Oxford University Press.

2 Cattell, J. M. 1900 On relations of time and space in vision. *Psychol. Rev.* **7**, 325–343. (doi:10.1037/h0065432)

3 Javal, L. É 1878 Essai sur la physiologie de la lecture. *Annales d'Oculistique* **80**, 240–274.

4 Erdmann, B. & Dodge, R. 1898 *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage.* Halle, Germany: Niemeyer.

5 McConkie, G. W. & Rayner, K. 1976 Identifying the span of the effective stimulus in reading: literature review and theories of reading. In *Theoretical models and processes of reading* (eds H. Singer & R. B. Ruddell), pp. 137–162. Newark, NJ: International Reading Association.

6 Bridgeman, B., Hendry, D. & Stark, L. 1975 Failure to detect displacement of the visual world during saccadic eye movements. *Vis. Res.* **15**, 719–722. (doi:10.1016/0042-6989(75)90290-4)

7 McConkie, G. W. & Zola, D. 1979 Is visual information integrated across successive fixations in reading? *Percept. Psychophys.* **25**, 221–224.

8 O'Regan, J. K. & Lévy-Schoen, A. 1983 Integrating visual information from successive fixations: does trans-saccadic fusion exist?. *Vis. Res.* **23**, 765–768. (doi:10.1016/0042-6989(83)90198-0)

9 Rensink, R. A. 2002 Change detection. *Annu. Rev. Psychol.* **53**, 245–277. (doi:10.1146/annurev.psych.53.100901.135125)

10 Grimes, J. 1996 On the failure to detect changes in scenes across saccades. In *Perception: Vancouver studies in cognitive science*, vol. 2 (ed. K. Atkins), pp. 89–110. New York, NY: Oxford University Press.

11 O'Regan, J. K., Deubel, H., Clark, J. J. & Rensink, R. A. 2000 Picture changes during blinks: looking without seeing and seeing without looking. *Vis. Cogn.* **7**, 191–211.

12 Rensink, R. A., O'Regan, J. K. & Clark, J. J. 1997 To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* **8**, 368–373. (doi:10.1111/j.1467-9280.1997.tb00427.x)

13 Levin, D. T. & Simons, D. J. 1997 Failure to detect changes to attended objects in motion pictures. *Psychon. Bull. Rev.* **4**, 501–506.

14 Wallis, G. & Bülthoff, H. 2000 What's scene and not seen: influences of movement and task upon what we see. *Vis. Cogn.* **7**, 175–190. (doi:10.1080/135062800394757)

15 Angelone, B. L., Levin, D. T. & Simons, D. J. 2003 The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception* **32**, 947–962. (doi:10.1068/p5079)

16 Hirose, Y., Kennedy, A. & Tatler, B. W. 2010 Perception and memory across viewpoint changes in moving images. *J. Vis.* **10**, 2.1–19. (doi:10.1167/10.4.2)

17 Simons, D. J. & Levin, D. T. 1997 Change blindness. *Trends Cogn. Sci.* **1**, 261–267. (doi:10.1016/S1364-6613(97)01080-2)

18 Tatler, B. W. 2001 Characterising the visual buffer: real-world evidence for overwriting early in each fixation. *Perception* **30**, 993–1006. (doi:10.1068/p3121)

19 Land, M. F. & Tatler, B. W. 2009 *Looking and acting: vision in natural behaviour*, Oxford, UK: Oxford University Press.

20 O'Regan, J. K. & Noë, A. 2001 A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* **24**, 939–973 (discussion 973–1031). (doi:10.1017/S0140525X01000115)

21 O'Regan, J. K. 1992 Solving the real mysteries of visual-perception: the world as an outside memory. *Can. J. Psychol.* **46**, 461–488. (doi:10.1037/h0084327)

22 Rensink, R. A. 2000 The dynamic representation of scenes. *Vis. Cogn.* **7**, 17–42. (doi:10.1080/135062800394667)

23 Simons, D. J. & Rensink, R. A. 2005 Change blindness: past, present, and future. *Trends Cogn. Sci.* **9**, 16–20. (doi:10.1016/j.tics.2004.11.006)

24 Fernandez-Duque, D. & Thornton, I. M. 2000 Change detection without awareness: do explicit reports underestimate the representation of change in the visual system? *Vis. Cogn.* **7**, 323–344. (doi:10.1080/135062800394838)

25 Hollingworth, A. & Henderson, J. M. 2002 Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 113–136. (doi:10.1037/0096-1523.28.1.113)

26 Irwin, D. E. & Zelinsky, G. J. 2002 Eye movements and scene perception: memory for things observed. *Percept. Psychophys.* **64**, 882–895.

27 Melcher, D. 2006 Accumulation and persistence of memory for natural scenes. *J. Vis.* **6**, 8–17.

28 Tatler, B. W., Gilchrist, I. D. & Rusted, J. 2003 The time course of abstract visual representation. *Perception* **32**, 579–592. (doi:10.1068/p3396)

29 Irwin, D. E. 1992 Visual memory within and across fixations. In *Eye movements and visual cognition: scene perception and reading* (ed. K. Rayner), pp. 146–165. New York, NY: Springer-Verlag.

30 Irwin, D. E. & Andrews, R. 1996 Integration and accumulation of information across saccadic eye movements. In *Attention and performance XVI: information integration in perception and communication* (eds T. Inui & J. L. McClelland), pp. 125–155. Cambridge, MA: MIT Press.

31 Kahneman, D. & Treisman, A. 1984 Changing views of attention and automaticity. In *Varieties of attention* (eds R. Parasuraman & D. R. Davies), pp. 29–61. New York, NY: Academic Press.

32 Hollingworth, A. 2004 Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 519–537. (doi:10.1037/0096-1523.30.3.519)

33 Hollingworth, A. 2005 The relationship between online visual representation of a scene and long-term scene memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 396–411. (doi:10.1037/0278-7393.31.3.396)

34 Hollingworth, A. 2007 Object-position binding in visual memory for natural scenes and object arrays. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 31–47. (doi:10.1037/0096-1523.33.1.31)

35 Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. 2008 Visual long-term memory has a massive storage capacity for object details. *Proc. Natl Acad. Sci. USA* **105**, 14 325–14 329. (doi:10.1073/pnas.0803390105)

36 Melcher, D. 2001 Persistence of visual memory for scenes. *Nature* **412**, 401. (doi:10.1038/35086646)

37 Tatler, B. W. & Melcher, D. 2007 Pictures in mind: initial encoding of object properties varies with the realism of the scene stimulus. *Perception* **36**, 1715–1729. (doi:10.1068/p5592)

38 Tatler, B. W., Gilchrist, I. D. & Land, M. F. 2005 Visual memory for objects in natural scenes: from fixation to object files. *Q. J. Exp. Psychol.* **58A**, 931–960.

39 Pertzov, Y., Avidan, G. & Zohary, E. 2009 Accumulation of visual information across multiple fixations. *J. Vis.* **9**, 2, 1–12.

40 Aginsky, V. & Tarr, M. J. 2000 How are different properties of a scene encoded in visual memory? *Vis. Cogn.* **7**, 147–162. (doi:10.1080/135062800394739)

41 Karn, K. S., Møller, P. & Hayhoe, M. M. 1997 Reference frames in saccadic targeting. *Exp. Brain Res.* **115**, 267–282. (doi:10.1007/PL00005696)

42 Castelhano, M. S. & Henderson, J. M. 2007 Initial scene representations facilitate eye movement guidance in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 753–763. (doi:10.1037/0096-1523.33.4.753)

43 Brockmole, J. R., Castelhano, M. S. & Henderson, J. M. 2006 Contextual cueing in naturalistic scenes: global and local contexts. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 699–706. (doi:10.1037/0278-7393.32.4.699)

44 Oliva, A., Wolfe, J. M. & Arsenio, H. C. 2004 Panoramic search: the interaction of memory and vision in search through a familiar scene. *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 1132–1146. (doi:10.1037/0096-1523.30.6.1132)

45 Hochberg, J. 1968 In the mind's eye. In *Contemporary theory and research in visual perception* (ed. R. N. Haber), pp. 309–331. New York, NY: Holt.

46 Colby, C. L. & Goldberg, M. E. 1999 Space and attention in parietal cortex. *Annu. Rev. Neurosci.* **22**, 319–349. (doi:10.1146/annurev.neuro.22.1.319)

47 Merriam, E. P. & Colby, C. L. 2005 Active vision in parietal and extrastriate cortex. *Neuroscientist* **11**, 484–493. (doi:10.1177/1073858405276871)

48 Henderson, J. M. 2003 Human gaze control during real-world scene perception. *Trends Cogn. Sci.* **7**, 498–504. (doi:10.1016/j.tics.2003.09.006)

49 Henderson, J. M. 2006 Eye movements. In *Methods in mind* (eds C. Senior, T. Russell & M. Gazzaniga), pp. 171–191. Cambridge, MA: MIT Press.

50 Henderson, J. M. & Hollingworth, A. 1998 Eye movements during scene viewing: an overview. In *Eye guidance while reading and while watching dynamic scenes* (ed. G. Underwood), pp. 269–293. Oxford, UK: Elsevier.

51 Hayhoe, M. M. 2008 Visual memory in motor planning and action. In *The visual world in memory* (ed. J. R. Brockmole), pp. 117–139. Hove, UK: Psychology Press.

52 Ballard, D. H., Hayhoe, M. M., Li, F., Whitehead, S. D., Frisby, J. P., Taylor, J. G. & Fisher, R. B. 1992 Hand eye coordination during sequential tasks. *Phil. Trans. R. Soc. Lond. B* **337**, 331–339. (doi:10.1098/rstb.1992.0111)

53 Ballard, D., Hayhoe, M. & Pelz, J. 1995 Memory representations in natural tasks. *J. Cogn. Neurosci.* **7**, 66–80. (doi:10.1162/jocn.1995.7.1.66)

54 Triesch, J., Ballard, D. H., Hayhoe, M. M. & Sullivan, B. T. 2003 What you see is what you need. *J. Vis.* **3**, 86–94.

55 Droll, J. A. & Hayhoe, M. M. 2007 Trade-offs between gaze and working memory use. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 1352–1365. (doi:10.1037/0096-1523.33.6.1352)

56 Hollingworth, A. & Henderson, J. M. 1999 Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychol.* (*Special Issue on Object Perception and Memory*) **102**, 319–343.

57 Castelhano, M. S. & Henderson, J. M. 2005 Incidental visual memory for objects in scenes. *Vis. Cogn.* (*Special Issue on Real-World Scene Perception*) **12**, 1017–1040.

58 Pelz, J. B. & Canosa, R. 2001 Oculomotor behavior and perceptual strategies in complex tasks. *Vis. Res.* **41**, 3587–3596. (doi:10.1016/S0042-6989(01)00245-0)

59 Hayhoe, M. M., Shrivastava, A., Mruczek, R. & Pelz, J. B. 2003 Visual memory and motor planning in a natural task. *J. Vis.* **3**, 49–63.

60 Mennie, N., Hayhoe, M. & Sullivan, B. 2007 Look-ahead fixations: anticipatory eye movements in natural tasks. *Exp. Brain Res.* **179**, 427–442. (doi:10.1007/s00221-006-0804-0)

61 Anstis, S. M. 1974 A chart demonstrating variations in acuity with retinal position. *Vis. Res.* **14**, 589–592. (doi:10.1016/0042-6989(74)90049-2)

62 Aivar, M. P., Hayhoe, M. M., Chizk, C. L. & Mruczek, R. E. B. 2005 Spatial memory and saccadic targeting in a natural task. *J. Vis.* **5**, 177–193.

63 Brouwer, A. & Knill, D. 2007 The role of memory in visually guided reaching. *J. Vis.* **7**, 1–12.

64 Montello, D. R. 1993 Scale and multiple psychologies of space. In *Spatial information theory: a theoretical basis for GIS* (eds A. U. Frank & I. Campari), pp. 312–321. Berlin, Germany: Springer-Verlag.

65 Hirtle, S. C. & Jonides, J. 1985 Evidence of hierarchies in cognitive maps. *Mem. Cogn.* **13**, 208–217.

66 Brockmole, J. R. & Wang, R. F. 2005 Spatial processing of environmental representations. In *Neurobiology of attention* (eds L. Itti, G. Rees & J. Tsotsos), pp. 146–151. New York, NY: Academic Press.

67 Wang, R. F. & Brockmole, J. R. 2003 Human navigation in nested environments. *J. Exp. Psychol. Learn. Mem. Cogn.* **29**, 398–404. (doi:10.1037/0278-7393.29.3.398)

68 Simons, D. J. & Wang, R. F. 1998 Perceiving real-world viewpoint changes. *Psycholog. Sci.* **9**, 315–320. (doi:10.1111/1467-9280.00062)

69 Wang, R. F. & Simons, D. J. 1999 Active and passive scene recognition across views. *Cognition* **70**, 191–210. (doi:10.1016/S0010-0277(99)00012-8)

70 Garsoffky, B., Schwan, S. & Hesse, F. W. 2002 Viewpoint dependency in the recognition of dynamic scenes. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 1035–1050. (doi:10.1037/0278-7393.28.6.1035)

71 Garsoffky, B., Huff, M. & Schwan, S. 2007 Changing viewpoints during dynamic events. *Perception* **36**, 366–374. (doi:10.1068/p5645)

72 Andersen, R. A., Snyder, L. H., Bradley, D. C. & Xing, J. 1997 Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* **20**, 303–330. (doi:10.1146/annurev.neuro.20.1.303)

73 Burgess, N. 2006 Spatial memory: how egocentric and allocentric combine. *Trends Cogn. Sci.* **10**, 551–557. (doi:10.1016/j.tics.2006.10.005)

74 Burgess, N. 2008 Spatial cognition and the brain. *Ann. N. Y. Acad. Sci.* **1124**, 77–97. (doi:10.1196/annals.1440.002)

75 Chang, S. W., Papadimitriou, C. & Snyder, L. H. 2009 Using a compound gain field to compute a reach plan. *Neuron* **64**, 744–755. (doi:10.1016/j.neuron.2009.11.005)

76 Land, M. F., Mennie, N. & Rusted, J. 1999 The roles of vision and eye movements in the control of activities of daily living. *Perception* **28**, 1311–1328. (doi:10.1068/p2935)

77 Waller, D. & Hodgson, E. 2006 Transient and enduring spatial representations under disorientation and self-rotation. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 867–882. (doi:10.1037/0278-7393.32.4.867)

78 Droll, J. & Eckstein, M. 2009 Gaze control and memory for objects while walking in a real world environment. *Vis. Cogn.* **17**, 1159–1184.

79 Wolbers, T., Hegarty, M., Büchel, C. & Loomis, J. M. 2008 Spatial updating: how the brain keeps track of changing object locations during observer motion. *Nat. Neurosci.* **11**, 1223–1230. (doi:10.1038/nn.2189)

80 Duhamel, J.-R., Colby, C. L. & Goldberg, M. E. 1992 The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* **255**, 90–92. (doi:10.1126/science.1553535)

81 Bays, P. M. & Husain, M. 2007 Spatial remapping of the visual world across saccades. *Neuroreport* **18**, 1207–1213. (doi:10.1097/WNR.0b013e328244e6c3)