Priorities for selection and representation in natural tasks

Benjamin W. Tatler, Yoriko Hirose, Sarah K. Finnegan, Riina Pievilainen, Clare Kirtley and Alan Kennedy

School of Psychology, University of Dundee, Dundee DD1 4HN, UK

Selecting and remembering visual information is an active and competitive process. In natural environments, representations are tightly coupled to task. Objects that are task-relevant are remembered better due to a combination of increased selection for fixation and strategic control of encoding and/or retaining viewed information. However, it is not understood how physically manipulating objects when performing a natural task influences priorities for selection and memory. In this study, we compare priorities for selection and memory when actively engaged in a natural task with first-person observation of the same object manipulations. Results suggest that active manipulation of a task-relevant object results in a specific prioritization for object position information compared with other properties and compared with action observation of the same manipulations. Experiment 2 confirms that this spatial prioritization is likely to arise from manipulation rather than differences in spatial representation in real environments and the movies used for action observation. Thus, our findings imply that physical manipulation of task relevant objects results in a specific prioritization of spatial information about task-relevant objects, possibly coupled with strategic de-prioritization of colour memory for irrelevant objects.

1. Introduction

The available visual information in a natural environment far exceeds the sampling capabilities of the human visual system. Foveated spatial sampling and punctuated temporal sampling combine to impose strict spatio-temporal limits on visual information gathering. The manner in which we gather and use sampled visual information is unsurprisingly under strict strategic control. We prioritize information in our surroundings both by selectively looking and selectively retaining information from where we look. This study considers the manner in which being engaged in an active real-world behaviour influences our priorities for selection and retention of visual information.

The eyes select locations that provide information relevant to the current behavioural goals [1–3]. Moment-to-moment decisions about where to fixate may reflect the outcome of competitive processes played out within attentional priority maps [4]. Both bottom-up and top-down sources of information contribute to these priority maps [5,6], which are likely to be implemented neurally in the lateral intraparietal area [7]. Tight coupling between vision and action in tasks in which objects are manipulated demonstrates that selection priorities are closely linked to ongoing motor acts and the moment-to-moment information requirements for the current actions [8–11]. As such any underlying priority map must reflect the priorities associated with the current motor action required to pursue our ongoing behavioural goals. Indeed, in active tasks, selection priorities are dominated by high-level factors associated with selecting the optimal place to fixate at any moment in time in order to best serve ongoing behavioural goals [12], and it may be appropriate to consider the priority maps as reflecting anticipated behavioural reward at each location in space and time [3,13,14].

Memory representations aid ongoing behaviour by providing supporting evidence for fixation selection [15], and this is particularly evident when
visual information supplied by the current retinal image is insufficient. Remembered information informs saccade programming when the intended saccade target appears in low-acuity peripheral vision [16], is no longer visible [17–19], or falls outside the current field of view [9,20]. These memory representations must minimally include information about where behaviourally relevant objects are located in the environment if they are to assist in deploying gaze during natural task performance. Indeed, objects that are relevant to an individual’s behavioural task are remembered better than objects that are not relevant [21,22]. Furthermore, encoded object information may be retained in memory if it is known to be required later in the task [23–25]. Task-based biases in memory representations may arise solely from task-based differences in fixation allocation [26,27]. However, such task-based memory representations may also involve additional strategic control of encoding and/or retention of viewed information [22,28–30]. Strategic control of information encoding and retention from fixations appears to involve prioritization of particular features for task-relevant objects, coupled with de-prioritization of the same object features for task-irrelevant objects [29].

While the above-mentioned studies argue that behavioural goals can influence how much, and what kind of, information is encoded and/or retained from fixations, it is important to note that in all cases the task manipulations were essentially visual. Even in the real-world environment of Tatler & Tatler [29], participants did not manipulate objects. If we are to understand the manner in which information is strategically prioritized for memory representations in natural behaviour, then it is important to consider memory processes in the context of a real task, because the act of physically manipulating an object might itself change the representational priorities for that object. Indeed, perceptual and cognitive processes are strongly influenced by actual or planned actions towards objects [31–33]. These action-based effects on cognition may not be an automatic consequence of viewing an object, but may be restricted to situations in which judgements are related to action [34]. As such, in a purely visual task such as memorizing objects or searching a visual scene, priorities in memory representations may not reflect those present in situations where active manipulations of objects are carried out.

Prioritization in active tasks may involve cross-modal effects associated with proprioceptive feedback from the manipulated object, or may arise primarily from activating representations of action plans for the object [35]. Thus, if we are to understand how real-world behaviour influences priorities for selection and memory it is important to consider the relative contribution of action and action observation to these priority settings.

Observing actions performed by another individual can result in the same proactive deployment of gaze as is found when performing the same actions [36]. However, this proactive action is reduced when observing unpredictable action [37], and gaze allocation becomes reactive when the hands of the actor are not visible and the objects therefore appear to move on their own [36]. While action observation may be sufficient to activate action plans and so influence gaze allocation, it is not clear whether object memory representations are equally shaped by action and action observation, or whether physical manipulation of objects results in different representational priorities.

In this study, our primary concern is whether strategic control of information encoding and retention from fixations [22,28–30] is influenced by whether or not we are engaged in active manipulations with objects. Given the strong link between how long an object is fixated and the efficacy of ensuing object memory [38,39], it is important to characterize how action influences the amount of time for which objects are fixated. Our measure of selection is therefore restricted to the total amount of time spent fixating each object. We can then consider whether any differences in representation identified can be accounted for by differences in fixation time, or whether they are indicative of different priorities in information extraction and retention.

Experiment 1 considers priorities for selection and memory during a real-world task (making tea) and during action observation, with another individual performing the same task. Observing actions from a third-person perspective is sufficient to activate motor representations that permit proactive allocation of gaze [36]. However, observing actions from a first-person perspective facilitates action imitation [40] compared with action observation from a third-person perspective. Moreover, viewing actions from a first-person perspective is associated with different underlying neural activation than viewing from a third-person perspective [41]. Thus, we compare real-world tea-making with action observation of first-person movies of tea making.

If priorities for selection and representation are set by task-relevance and the activation of action plans, then we might expect that fixation allocation and object memory to be rather similar for natural action and action observation. Indeed, previous research has shown that when observing action, gaze allocation of the observer can be extremely similar in space and time to that of the actor performing the actions [36]. However, the extent to which action observation leads to the same proactive allocation of gaze as an individual performing the action depends upon whether the observed action is predictable or not [37]. When observing a familiar over-learnt task such as making tea, it is unclear as to what extent the actions can be considered predictable or unpredictable and so it is unclear how acting and observing acting will differ in terms of selection. Importantly, it is not yet understood how action and action observation differ in terms of the priorities they set for object memory representations. This first experiment therefore not only allows us to describe the priorities for selection and representation in a natural task, but also to assess whether these priorities are shaped by object manipulation.

2. Experiment 1

(a) Method

(i) Participants

Sixteen participants were recruited at the University of Dundee. All had normal vision or corrected-to-normal vision. Half of the participants made tea in a real kitchen. The other eight participants watched first-person videos recorded from the eye trackers worn by the tea-making participants.

(ii) Eye movement recording

Natural task condition. Participants wore a mobile eye tracker (Positive Science LLC). This eye tracker comprises two video cameras mounted on the frame of a pair of spectacles, one forward-facing camera to record a head-centred view of
the scene, and second one facing towards the right eye. Movies from the two cameras were recorded separately in digital format and then synchronized and analysed offline, with gaze estimation using the YARBUS software package (v. 2.2.3) supplied by Positive Science LLC. Gaze direction was via pupil tracking and was estimated from nine-point calibration procedure at the start and end of each recording session. The accuracy of this system is sufficient to allow spatial estimates of gaze direction to within a degree of visual angle.

Action observation condition. Eye movements were recorded using the SR Research EyeLink 1000 eye tracker recording at 1000 Hz. Gaze direction was via pupil tracking and was estimated from a nine-point calibration procedure carried out at the start of the recording session. Calibration was repeated if a nine-point validation procedure carried out immediately after calibration indicated a mean spatial accuracy across the nine points that was worse than 0.5° or a maximum spatial accuracy at any single point that was worse than 1° of visual angle. No saccade detection algorithms were applied to the data.

(iii) Procedure

Natural task condition. Participants made tea in a large kitchen. Among other items typically found in a kitchen, there were five objects required to make tea (kettle, milk jug, mug, tea caddy, sugar bowl), and five objects not required for tea making but plausibly found in a kitchen context (water jug, bowl, dish towel, saucepan, toaster). The five non-tea objects were placed close to (within 20 cm and on the same surface as) the tea-making objects. Participants were asked to make tea with milk and sugar and were given no time restrictions for this task. After completing the tea-making task, participants were given a paper questionnaire testing memory for the five tea-making and five non-tea objects. For each object, memory was tested using four-alternative forced-choice questions with plausible foils. For identity questions, greyscale photographs were used with foils drawn from the same object class. For colour questions, verbal colour labels were used, with viable colour foils for the target object. For position questions, an outline plan view of the kitchen was shown with four possible locations; foil locations never coincided with the locations of the other target objects. The order in which the objects were tested was randomized, and the order of questions for each object was counterbalanced.

Action observation condition. Participants watched head-centred videos of another individual making tea. These videos were those recorded in the natural task condition described above and were taken from the head-mounted scene camera on the mobile eye tracker used in the natural task condition. The eight first-person videos recorded in the natural task condition were assigned to the eight participants in the action observation condition, such that each participant in the action observation condition watched a different first-person tea-making movie. Movies were displayed in 720 × 576 pixel format on a 19 inch monitor with a screen resolution of 800 × 600 pixels. Participants viewed movies with their head stabilized using a chin rest positioned approximately 60 cm from the monitor. After the end of the movie, participants completed the same paper questionnaire as was used in the natural task condition.

The first-person videos used in this study present a potential challenge for saccade planning. Volitional gaze changes in real-world tasks involve combined eye and head movement. For the video observation condition, participants did not move their heads, but the content of the video reflected the head movements made by participants in the real environment. It is also the case that observers will not necessarily know the precise order in which actors will use the objects in the tea-making task. When the objects that are to be acted upon are less predictable, gaze allocation is less proactive [37]. Thus, when watching the videos, saccade planning may be more reactive, or at least less proactive, in response to the continual head movements present in the videos.

(iv) Analysis

For both conditions, videos were prepared showing gaze estimation for each frame of the 30 Hz movie of the kitchen. For the natural task condition, these gaze movies were created using the YARBUS software. For the action observation condition, these movies were created using MATLAB to overlay gaze samples from the EyeLink 1000 on the frames of the observed first-person movies. Subsequent analyses of eye movement data were carried out manually and involved counting the number of frames in which each of the 10 target objects were fixated.

For analyses of the effects of natural task performance and action observation on selection and memory, linear mixed effects models were run using the lme4 package [42] in the R statistical programming environment [43]. Where necessary, p-values were estimated using Markov chain Monte Carlo sampling derived from the pval.fcn() function in the languageR library. The LME modelling approach has many advantages over traditional ANOVA models [44]; crucially, it allows between-subject and between-item variance to be estimated simultaneously. Separate LME models were run to predict fixation time and memory performance in each question type. For modelling the total time spent fixating an object (the summed duration of all fixations on the object hereafter referred to as total fixation time), action condition (real, observed) and object type (task-relevant, task-irrelevant) were treated as fixed effects. Subjects and items were included in the model as random factors. Following inspection of the distribution and residuals, total fixation time was log-transformed in order to meet LME assumptions. For modelling performance in the memory questions, logistic models were run with action condition (real, observed) and object type (task-relevant, task-irrelevant) as fixed effects and subjects, items and fixation time as random factors. By including total fixation time as a random effect, we were able to consider effects of our fixed effects on memory performance that were above and beyond those resulting from differences in fixation time on each object. In all analyses, interactions were broken down using planned contrasts, corrected for multiple comparisons.

(b) Results

(i) Selection

Total fixation time on objects was influenced by the action condition, $\beta = -0.241, \text{s.e.} = 0.108, t = -2.24, p = 0.027$; the type of object, $\beta = -1.098, \text{s.e.} = 0.219, t = -5.00, p < 0.001$ and the interaction between these two factors, $\beta = 0.401, \text{s.e.} = 0.130, t = 3.08, p = 0.003$ (figure 1). More time was spent fixating task-relevant (and therefore manipulated) objects in the natural task condition than in the action
observation condition, \( t = 2.24, p = 0.027 \). For task-irrelevant objects (which were not manipulated in the natural task condition), there was no difference in time spent fixating these objects in the natural task and action observation conditions, \( t = 1.40, p = 0.164 \).

(ii) Representation

For identity questions, performance was not influenced by the action condition alone, \( \beta = -1.011, \text{s.e.} = 0.706, z = -1.43, p = 0.152 \), but was influenced by object type, \( \beta = -3.030, \text{s.e.} = 1.275, z = -2.38, p = 0.017 \), and the interaction between object type and action condition, \( \beta = 1.937, \text{s.e.} = 0.933, z = 2.077, p = 0.038 \) (figure 1). There were no differences in identity memory between the natural task and action observation conditions for task-relevant objects, \( z = 1.42, p = 0.155 \) or task-irrelevant objects, \( z = 1.55, p = 0.121 \). Identity memory was better for task-relevant objects than task-irrelevant objects in the natural task condition, \( z = 2.39, p = 0.017 \), but not in the action observation condition, \( z = 0.89, p = 0.374 \).

For performance in colour questions, the only significant effect was that of object type, \( \beta = -2.238, \text{s.e.} = 1.012, z = -2.21, p = 0.027 \), with better memory for task-relevant objects (figure 1); there was no significant effect of action condition, \( \beta = 0.754, \text{s.e.} = 0.674, z = 1.12, p = 0.263 \), nor of the interaction between action condition and object type, \( \beta = 0.947, \text{s.e.} = 0.870, z = 1.09, p = 0.276 \). Because we were a priori interested in the comparisons across object types and action conditions, we ran planned comparisons despite the lack of significant interaction. There was no difference in colour memory for task-relevant objects in the natural task and action observation conditions, \( z = 1.11, p = 0.268 \). Colour memory was better for task-irrelevant objects in the action observation condition than in the natural task condition, \( z = 2.62, p = 0.009 \). Colour memory was better for task-relevant objects than task-irrelevant objects in the natural task condition, \( z = 2.29, p = 0.022 \), but not in the action observation condition, \( z = 1.27, p = 0.204 \).

For position questions, performance was influenced by action condition, \( \beta = -1.477, \text{s.e.} = 0.716, z = -2.06, p = 0.039 \), object type, \( \beta = -3.511, \text{s.e.} = 0.770, z = -4.56, p < 0.001 \) and the interaction between these two factors, \( \beta = 1.978, \text{s.e.} = 0.888, z = 2.23, p = 0.026 \) (figure 1). Memory for the positions of task-relevant objects was better in the natural task condition than in the action observation condition, \( z = 2.08, p = 0.037 \). Memory for the positions of task-irrelevant objects was not influenced by the action condition, \( z = 1.12, p = 0.261 \). Position memory was better for task-relevant objects than task-irrelevant objects both in the natural task condition, \( z = 4.74, p < 0.001 \) and in the action observation condition, \( z = 2.61, p = 0.009 \).

Binomial tests revealed that memory for task-relevant objects was above chance for all object properties in both action conditions, all \( ps < 0.001 \). In the natural task condition, memory for the position of task-relevant objects was better.
than memory for the colour or identity of task-relevant objects, $t_{112.25} = 2.55, p = 0.012$. Memory for the colour and identity of these objects was no different, $t_{75.41} = 1.31, p = 0.194$. For the same objects in the video condition, performance in all questions was similar, all $t < 1$. Thus, manipulating objects resulted in a selective benefit for position memory that was not found when observing manipulations from the first-person perspective.

For task-irrelevant objects, binomial tests revealed that in the natural task condition performance was above chance for identity questions, $p = 0.012$, but not for colour, $p = 0.211$, or position questions, $p = 0.336$. For the same objects in the action observation condition, performance was above chance for all question types, identity: $p < 0.001$, colour: $p < 0.001$, position: $p = 0.048$. For task-irrelevant objects in the natural task condition, performance on all question types was similar, all $t < 1.25$, all $p > 0.215$. In the action observation condition, memory for the position of task-irrelevant objects was worse than memory for their colour or identity, $t_{63.62} = 2.01, p = 0.049$. Memory for the colour and identity of these objects did not differ, $t < 1$.

(c) Discussion

Active engagement in a natural task resulted in more time spent fixating task-relevant objects than when an individual’s actions were observed from a first-person perspective. Active manipulation of (task-relevant) objects also increased memory for their position but did not change memory for their colour or identity compared with the same objects in the action observation condition. These findings suggest a specific advantage for position memory when manipulating objects, implying that manipulation selectively prioritizes spatial information about objects. This selective prioritization of position memory for manipulated objects is further supported by our finding that the position of task-relevant objects was remembered better than their colour or identity in the natural task condition, whereas position was remembered no better than colour or identity for the same objects in the action observation condition. Our finding that memory for the colour of task-irrelevant objects was better in the action observation condition than in the natural task condition could be used to suggest that physical manipulation of objects results in strategic de-prioritization of colour memory for objects that are task-irrelevant and therefore not manipulated.

The findings from experiment 1 therefore suggest that in natural tasks, interaction with objects biases representational priorities to favour spatial information over other object information and we can speculate that this advantage comes from the act of manipulating task-relevant objects rather than observing or planning action.

However, spatial priorities for representation could arise either as a result of direct manipulation of objects or as a result of physical presence in the environment. Movement through a real environment poses different representational challenges for spatial information and may result in different forms of spatial memory [20]. In particular, real environments in which an individual is free to move around might favour egocentric representational frames of reference [20,45,46], possibly implemented neurally in the precuneus [47]. By showing participants in the action observation condition first-person videos recorded in the natural task condition, we hoped to minimize differences in spatial challenges provided by the visual information. However, it is still possible that experiencing changes in egocentric viewpoints by physical movement through the environment—where viewpoint changes are generated by the observer—would result in differences in spatial memory for objects compared with experiencing the same changes in viewpoint without egomotion—where changes in viewpoint are not generated by the observer [48].

There are two approaches for considering the relative roles of physical manipulation and physical presence in the environment upon our observed benefit for spatial memory. First, we can compare memory for task-relevant objects—which were always manipulated—and task-irrelevant objects—which were never manipulated—in the natural task and action observation conditions. For task-irrelevant objects, position memory was poor in both action conditions. Thus, there was no evidence for spatial memory being better in the real environment than in the video observation task. This finding suggests that the observed position memory advantage for manipulated objects in the natural task condition did not reflect any spatial memory benefit from being present in the real environment rather than watching a first-person perspective movie. However, these objects are irrelevant to the task and task-relevance itself changes representational priorities, favouring spatial memory in task-relevant objects and de-prioritizing spatial memory in task-irrelevant objects even in the absence of physical manipulation [29]. If we wish to attribute the observed spatial memory benefit for task-relevant objects in natural tasks to factors associated with physical manipulation, it is therefore necessary to consider a comparison between presence in the environment and observation of the environment for task-relevant objects that are not manipulated. This opportunity is not provided by experiment 1, but is addressed in experiment 2 in which participants completed a memory task rather than an active manipulation task.

It is also important in experiment 2 to discount the possibility that the observed priority changes for selection and representation cannot be attributed to any difficulties participants face when viewing videos that result from another individual’s head movements (see experiment 1 method for more details).

By comparing the results of experiments 1 and 2, we will be able to assess the relative contributions of physical interaction with objects and physical presence in an environment as they influence priorities for the selection and representation of objects during real-world behaviour.

3. Experiment 2

(a) Method

(ii) Eye movement recording

As for experiment 1.
(iii) Procedure
Real-world condition. As for the natural task condition in experiment 1 except that participants were instructed to enter the kitchen and memorize as much as they could about all objects present in the room, without touching anything in the kitchen. After about 2 min, participants were asked to leave the room. This time limit was chosen to match the average time spent making tea in experiment 1 (\(M = 2\text{ min } 19\text{ s, s.d. } = 40.7\text{ s}\)). Subsequent checking of the eye-tracking videos recorded in experiment 2 showed that the average time participants spent in the kitchen in experiment 2 was 2 min 13 s (s.d. = 5.4 s).

Video observation condition. As for the action observation condition in experiment 1 except that participants watched first-person videos recorded from the real-world condition of experiment 2 and were instructed to memorize as much as they could about all objects present in the video that they watched.

(iv) Analysis
As for experiment 1 except that object type was not included as a fixed effect in the analyses. As a result, models had only one fixed effect, observation condition (real world, first-person video).

(b) Results
(i) Selection
There was no difference between the amount of time spent fixating objects in the kitchen and the amount of time spent fixating objects when watching the first-person movies, \(\beta = -0.133,\text{ s.e. } = 0.091, t = -1.45, p = 0.148\) (figure 2).

(ii) Representation
Memory was significantly better for objects when physically present in the kitchen than when watching the first-person videos for identity, \(\beta = -0.812,\text{ s.e. } = 0.307, z = -2.64, p = 0.008\), and (marginally so) for position memory, \(\beta = -0.549,\text{ s.e. } = 0.280, z = -1.96, p = 0.050\). There was no equivalent difference in colour memory for objects, \(\beta = -0.332,\text{ s.e. } = 0.297, z = -1.12, p = 0.263\) (figure 2).

Binomial tests revealed that performance was above chance for all question types both in the real-world and first-person observation conditions, all \(p s < 0.001\). There were no significant differences between question types in the real-world condition, all \(t s \leq 1.25\), all \(p s \geq 0.211\), or in the video observation condition, all \(t s \leq 1\).

(c) Discussion
Participants in the real environment and those watching the first-person movies spent the same amount of time fixating objects. However, those who were present in the real environment showed better position and identity memory for objects than those who observed the same objects in first-person perspective movies. Colour memory was no different for those in the environment from those observing videos. Participants performed equally well on identity, colour and position questions in each of the two observation conditions. The differences observed between our
observation conditions may be a result of whether or not the changes in viewpoint are self-generated. Alternatively, it may be that mere presence in a real-environment changes underlying representational priorities and processes to facilitate particular aspects of memory representations. From the present results, it is not possible to distinguish these possibilities.

The lack of difference in the allocation of fixation time for the real-world and video observation conditions suggests that this aspect of selection was similar in these two conditions. As such, we can suggest that watching the videos did not result in particularly different distribution of processing time with respect to the objects tested despite the challenges presented by viewing movies that result from the continual head movements of another individual. It would therefore appear that the total fixation time differences found between the natural task and action observation conditions in experiment 1 reflect differences associated with whether or not the object was manipulated rather than artefacts of viewing the first-person movies.

The findings from experiment 2 suggest that moving through—or, at least, being present in—an environment does change representational priorities, but this acts primarily via enhanced memory for object identities, together with a marginal improvement for object positions. However, these findings are not able to explain those from experiment 1, in which we found a specific enhancement of position memory for task-relevant objects, but not for other forms of memory for the same objects when physically manipulating them. Taken together with the results from experiment 1, our findings do not suggest that merely being present in and moving through an environment specifically biases representations of task-relevant objects towards better memory for position. Rather the selective prioritization of position information in representations of task-relevant objects seems likely to originate from the fact that these objects were the target of physical manipulation by the observers. Our interpretation assumes that the underlying memory processes for spatial representation in experiments 1 and 2 are the same, but with different priority settings depending upon the condition. However, an alternative explanation could be that different spatial memory processes underlie these two experiments, and that the spatial memory processes underlying action and action observation are facilitated by the real-world environment, thus indirectly resulting in better memory when manipulating objects than when observing the manipulation.

4. General discussion

Across two experiments, we considered how priorities for selection and representation of object information in natural tasks may arise. We are able to use these findings to speculate about the manner in which selection (indexed by foveal processing time) and representation reflect priority settings arising from three factors: task-relevance, physical presence in and movement through the environment, and active manipulation of objects.

As expected, when engaged in a natural task, priorities for selection and representation favoured task-relevant objects in the environment. We confirm previous suggestions that task-relevance not only results in selection prioritization, but also in more faithful memory given an equivalent amount of fixation time [22,28–30]. We found that the importance of task relevance for memory representations depended upon whether objects were physically manipulated or merely observed in first-person action observation movies (experiment 1). As such, comparisons of representational priorities for task-relevant and task-irrelevant objects in tasks that do not involve physical manipulation of objects [29] may not reflect accurately the influence of task relevance on representation in natural tasks.

Experiment 2 suggested that being present in, and moving through, a real environment did not influence the total amount of fixation time on an object but did improve memory for identity and position. Better memory for objects when viewpoint changes are made via egomotion is consistent with previous studies [48]. If we consider that the representations formed in both real-world settings and our first-person observation conditions are likely to be egocentrically organized [20,45,46], we can suggest that the formation and updating of these representations is facilitated by egomotion in a real environment. It may be that representational processes for scene and object memory are supported by sensorimotor evidence supplied by the vestibular system and proprioception [48], which are necessarily absent when observing movies from a first-person perspective.

Action and action observation can result in remarkably similar deployment of gaze in space and time when watching predictable actions [36]. When watching unpredictable actions gaze is still deployed proactively, but is done so later than when actually performing the action [37]. We showed that the amount of time for which objects are selected and the priorities for representation were different when manipulating objects in the context of a natural task than when observing these actions from a first-person perspective. Objects that were manipulated in the real-world task were fixated for longer than when these object manipulations were viewed on first-person movies. These selection differences for task-relevant objects were unlikely to be driven by whether or not the observer was physically present in the real environment, because no such differences were observed for task-irrelevant objects (experiment 1) or for objects in the memory task (experiment 2). Therefore, the observed increase in fixation time on task-relevant objects in the real-world environment is likely to reflect different selection priorities for objects that we physically manipulate. In natural tasks, there is tight spatio-temporal coupling between vision and manipulation [8,11,49]; in many natural tasks, the eyes typically arrive at objects around 0.5–1 s before a manipulation begins and remain largely on the object throughout most of the manipulation [9,50]. Any disruption of this spatio-temporal coupling when observing actions might result in the observed differences in fixation time. Observing action does not in itself appear to disrupt spatio-temporal coupling of vision and action [36], unless the observed action is unpredictable [37]. When observing tea-making aspects of the task will be unpredictable—participants will not necessarily know which object is to be selected next—and as such our finding of less time spent on task-relevant objects during action observation than during actual action is consistent with Rotman et al.’s suggestion that the eyes arrive later at objects during action observation if the actions are less predictable. It may also be the case that because participants are not generating the actions they observe in the videos,
they are less likely to look ahead to the targets of future (observed) actions [10,31,52].

Physical manipulation of objects resulted in changes to representational priorities compared with non-manipulated objects or first-person action observation, with a specific prioritization of object position information. Action-based changes in cognitive representation are plausible within a grounded account of cognition [32,33]. Grounded explanations of cognition, which have suggested that representations are grounded in the environment.

This research was supported by a research grant awarded to B.T. and A.K. by The Leverhulme Trust (Project No. F001430).

Acknowledgements

References

42. Bates DM, Maechler M, Bolker B. 2011 *lme4*: linear mixed-effects models using S4 classes (R package version 0.999375 – 42). See http://cran.r-project.org/web/packages/lme4.