PERGAMON

Vision Research 41 (2001) 1561–1574

Vision Research

www.elsevier.com/locate/visres

A model of eye movements and visual working memory during problem solving in geometry

Julie Epelboim [1], Patrick Suppes *

Center for the Study of Language and Information, Stanford University, Stanford, CA 94305-4115, USA

Received 23 August 1999; received in revised form 23 March 2000

Abstract

The Oculomotor Geometry Reasoning Engine (OGRE) was proposed to model eye movements and visual working memory during problem solving in geometry. OGRE postulates that geometrical elements from diagrams are added to visual working memory when they are scanned. Newly-added elements overwrite elements already in memory. The model was applied to eye-movement patterns of three subjects: two geometry experts and one non-expert. Their eye movements and verbal protocols were recorded as they solved geometry problems posed with diagrams. Subjects used highly redundant eye-movement patterns with multiple rescans of the same geometrical elements. OGRE's model of visual memory provided a good fit for the distribution of times between rescans. The model was used to estimate the size of visual working memory used in geometry. The estimates varied as a function of both problems and subjects, with means and standard deviations for each subject being: 5.3 ± 1.4, 4.0 ± 0.9 and 4.7 ± 1.6. © 2001 Published by Elsevier Science Ltd.

Keywords: Visual memory; Eye movements; Scanpaths; Problem solving

1. Introduction

Solving geometry problems is a complex process. The solver must (i) read the text; (ii) construct a diagram, if one is not provided; (iii) search this diagram for familiar patterns; (iv) retrieve relevant facts from memory; and (iv) make inferences, including numerical computations, that eventually lead to solution. This multifaceted process can proceed so quickly that it cannot be observed directly with currently available technology. Spoken or written protocols have limited value because only the inferences that reach 'awareness' can be reported. Some of the inferences reported in the protocol may be the result of several smaller steps on which the solver cannot verbally remark without disrupting the train of thought.

Fortunately, there is a type of protocol that has the potential of circumventing these obstacles. When problems are presented visually, as diagrams, the problem solver's eye movements may provide the experimenter with a window on the mind. Eye-movement protocols have some clear advantages over conventional written or spoken protocols. It does not require any additional effort or training on the part of the subject and rather than being disruptive, the eye movements are an integral part of the problem-solving process.

Using eye movements to infer cognitive and perceptual processes, however, is not without difficulties. Viviam (1990) discussed many common problems inherent in this type of research, and concluded that the only possibly useful approach to interpreting eye-movement data is to work within a specific theoretical framework. Here, we describe such a theoretical framework developed for the task of solving geometry problems posed with diagrams. Before our model is presented, it is useful to discuss some of the highlights of the relevant research on problem solving, visual working memory and eye movements. The following review is not meant to be comprehensive. It is meant to illustrate the variety of approaches used in the field.

* Corresponding author.
E-mail address: suppes@clsi.stanford.edu (P. Suppes).

[1] We sadly report that Julie Epelboim died 10 January 2001.

0042-6989/01/$ - see front matter © 2001 Published by Elsevier Science Ltd.
PII: S0042-6989(00)00256-X

1.1. Background

A popular modern approach to modeling human problem solving is the use of 'production systems'. The most developed of such systems is Anderson's ACT-R family of models (Anderson, 1993). Productions are rules, consisting of conditions and actions, which are organized in a hierarchical goal structure. Problem solving involves matching items in declarative memory with elements in the condition portion of the productions. When a match is found, the production 'fires', that is, the action part of the matching production is performed. This process is repeated until the solution is reached.

Production-based models do fairly well in predicting behavior under some circumstances. However, detailed analyses of protocols show that problem solving rarely proceeds in a hierarchical, goal-directed or deterministic manner of production-system models. Problem solving seems to be more probabilistic, as shown, for example, by Suppes and Sheehan (1981), using 1455 computer-based proofs in set theory.

The probabilistic nature of problem solving can also be supported with eye-movement data. For example, examination of the eye-movement patterns of subjects doing column arithmetic exercises shows that they did not follow the simple right-to-left, top-to-bottom algorithm exactly, despite being specifically reminded to do so (Suppes, Cohen, Laddaga, & Floyd, 1983). Over 30% of fixations were backtracks, skips ahead, and spurious operations that did not fit the algorithm at all. The frequency of such anomalous fixations was surprisingly high: 8–16% for adult subjects and 17–36% for children.

The eye-movement patterns observed during column arithmetic, reading (Epelboim, Booth, & Steinman, 1994), mental animation of mechanical diagrams (Hegarty, 1992), as well as other tasks, support the hypothesis that human reasoning, even during the execution of very simple algorithms, is highly probabilistic (Suppes, 1981). People forget their place in the algorithm, they forget the stimulus just observed, and they forget an intermediate result and must repeat a step or even start over from the beginning. Sometimes, they recognize a familiar pattern and skip using the algorithm altogether. Problem solving must be highly probabilistic because of the continually active nature of human memory and perception.

The modern concept of working memory that has a limited capacity was developed primarily by Baddeley (1986). According to Baddely, the working memory system temporarily stores information during performance of complex cognitive tasks. This concept of working memory is distinguished from long-term memory and also very short-term, or iconic memory (mostly attributed to visual persistence). The items in working memory remain long enough to be useful in the performance of at least one mental operation.

Working memory consists of the 'central executive', which performs the calculations, the 'phonological loop', which maintains speech-based information, and the 'visuospatial sketchpad', which sets up and maintains visual imagery. The mechanisms for forgetting in Baddeley's model is decay over time and modality specific interference. The phonological loop is the most extensively investigated component of this model. Much less is known about the visuospatial sketchpad and the central executive — the components that must be active in many kinds of problem solving, including geometry.

The estimated size of visual working memory depends on specific experimental conditions, but all estimates are relatively small, i.e. fewer than 10 items. For example, Luck and Vogel (1997) showed that humans can remember about four visual items regardless of whether the items represented single features, e.g. colors, or conjunctions of features, e.g. color + orientation + size. They concluded that visual working memory 'stores integrated objects rather than individual features'. A similar estimate was obtained by Lachter and Hayhoe (1995) who found that performance of subjects making judgments about the spatial arrangement of a sequence of dots dropped radically when more than four dots were used (see also Hayhoe, Bensinger, & Ballard, 1998).

A larger estimate for the size of visual working memory was obtained by Glassman, Garvey, Elkins, Kasal, and Couillard (1994) who found that both humans and rats remembered the locations of about 14 out of 17 arms of a radial maze. When the probability of guessing was taken into account, these results led to estimates of the size of visuo-spatial working memory that were on the high end of 'magic number' 7 ± 2 described by Miller (1956), in his classic review of findings on short-term-memory and attention.

There are also a few studies that estimate the size of visual memory to be just one item. Broadbent and Broadbent (1981), for example, showed that subjects can reliably remember only one item when stimuli are meaningless shapes and the subjects are prevented from phonological encoding, i.e. naming the shapes on the basis of their resemblance to familiar objects. They argued that studies which reported larger estimates of the size of visual working memory, reflected the subjects' use of phonological encoding. Walker, Hitch, and Duroe (1993), however, used a similar task to show that similarity between the most recent shape and earlier shapes had a deleterious effect on recall of the earlier items, suggesting that at least some information about previous items is retained in visual working memory.

Estimates of the size of visual working memory described so far used tasks in which subjects were specifi-

cally asked to remember sets of objects. In contrast, Ballard, Hayhoe, and Pelz (1995) estimated the size of visual working memory actually used in a visuomotor task. They asked subjects to copy meaningless models made of colored blocks and recorded their eye movements. They found that subjects tended to look at the model about twice per block, at least as the first couple of blocks were being put in place. The authors concluded that subjects used the visual display to extend their visual working memory. When preparing to add a block, subjects looked at the model once to decide the color of this block, and a second time to find where this block should go in the copy. The authors concluded that the subjects could remember only one feature — either the color or the location of one block, and that 'visual representations are limited and task-dependent'.

Despite the limitations of visual representations, many problems are much easier to solve when presented visually rather than verbally. Larkin and Simon (1987) showed the superiority of diagrammatic representations formally by comparing simulated problem-solving programs that used diagrammatic-like or verbal-like data structures as input. Simulations showed that in a number of tasks, including geometry, diagrammatic data structures led to programs with greater computational efficiency. Larkin and Simon (1987) concluded that diagrams 'can be better representations not because they contain more information, but because the indexing of this information can support extremely useful and efficient computational processes'. They were referring to human abilities to make perceptual inferences and to shift attention quickly and effortlessly.

Larkin and Simon (1987) suggested that mental images, although less detailed, can be used as effectively as external diagrams. The use of mental images may be possible for simple problems where the solution can be reached by focusing attention on only one element at a time (for example, simple flow charts). In more complex problems, such as geometry problems used in our experiment, the problem-solver must keep in mind not just a single feature of the diagram, but also a set of relationships among the various parts of the diagram. The latter types of problems should be more difficult to solve without a visible diagram because it has been shown repeatedly that humans do better when they can scan visual scenes than when they have to maintain mental images in memory. One example of this phenomenon is the block-copying task described above. Another example was observed by Epelboim et al. (1995), who found that when subjects looked at a sequence of targets, the subject who used visual search instead of remembering the locations of the targets, performed faster and benefited more from practice than the other three subjects who memorized target locations. Further evidence that mental images are unreliable comes from experiments that show that large changes in the visual scene can occur during blinks, saccades or other visual transients without the observer noticing the change (e.g. O'Regan, Rensink, & Clark, 1999).

1.2. Our model

The focus of our model is the part of memory that functions as short-term storage for intermediate results of visual perception, analogous to Baddely's 'visuospatial sketchpad'. Our version of visual working memory stores memory images of visual objects that are meaningful and relevant in the context of the current task. In the case of geometry diagrams, these objects are angles, line segments, figures (e.g. triangles) and text. The mechanism for adding memory images to visual memory is oculomotor scanning. The mechanism for forgetting is interference between the object being scanned and the objects already in visual memory. A detailed description of the model follows.

2. The oculomotor geometrical reasoning engine (OGRE)

The model consists of definitions and axioms about fixation duration, scanpaths and visual memory.

2.1. Axioms about fixation durations

The simplest assumption for the distribution of fixation durations is that the execution time of each fixation is a random variable independent of past processing or present perceptual state. If this assumption were true, fixation durations would be exponentially distributed. This is obviously not the case, because distributions of fixations observed in a wide variety of tasks are not maximum near 0, but reach the peak after about 200 ms, and then decay approximately exponentially.

A slightly more complex model for fixation durations has been used in the past (Suppes, Cohen, Laddaga, & Floyd, 1983; Suppes, 1990). This model assumes that each fixation is composed of some number of low-level eye control instructions. There are no proposed physiological or psychological processes that correspond to 'eye-control instructions'. These are simplified theoretical constructs that help model the data. The model assumes that during each fixation, n low-level eye control instructions are executed and that the execution times of eye-control instructions are identically distributed. Furthermore, it is assumed that execution times of eye-control instructions are exponentially distributed and that for each fixation $n = 1$, or $n = 2$. Under these assumptions, the distribution of fixation

durations can be described as the sum of two distributions: an exponential and a convolution of two exponential distributions with the same parameter. Although this model provided a reasonable fit for fixation durations observed while some subjects performed column arithmetic (Suppes, Cohen, Laddaga, & Floyd, 1983), it does not fit the present data, because of the shortage of short fixation durations.

Here, a different model is proposed. It uses the following axioms:

Axiom FD1. *Execution times of individual eye-control instructions are independent, identically distributed, memoryless, and, therefore, exponentially distributed.*

Axiom FD2. *In geometrical problem solving, the number of eye-control instructions per fixation, $n(s)$, is constant for a given subject s.*

Axiom FD3. *The eye-control instructions are performed sequentially: instruction $i+1$ begins immediately after instruction i terminates.*

These axioms describe fixations as a Poisson process — a renewal process in which the time between recurring events (here onsets of eye-control instructions) is exponentially distributed. In the case where all exponential distributions have the same parameter, such as assumed here, the total time is distributed as the sum of $n = n(s)$ exponential distributions, which is the gamma distribution (see Luce, 1986, p. 500, for a derivation). Its probability density function, $f_n(t)$, is:

$$f_n(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}, \quad (1)$$

where n is the number of theoretical eye-control instructions, and λ is the parameter of the exponential distributions.

A different type of a renewal process, which could be used to model the fixation duration data, is a parallel process, in which n events, whose execution times are identically and exponentially distributed, are executed in parallel. The total time of the process is the time when all events are completed. One type of such process can be modelled by an Extreme Value, Type I distribution (Luce, 1986, p. 503). The Gamma distribution and the Extreme Value, Type 1 distribution are similar in shape and tend to provide comparable fits to the data. There are theoretical difficulties in differentiating between parallel and serial models of reaction times when no physical limits on processing speed can be set. These difficulties have been studied and reported by Townsend and colleagues (e.g. Townsend & Thomas, 1994). In this study we will limit our analysis of the distribution of fixation durations to the serial model described in Axiom FD3.

2.2. Definitions for scanpaths

A scanpath is a sequence of eye movements used to solve one geometry problem. It can be written as: $f_{1,s_1}, f_{2,s_2}, \ldots, f_{n,s_n}$, where f_i is the ith fixation (a period during which gaze remains in the same location on the diagram), and s_i is the ith saccade (a ballistic eye movement that brings gaze to a new location) in the sequence.

Duration of fixation f_i is the length of time between the offset of saccade s_{i-1} and the onset of saccade s_i. Likewise, duration of saccade s_i is the length of time between the offset of fixation f_{i-1} and the onset of fixation f_i.

Each diagram is a configuration of geometrical elements (g)-angles, line segments, regions (e.g. inside circles or triangles), and text (all explanatory text included with the diagram is treated as a single element). The gs are indexed as follows: line segments — $s_1, \ldots, s_n$, angles — $a_1, \ldots, a_m$, regions — $r_1, \ldots, r_k$, text — t. The angle in question (marked with a '?' on the diagram, see Fig. 1) is treated separately from the other angles and labeled q. Since the problem solver's visual field is not restricted to the diagram, it is useful to introduce another geometrical element, labeled 'o', for other, which corresponds to any location outside of the diagram.

Each fixation, f_i, is associated with exactly one g, which we note as $f_i(g)$.

A scan, $F_j(g)$, is a sequence $f_{i,s_i}, f_{i+1}, s_{i+1}, \ldots, f_{n,s_n}$, in which all fixations are associated with the same geometrical element g. We introduce a second subscript on f_i, namely, $f_{i,j}$, to show that this fixation is the ith fixation of scan F_j.

2.3. Axioms about visual memory

To begin with, visual working memory (V) is a set of registers for storing memory images, $I(g)$, of geometrical elements. The contents of V are not ordered.

Axiom V1. *All registers in V are quickly filled with images from the visual presentation of the new problem.*

Axiom V2. *The size of V, M, is constant for a given subject and problem.*

We show later that M varies with subject and problem, but we do not test directly our assumption that M is constant for a given subject and problem. This is not practical with our data set. Consequently, we cannot empirically distinguish between the axiom as formulated and the assumption that even for a given subject and problem, M is a random variable with positive variance. From the standpoint of this latter assumption, the estimates used later depend on the mean values of M for given subject and problem.

Axiom V3. *During each scan $F_j(g)$ the image of g, $I(g)$, is added to V.*

With this apparatus, but before making more formal assumptions, we sketch the time sequence of events, both observable and unobservable. On the other hand, we restrict ourselves to the visual and oculomotor processing and exclude in our formal framework the details of mental computations and of the generation and production of the running verbal protocol. We do use the protocol to provide confirming evidence about the visual processing, as will be evident later.

Here is the time-sequence sketch, where:

V_j State of visual memory on scan F_j,
s Saccade,
g_R Geometrical element that has been previously scanned and is now rescanned,
g_N Geometrical element that is new, i.e. it is being scanned for the first time during this problem, or after many scans since it was last in visual memory.

The diagram below starts with a memory state, V_j, and a fixation, $f_{i,j}(g)$ of element g. The arrows ($\Rightarrow$) show transitions between states and events. The third step in the diagram shows all possible outcomes following the saccade, s.

$$V_j, f_{i,j}(g) \Rightarrow s \Rightarrow \begin{cases} (i)\ f_{i+1,j}(g) & \Rightarrow V_j \\ (ii)\ f_{1,j+1}(o) & \Rightarrow V_{j+1} = V_j \\ (iii)\ f_{1,j+1}(g_R) \\ (iv)\ f_{1,j+1}(g_N) & \Rightarrow V_{j+1} \neq V_j \end{cases} \tag{2}$$

Keeping this time sequence in mind, here are the additional axioms on visual working memory and scanning.

Axiom V4. *At the end of fixation $f_{i,j}(g)$, a saccade s occurs and then one of the four possibilities shown in* Eq. (2) *is realized on the next fixation*:

1. $f_{i+1,j}(g) \in F_j$ — *a fixation of the same g and therefore no change in scan F_j.*
2. $f_{1,j+1}(o) \in F_{j+1}$ — *a fixation outside (o) the diagram and no change in memory, $V_{j+1} = V_j$.*
3. $f_{1,j+1}(g_R) \in F_{j+1}$ and $I(g_R) \notin V_j$ — *rescan of g_R.*
4. $f_{1,j+1}(g_N) \in F_{j+1}$ and $I(g_N) \notin V_j$ — *scan of new element g_N.*

This axiom implies that if the next fixation is on an element that is different from the currently fixated element, two things happen. First, a new scan begins, and second, the content of memory is changed. The exception is when the fixation is outside the diagram (o), in which case a new scan begins, but the contents of memory remain the same. Note, that once the scanned element enters visual memory, content of memory remains the same for the duration of the scan.

Axiom V5. *In cases* (iii) *and* (iv) *of* Axiom V4, *visual memory is changed*:

(iii) $V_{j+1} = V_j - I(g') + I(g_R)$. *That is, $I(g_R)$ is added to the contents of V_j after another image $I(g')$ that was already in V_j is overwritten.*

(iv) $V_{j+1} = V_j - I(g') + I(g_N)$. *Same as (iii) just above, except that the added image is of a new element, g_N, that has not already been scanned.*

Axiom V6. *Let M be the integer size of V_j. If an image $I(g)$ is added to V_j ($I(g) \notin V_j$ & $I(g) \in V_{j+1}$), then the probability of selecting an image in V_j to be overwritten has a uniform distribution on the images in V_j.*

Axiom V7. *If $I(g)$ is overwritten on scan F_{j+1}, it is scanned again on the following scan F_{j+2} with probability $1-\varepsilon$, such that $I(g) \in V_j$ & $I(g) \notin V_{j+1}$ & $I(g) \in V_{j+2}$.*

These axioms can be summarized intuitively as follows: During each scan, the memory image of the element being scanned is added to V, and another element that was already in V is overwritten. This scheme maintains a constant number of images in V during a given problem, and we assume that V is filled at the beginning of the problem (Axiom V1).

The next definition ties the functioning of V to a measurable quantity in eye-movement data.

Let g be the element scanned on F_j and again for the first time on F_{j+k}. Then the number k is the rescan time. Since consecutive scans cannot be associated with the same g, $k \geq 2$.

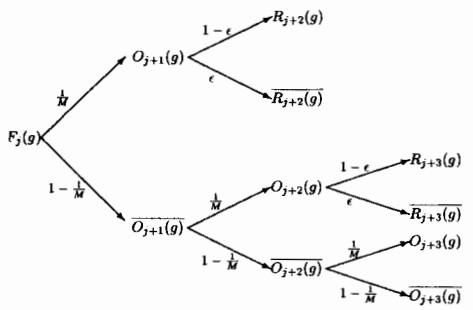
Axiom V7 implies that a memory image of an element g is added on the scan following the scan during which it was overwritten, with probability $1-\varepsilon$, which is close to 1 for small ε. Note however, that occasionally (with probability ε), an element will not be rescanned immediately after being overwritten, perhaps because g is no longer needed for the current mental operation. It may be rescanned much later, if it becomes needed for a different mental operation. We use these very long rescan times to estimate ε.

In order to generate the theoretical distribution of rescan times, it helps to produce the following probability tree. We introduce new notation:

$R_j(g)$ — g is rescanned on scan F_j. This implies $g \notin V_{j-1}$ & $g \in V_j$. Note that a rescan R_j is just a special case of scanning, so $R_j(g) = F_j(g)$ if g had been scanned earlier in the problem.

$O_j(g)$ — g is overwritten on scan F_j. This implies $g \in V_{j-1}$ & $g \notin V_j$.

The bar of $\overline{O_j}$ or $\overline{R_j}$ implies negation. Thus $\overline{O_{j+1}(g)}$ means that the image $I(g)$ was not overwritten during scan F_{j+1}.



(3)

Based on this tree, it can be seen that if g was scanned on scan F_j then the probability of its being rescanned on scan F_{j+3} is $(1-1/M)(1/M)(1-\varepsilon)$ Using this tree it is easy to show that the distribution of rescan times is given by Eq. (4). To simplify notation, we omit the argument g and write R_j and O_j to mean $R_j(g)$ and $O_j(g)$.

$$P(R_{k+j} | O_{k-1+j}, \overline{R_{k-1+j}, O_{k-2+j}}, \cdots, \overline{R_{j+1}}, \overline{R_j}) = \left(1-\frac{1}{M}\right)^{k-2}\left(\frac{1}{M}\right)(1-\varepsilon), \quad k \geq 2 \tag{4}$$

The estimate of visual memory size is the value of parameter M that produces the best fit of Eq. (4) to the eye-movement data, where M can vary for different subjects and problems.

2.4. Approximate independence of path

Axiom IP. *Within ε, the probability of a scan $F_j(g_j)$ depends only on the immediate prior scan, $F_{j-1}(g_{j-1})$ and contents of visual working memory, V_{j-1}. The rare events with probability ε can depend on the distant past.*

What we may prove from this last axiom and the earlier axioms is that within ε the sequence of random variables V_1, F_1, V_2, F_2,..., V_n, F_n, is a first-order Markov process. This means that what happens on scan n depends only on what happens on scan $n-1$, not on any earlier scan. This independence of scans before $n-1$ is what justifies the description of this axiom as one about independence of path. An example of a violation of this axiom, and the next simplest case, is if the sequence was a second-order Markov chain, which would mean that V_n, F_n, depended not only on V_{n-1}, F_{n-1}, but also on V_{n-2}, F_{n-2}. Our model would have to be much more complicated to account for this dependence.

We now turn to the experiment.

3. The experiment

3.1. Method

3.1.1. Data collection

Subjects. Three subjects participated. Two of the subjects (Experts, ME and MS) were skilled at solving geometry problems. They had graduate training in physics, and encountered problems similar to those used in the experiment in their professional life. The third subject (Non-expert, RS) had last solved geometry problems in high school, over 50 years prior to the experiment. He reported that he had little idea as to what to do on most of the problems, but tried hard to apply the little geometrical knowledge that he had to achieve a solution.

Problems. Each problem consisted of a diagram in which some angles were labeled with letters and numerical values. Some problems also contained brief text stating initial conditions. For each problem the subject was asked to find the value of the unknown angle, labeled with a '?'. Two of the problems are shown in Fig. 1. Each subject solved 10 problems.

Problems were presented on a high-quality LCD screen of a laptop computer. Subjects adjusted their distance to the screen to 'a comfortable reading distance'. As a result, one character subtended 22–29 min of arc, and the whole screen was 17–22° wide, and 12–18° high. The area of the problem never exceeded 0.8 of the screen in width and 0.9 of the screen in height. The head was stabilized from above. Viewing was monocular, with the non-viewing eye patched.

Eye movement recording. The Maryland Revolving-Field monitor (MRFM) was used to record horizontal and vertical eye orientations of the subjects. This apparatus uses sensor-coil/magnetic-field technique and phase detection on both meridians. The accuracy of the instrument is better than 1 min of arc. The sampling rate was set at 488 Hz (effective bandwidth = 244 Hz).

The MRFM is capable of measuring accurate and precise gaze (line-of-sight in space) with seated, but otherwise unrestrained subjects. However, the head movements of subjects were restricted in this experiment, to simplify data processing. This was done with the aid of a bicycle helmet attached to the frame of the MRFM. This method of restraining the head, in contrast to more conventional bite board methods, allowed the subjects to talk while they were solving problems (see Epelboim, Booth, & Steinman, 1994 for a more detailed description of the MRFM apparatus used in a similar setup).

Procedure. Calibration and problem trials alternated. During calibration, the subjects fixated each of 9 pluses

(+) that were presented in a 3 × 3 grid on the screen. Subjects fixated each + for 2 s and made a saccade to the next cross when prompted by a beep.

At the end of a calibration trial, a fixation + appeared at the upper left corner of the screen. The subjects fixated the +, and then started a trial, when ready, by pressing a button. The problem appeared and the subject started solving it. When he was finished, the subject pressed the same button again, after which the screen was cleared, and a fixation + appeared. The subject pressed the button to start the next calibration trial. There was a time limit of 5 min for each problem.

Subjects were not allowed to write or sketch anything, but the problems were selected to be simple enough to solve mentally. Subjects were asked to reason aloud, and their speech was recorded.

3.1.2. Analyses of data

Saccade detection. Saccades were detected with a computer program that uses an acceleration criterion. The criteria were established empirically for each subject by looking at eye movement traces with saccades flagged and adjusting the criteria until all the saccades were detected. Fixations were defined as periods of relatively stable gaze between two saccades.

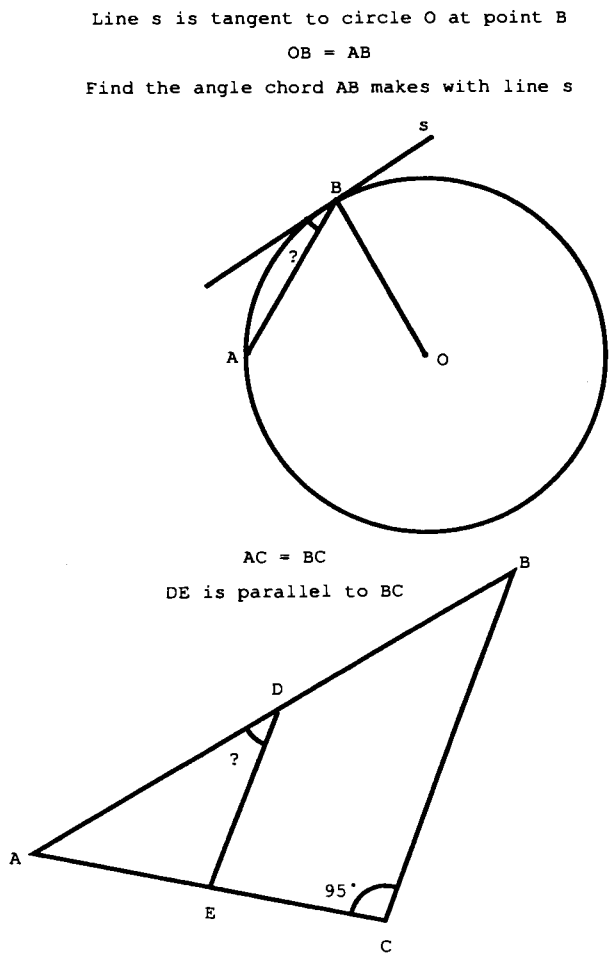


Fig. 1. Examples of geometry problems used in this study.

Blinks were detected manually. Fixations that were interrupted by a blink and did not contain a total of 50 ms of stable gaze were not used in the analyses. The frequency of blinking varied among subjects. Fewer than 1% of fixations were discarded for RS, < 3% for MS and < 7% for ME.

Assignment of fixations to geometrical elements. The locations of fixations on the diagram for a given problem were calculated using data from the calibration trials before and after that problem. The precision and accuracy of this location on the diagram was better than the size of one character in the accompanying text.

Each fixation was assigned to a geometrical element of the diagram. Fixations that fell more than 2 character-widths outside the diagram or text were labeled *o*, for other.

Text that accompanied the problem was considered a single element (not broken into words). A more realistic treatment of text would require a model of reading to be embedded into the model for geometry. At this stage, the simplification of considering text a single element seems reasonable, especially since text that accompanied the diagrams was kept brief.

Note that the assignment of fixations to geometrical elements depended on the definition of the borders between elements, which, in some instances, had to be selected subjectively. In order to assess the amount of uncertainty in assigning fixations to geometrical elements, 1/3 of all fixations were assigned to elements by two observers. The agreement between the two sets of assignments was > 95%.

3.2. Results

3.2.1. Global analysis of scanpaths and verbal protocols

Before testing specific axioms, it is useful to take a global look at scanpaths and their relationship with verbal protocols of the subjects. This analysis will show that eye movements do not simply reflect the protocol, but carry additional information that can be useful for modeling cognitive and perceptual processes used to solve the problem. The following example of a subject solving a problem is a representative case study for the process.

Consider ME's protocol for the problem in Fig. 1:

> Line *s* is tangent to circle *O* at point *P*. OB = AB, find the angle chord AB makes with line *s*. Ok, well... so, the unknown angle is the complement of angle ABO. Ah, so... OB = AB, ah... ok, that means that a triangle formed by connecting points *O* and *A* would have to be an isosceles triangle. Ah, in fact it would have to be an equilateral triangle. So that means that the angle ABO is 60° and the unknown angle is 30°.

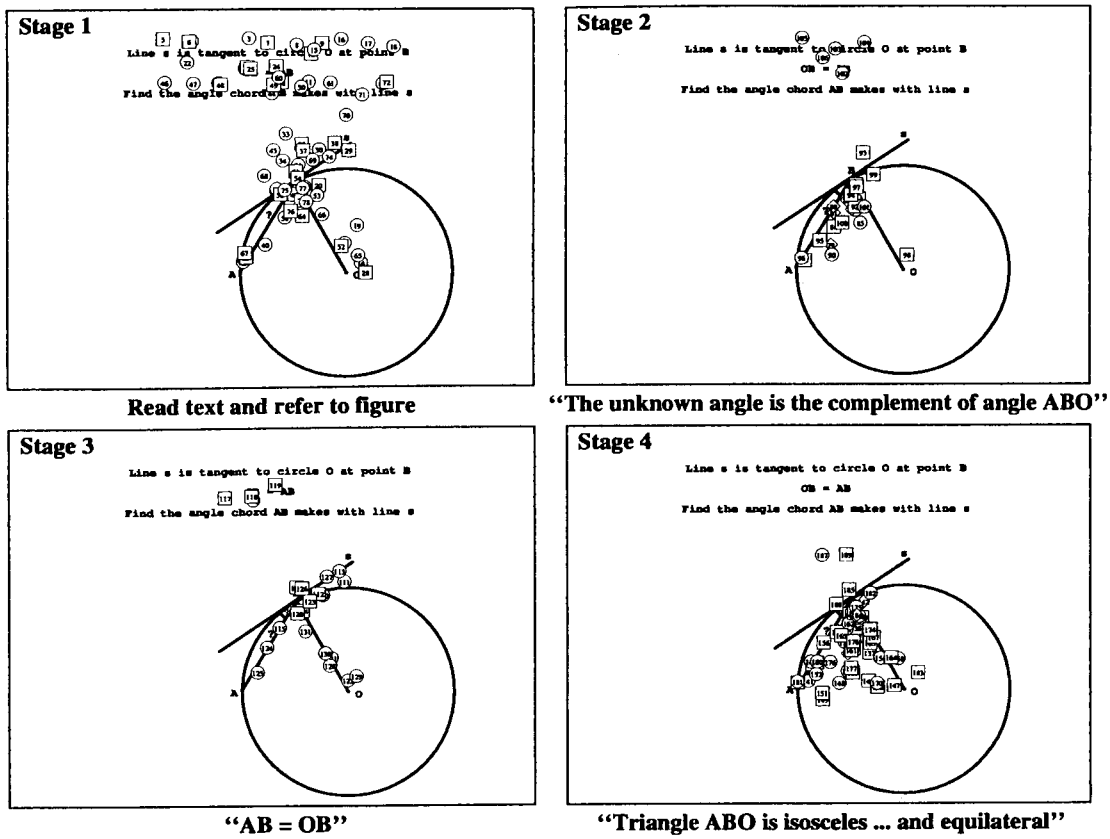


Fig. 2. Distributions of fixations during the four stages of the protocol for subject ME. Each symbol represent 1 fixation. Circles show fixations shorter than 300 ms, squares show fixations 300–600 ms in duration, and diamonds show fixations over 600 ms long. The number inside each symbol shows the sequential number of that fixation in the scanpath.

This protocol was used to divide ME's path to solution into 4 stages. Fixations that took place during each of the stages are shown in Fig. 2. It is clear that as ME talked about different parts of the diagram, he tended to fixate the relevant elements more frequently than other elements.

An interesting observation can be made about ME's eye movements in Stage 4. Here, ME made many fixations inside the triangle ABO, which is central to solving the problem. This triangle, however, is not completely shown on the diagram — only sides AB and BO are actually shown. But ME constructed the triangle mentally and scanned inside this partially visualized figure as he solved the problem. The other expert subject, MS, also scanned this imaginary triangle. The detection of the triangle was not simply a perceptual process (Gestalt closure, for example), but reflected the higher-level reasoning about the problem. The non-expert, RS, neither mentioned triangle ABO in his protocol, nor scanned it. Fig. 3 shows this difference in scanning patterns between the experts and the non-expert. This pattern of differences was typical. All three subjects started working on each problem by reading the text, if any, and referring to the relevant elements on the figure (Stage 1, in Fig. 2). After this brief information gathering stage, the experts often looked at constructed elements, such as the triangle ABO in ME's Stage 4.

A closer examination of locations of scans that took place during each utterance, tabulated in Table 1, shows that ME did not simply scan the elements as he mentioned them. His scanpath was very redundant — he kept returning to elements already seen. This redundancy, which was typical for all subjects on even the simplest problems, was not present in the protocol.

Does the oculomotor redundancy reflect operation of limited visual working memory that requires constant refreshing, as the OGRE model proposes, or do subjects simply shift gaze within the diagram to give the eyes something to do while the problem is being solved internally, without the need for continuous visual input? This question can be explored by looking at what happened when the subjects in this experiment performed mental arithmetic, as they occasionally had to do in order to solve the problem. The mental arithmetic process should not require visual input. If the purpose of the fixations is to acquire or update visual information about the diagram that is needed for the current mental operation, then the scanning during mental arithmetic should be different from the pattern observed during the rest of the problem solving. It should

either be unrelated to the structure of the diagram, or limited to the areas that contain the numbers that are being processed.

Most of the arithmetic needed to solve the problems was simple enough to be solved during one or two scans. Occasionally, however, the subjects got stuck on a particular mental arithmetic operation, for example adding the sizes of two angles and subtracting the result from 180 to find the third angle. When that happened the eye movement pattern was obviously distinct from the normal pattern. The subjects continued to shift gaze at the same rate, but instead of looking from one element to the next, they either looked outside the diagram (up at the ceiling or or down at their shoes, for example), or repeatedly fixated a region near the center of the screen. Two typical examples of the eye movement pattern during mental arithmetic are shown in Fig. 4. When mental arithmetic was not being performed, the subjects made very few fixations outside of the diagram (fixations of type 'other'), and rarely remained within the same region for more than 3 fixations ($< 3\%$).

Figs. 2, 3 and 4 show that global eye-movement patterns of the subjects depended to some extent on the stage and quality of their reasoning process, as determined by the protocol. Although there was no systematic relationship between individual scans and concurrent utterings, the evidence from mental arithmetic suggests that the repetitive scanning of diagram elements served an important role in acquiring and updating visual information about the diagram.

Next, we examine in some detail the axioms of the OGRE model, starting with axioms about fixation durations.

3.2.2. Distribution of fixation durations

Axiom FD3 proposes a serial model of fixation duration, in which a fixation terminates when n eye-control instructions are completed. The assumption is that the execution times for the eye-control instructions are identically and exponentially distributed. This describes a Poisson process, which is modeled by a Gamma distribution. Fig. 5 shows Gamma probability density functions fitted to the histograms of fixation durations of the 3 subjects. Statistically, the fits are good ($\chi^2 < 1$) although not perfect. The best maximum likelihood fit value for n was 3 for all subjects. The values for λ were very similar for the 3 subjects: 0.0098 for ME, 0.0090 for MS and 0.0084 for RS.

3.2.3. Statistical properties of sequences of scans

In order to test the independence-of-path assumption of Axiom IP for sequences of g's scanned, χ^2-tests were used to determine the Markov order of these sequences (Anderson & Goodman, 1957). A separate χ^2 was calculated for each problem and each subject. First the hypothesis that the sequence has no dependencies (zero-order process) was tested against the hypothesis that

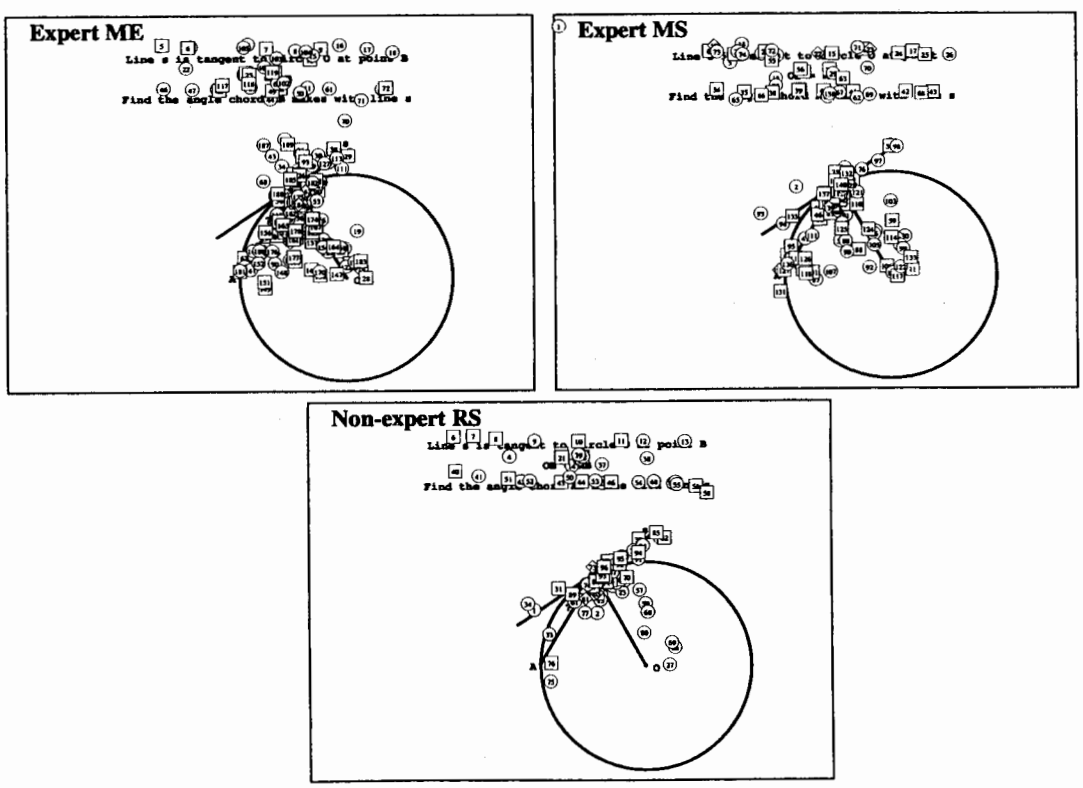


Fig. 3. Comparison of fixation distributions of two expert subjects and the non-expert. All fixations for each subject are shown.

Table 1
Scans and utterances for subject ME while he was solving the problem shown in Fig. 1 (top)[a]

Fixations	Scan	Utterance
3–6	'Line s is'	
7–9	'tangent to circle O'	'Line s is tangent to circle'
10	∠B	
11–12	'circle O'	
13–14	O	
15–18	'circle O at point B'	'O'
19–20	Inside circle → ∠B	'at point B'
21–25	'OB = AB'	'OB'
26–43	∠B → A → O → s → toward text ∠B → AB	'equals AB' (35–40)
44–50	'find the angle chord B makes'	'Find the angle chord AB'
51–59	∠B → OB → ∠B → A → △ABO → ∠? → ∠B	
60–62	'AB makes with line'	'makes'
63–69	△ABO → O → OB → A → A → ∠B → s	'with line'
70–72	'with line s'	
73–78	BO → s → ∠? → ∠B	's' (77)
79–90	△ABO → A → △ABO	'ok, well' (81–82)
91–97	∠B → s → ∠B → AB → ∠B	'the unknown angle is the'
98–101	O → s → △ABO → BO	'complement of angle ABO'
102–106	'tangent to circle O'	
107–115	∠B → ∠? → s → ∠B → AB	
116–119	'OB = AB'	
120–123	∠B → OB → s	
124–131	AB → s → OB	'OB equals AB' (125)
132–140	AB → △ABO → AB	'ah, ok!'
141–147	AB → ∠B → AB → ∠B → AB → A → O	
148–157	△ABO → AB	'triangle formed by connecting points'
158–164	BO → △ABO → BO	'O and A'
165–173	△ABO → ∠B → BO → ∠B → △ABO → ∠B	'would have to be an isosceles triangle'
174–183	BO → ∠B → △ABO → ∠B → AB → A → s → O	'ah! In fact it would have to be an equilateral triangle'
184–189	∠B → outside figure	'so that means that angle B is 60° and the unknown angle is 30°'

[a] Scanpaths for the four stages (separated by horizontal lines in the table) are shown in Fig. 2.

the sequence can be modeled as a first-order Markov chain. The χ^2 values were calculated for each problem, and were statistically significant ($P < 0.01$) for all problems for subjects ME, for 8 out 10 problems for subject MS and for 8 out of 10 problems for subject RS. This outcome means that the prediction for g of scan F_j can be improved significantly if g of scan F_{j-1} is taken into account, for most of the problems.

First-order vs. second-order dependence was tested next. In this case none of the χ^2 values were significant ($P > 0.1$). This means that there is no statistically significant improvement for predicting the state at scan F_j if g's of two previous scans are taken into account, as opposed to just one. In short, the element g scanned on each scan depends solely on the g of the scan that just precedes it, and the sequences of scans can be modeled with a first-order Markov process. This analysis is consistent with Axiom IP even though the test could not use the necessarily unobservable contents of memory, V_j. The results are summarized in Table 2. As remarked after the statement of Axiom IP, this negative result for second-order effects supports the strongly simplifying assumption of path independence. A positive result would have forced us to introduce a second-order theory, necessarily much more complex.

We also applied the same χ^2 tests to sequences of individual fixations, as opposed to scans. The results are also summarized in Table 2. As with scans, fixations for all but three of the problems (two for MS and one for RS) could be modeled by a first-order Markov chain.

3.2.4. *Estimates of the size of visual working memory*

Estimates of ε. According to Axiom V7, ε is the probability that g is overwritten on scan F_j and is not rescanned on scan F_{j+1}. This g may be rescanned much later, resulting in a very long rescan time. In other words, suppose an element is scanned on scan F_j and the next time it is scanned on scan F_{j+k}, and k is large. It is not likely that this element stayed in memory for k scans and was rescanned because it was overwritten on scan F_{j+k-1}. It is more likely that it was overwritten some time during the k scans, and was not rescanned after being overwritten. Given this reasoning, we estimated ε by looking at the extreme part of the tail of histograms of rescan times.

A separate estimate of ε was determined for each subject by visually examining the histograms of rescan times (the number of scans between consecutive scans of the same g), summed over all the problems (see Fig. 6). We used the cutoff points of 70 for ME, 55 for MS and 45 for RS. Based on these cutoffs, the estimates of ε, calculated as the number of rescan times greater than the cutoff divided by the total number of rescans, were 0.008 for ME, 0.011 for MS and 0.009 for RS. To give a sense of the variability of ε as a function of cutoff

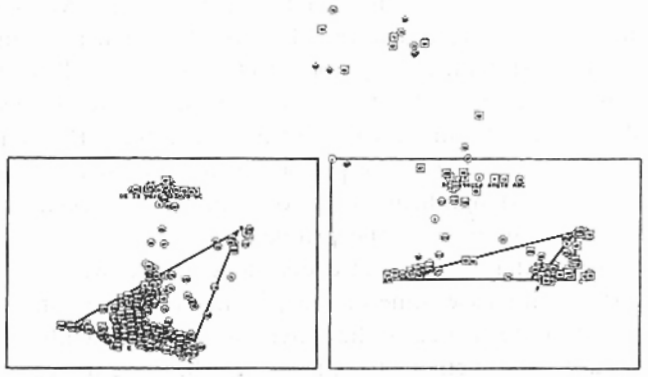


Fig. 4. Examples of eye movement patterns during mental arithmetic for subjects ME (left), and MS (right). All fixations for each problem are shown. The high concentration of fixations in the center of the screen for ME, and the fixations above the display for MS, occurred while the subjects were performing mental arithmetic. The rest of the fixations represent the normal problem-solving pattern.

point, for subject ME, a cutoff point of 60 results in $\varepsilon = 0.017$, and a cutoff point of 80 results in $\varepsilon = 0.004$.

Estimates of M. Axiom V2 states that M is constant for each subject and problem. Our initial evaluation of M assumed that M is constant across problems, but may vary among subjects. This assumption allowed us to pool the data over all the problems, resulting in more data points and more robust fit. It also gave a sense of variability of average M among the subjects.

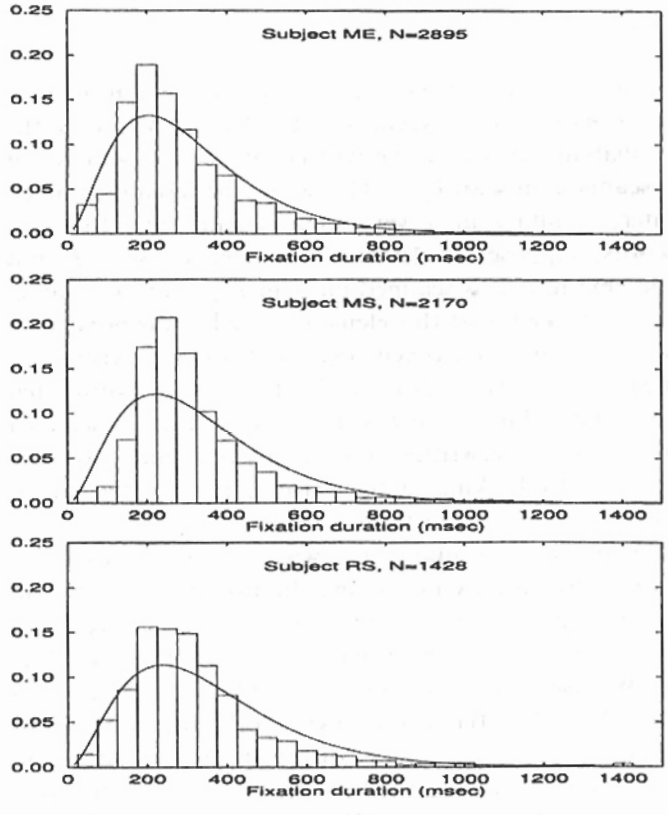


Fig. 5. Fits of the Gamma probability density function to the distributions of fixation durations. Bin size in the histograms is 50 ms.

The M parameters that produced the best fit of Eq. (4) to the whole set of data (summed over all problems) are: 5.5 for ME, 4.0 for MS, and 4.1 for RS. The estimates of M were not sensitive to the exact value of ε, as long as ε was of the order of 0.01 or less. Curves fitted to the histograms of rescan times summed over all the problems are shown in Fig. 6.

Eq. (4) was also fitted to the rescan time histograms calculated for individual problems. The results are shown in the rightmost column of Table 2. Consistent with Axiom V2, there was some variability in estimates for M for individual problems. ME's estimates ranged from 4.3 to 8.4 (mean = 5.8, SD = 1.4)-MS's estimates ranged from 2.4 to 5.3 (mean = 4.0, SD = 0.9); RS's estimates ranged from 2.4 to 7.8 (mean = 4.7, SD = 1.6). All fits were statistically reliable ($\chi^2 < 1$).

The values of M estimated for individual problems were smaller than and did not correlate with the number of different g's scanned in a given problem ($\rho^2 = 0.1$ for ME, 0.1 for MS, and 0.22 for RS). This supports the proposition that the variability of the size of visual working memory was independent of problem complexity, as measured by the number of geometrical elements. The estimates M fell somewhat short of the 'magic number' 7 ± 2. Twelve of the 30 estimates are within the range. Twenty-five of them are within the range 5 ± 2.

4. Discussion

A model-theoretic approach, based on eye movement data, was used to estimate a cognitive variable, viz. the size of visual working-memory. The use of eye movements made it possible to measure this variable during a realistic, complex cognitive task. All prior quantitative estimates of the capacity of visual working-memory have been based on simpler memory tasks, for example, recall of a series of objects presented on a display.

Our estimates of the size of visual working memory are similar to some of the prior estimates obtained under a variety of conditions (e.g. Walker et al., 1993; Lachter & Hayhoe, 1995). They are somewhat lower than the range of 7 ± 2 (Miller, 1956). They support the idea that although the visual memory size is relatively small, more than just one item is stored, as has been postulated by some theories (e.g. Broadbent & Broadbent, 1981; Ballard et al., 1995). Indeed, we are skeptical that the kind of complex problem solving in geometry that makes substantial use of a diagram can be adequately modeled psychologically with a visual working memory of size one. On the other hand, it is likely that the range of estimates of the size of visual working memory will vary even more as models like the one proposed here are applied to a wider variety of visual tasks. An important problem for future theory is

Table 2
Summary of analyses for individual problems[a]

Problem	Trial length (s)	Number of g's scanned	Number of fixations	Markov order for fixations	Number of scans	Markov order for scans	Estimate of M
1	86.5	13	165	1	134	1	5.7
2	74.4	10	137	1	102	1	5.8
3	102.3	15	205	1	158	1	8.4
4	52.7	16	130	1	98	1	5.7
5	60.4	14	185	1	147	1	8.0
6	111.2	18	240	1	183	1	4.3
7	24.7	8	54	1	40	1	5.8
8	73.7	12	125	1	99	1	4.8
9	35.8	21	109	1	95	1	5.0
10	106.1	19	251	1	187	1	4.3
	Overall		1601		1243		5.5
1	39.5	13	91	1	61	0	4.3
2	62.5	10	108	0	87	1	5.0
3	66.8	12	123	1	75	0	4.1
4	74.6	14	192	1	144	1	4.7
5	15.5	7	32	1	19	1	3.0
6	75.7	15	176	1	122	1	2.4
7	16.1	7	33	1	24	1	3.2
8	41.6	10	114	1	78	1	3.5
9	24.2	14	65	1	47	1	4.4
10	82.9	15	244	1	185	1	5.3
	Overall		1180		842	1	4.0
1	43.6	12	80	1	54	0	2.4
2	42.7	7	75	1	54	1	3.1
3	29.0	12	65	1	43	1	7.8
4	46.4	14	114	1	82	1	5.2
5	51.8	11	129	1	83	1	5.7
6	39.7	14	93	1	67	1	6.9
7	21.3	6	46	0	36	0	4.4
8	81.8	11	170	1	93	1	3.6
9	36.2	10	91	1	68	1	3.8
10	26.7	9	81	1	54	1	3.8
	Overall		944		646		4.1

[a] See text.

to model in detail the interaction between the nature of the task and the size of visual working memory needed or actually used.

The OGRE model, unlike most other models of cognitive processes based on eye-movement data, emphasizes the role of stochastic processes in the control of eye movement. This emphasis does not imply that higher-level cognitive processes do not have influence over eye movements. On the contrary, as can be seen in Fig. 3, which compares scanpaths of expert and non-expert subjects, the inferencing process has a large effect on the global eye-movement pattern. According to OGRE, however, the inferencing process does not control gaze directly. It determines what visual information is required for the current computation and delegates the details of placing this information in visual working-memory, and maintaining it there, to a lower-level visuomotor agent. Inferences are made at a higher level with the agent simply being required to act efficiently when required to perform one or another visuomotor action. A simple stochastic process is probably the most efficient solution for freeing the more intelligent inferencing process from dealing with details of oculomotor control. This dichotomy between planning and doing is well accepted in the literature on motor control (see Sternberg, Monsell, Knoll, & Wright, 1978 for a general discussion, and Zingale & Kowler, 1987 for application of this dichotomy to eye-movement control).

The assumption of a dichotomy between mental operations and oculomotor control must also assume that visual working-memory, with a capacity greater than one item, must be capable of storing information obtained during the prior few fixations and make this information available to the higher-level cognitive process. This assumption, however, is not convenient for making deterministic models of cognitive processes on the basis of eye-movement data, because it does not allow a simple mapping between any individual eye

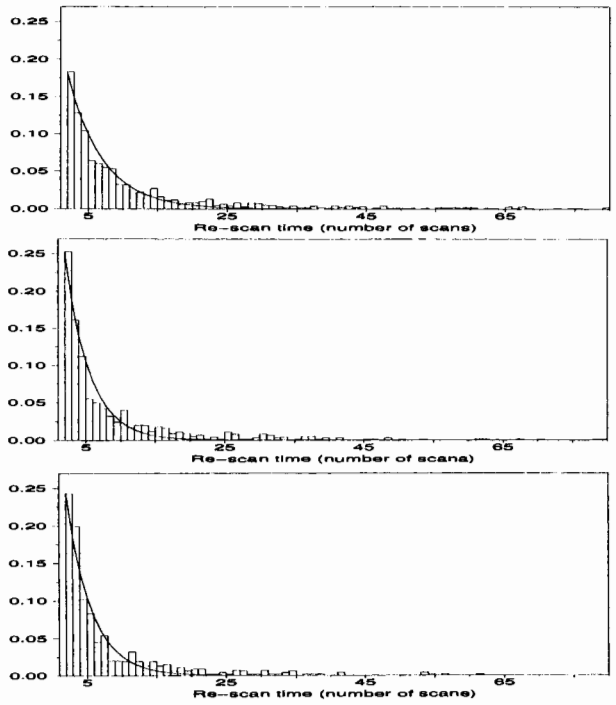


Fig. 6. Histograms of rescan times. Bin size is 1 scan. Plots of Eq. (4) with the best-fit parameter M are shown on each graph.

movement and a specific mental operation. Many models assume direct cognitive control over eye movements and do not consider the contents of visual working-memory. This assumption is unfortunate because it permits unrealistically simplistic models.

The OGRE model could be modified and extended to apply to other cognitive tasks that use visual information. For example, it seems almost certain that visual working memory is used during reading. This hypothesis would naturally lead to taking clauses, or short phrases, as possible units of reading, instead of just single words, as is the case in most recent eye-movement based theories of reading. Distributions of regressions to previously fixated words could be used to estimate the size of visual working memory used in reading in the way distributions of rescan times were used to make this estimate for geometry.

Acknowledgements

This research was partially supported by NIMH 5-F32-MH11282-03; AFOSR 01-5-28320.

References

Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum.

Anderson, T. W., & Goodman, L. A. (1957). Statistical infer about Markov chains. *Annals of Mathematical Statistics*, 89–110.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford Univer Press.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Mem representation in natural tasks. *Journal of Cognitive Neuroscie 7*, 66–80.

Broadbent, D. E., & Broadbent, M. H. P. (1981). Recency effect visual memory. *Quarterly Journal of Experimental Psychol 33A*, 1–15.

Epelboim, J., Booth, J., & Steinman, R. M. (1994). Reading unspa text-implications for theories of reading eye movements. *Vi Research, 34*, 1735–1766.

Epelboim, J., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Erkelens, C. J., & Collewijn, H. (1995). The function of vi search and memory in sequential looking tasks. *Vision Resea 35*, 3401–3422.

Glassman, R. B., Garvey, K. J., Elkins, K. M., Kasal, K. L. Couillard, N. L. (1994). Spatial working memory score of hum in a large radial maze, similar to published score of rats, imp capacity close to the magic number 7 ± 2. *Brain Research Bull 2*, 151–159.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). T constraints in visual working memory. *Vision Research, 38*, 1 138.

Hegarty, M. (1992). Mental animation: inferring motion from st displays of mechanical systems. *Journal of Experimental Psyc ogy: Learning, Memory and Cognition, 18*, 1084–1102.

Lachter, J., & Hayhoe, M. (1995). Capacity limitations in memory visual locations. *Perception, 24*, 1427–1441.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometim worth ten thousand words. *Cognitive Science, 11*, 65–99.

Luce, R. D. (1986). *Response times*. New York: Oxford Univer Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual work memory for features and conjunctions. *Nature, 390*, 279–281.

Miller, G. A. (1956). The magical number seven plus or minus t some limits in our capacity for processing information. *Psychol ical Review, 63*, 81–97.

O'Regan, J., Rensink, R., & Clark, J. (1999). Change-blindness a result of 'mudsplashes'. *Nature, 34*, 6722.

Sternberg, S., Monsell, S., Knoll, R., & Wright, C. (1978). latency and duration of rapid movement sequences: compari of speech and type writing. In G. E. Stelmach, *Informa processing and motor control and learning*. New York: Acade Press.

Suppes, P. (1981). Future educational uses of interactive theor proving. In P. Suppes, *University-level computer assisted struction at Stanford: 1968–1980* (pp. 165–182). Stanford Univ sity.

Suppes, P. (1990). Eye-movement models for arithmetic and read performance. In E. Kowler, *Eye movements and their role in vi and cognitive processes* (pp. 455–478). Amsterdam: Elsevier S ence (Biomedical Division).

Suppes, P., Cohen, M., Laddaga, R., & Floyd, H. (1983). A pro dural theory of eye movements in doing arithmetic. *Journa Mathematical Psychology, 27*, 341–369.

Suppes, P., & Sheehan, J. (1981). CAI course in axiomatic set theo In P. Suppes, *University-level computer assisted instruction Stanford: 1968–1980* (pp. 3–80). Stanford University.

Townsend, J. T., & Thomas, R. D. (1994). Stochastic dependencie parallel and serial models: effects on systems of factorial inter tions. *Journal of Mathematical Psychology, 38*, 1–34.

Viviam, P. (1990). Eye movements in visual search — cognitive, perceptual and motor control aspects. In E. Kowler, *Eye movements and their role in visual and cognitive processes* (pp. 353–394). Amsterdam: Elsevier Science (Biomedical Division).

Walker, P., Hitch, C. J., & Duroe, S. (1993). The effect of visual similarity on short-term memory for spatial location: implicati for the capacity of visual short-term memory. *Acta Psycholog 83*, 203–224.

Zingale, C. M., & Kowler, E. (1987). Planning sequences of sacca *Vision Research*, *27*, 1327–1341.