



# In what ways do eye movements contribute to everyday activities?

Michael F. Land<sup>a,\*</sup>, Mary Hayhoe<sup>b</sup>

<sup>a</sup> School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK

<sup>b</sup> Center for Visual Sciences, University of Rochester, 274 Meliora Hall, Rochester, NY 14627-0270, USA

Received 25 October 2000; received in revised form 30 January 2001

## Abstract

Two recent studies have investigated the relations of eye and hand movements in extended food preparation tasks, and here the results are compared. The tasks could be divided into a series of actions performed on objects. The eyes usually reached the next object in the sequence before any sign of manipulative action, indicating that eye movements are planned into the motor pattern and lead each action. The eyes usually fixated the same object throughout the action upon it, although they often moved on to the next object in the sequence before completion of the preceding action. The specific roles of individual fixations could be identified as locating (establishing the locations of objects for future use), directing (establishing target direction prior to contact), guiding (supervising the relative movements of two or three objects) and checking (establishing whether some particular condition is met, prior to the termination of an action). It is argued that, at the beginning of each action, the oculomotor system is supplied with the identity of the required object, information about its location, and instructions about the nature of the monitoring required during the action. The eye movements during this kind of task are nearly all to task-relevant objects, and thus their control is seen as primarily ‘top-down’, and influenced very little by the ‘intrinsic salience’ of objects. © 2001 Elsevier Science Ltd. All rights reserved.

*Keywords:* Eye movement; Everyday settings; Vision

## 1. Introduction

Very little of the eye movement research over the past century has been concerned with ordinary activities in everyday settings. Thus we have a great deal of information about single eye movements in laboratories, and also about some continuous tasks such as reading and driving a car. Most of the things we do, however, are not like this. Housework, food preparation, carpentry or gardening all involve a succession of actions, each different from the previous one. Usually these subtasks involve an action performed on one or more objects, for example taking the lid off a pan or hitting a nail with a hammer. Recently two studies have addressed the nature of the involvement of vision in coordinating such actions, and specifically the roles of eye movements in moving gaze to places where information is needed for the execution of the task. One

study involved tea-making which, if done the English way, requires 40–50 separate actions (Land, Mennie, & Rusted, 1999); in the other, students made peanut butter and jelly sandwiches (Hayhoe, 2000). In both cases the location of foveal gaze was monitored continuously using a head-mounted eye tracker with an accuracy of about 1°, and the head was free to move. In the tea-making study the three subjects had to move about the room to locate the objects required for the task; in the sandwich-making task the seven subjects were seated in one place, in front of a table. This paper compares the two studies, and looks for common principles that may apply to all such multi-component activities. For further details the reader is referred to the two studies mentioned.

The central question in this review is the relation between ongoing motor actions and the eye movements that accompany them. One possibility is that the eyes are mainly involved in ‘forward planning’ seeking out objects for future use and setting up the operations to be performed on them. At the other extreme, the eyes

\* Corresponding author.

E-mail address: m.f.land@sussex.ac.uk (M.F. Land).

might be essentially passive until summoned by the motor system to provide it with some specific piece of information, such as the location of a knife or the depth of water in a kettle. Neither of these extremes fits our observations. Our general conclusions are that the eyes do provide information on an ‘as needed’ basis (as Fig. 1 clearly shows), but that the relevant eye movements usually precede the motor acts they mediate by a fraction of a second. They are thus in the vanguard of each action plan, and are not simply responses to circumstances.

We shall consider three kinds of question. First, we review what the two studies have to say about the way the eyes are involved in the control of each action. To what extent is the way that the eye movement system directs gaze an integral part of the motor programme of the action? Second: can one classify the types of function that are performed by vision at different stages of

an action? Third: what information does the eye movement system itself need, in order to put the eyes in a position to supply the information that the motor system needs?

## 2. The involvement of eye movements in an action sequence

There are (at least) two levels of organization to be considered in these kinds of tasks. There are the individual actions themselves (pick up knife, put a teabag in the pot etc.), and there is the sequence of actions — the ‘script’ of the task as a whole. Here we are concerned mainly with the roles of vision in individual actions, although the transition from one action to another, which is under the control of the script, is also of interest. Schwartz, Reed, Montgomery, Palmer and Mayer (1991) and Schwartz, Montgomery, Fitzpatrick-DeSalme, Ochipa, Coslett and Mayer (1995) have developed a useful system for categorizing the individual actions. These basic object-action conjunctions they call ‘A1s’, defined as ‘simple actions that transform the state or place of an entity through manual manipulation’. Larger units of action incorporating a number of A1s Schwartz describes as A2s (e.g. fill the kettle) but these refer to sub-goals of the whole task rather than the actions themselves.

In the study of tea-making (Land et al., 1999) we found that the A1 description fitted the pattern of eye movements associated with the actions very well. We have called the combination of A1 actions and the eye movements that go with them ‘object related actions’ (ORAs). For our purposes an ORA comprises all the acts performed on a particular object without interruption (e.g. the sequence: pick up mug, move it to new location, set the mug down, would constitute one ORA). Thus in Fig. 1a the ORA sequence would be ‘inspect and pick up kettle’, ‘remove kettle lid’, ‘turn on taps’, and ‘put kettle in water stream’. An ORA usually began with a ‘defining moment’, when gaze moved from the last object to be manipulated to the next in the sequence (Figs. 1 and 2). In both the tea-making and sandwich-making studies the eyes typically fixated each object before any sign of manipulative activity occurred. In the tea-making task the average lead time was 0.56 s, but for the sandwich-making it was much shorter, 0.09 s. In both studies the standard deviations were large (averages 1.1 s for tea-making, 0.4 s for sandwich-making) meaning that there were a considerable number of instances which went against the rule, with manipulative action initiated before the first eye-movement. At the end of each action there was a similar lead time, with the eyes moving on to the next object in the script on average 0.61 s before the previous action was completed (in the tea-making task).

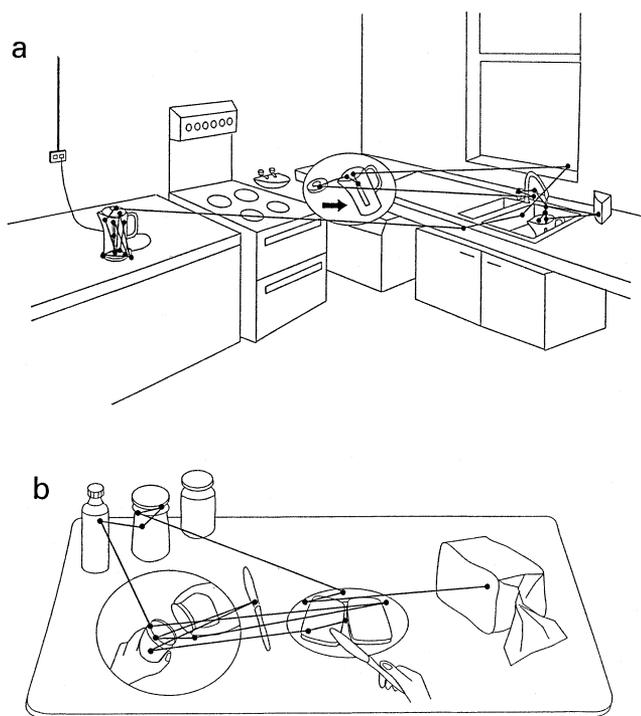


Fig. 1. The first series of fixations in the tea-making and sandwich-making tasks. In both the sequence is semi-diagrammatic because of the movements involved, but the targets of all detected saccades ( $> 1^\circ$ ) are shown. (a) Tea-making: 11 fixations are made on the kettle, which is then picked up as the sink is viewed briefly. The lid is removed from the kettle in transit (4 fixations). The taps are fixated and the hand goes to the R tap. Vision then shifts to the water stream. There is one task-irrelevant fixation, on the sink-tidy (right). The sequence lasts about 10 s, and the field depicted is about  $100^\circ$  wide. (b) Sandwich-making. This begins with the movement of a slice of bread from the bag to the plate. The peanut butter jar is then fixated, lifted and moved to the front as its lid is removed. The knife is fixated and picked up, then used to scoop out the peanut butter. This is spread on the bread, the jar is revisited for more butter, and this too is spread. There is one possible task-irrelevant fixation, on the bottle (left). The field of view depicted is about  $40^\circ$  wide.

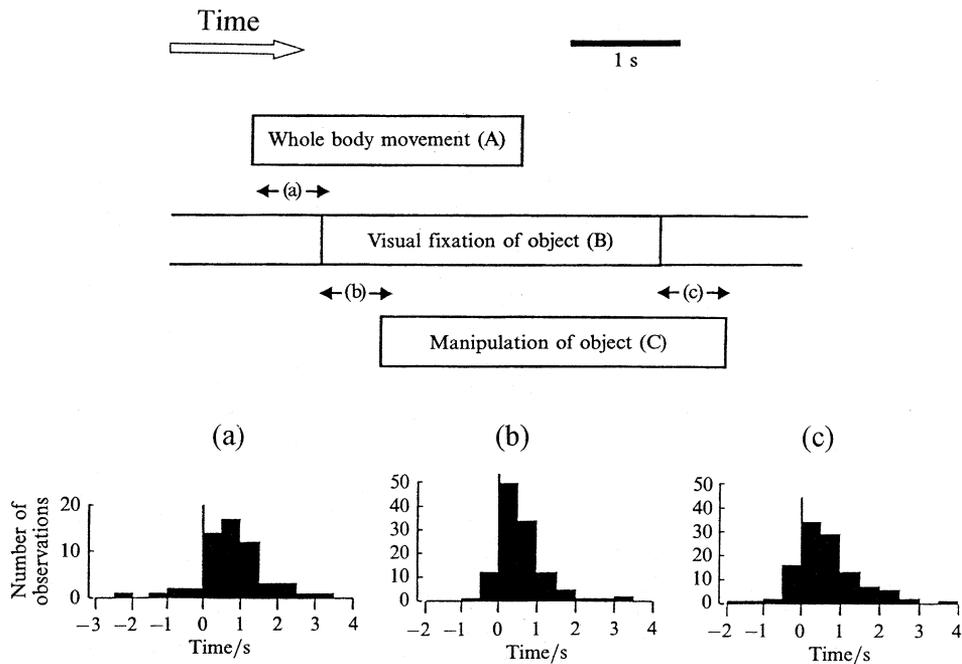


Fig. 2. The time relations of body, eye and hand movements during a total of 137 ORAs, made by three subjects while making tea. Movements of the whole body (A) precede the first fixation (B) by an average of 0.61 s (a), and these precede the first signs of manipulation (C) by 0.56 s (b). As each action draws to a close, fixation moves to the next object on average 0.61 s before the end of the preceding action (c). All times are indexed to the beginning of the first fixation of each ORA. Details and statistics in Land, Mennie and Rusted (1999).

Presumably this means that the motor system has half-a-second's worth of information available to it from a visual buffer. This is somewhat shorter than the buffer times involved in other types of action (reading aloud and typing, ca. 1 s; steering a car, 0.8 s; musical sight reading, 0.7–1.4 s; Furneaux & Land, 1999). The pattern of timing of the components is shown in more detail in Fig. 2.

The average duration of an ORA in the tea-making study was 3.04 s for the period (B) that the eyes fixated the object, and 3.25 s for the period of manipulation (C) (Fig. 2). These values exclude ORAs which involved long periods of waiting (for the kettle to fill, etc.), so they are essentially 'self-paced' actions. It is intriguing that Schleidt (1988) found that spontaneous repeated acts by individuals from four different cultural backgrounds had modal durations of 3 s. Perhaps this similarity of timing is pure coincidence, but it is also possible that it reflects the duration of some more fundamental cognitive process.

The attachment of gaze to particular objects was strikingly consistent in both studies: particular objects were generally fixated for the duration of the actions performed on them. Occasionally the eyes would look elsewhere during an action, especially if the action involved waiting (e.g. for the jug to fill or the kettle to boil), but this was exceptional. In both studies there were periods of search, particularly before the task proper began, during which objects were located and at

least some of their positions were memorized, but without any manipulative action. In a few cases the two hands had separate roles. Rarely this involved actions on different objects (in one such instance the lid was put back on a milk container while the teapot was swirled with the other hand, and gaze alternated from one sub-action to the other). More commonly both hands were engaged with the same object (e.g. the left hand held the bread while the right spread peanut butter on it with a knife).

In both studies we were struck by the rarity of eye movements to objects that were irrelevant to the task (see Fig. 1). The proportion of task-irrelevant objects viewed (other than during periods of waiting) was under 5% in both studies, even though, particularly in the tea-making task, there were numerous potential distractors. We conclude that — in real tasks — the eyes are driven much more by top-down information from the script, and rather little by the 'intrinsic salience' of objects (see Findlay & Walker, 1999). In one sandwich-making experiment involving four subjects 50% of the objects on the table were task-irrelevant (pliers, scotch tape, forks, etc.). In the interval before the task commenced, while the eyes were scanning the table, the proportion of irrelevant objects fixated was 52%. When the task started this reduced to 18%. Presumably this represented a shift from a salience-driven to a task-driven basis for selecting fixation targets.

We anticipated considerable inter-subject variation, but there was much less than we expected. There were variations in the order in which the actions were performed, especially in the tea-making task, but in the details of the eye movements made during each ORA there was much more agreement. For example, when three subjects filled the kettle, as in Fig. 1a, the numbers of fixations on the kettle were 11, 9 and 7, on the kettle and lid 4, 3 and 2, and on the taps 3, 4 and 7, the last figure involving a repeat because the hand was withdrawn from the taps to put down the kettle lid (details in Land et al. (1999) Fig. 2). In relation to the sandwich-making task, which also included pouring a drink, Hayhoe commented: ‘The locations of the fixations were also very reproducible between subjects, for example, subjects fixate the mouth of the bottle when pouring and then transfer gaze to the level of cola in the glass when about half-way through. Thus many details of the fixations, and by inference the ongoing visual computations, are governed by the task goals, together with the physical constraints of the world.’ (Hayhoe, 2000). It seems that the way the human visual system is constructed ensures that competent subjects acquire very similar oculomotor techniques when they interact with objects.

There were three main differences between the results of the two studies. First, as already mentioned, the eye-hand latency was much shorter in the sandwich making task, although the eyes still generally led manipulation. Second, there were many short duration fixations (< 120 ms) in the sandwich-making task, approximately 20% of the total. In tea-making, there were fewer than 1%. Third, about 13% of reaching actions

— mostly the setting down of objects — were ‘un-guided’ during sandwich making, that is, there was no fixation on the object or the set-down point prior to or during the action. For the tea making the proportion was less than 5%. It seems likely that all these differences may be explained by the fact that the sandwich-making was a sit-down task with all the objects needed for the task within reach, whereas in the tea-making the subjects were standing and mobile, with the objects located in different parts of the room. This inevitably imposed a slower tempo on the tea-making task, and also provided a less secure framework for dealing with objects ‘blind’.

The difference between the tasks is illustrated by the statistics of gaze-saccade sizes. Fig. 3 presents histograms of the sizes of gaze saccades made within ORAs (above) and between them (below). They show that within-action saccades have remarkably similar distributions for the two tasks, with a peak between 5 and 10° and few saccades less than 5° or greater than 20°. The between-action saccades, which transfer gaze from one object to the next, are similar for the two tasks in that they show no clear peak, but differ in that they do not exceed 30° in the sandwich-making, but reach up to 90° in tea-making. This difference is to be explained by the fact that the tea-making required movements around the room, some of which were accompanied by combinations of saccades that changed gaze by up to 180°. The largest single saccades that accurately targeted new objects (without correction saccades) had amplitudes of about 50°. The striking differences between the shapes of the between- and within-action histograms suggests that they are pro-

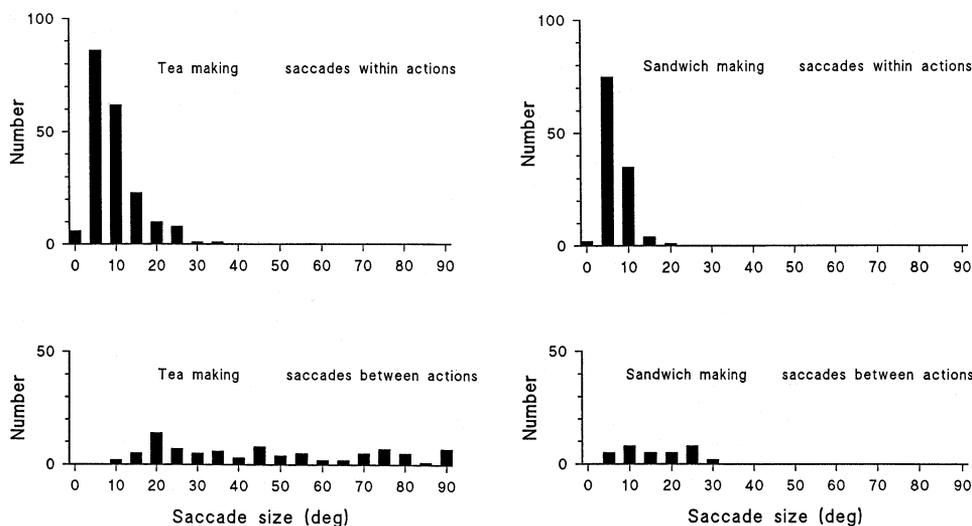


Fig. 3. Distributions of gaze saccade sizes during tea-making (one subject, left) and sandwich making (two subjects, right). ‘Within action’ saccades have a similar distribution in both tasks, with a mean size of about 8°. ‘Between action’ saccades (i.e. the first saccades to each new object in the task) are distributed more widely, with no clear peak. This reflects that objects involved in the tasks are scattered more or less at random over the field of action. This is the whole room in tea-making, but only a small table in sandwich-making, hence the difference in the widths of the distributions.

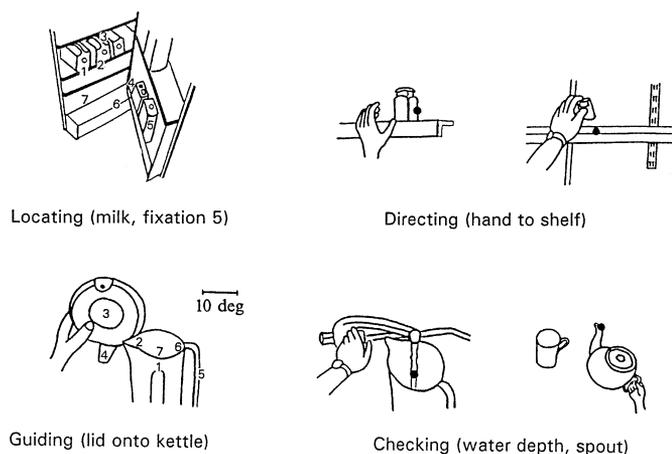


Fig. 4. Illustrations of the four identified roles of fixations: locating, directing, guiding and checking. Details in the text.

grammed in different ways: between-action saccades are largely determined by instructions from the script specifying the next object in the sequence. Within-action saccades, by contrast, may involve more directly visual cues, or simply an itch-like need to move the eyes a short distance, two or three times a second.

### 3. The functions of fixations during an object related action

By studying where the eyes fixate, and observing what happens in the period immediately after each fixation, one can get a reasonably clear idea of the roles of particular fixations in ongoing behaviour. In the tea making study we found that about one-third of all fixations could be clearly linked with subsequent actions. The remaining two-thirds were made after an action had been initiated so that although they may well have had similar functions in guiding action, it was less clear how they were related to changes in motor behaviour. In the sandwich-making task, which was executed faster, an even higher proportion of fixations had clear functions (Fig. 1).

Where the functions of fixations could be established we found — in both studies — that they could be classified into just four categories, which we designated as locating, directing, guiding and checking (Fig. 4).

1. *Locating*. On entering the kitchen each of our three tea-making subjects spent an initial period of about half a minute looking around the room, with gaze resting on objects that would subsequently be used in the task. During the tea-making task, which typically lasted about 4 min, there were several other periods of similar behaviour each lasting a few seconds. During this ‘surveying’ behaviour the location of some objects to be used in the future was established. This was evident from the fact that

subjects oriented accurately to objects behind them, when these had been fixated a minute or more earlier. In another example, the milk was located in the fridge, although it was not removed until over a minute later. Thus some fixations are concerned with establishing the locations of objects, even though there is no associated motor activity at the time of the fixation.

2. *Directing*. Many actions begin with a movement of the hand to contact an object. These are nearly always preceded by a fixation on the object (there were one or two cases where an object was contacted while the eyes were looking elsewhere; presumably this was done from memory). Typically only a single fixation is involved, and the eye usually moves away from the object just before the hand reaches it. Thus the grasp itself is often not executed under visual feedback. It seems that the main function of the directing fixation is to provide fovea-centred goal-position information for the motor system of the arm, which then concludes the movement in a (visually) open-loop manner. Some information about the shape of the object to be grasped is probably also obtained, as the hand ‘preshapes’ on its way to the target. Another example of a directing movement is putting an object down. As in grasping it is the destination that is fixated, in this case the place on the bench or shelf where the object will be put down.
  3. *Guiding*. Manipulations often involve two objects, for example a kettle and its lid, where both objects have to be guided relative to each other so that they make contact in an appropriate way. Most tool use is of this nature (e.g. spanner and nut). It is more complicated than simple directing, and this is reflected in the associated eye movements. As the two objects approach a number of fixations are made, often alternating between the two objects, and the action is usually completed under visual control. Some guided actions may involve more than two objects, for example knife, bread and peanut butter.
  4. *Checking*. Actions frequently terminate when some condition is met: the kettle is full, the water boils, the top is off the bottle. These checking operations require the eye to dwell on some appropriate region of an object, either in one long fixation or in a series of repeated fixations. When the particular condition is met, for example the water reaches the right level, an action is triggered that is not necessarily directly connected with the site being checked. In the example given the tap is turned off although gaze is directed into the kettle.
- The locations that are not fixated during a manipulative task are of almost as much interest as those that are. It seems that the oculomotor system has a strategy

involving a set of priorities, which insure that vision is used as economically as possible. The impression one obtains from the records is that vision is a scarce and valuable resource, and it is disengaged from a particular aspect of an action as soon as another sense is available to take over. Three types of restriction were universal.

1. The hands are hardly ever fixated. When an object is to be picked up or set down it is the object or its destination that is fixated, not the approaching hand (Fig. 4). Presumably proprioceptive information is adequate to locate the hand and fingers precisely.
2. Once the hands have acquired an object, it is no longer fixated (e.g. the taps in Fig. 1). It seems that touch and proprioception take over from vision as soon as grasp is established.
3. When a task involves pouring liquid it is always the destination vessel that is fixated, not the source. This seems to be common sense: flow can be judged from either vessel, but the receiving vessel has to be monitored to terminate the pouring.

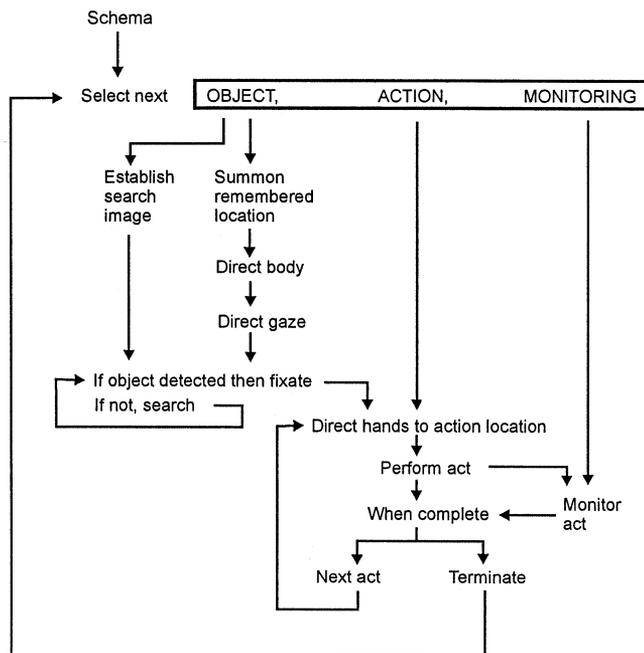


Fig. 5. Scheme illustrating the flow of information during an ORA. At the beginning of each ORA the schema (or script) must supply the visual system with information about the identity of the next object, the oculomotor system with information about its location, and the motor system about the action(s) to be performed. The visual and oculomotor systems also require information about the monitoring actions to be performed. The action commences once the object is located, and terminates when the monitoring indicates that the appropriate condition is fulfilled. Further actions on the same object may be initiated, or the action terminated and the script consulted again for the instructions for the next action.

#### 4. What information does the eye movement system need?

Since eye movements usually precede hand movements, the oculomotor system must get the new operational instructions before the motor system of the arm and hands (although later than the locomotor system of the trunk, see Fig. 2). To perform an ORA the relevant object needs to be specified and located. For the visual system this means that the appropriate search image needs to be invoked so that recognition can occur. If the object has been located previously its remembered position in space needs to be translated into an appropriate combination of body and eye movements. The spatial precision of this system is unknown, but no doubt the length of time since the last locating fixation has some effect. We have one or two clear examples where an object was fixated a minute prior to its ultimate use, and the ‘memory’ saccade got to within about  $20^\circ$  of the object, to be followed by one or two visually guided saccades which accurately fixated the target. Thus object localization is often a two-stage process, with place memory being used to get the eyes close enough for direct object detection. In some cases, where an object has not been located previously, a simple scanning strategy was also invoked, but this was fairly uncommon.

With the object located and transferred to the fovea, manipulation can begin. The motor system of the arms and hands needs to know what it is to do — the verbs that go with the noun. As far as the visual system is concerned this may involve directing and guiding operations, but more especially most manipulations require some sort of visual checking (not all; however, some simple operations such as the removal of a lid can be done without any visual involvement). Thus the visual system needs to know, as part of its information package, where it must look and what it must find, as it establishes whether or not an action is proceeding properly, or is complete.

We have tried to represent the information flow during a single ORA in Fig. 5. On (or before) the end of the preceding action the central executive (supposedly in the frontal lobes) supplies the oculomotor system and the motor systems of the trunk and arms with information about the object (identity and location), the manipulations to be performed and the checking operations required. First the eyes need to locate the object — either by a combination of place memory and direct vision, or by a search routine. When the object is located manipulation is initiated, watched over by the eyes which check that key operations are completed. When each sub-act is complete, a further manipulation on the same object may follow (e.g. take down sweeteners, add one to cup, replace on shelf), or if there are no further sub-acts to perform the overall action is termi-

nated, the present instructions are abandoned and new ones for the next object related action are sought.

## 5. Conclusions

Although object related actions, or AIs, are the lowest level in the hierarchy of ‘schemas’ that contribute to the structure of a complex task (Norman & Shallice, 1986), our eye movement studies show that they have unexpected complexity. Rather than simply being stereotyped components in an automated sequence, these actions involve object identification, place memory, task execution and a surprising amount of monitoring. As those of us who make tea in the mornings know, this involves minimal conscious awareness or intervention: typically we are listening to the radio. A distinction is commonly made between unconscious automatic actions that proceed without feedback, and consciously controlled actions that are subject to monitoring (e.g. Underwood & Everatt, 1996). In our experience this doesn’t seem to be a useful distinction in this context, since the kinds of object related actions we observe are barely conscious but nevertheless monitored by the eyes in a variety of ways. Where the conscious/unconscious distinction may have a much more important role is in the learning of new motor skills (such as learning to type or play the piano), and it will certainly be of interest to see how the patterns of eye movement change as such skills are acquired.

We were surprised by how explicitly eye movements were related to actions. Hayhoe (2000) concluded ‘The role of vision from moment to moment is determined almost exclusively by the current stage in accomplishing the task. There appears to be little room for other functions.’ This ‘do it where I’m looking’ strategy also applied in the somewhat less natural block-copying task of Ballard, Hayhoe, Li, and Whitehead (1993), where every movement and checking operation was achieved via an eye movement, with little or nothing held over in memory across fixations. In the current studies there is some carry over of information from earlier to later parts of an extended task, as shown by the accuracy with which previously fixated task-relevant objects

(which may be outside the subjects’ immediate field of view) are re-fixated. However, our studies lend no support to the idea that the visual system builds up a detailed model of the surroundings and operates from that. Most information is obtained from the scene as it is needed.

## Acknowledgements

MFL is grateful to the Gatsby Foundation (UK) for support, and to Neil Mennie for help with data analysis.

## References

- Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1993). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London B*, *337*, 331–339.
- Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*, 661–721.
- Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society of London B*, *266*, 2435–2440.
- Hayhoe, M. (2000). Vision using routines: a functional account of vision. *Visual Cognition*, *7*, 43–64.
- 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.
- Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro, *Consciousness and self-regulation: advances in research and theory*, vol. 4 (pp. 1–8). New York: Plenum.
- Schwartz, M. F., Reed, E. S., Montgomery, M. W., Palmer, C., & Mayer, N. H. (1991). The quantitative description of action disorganisation after brain damage: a case study. *Cognitive Neuropsychology*, *8*, 381–414.
- Schwartz, M. F., Montgomery, M. W., Fitzpatrick-DeSalme, E. J., Ochipa, C., Coslett, H. B., & Mayer, N. H. (1995). Analysis of a disorder of everyday action. *Cognitive Neuropsychology*, *12*, 863–892.
- Schleidt, M. (1988). A universal time constant operating in human short-term behaviour repetitions. *Ethology*, *77*, 67–75.
- Underwood, G., & Everatt, J. (1996). Automatic and controlled information processing: the role of attention in the processing of novelty. In O. Neumann, & A. F. Sanders, *Handbook of perception and action*, vol. 2 (pp. 185–227). London: Academic Press.