
The roles of vision and eye movements in the control of activities of daily living

Michael Land, Neil Mennie, Jennifer Rusted

Sussex Centre for Neuroscience and Laboratory of Experimental Psychology, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK; e-mail: M.F.Land@sussex.ac.uk

Received 4 May 1999, in revised form 9 August 1999

Abstract. The aim of this study was to determine the pattern of fixations during the performance of a well-learned task in a natural setting (making tea), and to classify the types of monitoring action that the eyes perform. We used a head-mounted eye-movement video camera, which provided a continuous view of the scene ahead, with a dot indicating foveal direction with an accuracy of about 1 deg. A second video camera recorded the subject's activities from across the room. The videos were linked and analysed frame by frame. Foveal direction was always close to the object being manipulated, and very few fixations were irrelevant to the task. The first object-related fixation typically led the first indication of manipulation by 0.56 s, and vision moved to the next object about 0.61 s before manipulation of the previous object was complete. Each object-related act that did not involve a waiting period lasted an average of 3.3 s and involved about 7 fixations. Roughly a third of all fixations on objects could be definitely identified with one of four monitoring functions: *locating* objects used later in the process, *directing* the hand or object in the hand to a new location, *guiding* the approach of one object to another (eg kettle and lid), and *checking* the state of some variable (eg water level). We conclude that although the actions of tea-making are 'automated' and proceed with little conscious involvement, the eyes closely monitor every step of the process. This type of unconscious attention must be a common phenomenon in everyday life.

1 Introduction

Most of the actions that make up our lives involve vision, if we are normally sighted. We have to locate objects, change their positions, manipulate them in various ways, all presumably under visual control. We have little or no conscious knowledge of where our eyes are fixating at any instant, and because of this we tend to think of the eyes as passive receivers, rather like cameras, taking up and passing on the information required for the particular task. However, scrutiny of the eyes of someone engaged in a complex motor task shows that this is not the case. The eyes dart from one place to another, two or three times a second. In this study we ask the question: are these eye movements essentially random, or are they intimately related to the requirements of the motor task? If the latter is true, are fixations directed specifically to the places from which information is needed, and can the eye-movement pattern thus be thought of as an integral part of the motor program itself?

In activities where movements of the eyes have been much studied, such as reading (O'Regan 1990; Rayner 1995), music reading (Weaver 1943; Land and Furneaux 1997), and steering a car (Land and Lee 1994), the strategy of the oculomotor system is to keep the centre of gaze very close to the point at which information is extracted—within a few letters or notes in text and music reading, and within a few degrees of the inside of the bend (the tangent points) when steering on a winding road. In an artificial task that involved the assembly of a copy of a pattern made of coloured blocks, Ballard et al (1992) and Hayhoe et al (1998) found that every action—choosing a block, checking its colour, finding its proper position—involved a new fixation, with eye movements generally preceding motor acts by a fraction of a second. Eye movements thus seem to be quite tightly coupled, temporally and spatially, to the motor actions of the particular task.

In this paper we ask to what extent this is true for everyday activities, and we have chosen to study the rather archetypal task of making tea (Rusted et al 1995).

This differs from the tasks discussed above in that it is not repetitive. Each of the 40–50 acts involved in tea-making is unique, and so requires both a different motor program and presumably a different pattern of eye movements as well. Thus we can expect to learn not only how eye movements are involved in single acts, but how these acts join together. A key question here is to what extent the eyes are proactive and lead the motor program, seeking out information that will be required in the near future, or whether they are essentially reactive, being called up by the motor program when specific information is required. We will show that visual fixation does indeed precede motor manipulation, putting eye movements in the vanguard of each motor act, rather than as adjuncts to it.

To our knowledge there have so far been no studies of the eye movements involved in activities of daily living, in a natural setting. This is no doubt because, until recently, most eye-movement recording devices were not suitable for use outside the laboratory. Here we have employed a lightweight head-mounted eye camera used previously for driving studies (Land and Lee 1994). It provides a view of the scene ahead, with a dot that indicates foveal direction, as well as the coordinates of eye direction relative to the head (figure 1).

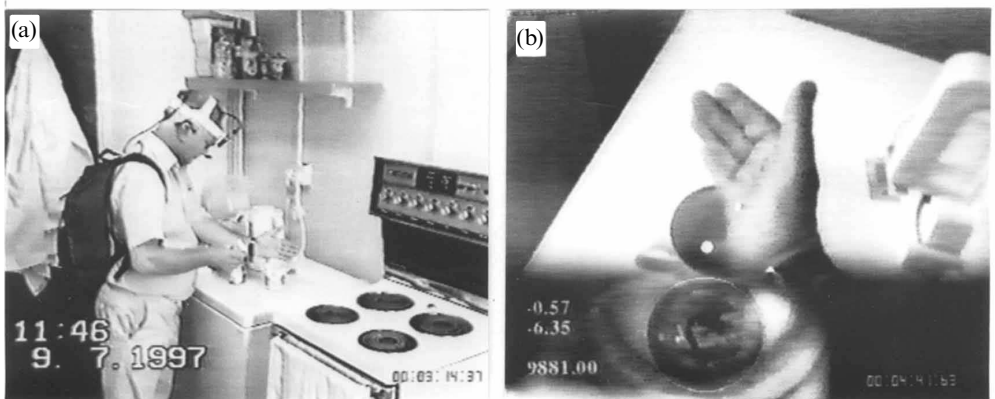


Figure 1. Prints from (a) the activity video, and (b) eye-movement video of the same instant, when the sweetener is dropped into the mug (3.14 on figure 3). The head-mounted camera and backpack are shown in (a). In (b) the white dot is the direction of regard of the fovea (into the mug). The eye can be seen in the bottom third of the frame, with a bright ellipse fitting the iris. The angular width of upper part of the frame is approximately 35 deg. Note that (b) is right–left reversed compared with (a) because of the mirror in the camera system.

Tea-making, at least for the British population, is a very familiar, overlearned activity involving a sequence of actions which typically occur in a particular order. Such automated activities, comprising ‘stereotypic sequences of actions’ (Schank and Abelson 1977), are sometimes supposed not to require feedback for their execution (Underwood and Everatt 1996). Once learned, they cease to require on-line monitoring or supervisory attention (Norman and Shallice 1986; Shallice 1988), and run directly from a memory ‘script’. The principal conclusion from the study presented here is that almost every act in the tea-making sequence is guided and checked by vision, with eye movements usually preceding motor actions. Thus in a literal sense it is not true that these supposedly automated acts are under open-loop control, and it opens up the question of what is meant by ‘supervisory attention’, when the eyes are manifestly attending to the objects involved in automated behaviours.

Preliminary accounts of this study have appeared elsewhere (Findlay 1998; Land et al 1998; Hayhoe and Land 1999).

2 Methods

Three subjects (ML, male, aged 55; SF, female, aged 28; JB, male, aged 46) each made a cup of tea in a small rectangular kitchen in the University. The room had a worktop (counter) on the left of the door, a sink straight in front, and another worktop and refrigerators on the right (see figures 1 and 2). The subjects had seen the kitchen once on a previous day, but the positions of the various objects and utensils—kettle, teapot, etc—had been changed, so that some search was needed before each could be used.

Actions were monitored by a video camera to the right of the sink. Eye movements were recorded with a device previously used in driving studies (Land 1993; Land and Lee 1994). It consisted of a small video camera (ELMO) mounted on a band from a construction helmet clamped firmly to the head. The camera's field of view was split optically

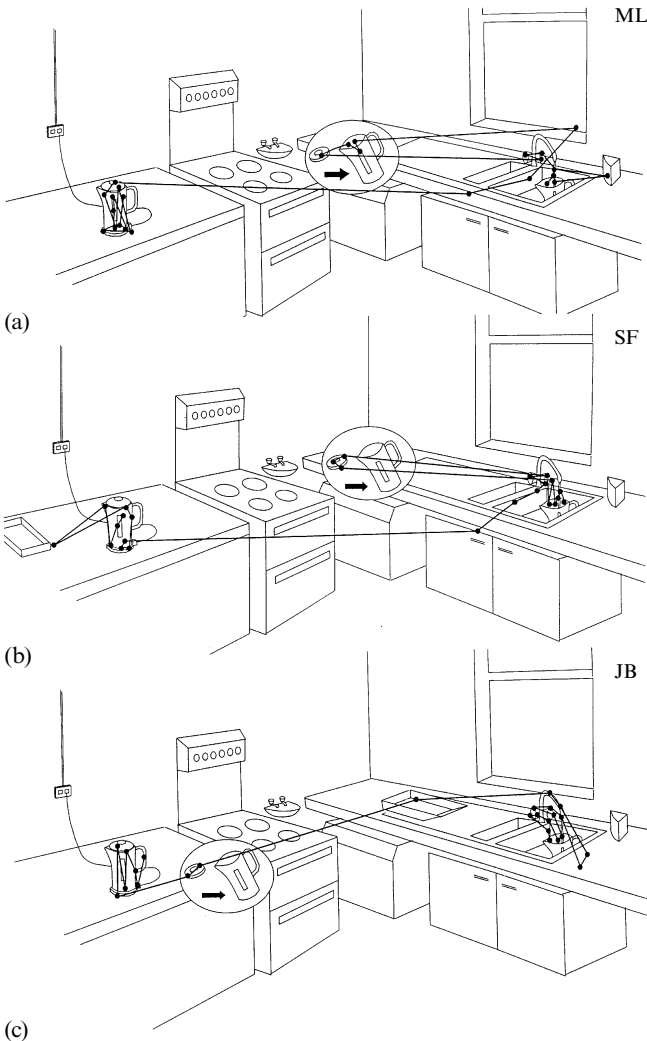


Figure 2. Record of the fixations made by three subjects during the first sequence after the kettle is first detected (0.05–0.20 on figure 3), and during which the kettle is moved from the worktop (left) to the sink (right). Because of the changing viewpoint, the angular relations are only approximate, but fixation positions relative to the objects of regard are accurately represented. Note the associations of fixations with particular objects or other entities—kettle, sink, kettle and lid, taps, water stream—which correspond in time to the actions that relate to them. Note also the rough correspondence in the numbers of fixations that are devoted by each subject to corresponding objects.

so that the upper two-thirds imaged the view ahead via a part-silvered mirror in front of (but not obscuring) the left eye (figure 1b). The eye and camera lens had the same virtual location, giving a parallax-free image. The lower third of the field imaged the eye itself via a concave mirror (10 cm focal length) in front of and slightly below the eye. In this way the scene ahead and the eye with its various movements were recorded together onto a portable video recorder (SONY Video Walkman) housed in a backpack. Later, the tape was played back frame by frame, via a mixer that added a computer-generated model of the eye in which the outline of the modelled iris was marked with a bright ring. This model could be manipulated with a tracker ball, so that the model iris fitted the real one. With proper calibration this fit meant that the angular coordinates of the model corresponded to the direction of view of the eye itself, relative to the head. These were stored, and also used to generate a spot that was superimposed onto the 'head's eye' view, which showed the direction of gaze on the scene to an accuracy of about 1 deg (roughly the width of the spot on figure 1b). Each frame, with the spot on, was re-recorded onto a second videotape with a single-shot recorder (Panasonic AG-6720), and this second-generation tape was used for all subsequent measurements.

The two tapes (activities and eye movements) were synchronised, and examined frame by frame to determine the timings of all actions of the body and limbs and manipulations of objects, and all eye movements, to an accuracy of better than 0.1 s, during the 4 min it took to make the tea. In this paper one session from each subject has been analysed in full, a total of about 36 000 frames. One video is available on the web site: http://www.biols.sussex.ac.uk/Home/Jenny_Rusted/ as well as the journal web site <http://www.perceptionweb.com/perc1199/land.html>, and will be archived on the CD-ROM distributed with issue 12 of the journal.

3 Results

3.1 *Levels of description*

It is possible to divide up the tea-making task in many different ways. The operation can be thought of as a control hierarchy in which the largest units describe the goals and subgoals of the operation. For example the overall (level 1) goal 'make the tea' can be divided into subgoals (level 2) 'put the kettle on', 'make the tea', 'prepare the cups'. These comprise smaller (level 3) acts such as 'fill the kettle', 'warm the pot', which themselves can be broken down into component actions. Taking 'filling the kettle' as an example, the individual actions involved are 'find the kettle', 'lift the kettle', 'remove the lid', 'transport to sink', 'locate and turn on tap', 'move kettle to water stream', 'turn off tap when full', 'replace lid', 'transport to worktop'. These 4th level acts seem to be irreducible, in a functional sense.

There is, however, a fifth level of subdivision, represented by eye fixations (figure 2). Our eye-movement recordings show that each of the nine level-4 acts, just listed, involves an average of 5.4 fixations (range 4.7–6.3). The fixations are not, in general, directly synchronised to specific actions of the limbs (although certain fixations may have quite specific functions within each level-4 act—see 'specific roles of fixations' below). What is clear, however, is that almost all fixations that are made while a level-4 act is in progress are directed to the object or objects involved in that act. Thus in figure 2a the subject spends 12 fixations on the kettle, 3 on the area of the sink before approaching it, 4 on the kettle while removing the lid in transit, 3 on the taps, and 4 on the water stream, with one apparently 'irrelevant' glance to the sink tidy. In spite of minor differences in the order of the component acts, the other two subjects showed remarkable similarities in the way their eyes covered the same operations. In particular the numbers of fixations devoted to each object were similar: kettle 7–12, lid 2–4, taps 3–4 (visited twice by JB). The two apparently irrelevant fixations to the right on figure 2c were in fact to the spot at which JB was about to put the lid down.

To the extent that one can speak of 'natural' units of behaviour, the best candidates seem to be these conjunctions of visual fixation and manipulation, linked to objects. Within each unit, visual involvement and action last about the same time, although there are slight differences in timing at the beginnings and ends of each unit (vision typically leads action by a second or less—see figure 4). At either end of each unit there is nearly always a clearly identifiable large saccade that switches gaze from one object to another, but within the unit gaze rarely strays from the object of interest; fewer than 5% of fixations were to 'irrelevant' parts of the visual field (examples are the saccade to the sink tidy on the far right of figure 2a, and to the tray at the left in figure 2b). During each unit, all one's sensory-motor equipment appears to be involved with a single object-centred task. In figure 3 the complete record for ML has been divided in alternating greys in a way that emphasises these units. We will refer to them as 'object-related actions' (ORAs) from now on, and define them as the sets of acts (including eye movements) associated with the current object of fixation.

ORAs are generally the same as the level-4 units discussed earlier, but may sometimes be more like level-3 actions. This happens when several things are done to an object without re-fixation onto a new object, as for example when the sweeteners were found, used, and returned to the shelf (figure 3, 3.09 to 3.16). They also correspond quite closely to the A-1 units of action described by Schwartz et al (1991, 1995), which were simple manipulations related to objects (see section 6). Our ORAs are not identical to A-1s because they include fixations as well as actions (the beginnings and ends of ORAs are defined by the timings of the saccades to and from particular objects), but ORAs share with A-1s the sense of linkage to objects.

Whilst the majority of ORAs are simple visuo-motor conjunctions as described, a few are more complex. Some actions are 'embedded' in others. For example, between 0.11 and 0.14 (figures 2a and 3) the kettle lid is removed as the kettle is being taken to the sink, with the lid being viewed between fixations on the sink. On a very few other occasions two actions really are performed at the same time. For example, between 3.30 and 3.40 one hand puts the top on the milk while the other swirls the teapot (see figure 8k). In this case, gaze alternates between the teapot spout and the milk. Clearly, time-sharing is possible, but it is not very common. There were also occasional instances, especially in JB's record, where the eyes would engage an object and the hands contact it, but where fixation was interrupted by a 1–2 s spell of search before manipulation was resumed. Thus the ORA classification of acts is not entirely without problems, but it is certainly the dominant pattern to emerge from the data.

In contrast to the clear temporal boundaries of the ORAs, there is nothing special in either the visual or motor domains that distinguishes the beginnings and ends of the larger (level 2 and 3) units from each other. The logical status of these operations in the goal structure of the task does not seem to be reflected in the way they are executed or terminated.

Three classes of activity involve vision but no action. The eyes may move across the scene simply to relocate gaze from one place to another, and where the total movement is large enough (eg 180° from one counter to the other) this may be done either in a single large (gaze) saccade, or a series of somewhat smaller ones. There is no indication here that the visual system is looking for anything, and often the vestibulo-ocular reflex is suspended during these large saccades, with gaze direction being carried passively by the head rotation (this conclusion is based on patterns of eye-head coordination observed in this study). In other cases a very similar pattern of eye movements is associated with true search, as in the search between 0.48 and 0.52 (figure 3) when the teapot is located, and then between 0.53 and 0.57 when the tea caddy is found. Thirdly, the eyes may be looking 'aimlessly' around, as for example in two episodes of looking out of the window while waiting for the tap water to warm up (1.16 to 1.27).

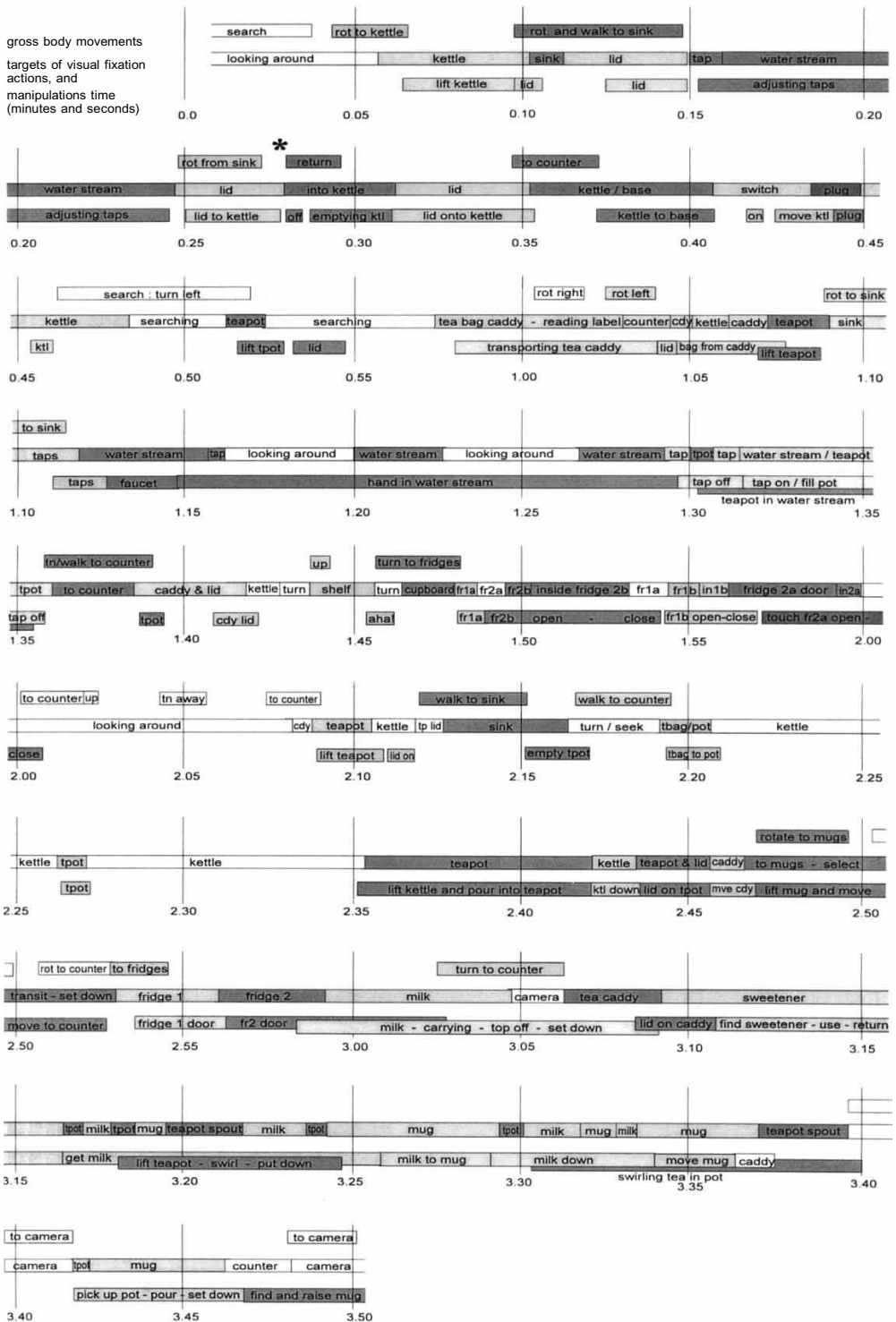


Figure 3. Complete sequence of visual and motor events during a single tea-making session by subject ML. In each set of three rows the uppermost row shows the durations of gross body movements, the middle row shows the object or objects fixated by the eyes, and the bottom row the objects being manipulated by the hands. There is a close correspondence throughout between the object that is fixated and that manipulated, as indicated by the corresponding shading in the lower two rows, or all three rows when a gross movement is also involved. Unshaded regions indicate search, or other forms of looking around, where no action is involved. The asterisk at 0:28 indicates a vocal instruction (see text); otherwise the whole task is self-paced. The time scale is used in the text to refer to particular actions.

4 Time relationships of vision and motor acts

Each ORA, as defined above, involves visual engagement with an object, and manipulation of that object. It may, in addition, require gross body movement if, for example, the object is in a different part of the room. What comes first? Do the eyes seek the object before action starts, or are vision and action simultaneous? How do the body movements fit in? We examined all ORAs in the three records (137 in total), excluding only periods when two actions were occurring simultaneously (eg between 3.15 and 3.40 in figure 3). For the others, the times of the first movement of the body, of the first saccade to the relevant object, and of the first indication of a hand or limb movement related to that object were all determined. The average results are shown in figure 4. 42% of the ORAs involved whole-body movements, and only these contribute to the upper record. About 30% of the 43 ORAs were excluded from the histogram showing the timings at the end of the units, because they involved waiting for something to happen (eg for the kettle to fill). This prolonged them beyond the ‘natural’ time-course of the action itself.

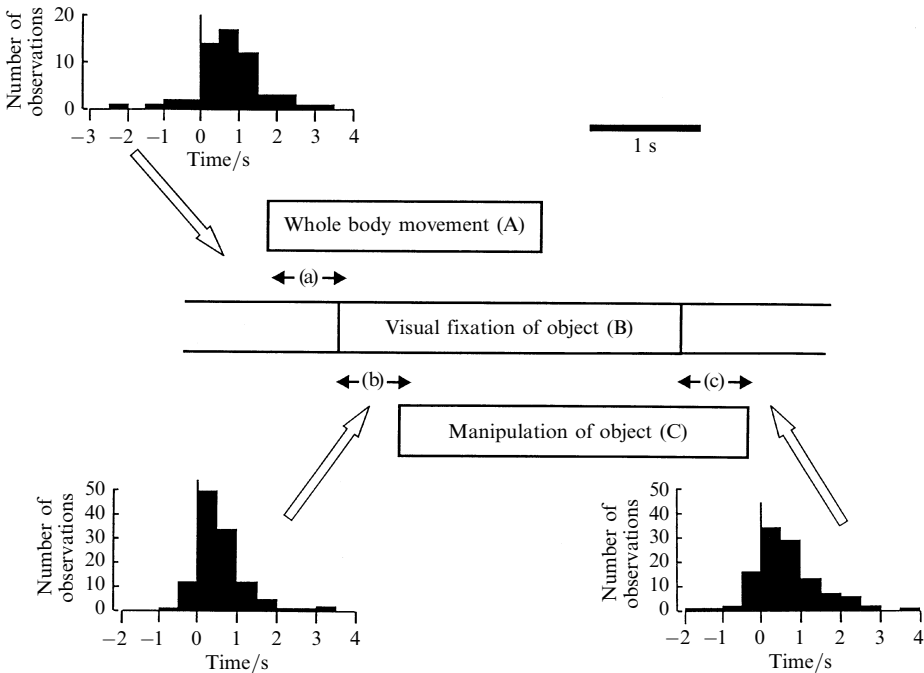


Figure 4. Average relative timings of body movements (A), visual fixation of particular objects (B, which may involve several individual fixations), and manipulation of those objects (C). Note that body movements lead vision (a), which leads manipulative actions (b), in each case by an average of about half a second. Typically, gaze moves on to the next object before completion of manipulation of the current object (c). All times are taken from the centre point of the first saccade to the new object (line 2), as this is the best-defined instant in the sequence. Pooled data from all three subjects; individual statistics are provided in table 1.

Figure 4 shows that the average duration of the visual component of an ORA is 3.04 s and the motor component 3.25 s. The first sign of a progression from one unit to the next is often a gross movement of the body; the trunk, it seems, gets the new instructions first. On average, the beginning of the trunk movement precedes the first saccade to the object to be manipulated by 0.61 s, and that saccade precedes the first sign of manipulation (eg the beginning of a movement of the arm towards the object fixated) by 0.56 s. As the histogram shows, most of these delays lie between 0 and 1 s.

At the end of each ORA, gaze typically moves on to the next object between 0 and 1 s before the last motor act has been completed (mean 0.61 s). These observations are consistent with the idea that vision starts to supply information to a buffer up to a second before that information is used, and that the buffer continues to disgorge information for up to a second after visual input to it has ceased (Land and Furneaux 1997). In single acts of pointing and reaching discussed by Jeannerod (1988) the eyes led the hand, but by a much shorter time (< 0.2 s), possibly because little or no buffering is involved in simple acts of this kind, or because repeated tasks require less visual processing.

Table 1 gives details of these timings for each of the three subjects. Again there is remarkable consistency in the durations of the components of the acts, and in the leads and lags involved. In particular, all three sets of timings conform to the sequence depicted in figure 4. Statistically, none of the lead and lag differences between individuals reached a 5% level of significance (Student's *t*-test), which would require a difference in the means of about 0.3 s. However, in the durations of the components, ML and SF differed significantly from each other ($p < 0.05$) on one of the three pairings, but both ML and SF differed from JB on two of the three pairings. JB's ORAs tended to be somewhat longer, partly because of his tendency to insert brief episodes of search, as mentioned earlier.

Table 1. Durations of components of object-related acts and their relative timings [mean \pm standard deviation (number of observations)]. See figure 4 for meanings of (A), (a) etc.

	Duration of components/s		
	ML	SF	JB
Whole body movement (A)	2.73 \pm 1.17 (15)	1.82 \pm 0.71 (21)	2.86 \pm 1.38 (22)
Visual fixation of object (B)	2.60 \pm 1.37 (37)	2.80 \pm 1.80 (44)	3.83 \pm 2.09 (34)
Manipulation of object (C)	2.83 \pm 1.68 (32)	3.01 \pm 1.66 (41)	3.96 \pm 2.14 (33)
	Leads and lags of components/s		
	ML	SF	JB
Body leads fixation (a)	0.63 \pm 0.55 (16)	0.51 \pm 1.59 (21)	0.69 \pm 1.19 (22)
Vision leads manipulation (b)	0.57 \pm 0.69 (42)	0.43 \pm 0.57 (41)	0.68 \pm 0.97 (37)
Manipulation offset lags vision (c)	0.56 \pm 0.92 (41)	0.57 \pm 0.69 (41)	0.71 \pm 0.88 (36)

One episode is of particular interest because it involved a definite conscious intrusion into what otherwise seemed to be a well automated task. At 0.28 on figure 3 (asterisk), as the filled kettle was having its lid put on, it was pointed out to the subject that he was making tea for one only, not an army. After a short latency (it was not possible to tell exactly which word was the trigger) there was a change in the direction of body movement, a fixation shift back to the water in the kettle, and the beginning of lid removal. In contrast to the half-second asynchronies in ordinary self-paced acts (figure 4), the body movement, fixation, and manipulation change were all simultaneous here, to within 0.1 s.

5 Eye movements

5.1 Overall pattern and statistics

The distribution of saccade sizes for the whole of the three records is shown in figure 5a. In each there is a peak between 2.5 and 10 deg, with very few saccades smaller than this (1 deg saccades were readily detectable, so the absence of small saccades is real; microsaccades much smaller than 1 deg would have been missed, however, if present).

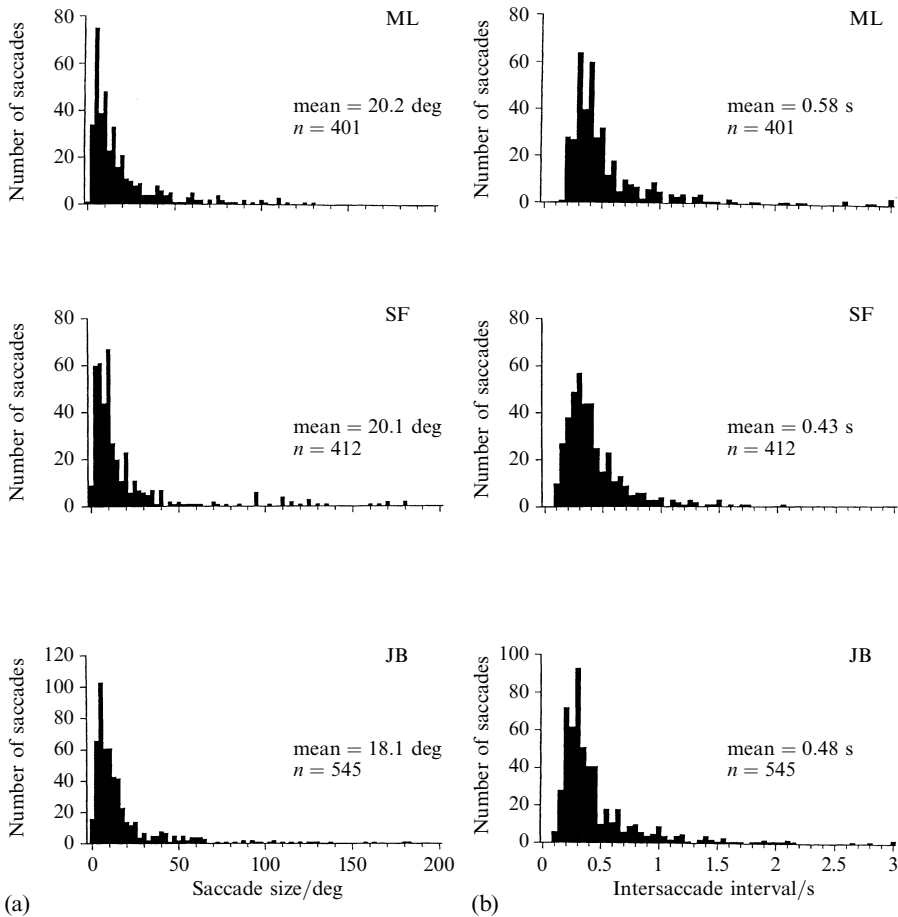


Figure 5. (a) Distribution of gaze saccade sizes (head + eye, ie the total movement of the line of sight) during each whole session. Bin width 2.5 deg. All three subjects show a peak in the range 2.5–10 deg, a long tail of large saccades, and a high mean value. Saccades of 1 deg or larger were detectable and are included, smaller saccades are not. (b) Histograms of intersaccade intervals throughout the three sessions. Although the mean values are close to 0.5 s, the modal values are about 0.3 s. The three distributions are very similar except that SF shows fewer very long fixations (> 2 s), and ML fewer very short ones (< 150 ms). Bin width 50 ms.

The distributions have very long tails of large saccades that take the mean size for the population to between 18.1 and 20.2 deg. Again, there is remarkable consistency in this figure. Compared with tasks such as text and music reading or picture search, this is a very high value [see Bahill et al 1975; typical saccades in reading are 7 letters long (O’Regan 1990), which for standard text at a reading distance of 40 cm is between 1 and 2 deg]. Roughly speaking, the smaller saccades in the 2.5–20 deg range deal with objects within an ORA, and the larger saccades transfer gaze between objects, or are involved in search (see figure 2; the distributions are not bimodal because the large relocating saccades have a wide range of amplitudes with no clear peak). It does seem that tasks involving substantial unrestrained motor activity make use of much larger saccades than tasks of passive scrutiny.

The distributions of intersaccade intervals (from the onset of one saccade to the onset of the next) is shown in figure 5b for all the saccades made during the three sessions. They have a profile typical for such distributions from other tasks (see Viviani 1990) with the majority of intervals in the range 0.2 s to 0.5 s, means of 0.58 s, 0.43 s,

and 0.48 s, respectively (0.49 s overall), and with long, roughly exponential tails. Since a typical saccade (20 deg, see figure 5a) lasts about 70 ms (Carpenter 1988), fixation durations are on average shorter by this amount (mean 0.42 s). These are still long by comparison with fixation durations during reading (0.2–0.25 s; Rayner 1995), but not far from the value 0.37 s found by Furneaux (1996) for music reading.

A factor tending to increase the mean is the comparatively large number of very long fixations that occur when the eyes are involved in checking actions. This is illustrated in figure 6, which shows the way intersaccade intervals are distributed in time during the session shown in figure 3. The most obvious features are the particularly long intervals associated with waiting tasks in which the state of some variable is being monitored. It seems that eye movements (1 deg or larger) are rarely made under these conditions, although we cannot rule out microsaccades (< 1 deg) during these long fixations. Otherwise there is little discernible structure in the record.

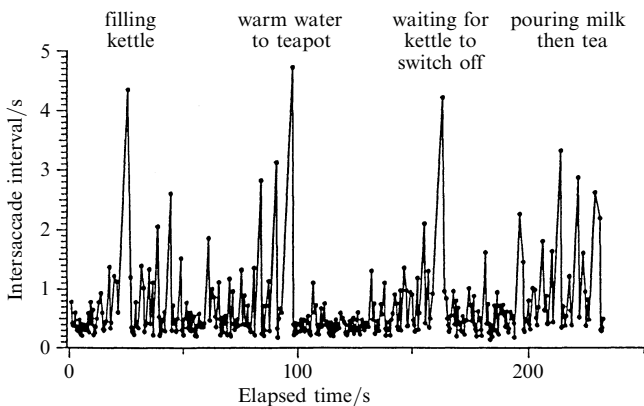


Figure 6. Distribution of intersaccade intervals during the whole session for subject ML (figure 3). The main feature is the small number of very long intervals associated with the monitoring of particular states, such as the amount of water in the kettle.

5.2 Specific roles of fixations

Figure 7 shows for the same subject as figure 3 (ML) the temporal pattern of saccades and fixations associated with the first part of the task—filling the kettle and switching it on. Careful inspection of the locations of the corresponding 80 fixations showed that 32 of them could be directly related to aspects of the task, because they preceded them by a second or less. These relations were of one of four types: ‘locating’, ‘directing’, ‘guiding’, and ‘checking’. They are defined as follows. ‘Locate’ is used for a fixation on an object (eg teapot, milk—see figure 8f) that will be used later in the process. It is assumed that the identity and location of the object are established in such fixations, but, although this is often demonstrably true, it may not always be the case. ‘Direct’ means that a fixation is to a location or a part of an object which is about to be approached and contacted by the hand, or by an object held by the hand. This kind of fixation appears to establish a vector for the motor system to use to control the approach (figures 8d, 8e, 8j). In many cases contact is not made until gaze has moved on from the target, implying that the final stages are not under closed-loop visual control. ‘Guide’ implies one, or frequently several, fixations between two objects (eg kettle and lid—figure 8c) that are approaching each other. Unlike ‘direct’ fixations, these provide the more or less continuous feedback required for successful docking, and are usually concerned with matching orientations as well as decreasing distances. ‘Check’ is used for fixation at a location where the state of some variable (eg water level) is being assessed (figures 8b, 8g, 8k, 8l). Presumably what is measured is the nearness of the value of the variable to a pre-established criterion level.

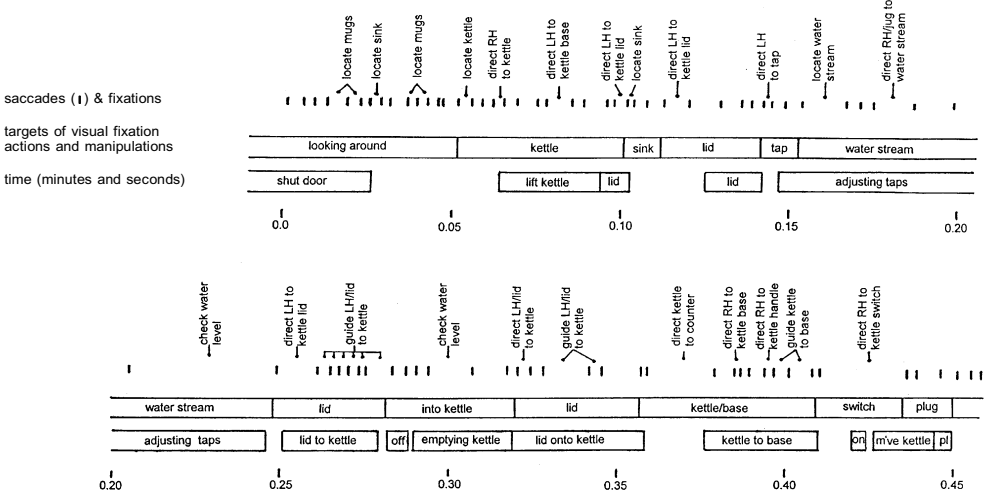


Figure 7. The specific roles of individual fixations in the first level-2 subtask ('fill the kettle') shown in figure 3. The pattern of saccades and intervening fixations are shown, and labels are given to those fixations which appear to have a clear function: for example, fixation on an object (eg mug, teapot) that will be used later in the process (= 'locate'); single fixation on a location or a part of an object that is immediately followed by an approach of the hand or object in the hand (= 'direct'); fixation—often multiple—on or between two objects (eg kettle and lid) that are approaching each other (= 'guide'); fixation at a location where the state of some variable (eg water level) can be assessed (= 'check'). Multiple fixations are made on every object while it is being manipulated (the exception is the kettle switch at 0.42 which only requires one fixation), but gaze rarely strays from the current object of attention.

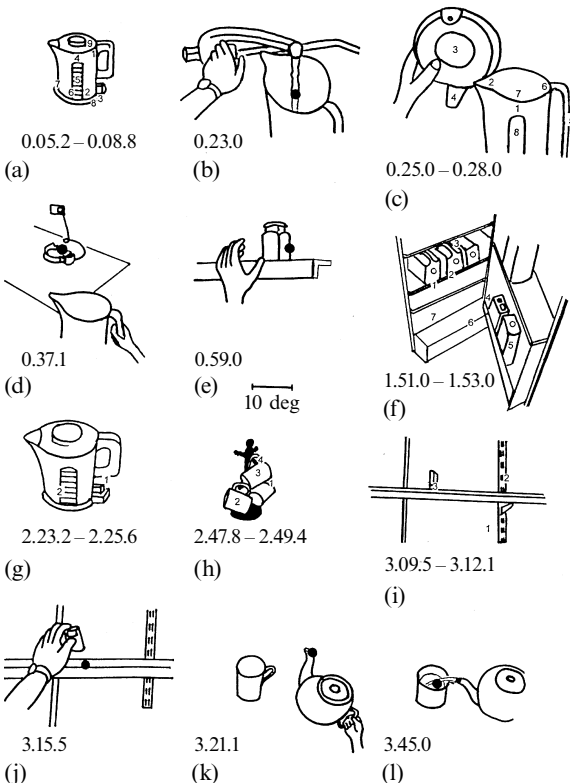


Figure 8. Examples of fixation patterns drawn from the eye-movement videotape. Sequences of successive fixation positions are indicated by numbers on the figures, and single fixations by single black dots. Numbers beneath each figure refer to timings in figure 3. 10 deg scale in centre applies to all figures. (a) Initial examination of kettle. (b) Tap control via water stream. (c) Fitting lid to kettle (drawing made at fixation 4). (d) Moving kettle to base: base is fixated. (e) Hand being directed to the tea-caddy. (f) Search around the inside of fridge 2. The tea-making milk is located at fixation 5. (g) Fixations checking the switch and gauge of the kettle when waiting for it to boil. (h) Selecting a mug. Hand goes to fixation 4. (i) Relocating sweetener prior to use requires 3 fixations. Sweetener last seen 68 s earlier. (j) Replacing sweetener 5 s after (i). Location on shelf is fixated first. (k) Swirling teapot: checking spout. (l) Pouring tea: receiving vessel fixated.

About another 25 fixations seem to be concerned with objects in similar ways to the 32 identified, although the links are less clear. Most of the remainder appear to be exploratory in the sense that they examine parts of objects that are not directly involved in actions (figure 8a). These proportions were similar for the other two subjects. During search behaviour many fixations are made to objects that are not subsequently involved in the overall task, and others are simply en passant stops made as gaze moves rapidly from one part of the scene to another. Overall, however, one has the strong impression that most eye movements are closely and purposefully linked to the ongoing actions. The oculomotor system clearly has a set of well-rehearsed and efficient strategies that enable the eyes to seek out the information required in the performance of every stage of the tea-making task.

5.3 Relocating objects

How does the eye find the next object to be dealt with in the overall sequence? In a number of instances that object had already been located some time earlier, so that, given perfect place memory, it should have been possible for gaze to be redirected to that object in a single saccade. This seemed to happen only for nearby objects and those that been fixated in the very recent past. For other objects it seemed that a two-stage process was involved, with place memory providing a coarse position, and local recognition completing the process. For example, when the milk was relocated at 2.58 (figure 3), having been first found at 1.52, the eyes and hand first went to the wrong fridge (at 2.53) and it took a total of 14 saccades from the time of the first relevant body movement before the milk was found.

The box of sweeteners was more easily found. This had been located twice during undirected searches (at 1.44 and 2.02 in figure 3) before being contacted and used (3.10) 68 s later. In fact, the final return at 3.10 was not quite exact (figure 8i): the eyes made a large (70 deg) saccade up from the counter, landing just below the shelf (i) followed by a saccade to the support (ii), and a final 18 deg saccade which landed accurately on the sweetener container (iii). In this case it seems that remembered coordinates were adequate to get the target within range of central vision, which was then able to complete the action using an appropriate search image. Interestingly, when the sweeteners were put back on the shelf, after an interval of only 5 s, gaze preceded their return to the exact point on the shelf from which they had come (figure 8j). This suggests a fast-decaying spatial memory of considerable initial accuracy. These initial observations suggest that a more complete analysis of all instances of returns to pre-located objects will be useful in exploring the interactions between place memory and direct visual detection.

5.4 What is not fixated?

The eyes cannot fixate everything relevant to the immediate task, and there appear to be some definite rules for determining which objects are viewed, and which are not. Three such rules are as follows.

1. The hands themselves are rarely fixated. If the hand is to contact an object, then that object is fixated, often at or close to the point of intended contact (figures 8e, 8h). Frequently, gaze has moved to a new object before the final part of the hand's trajectory is complete, so that the hand itself may not be fixated at all. Presumably, once direction and range are established visually, mechano-sensory information (proprioception and touch) are adequate to guide the hand to its target.
2. Objects that the hand has made contact with are rarely fixated again. For example, contact between the left hand and the cold tap (0.15; figure 3) is made while the taps are fixated, but thereafter, until the tap is finally turned off 10 s later, gaze is directed entirely at the water stream into the kettle (figure 8b). Interestingly, the water flow to

the kettle from the tap is altered several times in this period, as the kettle fills, but this is done without further visual reference to the tap.

3. Manipulation of certain familiar objects can be accomplished with no visual involvement at all. For example, after the teapot has been encountered and lifted at 0.52 (figure 3), the lid is contacted with the left hand and lifted from the pot while the eyes are looking around the room, as the body turns from one counter to another. In this and the previous example it appears that touch on its own is adequate to accomplish the task, without visual corroboration.

6 Discussion

6.1 *The monitoring role of the eyes*

The most surprising feature of the eye-movement record was the extent to which the eyes guide and monitor the performance of almost every component of the overall activity (see figures 8b–8e, 8j–8l). In tea-making, though not necessarily in all other activities, these monitoring actions could be divided into those concerned with finding objects or manipulating them ('locating', 'directing', and 'guiding'), and 'checking', where the state of some feature of an object (such as the water level) is measured. We have seen (figure 7) that about one third of all fixations could be clearly identified with the performance of one or other of these monitoring actions, and many other fixations were more loosely associated with them. It was possible to be sure about this because of the precision with which the eyes were directed. Objects that were about to be manipulated were nearly always foveated, with the centre of gaze moved, by a saccade, to within a degree or two of the appropriate part of the object or point in the scene. It was clear that the eyes do more than just register the scene passively in the manner of a camera; they seek out the places where the information they obtain will be of most value. Ballard et al (1993) described eye-movement strategies in a block-copying task as 'do-it-where-I'm-looking', and this epitomises the relationship between vision and action during tea-making.

The acts involved in a well-rehearsed activity like tea-making are generally regarded as having become automatic, that is their performance requires little or no supervision by higher, more conscious mechanisms (Norman and Shallice 1986; Underwood and Everatt 1996). This corresponds well with the subjective impression of the performance of the tea-making task; it seems to require very little conscious involvement for its successful completion, and one can indeed think about other matters whilst making tea. While this is perhaps less true in the present unfamiliar kitchen than in one's own kitchen, a novel environment (for such a familiar activity) will affect only the ease and speed of location of objects, and not the fluidity of the actions performed upon them. What we have found here is a large amount of detailed monitoring of actions, of which subjects were not consciously aware. Every component action was accompanied by appropriate visual feedback of some kind (figure 7), and this feedback is important. Closing the eyes at certain moments would—for a normally sighted person—be disastrous. A distinction is sometimes made between new activities that require closed-loop control, in the sense that the performance of individual components requires much individual checking, and automated activities where the control is open-loop, via memory, minimising the need for feedback (Underwood and Everatt 1996). We have found here that even highly automated activities involve both checking and closed-loop control. The difference between new and automated activities thus seems not to lie in the presence or absence of closed-loop control, but in the extent to which this control is consciously imposed. This study has shown that when an action has become 'automatic', it is not just the motor acts themselves that have become automated, it is the complete control systems responsible for their execution, which include sensory elements such as proprioceptors, touch receptors, and eyes.

6.2 *Natural units of action*

Memory for activities such as tea-making is generally assumed to be hierarchically structured (Brewer and Dupree 1983; van den Broeck 1988; Schwartz et al 1991, 1995). Individual actions in an event feed into higher-level units, which reflect the goal-directed nature of these actions. However, it is not clear how these higher-level units drive the execution of more basic subunits. The 'supervisory attention system' (SAS) of Norman and Shallice (1986) describes a hierarchical model in which high-level schemas represent and administer overall goals (such as 'make the tea') and the subgoals they comprise (such as 'boil the kettle'). These then structure the spatiotemporal pattern of lower-level schemas such as 'fill the kettle', 'warm the pot', etc via competitive interactions known as contention scheduling. Lower still are the constituent actions that execute these low-level schemas. These might be 'remove lid', 'turn on the tap', and so forth, which cannot be decomposed further functionally (they have to be completed to achieve anything) but which may nevertheless each involve a quite complex series of coordinated muscle movements. In addition to movements of the limbs and trunk, eye movements relevant to the task are made at rates of several per second. Thus a full account of an activity might require description at five levels or more. At the time of their production, are all these levels functional? What are the real sizes of the 'automatically activated' schemas?

In this study we have seen that for the most part the eyes look directly at the object being dealt with by the hands. Furthermore, the time relations between fixating and manipulating are quite tight and predictable (figure 4), with vision leading action by about half a second. Thus for periods of time lasting for a few seconds it appears that motor and sensory systems are linked to each other via single objects, or sometimes pairs of objects. The ubiquity of these sensory-object-action conjunctions leads us to believe that there is something basic about them in terms of the way the brain organises behaviour. A somewhat similar conclusion is implicit in the 'action coding system' used by Schwartz et al (1991, 1995), where units that they call A-1s are "simple actions that transform the state or place of an entity through manual manipulation". This description fits what we are proposing as 'object-related actions' (ORAs) rather well. Indeed we suggest that A-1 units are more than just actions, but, if examined in detail, would include equally well-defined systems of sensory monitoring, mediated predominantly, but by no means exclusively, by vision.

It is harder to define higher-order units of behaviour. Schwartz et al (1991) group A-1s into more inclusive A-2s, which are related to the subgoal structure of the particular task ('fill the kettle' describes a goal rather than the actions that achieve it). A-2s are likened to phrases in which the A-1s are the words, and they are rather similar to the low-level schemas of the Norman and Shallice (1986) formulation. However, in contrast to the object-related A-1s, there is nothing very obvious in the record of eye-movements and actions to distinguish A-2s. The transition from one object-related action is much the same whether or not a boundary is crossed from one subgoal to another. There are no obvious pauses as subgoals are achieved: there seem to be no full stops, although there are plenty of commas.

A similar conclusion was reached by Barsalou and Sewell (1985) in their analysis of action-based schemas. Hierarchical organisation may be imposed on the memory trace, but Barsalou and Sewell, amongst others, have argued that actions which comprise 'scripts' must be sequentially tagged in memory, since these temporal links and not the subgoals are prominent in verbal reconstruction of the script. In this study we have shown that higher-order units are not defined in the behavioural sequence either.

Is this always the case? Although there is the appearance of automaticity in the production of a routine sequence, there is evidence to suggest that the subgoal units are discrete and independent and may drive production sequences. Schwartz et al (1991)

described two patients with frontal head injuries whose injury resulted in loss of ability to perform simple routine activities with any degree of accuracy. In addition to gross object misuse (eg using a razor as a toothbrush) they exhibited severe disorganisation of level-2 units (subgoals). As recovery proceeded, subgoal organisation normalised. Subgoal disorganisation of this sort is not observed in the deteriorating performance of volunteers with Alzheimer-type dementia (Rusted et al 1995).

6.3 *What directs gaze?*

We have seen that in object-related actions the direction of gaze is very closely connected with the particular act. There were very few occasions where it seemed that the eyes were drawn to an object simply because of its 'salience' (ie its intrinsic ability to stimulate early parts of the visual pathway). Objects were fixated because of their relevance, not because they were big, bright, or visually exciting in other ways. This is not to deny a role for salience altogether: someone entering the room, for example, would certainly draw attention away from the task. Nevertheless, where one looks, in this kind of task where the visual environment is generally static, seems to be driven principally by the retrieved memory 'script' for the activity. There may be a real difference here between the way eye movements are organised in an unconstrained activity such as the free viewing of a picture and in a purpose-driven task such as tea-making. Salience is certainly important in free viewing, as in the famous 'Girl from the Volga' recording of Yarbus (1967), where the eye seems drawn to particular facial features, but a cognitively driven structure takes over when there is a job to be done—as in Yarbus's 'The Unexpected Visitor', where asking different questions produced radically different, task-related, patterns of eye movement. Stochastic salience-driven models may be appropriate for free viewing (eg Harris 1989), but for purposeful activities something more akin to the program-driven 'deictic' models proposed by Ballard et al (1992), or cognitively driven 'scan-path' models (Stark and Ellis 1981; Stark et al 1993) seem more appropriate (although the early idea that perception occurs as a direct consequence of the scan path is no part of the present proposal).

6.4 *The time scales of eyes and actions*

As pointed out by Newell (1990) and by Ballard et al (1998), simple tasks, such as saying a short sentence or playing a musical phrase, have a time scale of the order of 2 s. Schleidt (1988) found that spontaneous repeated acts by individuals from four different cultural backgrounds had modal durations of 3 s. In the block-matching task of Ballard et al (1982) the average time taken to choose and relocate each block was 1.7 s, and in a self-paced text-processing task involving a succession of read–type–check cycles the average duration of each cycle was 2.3 s (Fleischer and Strauss 1988). Here we find a very similar time for completing an object-related act (about 3 s, figure 4). It is thus tempting to think that there is some universality to this figure: that it has something to do with the intrinsic time scale over which the brain prefers to operate. The timings of all these tasks inevitably have some element imposed by the nature of the task itself, but it is more than suspicious that they seem to converge on this particular duration range.

Watching a videotape of a subject making tea, taken by an external observer, one sees the various actions merging into each other every few seconds, and this looks eminently normal. By contrast, the eye-movement video has events (saccades) roughly every half second (figure 7). It has the frenetic appearance of a movie that has been greatly speeded up. As a friend of ours, not involved with the study, put it after watching such a video: "No wonder I'm tired at the end of the day!" Whilst we recognise our actions easily in the observer video, we do not see ourselves as the actors in the oculomotor drama. This emphasises the fact that we really have no subjective insight into the way the oculomotor system operates on our behalf. Without our being aware of it, it employs an impressive knowledge base to superintend the operations that 'we'

are engaged in, and at a speed 5 or 6 times faster than the slow succession of actions that our consciousness seems to register.

6.5 *Unconscious attention?*

There is mounting evidence that eye movements and attention are closely linked at a neurophysiological as well as a psychological level. Attention shifts precede saccadic eye movements, are associated with their preparation, and involve some of the same neuronal machinery (Posner and Petersen 1990; Husain and Kennard 1996). Presumably the eye movements involved in the tea-making task have a similar relation to attention. However, these eye movements are made unconsciously in the monitoring of an essentially automated activity. In the schema system of Norman and Shallice (1986), automated activities are controlled by schemas which are “routine programs for the control of overlearned skills”. Conscious attention, in contrast, is a feature of the ‘supervisory attention system’, associated with the frontal lobe of the cortex, which has the function of handling nonroutine behaviours and intervening when problems arise in otherwise routine activities. “It functions by top-down activation or inhibition of schemata” (Stuss et al 1995).

Thus there is a problem in knowing whether or not to consider the monitoring that accompanies automatic and semi-automatic actions as being a kind of attention. We have no desire to become embroiled in what is already a hugely complicated field, but it does seem that unconscious monitoring has at least a claim to be considered as attentive (to the requirements of the task) and thus perhaps a form of attention. It seems that, even if the subjects themselves are not attending to the task, their oculomotor systems are.

We are probably not saying anything new. Posner and Petersen (1990) argued for a double system where a parietal attention system is concerned principally with visual orienting, and a frontal system that is more concerned with purposeful processing. It is possible that these two systems correspond in some way with the systems we see here, but this is not yet clear. An involvement of the parietal cortex in routine activities seems likely, as this is the region where sensory-motor transforms involved in the planning of action are known to occur (Andersen et al 1997). Nevertheless, the organisation of the overall task is under the control of the prefrontal cortex, as revealed by the effects of injury to this region (Schwartz et al 1995).

In summary, the results of this study clearly show that even automated routine activities require a surprising level of continuous monitoring. The picture we now have of the sensory-motor program of each ORA is quite complex. Activation of the program usually occurs on the termination of the previous activity in the low-level schema, although the variability that we observe between sessions indicates that the execution of the task is not in any absolute sense a “stereotypic sequence of actions”. To accommodate this, we have suggested that actions are tagged with relational probabilities that reflect the habitual ordering of a particular subject’s routine (Rusted and Sheppard, in preparation), but that this is not an essentially invariant component of even lower-order schemas. Once the ORA program is activated, memory must supply the identity of the relevant object, some information about its location (which may come from long-term memory in a familiar situation, or very recent memory in a novel one), details of the motor pattern to be executed, and details of the monitoring arrangements that are needed for its execution, including the places to which the eyes must be directed to provide the relevant information. In terms of current neurophysiological thinking (eg Milner and Goodale 1995) this will involve many parts of the cortex: the frontal lobes for overall coordination, the parietal lobes for eye–hand coordination, the temporal lobes for object identification, and the hippocampus for place memory. This illustrates the complexity of the coordination required for even the most overlearned activities.

Acknowledgements. We are grateful to the Wellcome Foundation for support for behavioural research into everyday routines (JR), and the BBSRC (UK) and the Gatsby Foundation for funding for the development of the eye-movement monitor (MFL).

References

- 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" *Annual Review of Neuroscience* **20** 303–330
- Bahill A T, Adler D, Stark L, 1975 "Most naturally occurring human saccades have magnitudes of 15° or less" *Investigative Ophthalmology & Visual Science* **14** 468–469
- Ballard D H, Hayhoe M M, Li F, Whitehead S D, 1992 "Hand-eye coordination during sequential tasks" *Philosophical Transactions of the Royal Society of London, Series B* **337** 331–339
- Ballard D H, Salgian G, Rao R, McCallum A, 1998 "On the role of time in brain computation", in *Vision and Action* Eds L R Harris, M Jenkin (Cambridge: Cambridge University Press) pp 82–119
- Barsalou L W, Sewell D R, 1985 "Contrasting the representation of scripts and categories" *Journal of Memory and Language* **24** 646–665
- Brewer W F, Dupree D A, 1983 "Use of plan schemata in the recall and recognition of goal-directed sections" *Journal of Experimental Psychology: Learning, Memory and Cognition* **9** 117–129
- Broeck van den P, 1988 "The effects of causal relations and hierarchical position on the importance of story statements" *Journal of Memory and Language* **27** 1–22
- Carpenter R H S, 1988 *Movements of the Eyes* 2nd edition (London: Pion)
- Findlay J, 1998 "Visual activity in everyday life" *Current Biology* **8** R640–R642
- Fleischer A G, Strauss J, 1988 "Predictive strategies in eye-head coordination during text processing" *Ergonomics* **31** 1467–1475
- Furneaux S, 1996 *The Role of Eye Movements during Music Reading* D Phil thesis, University of Sussex, Brighton, UK
- Harris C M, 1989 "The ethology of saccades: a non-cognitive model" *Biological Cybernetics* **60** 401–410
- Hayhoe M M, Bensinger D G, Ballard D H, 1998 "Task constraints in visual working memory" *Vision Research* **38** 125–137
- Hayhoe M M, Land M F, 1999 "Coordination of eye and hand movements in a normal visual environment" *Investigative Ophthalmology & Visual Science* **40**(4) S380 (ARVO Abstract 2005)
- Husain M, Kennard C, 1996 "The role of attention in human oculomotor control", in *Visual Attention and Cognition* Eds W H Zangemeister, H S Stiehl, C Freksa (Amsterdam: Elsevier) pp 165–175
- Jeannerod M, 1988 *The Neural and Behavioural Organization of Goal-directed Movements* (Oxford: Oxford Science Publications)
- Land M F, 1993 "Eye-head coordination during driving" *IEEE Systems, Man and Cybernetics Conference Proceedings, Le Touquet* volume 3, pp 490–494
- Land M F, Lee D N, 1994 "Where we look when we steer" *Nature (London)* **369** 742–744
- Land M F, Furneaux S, 1997 "The knowledge base of the oculomotor system" *Philosophical Transactions of the Royal Society of London, Series B* **352** 1231–1239
- Land M F, Mennie N, Rusted J, 1998 "Eye movements and the roles of vision in activities of daily living: making a cup of tea" *Investigative Ophthalmology & Visual Science* **39**(4) S457 (ARVO Abstract 2094)
- Milner A D, Goodale M A, 1995 *The Visual Brain in Action* (Oxford: Oxford University Press)
- Newell A, 1990 *Unified Theories of Cognition* (Cambridge, MA: Harvard University Press)
- Norman D A, Shallice T, 1986 "Attention to action: willed and automatic control of behaviour", in *Consciousness and Self-regulation: Advances in Research and Theory* volume 4, Eds R J Davidson, G E Schwartz, D Shapiro (Plenum: New York) pp 1–18
- O'Regan J K, 1990 "Eye movements and reading", in *Eye Movements and Their Role in Visual and Cognitive Processes* Ed. E Kowler (Amsterdam: Elsevier) pp 395–453
- Posner M I, Petersen S E, 1990 "The attention system of the human brain" *Annual Review of Neuroscience* **13** 25–42
- Rayner K, 1995 "Eye movements and cognitive processes in reading, visual search, and scene perception", in *Eye Movement Research: Mechanisms, Processes and Applications* Eds J M Findlay, R Walker, R W Kentridge (Amsterdam: North-Holland) pp 3–22
- Rusted J, Ratner H, Sheppard L, 1995 "When all else fails, we can still make tea: a longitudinal look at activities of daily living in an Alzheimer patient", in *Broken Memories: Case Studies in Memory Impairment* Eds R Campbell, M A Conway (Oxford: Blackwell Publishers) pp 397–410

-
- Schank R C, Abelson R P, 1977 *Scripts, Plans, Goals and Understanding* (Hillsdale, NJ: Lawrence Erlbaum)
- Schleidt M, 1988 "A universal time constant operating in human short-term behaviour repetitions" *Ethology* **77** 67–75
- 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
- Shallice T, 1988 *From Neuropsychology to Mental Structure* (Cambridge: Cambridge University Press)
- Stark L, Ellis S R, 1981 "Scanpath revisited: cognitive models direct active looking", in *Eye Movements: Cognition and Visual Perception* Eds D F Fisher, R A Monty, J W Senders (Hillsdale, NJ: Lawrence Erlbaum) pp 193–226
- Stark L, Yamashita I, Tharp G, Ngo H X, 1993 "Search patterns and search paths in human visual search", in *Visual Search 2* Eds D Brogan, A Gale, C Carr (London: Taylor and Francis) pp 37–58
- Stuss D T, Shallice T, Alexander M P, Picton T W, 1995 "A multidisciplinary approach to anterior attention functions" *Annals of the New York Academy of Sciences* **769** 191–211
- Underwood G, Everatt J, 1996 "Automatic and controlled information processing: the role of attention in the processing of novelty", in *Handbook of Perception and Action* volume 2, Eds O Neumann, A F Sanders (London/San Diego: Academic Press) pp 185–227
- Viviani P, 1990 "Eye movements in visual search: cognitive, perceptual and motor control aspects", in *Eye Movements and Their Role in Visual and Cognitive Processes* Ed. E Kowler (Amsterdam: Elsevier) pp 353–393
- Weaver H E, 1943 "A study of visual processes in reading differently constructed musical selections" *Psychological Monographs* **55** 1–30
- Yarbus A, 1967 *Eye Movements and Vision* (New York: Plenum)