Top-down control of eye movements: Yarbus revisited

Article *in* Visual Cognition · August 2009

DOI: 10.1080/13506280902793843

Jeff Pelz
Rochester Institute of Technology

VISUAL COGNITION, 2009, 17 (6/7), 790–811

Psychology Press
Taylor & Francis Group

Top-down control of eye movements: Yarbus revisited

Marianne DeAngelus and Jeff B. Pelz

Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA

Alfred Yarbus (1967) reported that an observer's eye movement record varied based on high-level task. He found that an observer's eye movement patterns during freeview were dramatically different than when given tasks such as "Remember the clothes worn by the people." Although Yarbus' work is often cited to demonstrate the task-dependence of eye movements, it is often misrepresented; Yarbus reported results for only one observer, but authors commonly refer to Yarbus' "observers". Additionally, his observer viewed the painting for 21 minutes with optical stalks attached to the sclera and with his head severely restricted. Although eye movements are undoubtedly influenced by high-level tasks, it is not clear how Yarbus' results reflect his unique experimental conditions. Because of Yarbus' role in the literature, it is important to determine the extent to which his results represent a sample of naïve observers under more natural conditions. We replicated Yarbus' experiment using a head-free eyetracker with 17 naïve observers. The presentations were self-paced; viewing times were typically an order of magnitude shorter than the times Yarbus imposed. Eye movement patterns were clearly task dependent, but some of the differences were much less dramatic than those shown in Yarbus' now-classic observations.

Key words: Attention; Eye movements; Visual cognition.

The doctoral work of Alfred Lukianovich Yarbus (alternatively spelled Iarbus) was published in book form as *The Role of Eye Motion in Vision Processes* in 1965 in Moscow (Iarbus, 1965). In 1967, it was translated into English by Basil Haigh and Lorrin Riggs (editor) and published as *Eye Movements and Vision* (Yarbus, 1967). The book covered both the ingenious eyetracking devices he devised to perform his experiments and the research he performed with those devices. Yarbus examined and characterized the mechanics of the oculomotor system, including the velocities and durations

Please address all correspondence to Jeff B. Pelz, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY 14623, USA. E-mail: pelz@cis.rit.edu

The authors wish to acknowledge Jason Babcock and Constantin Rothkopf for their valuable contributions.

http://www.psypress.com/viscog DOI: 10.1080/13506280902793843

of different types of eye movements. His book includes photographic records of corrective saccades, curved saccades, and eye movements of patients with disorders such as nystagmus and glaucoma. Figure 1 shows a number of the miniature optical devices, or suction "caps", designed by Yarbus that were attached to the eye to record eye movements or project stabilized retinal images. An entire chapter is devoted to the perceptual effects of stabilized retinal images, including blank fields, fields of high luminance and colour contrast, and flickering objects.

Eye movement patterns were recorded for observers viewing both stationary and moving objects, including simple stimuli, text, and optical illusions. The last chapter of *Eye Movements and Vision* shows records of observers viewing complex objects such as paintings and photographs. Although he was not the first researcher to conduct this sort of analysis (see, e.g., Buswell, 1935), he was one of the first to investigate the relationship between eye movement patterns and high-level cognitive factors. In Figures 107–112 of his book, Yarbus shows photographic records of eye movement patterns while viewing I. E. Repin's painting, "They Did Not Expect Him" (1884). The politically significant painting, portraying a Russian revolutionary returning from exile, was well known at the time. In Figure 2 (Yarbus' Figure 109), Yarbus showed a set of seven eye movement patterns for one observer as he or she viewed the painting under different instructions. In the first viewing, the observer was not given a specific instruction, but only asked to look at the painting. Before each of the subsequent six viewings, the instructions were: "Estimate the material circumstances of the family in the picture", "Give the ages of the people", "Surmise what the family had been doing before the arrival of the 'unexpected visitor'", "Remember the clothes worn by the people", "Remember the position of people and objects in the room", and "Estimate how long the 'unexpected visitor' had been away from the family".

There are striking differences in the eye movement patterns among the different conditions. These differences are more pronounced than among records of seven different observers freely viewing the painting without instruction, shown in Figure 107 of Yarbus' book. From this experiment, Yarbus concluded that the eyes "fixate on those elements of an object which

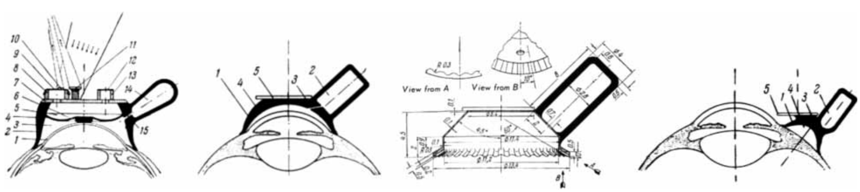

Figure 1. Diagrams of suction "caps". The rightmost cap is designed to record eye movements. From Yarbus (1967, pp. 30–33).

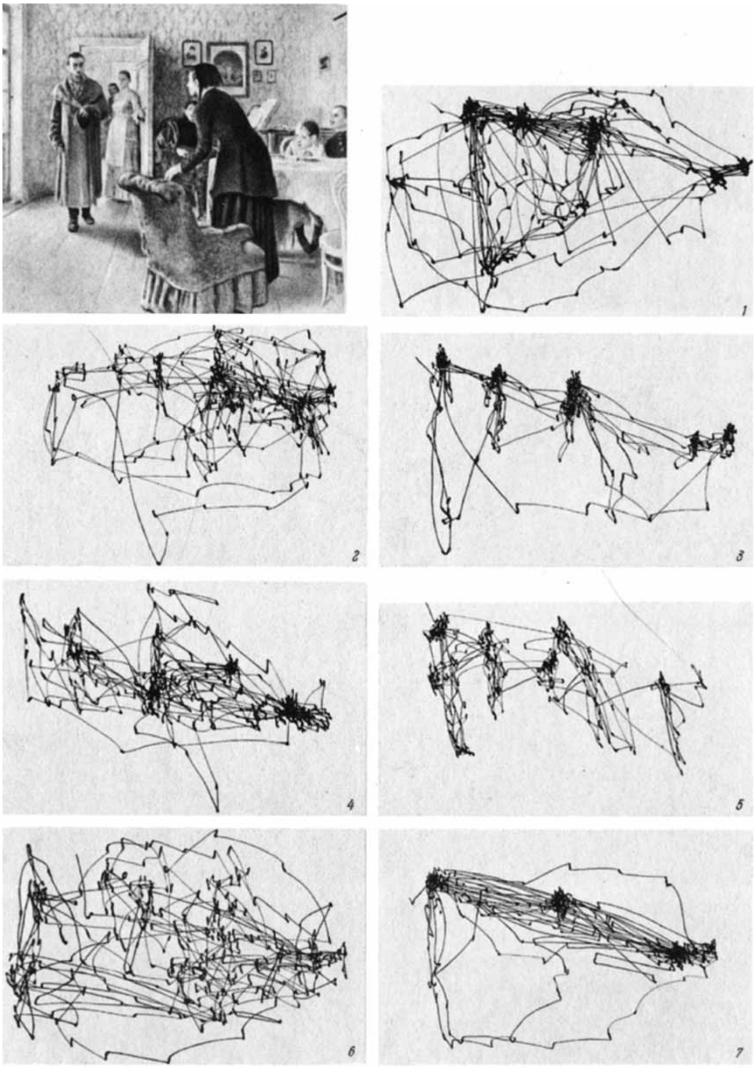

Figure 2. Figure 109 from Yarbus (1967, p. 174): Seven records of eye movements by the same subject. Each record lasted 3 minutes. The subject examined the reproduction with both eyes. 1) Free examination of the picture. Before the subsequent recording sessions, the subject was asked to: 2) estimate the material circumstances of the family in the picture; 3) give the ages of the people; 4) surmise what the family had been doing before the arrival of the "unexpected visitor"; 5) remember the clothes worn by the people; 6) remember the position of the people and objects in the room; 7) estimate how long the "unexpected visitor" had been away from the family.

carry or may carry essential or useful information" (Yarbus, 1967, p. 211). The eyes are not reactively drawn to salient, low-level properties of the image such as bright regions or edges. Instead, the elements fixated are those that provide the most information for the task at hand. As the task changes, so does the "informativeness" of certain regions, thereby changing the observer's viewing behaviour. Furthermore, the patterns and locations of eye movements give insight into what the observer was thinking.

Yarbus also noted a "cyclic" pattern of eye movements. Acknowledging that the 3-minute viewing time was more than long enough to fixate on the important regions of the picture, he noted the fact that once these regions were fixated, the observer did not move on to examine the secondary elements and details in the picture. To further investigate the temporal order of regions fixated, he repeated the 3-minute freeview task while changing the photographic plate every 25 s. From these records he noted similar scanpaths in which the observer looked at these primary regions again and again throughout the full viewing time. Yarbus concluded from other freeview experiments with additional complex images that this "cycle" can last from a few seconds to many tens of seconds (Yarbus, 1967, pp. 175, 193, 194).

IMPLICATIONS

The results from this experiment are significant in that they demonstrate a "top-down" component of visual selection, demonstrating the active nature of the human visual system. An observer's cognitive goal and past experiences interact with the visual stimulus in order to execute an appropriate behaviour. The system is not passive; it does not randomly or uniformly sample the visual environment, nor does it simply react to the stimulus. Guy T. Buswell had reported earlier that, "The directions given prior to looking at a picture have a marked influence upon the character of perception" (1935, p. 144). He based that statement on experiments comparing observers' eye movements when they viewed images without instruction to those made during (1) visual search, (2) after reading a page of descriptive material about the picture, or (3) when asked which of two images was preferred. Buswell's work was likely influenced by his thesis adviser, Charles H. Judd, who investigated the relationship between eye movements and the perception of visual illusions (Judd, 1905; Judd & Courten, 1905). Buswell's work predated Yarbus' by three decades; Yarbus took the investigation of the relationship between eye movement patterns and high-level cognitive factors beyond Buswell's pioneering work and has received far more attention in recent years.

YARBUS' METHODS

Yarbus' Figure 109 is very well known and often cited in literature concerning eye movements, behaviour, top-down control of oculomotor function, and high-level scene perception. The fact that is often overlooked is that the records published are for only one, unidentified observer. There is no reference in *Eye Movements and Vision* (1967) to other observers performing these tasks or producing similar results, and the original Russian journal article (Iarbus, 1961) contains the same set of records as shown in Figure 109 of the translated text.

It must also be noted that each of the photographic records was made as the observer viewed the picture for a full 3 minutes. During the recording, the observer's eye was anaesthetized and his eyelids were taped open with heated strips of adhesive plaster as seen in Figure 3. Figure 1 shows the small suction device ("cap") holding a small mirror that was firmly attached to the sclera. Light projected onto the mirror and reflected onto a piece of photographic film created the eye movement record. The observer's head was constrained using the chin and forehead rests seen in Figure 3.

The results of Yarbus' experiment are important and have been widely referenced as among the first to demonstrate the task-dependent nature of saccadic eye movements. The fact that his result was based on a single,

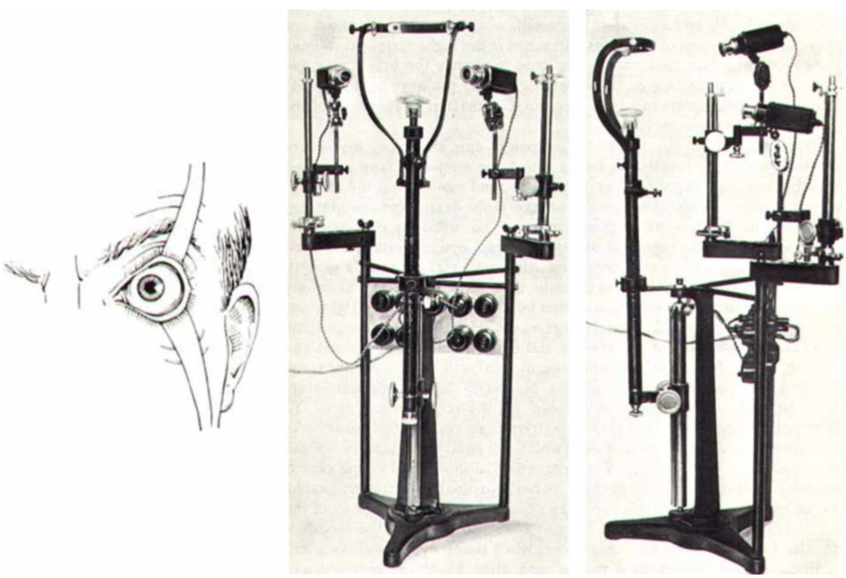

Figure 3. Left: Configuration of eyelids and strips of adhesive plaster during recording of eye movements. Right: Apparatus used in recording eye movements. The setup contains chin and forehead rests, light sources, and a control panel. From Yarbus (1967, pp. 41, 44).

unidentified observer was sufficient reason to replicate the experiment with a larger sample of naïve observers. The restrictive, intrusive, (and likely painful) nature of the instrumentation available to Yarbus was further motivation to repeat the experiment with contemporary eyetracking systems. We were also interested in examining the eye movement patterns of observers who were free to view the painting for only as much time as they thought necessary to answer the questions posed, rather than being forced to continue viewing for three minutes on each of the seven presentations. Finally, the relatively high-resolution, 2-D data output by modern eyetracking instrumentation allows the application of new analytical tools that were not available to Yarbus when he performed his pioneering research.

METHODS

Gaze tracking

Gaze position was determined by an integrated eye/head-tracking system consisting of an Applied Science Laboratories Series 501 video-based eyetracker ("ASL") and a Polhemus 3-Space Fastrak Magnetic Head Tracker ("MHT"). The MHT uses a fixed transmitter mounted just behind the observer and a small receiver attached to the ASL headgear. Position (x, y, z) and orientation (azimuth, elevation, roll) of the receiver are reported with respect to the transmitter. Combining the eye-in-head signal from the ASL with the head-in-space signal from the MHT provides a gaze-in-space vector that is used to determine the intersection of gaze on the display. The system allows image-centric gaze monitoring without constraining natural head movements. Position and orientation of the MHT transmitter is critical to the system calibration. To ensure repeatability, a three-axis laser level was affixed to the transmitter stand and three reference marks were placed on the ceiling, the base of the stimulus display, and an adjacent wall. Before each trial, the lasers were checked to ensure that the relative position and orientation of the transmitter and display were correct.

Calibration

The eye/head gaze tracking system was calibrated at the start of each session, and monitored every 10 images throughout a trial. Calibration was repeated if the reported gaze position fell outside circles of 1° radius surrounding the calibration check points.

Calibration consisted of optical alignment of the cameras and illumination source with the observer's eye and scene, defining nine calibration points on the image display, and capturing the vector difference between pupil centre and the first-surface corneal reflection at each of the nine points. The MHT system corrects for head movements, but initial calibration was performed with the head stable. Rather than use a mechanical constraint for calibration, observers held their head stable by maintaining a point projected from a semiconductor laser mounted to the ASL headgear within the circle about the central calibration point. Once the calibration was completed and verified, the laser was turned off and the observer was free to move his/her head naturally.

Gaze accuracy

The accuracy of the output fixation locations is dependent on several factors. First, because the video system tracks the retroreflection from the retina and the corneal reflection, differences between observers can affect performance. Using the MHT in conjunction with the eyetracker can introduce error if the headgear moves with respect to the head during the experiment. In a previous study in which this system was used (Babcock, 2002), angular deviation between fixation points and target calibration points (during calibration checks) was found to range between 0.4° and 1.1° across 26 observers, with an average deviation of 0.7° of visual angle. Additional uncertainty of fixation location may be introduced by the algorithm used to classify fixations and saccades, discussed later. An eye movement record was discarded if the final angular deviation was greater than 1.0°.

Fixation, saccade, and blink classification

Gaze position in the image plane was captured at 60 Hz, with no field averaging. Fixations, saccades, blinks, and any track losses were extracted from this data stream using an adaptive-velocity threshold; the threshold used to distinguish between fixations and saccades changes dynamically with the amount of noise in the signal, which varies between observers. See Rothkopf and Pelz (2004) for details on the detection algorithm.

Observers

Twenty-five observers (16 male, 9 female, ages 18–45 years, mean = 24 years, $SD = 5.9$) were compensated for participating in the experiment. All had normal or corrected-to-normal vision (only contact lenses were allowed) and were naïve to the purpose of the experiment. The Rochester Institute of

Technology Institutional Review Board approved the study, and informed consent was obtained from each observer after an explanation of the nature and possible consequences of the study.

Stimulus display

Images were displayed on a 50-inch Pioneer 503CMX Plasma display, driven by 1280×768 digital RGB input via a Pioneer PDA-5002 video card in linear display mode. While observers were free to move their heads, they were seated approximately 40-inches from the display, at which distance it subtended 50° of visual angle horizontally and 35° vertically. Images were presented in 24-bit colour at a resolution of 1280×768 pixels. Figure 4 shows the experimental setup. Repin's painting did not fill the available screen size and was presented at a resolution of 870×768 pixels, subtending a visual angle of $40^\circ \times 35^\circ$.

Figure 4. Experimental setup showing the plasma display, the eyetracking headgear, the magnetic head tracker transmitter, and receiver.

Replication of experiment

Yarbus' tasks were replicated within a larger experiment in which observers viewed a set of 57 digital images of paintings, photographs, and drawings. Observers were told to simply view the remaining artwork. The viewing was self-paced; the observer pressed the spacebar to move on to the next image. Between every 10 images, the observer was asked to look at a sequence of nine points to check the accuracy of the track. Observers were recalibrated if needed. Following the calibration check, a screen with a written instruction was presented. These instructions are the same as those reported by Yarbus (1967), and were also presented in the same order. (For the no-instruction "Freeview" task, Repin's painting was shown randomly within the first 10 images without any instruction; it was simply another painting within the set of random images.) The only exception was that "Estimate the material circumstances of the family in the picture" was reworded as "Estimate the financial circumstances of the family in the picture" because during a pilot experiment some observers did not readily understand the meaning of the original wording. When the observer had read the instruction, he or she pressed the spacebar to view Repin's "They Did Not Expect Him". The observer then performed the task, which in most cases involved answering questions out loud. When the observer completed the task, the spacebar was pressed to freely view the next random image. Including the calibration checks, instruction screens, and repetitions of Repin's painting, there were 78 images displayed during the experiment. The experiment lasted between 11 and 20 minutes. Of the 25 observers who performed the experiment, 17 met the one-degree calibration criterion throughout the full trial and were included in the subsequent analysis.

Although the focus of the study was to replicate Yarbus' tasks under more natural viewing conditions with a larger set of observers, we did replicate Yarbus viewing durations with one observer (male, age 28); eye movements were recorded as he viewed Repin's painting in 3-minute intervals for each of the seven tasks. For this experiment, no other images were shown between the tasks. This experiment lasted 21 minutes, plus the time the observer spent reading the instruction screens.

RESULTS

Given that the results published by Yarbus were very qualitative, and lacked any information about the temporal sequence of viewing, it is difficult to make a direct, quantitative comparison between the results of the present experiment and those reported by Yarbus. In the following sections, several qualitative and quantitative methods are used to compare the eye movement patterns of observers across the different tasks and viewing conditions.

Self-terminated viewing

Scanpaths. Figure 5 shows eye movement records of two observers as they performed each of the seven tasks during this experiment. For Observer A, shown on the left, the viewing times ranged from 6 to 92 s. Although the times are all well under 3 minutes, the eye movement patterns are remarkably similar to those published by Yarbus. Fixations in the "Freeview", "Ages", and "How long away" tasks fell primarily on faces and figures. The "Financial" and "Position" tasks elicited more spatially distributed patterns of fixations. Observer B represents an atypical observer, whose eye movement patterns are not drastically different among tasks. The view times were

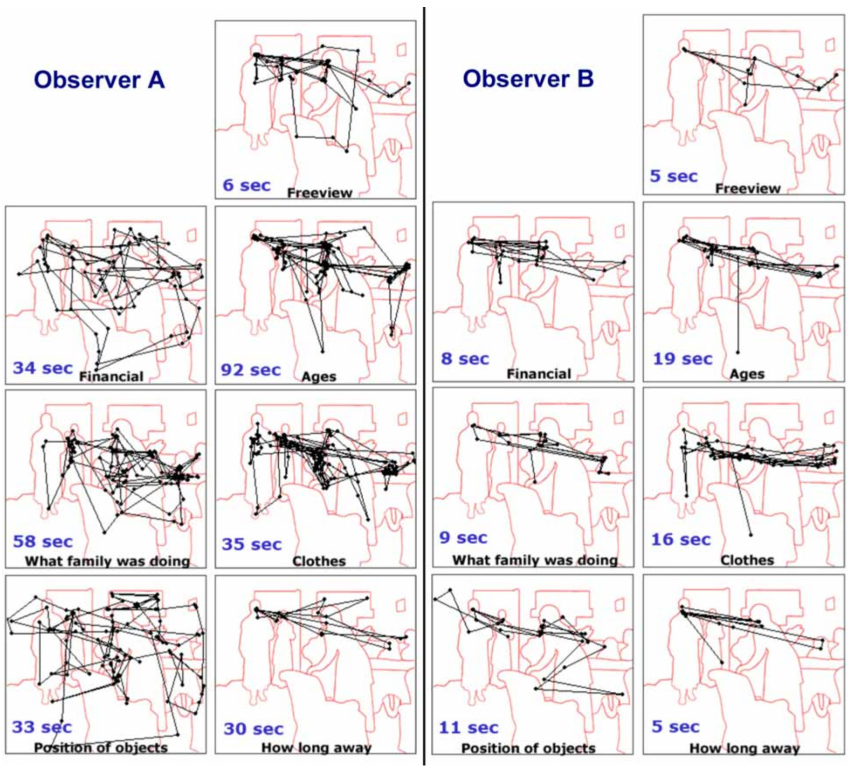

Figure 5. Eye movement records from two observers as they performed each of the seven tasks. Observer A represents an observer whose view patterns resembled those published by Yarbus. Observer B represents an atypical observer, whose view patterns did not differ significantly among tasks.

also shorter, ranging from 5 to 19 s. Both observers completed all tasks and answered questions out loud; Observer A gave more thoughtful and detailed answers compared with Observer B, which accounts for the differences in total task time.

Total viewing time. Whereas Yarbus had an enforced 3-minute viewing time, the self-terminated experiment allows us to examine how much time an observer spent performing each of the seven tasks. Figure 6 shows the total viewing time for each observer in each task. Viewing times were for the most part an order of magnitude less than in Yarbus' experiment. For the "Freeview" task, observers viewed the painting for only nine s on average. Nineteen seconds was the average time observers spent answering the question, "Estimate the financial circumstances." When asked to give the ages of the people, the painting was viewed for an average of 50 s, which is significantly longer than any other task. For the tasks "Surmise what the family had been doing", "Remember the clothes", and "Remember the

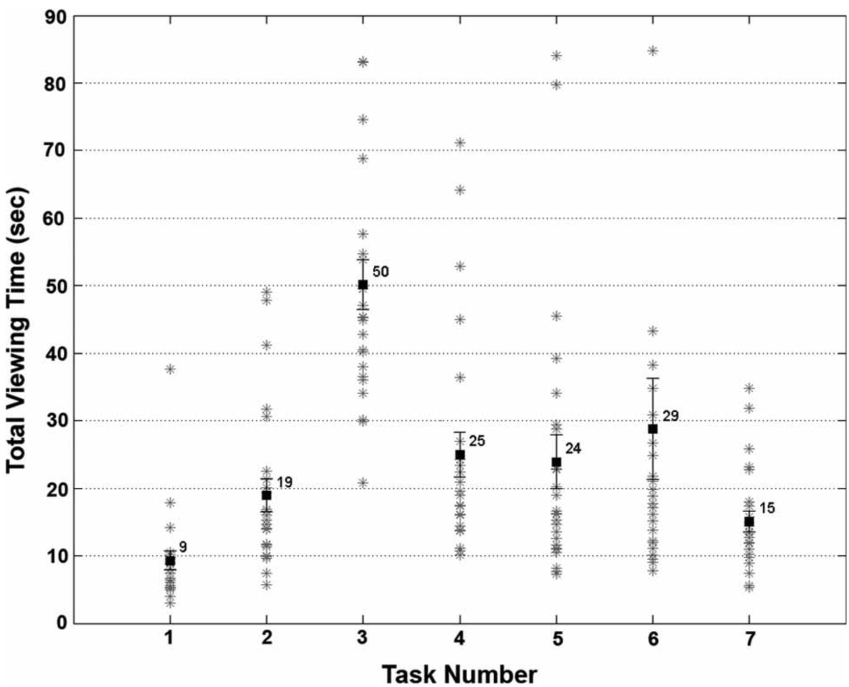

Figure 6. View times for each of the seven tasks. Black rectangles mark the average across all observers. Error bars represent one standard error of the mean. Tasks: 1: Freeview; 2: Financial circumstances; 3: Give the ages; 4: Surmise what family was doing; 5: Remember the clothes; 6: Remember the position of people and objects; 7: Estimate how long the visitor was away.

position of people and objects", the image was viewed for an average of 25, 24, and 29 s, respectively. The last task, "Estimate how long the visitor has been away", was completed in an average of 15 s.

Fixation duration. Another measure that was not reported by Yarbus was the duration of fixations in each task. Figure 7 shows histograms of fixation durations pooled across observers for the seven tasks. Listed on

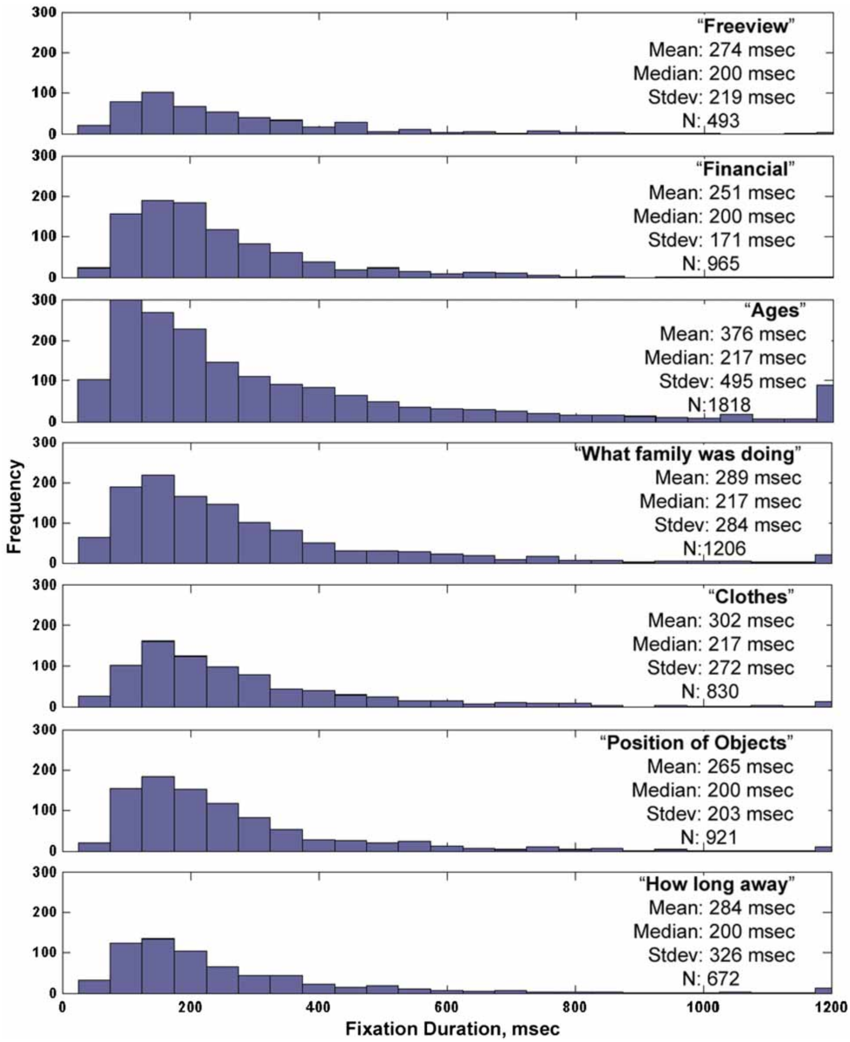

Figure 7. Histograms of fixation durations for each task across all observers.

each graph are the mean, median, and standard deviation of the distribution, and number of fixations. All the tasks had median fixation durations of near 200 ms. Additionally, all but one task showed a peak in the histogram near 150 ms; the histogram for the task "Give the ages of the people" peaks at only 100 ms, with 22% of fixations lasting 100 ms or less. Interestingly, this task also elicited the longest fixations; about 7% (133) fixations were longer than 1 s in duration. Yarbus noted that his observer had few saccades between faces, suggesting long fixation times.

Regions of interest. Yarbus' results focused entirely on the spatial distribution of fixations, in the form of photographic records shown in Figure 2 and comments derived from those records. Specifically, Yarbus often commented on the attention given to primary (faces and figures), secondary (furniture and objects in the foreground), and background elements of the image. We sought to quantify the amount of time an observer viewed these regions and to compare our results with Yarbus' commentary.

The image of Repin's painting was segmented into 22 different regions (e.g., the man's face, the man's figure, chairs, floor, etc.). The regions and associated labels are shown in Figure 8. The total gaze duration in each of these regions was found for every observer. These durations were then normalized by the viewing time for each observer and task to produce the percentage of time spent viewing each region. The results for each task are also shown in Figure 8, averaged across all observers.

During the freeview task, 20% of the time was spent fixating on the man's face, followed by the faces of the mother, wife, maid, and children. This result agrees with Yarbus' observer's behaviour; the most fixations fell on the faces of the people followed by the figures.

The distribution of gaze for the task "Estimate how long the 'unexpected visitor' was away" is very similar to that of the freeview task. Again the faces received the most fixations. More of the viewers' time, around 35%, was spent looking at the man's face for this task. His figure and the faces of the mother and children were also important. For this task, Yarbus noted particularly intensive movements between the faces of the children and man.

The task "Give the ages of the people" resulted in a more uniform distribution across the faces, with very little attention given to background elements. Yarbus reported that for this task, all of the observer's attention was concentrated on the faces, with few saccades between faces.

For the task, "Estimate the financial circumstances of the family", the faces were again the most attended-to regions. However, the clothing and furniture received more fixations than in previous tasks. Yarbus' observer paid particular attention to the women's clothing, armchair, and tabletop.

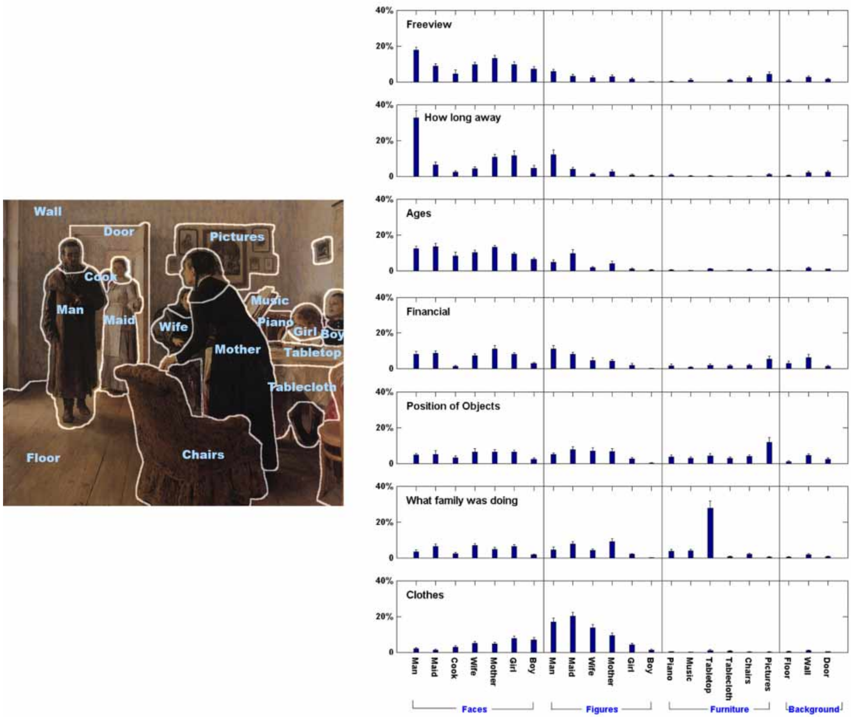

Figure 8. Left: Image segmented into 22 regions with associated labels. Right: Percentage of time spent viewing each region. Error bars represent standard error of the mean across 17 observers.

For the task, "Remember the position of people and objects in the room", every region received fixations in most cases, producing a more uniform distribution of gaze durations throughout the image. The pictures on the wall were examined for a larger fraction of time (12%) than in other tasks. The more uniform distribution across primary, secondary, and background elements agrees with Yarbus' report that his observer examined the whole room and all of the objects.

When asked to "surmise what the family had been doing before the arrival of the 'unexpected visitor'", the tabletop was evidently the most informative, as viewers spent 30% of their time looking at it. The piano and sheet music also received a larger percentage of viewing time than in other tasks. These are also the regions that Yarbus' observer attended to the most: "[T]he observer directed his attention particularly to the objects arranged on the table, the girl's and the woman's hands, and to the music" (Yarbus, 1967, p. 192).

Last, when asked to "remember the clothes worn by the people", viewers spent the majority of time examining the figures in the image, as did Yarbus' observer. The faces of the people in the image were invariably fixated, regardless of the task.

Between-observer variability. To measure the influence of task on viewing behaviour, the between-observer variability and between-task variability were compared. The fraction of time spent viewing each of the regions in the image can be thought of as a 22-element feature vector. As a measure of between-observer variability within each task, the Euclidean distance between each possible pair of vectors was found. The average distances within each task are shown in Figure 9. Also shown is the between-task (within-observer) distance, averaged across all observers; 0.39. For the first six tasks, the between-observer variability is below the between-task variability, indicating that the eye movement patterns across observers for that task were more similar than the patterns of one person performing different tasks. "Give the ages" task was the most similar between observers, with an average distance of 0.25, suggesting similar viewing strategies (e.g., most attention given to the faces). The average distance for the "How long

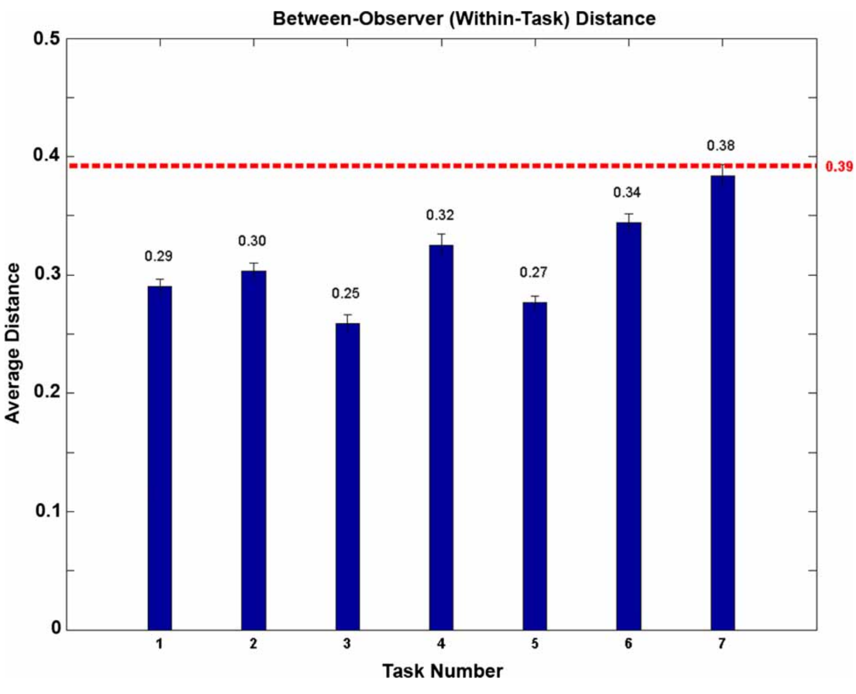

Figure 9. Between-observer variability for each task defined by the average distance between region histogram vectors. Error bars represent one standard error of the mean. The dotted line represents the within-observer (between-task) variability, averaged across 17 observers.

away" task, 0.38, was very close to the within-observer distance, suggesting that observers attended to different regions when performing that task. This was the last and shortest task; it is possible that some observers did not refixate on informative regions and instead used information gathered during previous viewings.

Enforced three-minute view, all tasks

Scanpaths. Yarbus' enforced 3-minute view time was replicated for a single observer. By the end of the experiment, he had been looking at the painting for 21 minutes. Figure 10 shows the eye movement patterns as well as percentage of time spent per region of interest. When compared to the average behaviour during the self-terminated condition, shown in Figure 8, there are many similarities. This may support Yarbus' report of "cyclic" eye movement patterns in which the same informative regions are re-fixated

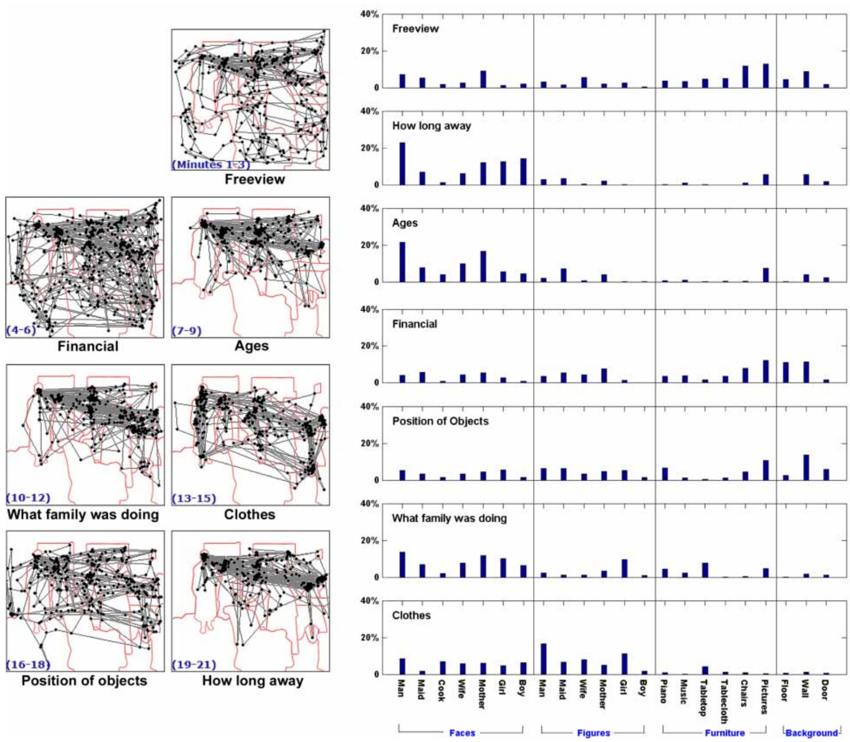

Figure 10. Left: Eye movement records of one observer. For each task, the observer viewed the painting for 3 minutes. The minutes in which each task was performed are labelled in the corner of each record. Right: Percentage of time one observer spent viewing each region as he viewed the image for 3 minutes per task.

throughout the extended view time. For example, in the "How long away" task, the observer spent most of his time looking at the man's face, and the expressions on the faces of the mother, girl, and boy; these were also the most fixated regions in Figure 8. The observer reported that he was trying to guess whether both of the children recognized the man. For the "Financial" and "Position of Object" tasks, the histograms are more uniform, as they were for the self-terminated condition. For the "Ages" task, this observer's behaviour was also similar in that almost all fixations fell on the faces, although he did spend some time examining the pictures on the wall.

There were also some dissimilarities between this observer's viewing behaviour and the results of the self-terminated condition. For example, the histogram for "Clothes" shows that the observer spent almost an equal amount of time looking at the people's faces as he did at their clothing. When guessing "What the family was doing", the observer did examine the tabletop, but unlike the self-terminated task where observers spent almost 30% of their time studying it, this observer spent less than 10% of the time fixating the tabletop. Instead, more time was spent examining the faces of the people in the room; the observer reported that by the end of the task, he was developing a story about the characters in the scene, their personalities, and what they were doing.

For the "Freeview" task, although the observer did spend time looking at the faces in the scene, many of the other objects, or "secondary details" were attended to, particularly the furniture and pictures on the wall. This result differs from both the self-terminated results as well as the behaviour of Yarbus' observer. It is possible that this observer's behaviour was similar to Yarbus' at the beginning of the experiment, and changed as time went on. In the next section we will investigate the effect of the enforced 3-minute view time on this task.

Temporal order. Yarbus noted that over several minutes, the eye movement "record obtained will clearly show that, when changing its points of fixation, the observer's eye repeatedly returns to the same elements of the picture. Additional time spent on perception is not used to examine the secondary elements, but to reexamine the most important elements" (Yarbus, 1967, p. 193). In an additional experiment, he found that an observer's eye movement patterns were similar during each of 25-s intervals during a 3-minute free examination of the painting.

To determine if the observer's viewing behaviour was indeed cyclic, or whether it changed over time, we examined the temporal order of regions fixated (i.e., shifts of gaze) for the "Freeview" task, shown in Figure 11. These regions are grouped into three areas of interest: Faces, figures, and background/furniture. This grouping corresponds to the elements Yarbus referred to as primary (faces and figures) and secondary (other foreground

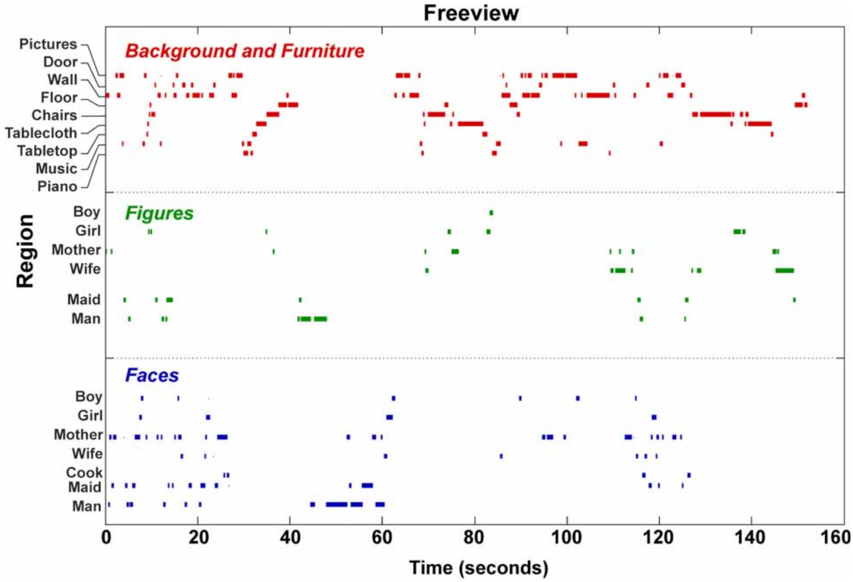

Figure 11. Temporal order for one observer during the "Freeview" task. The length of each line represents fixation duration. Each vertical level represents a different region. The bottom section is all of the Face regions. The regions in the middle section are the Figure regions. The top section is all other regions. (The time does not extend to the full 3 minutes, or 180 s, because the time during blinks and saccades has been removed.)

items and background). During the first 20–30 s of viewing, the observer made many short fixations on each of the faces, most of the figures, and almost all the background elements. For the next 30 s, the observer seemed to change behaviour. Gaze durations become longer, and were spent on the background elements for 10 s, then to the man's figure, and then to the man's face. During the next 30 s, the background elements were again examined. After that, there were again short fixations distributed among the faces, figures, and background, which is similar to the behaviour at the beginning of the task. For the rest of the viewing time, the background elements and figures were examined with long gaze durations.

In the "Clothes" task, again we see a change in viewing behaviour, illustrated in Figure 12. During the first half of the viewing time, the faces and clothes are each examined in turn with long gaze durations. During the second half, the background elements are examined, and gaze shifts frequently and more quickly among different regions.

The "Financial" task, shown in Figure 13, represents a task in which the behaviour is more consistent over time: Short fixations are distributed among all the areas throughout the viewing period.

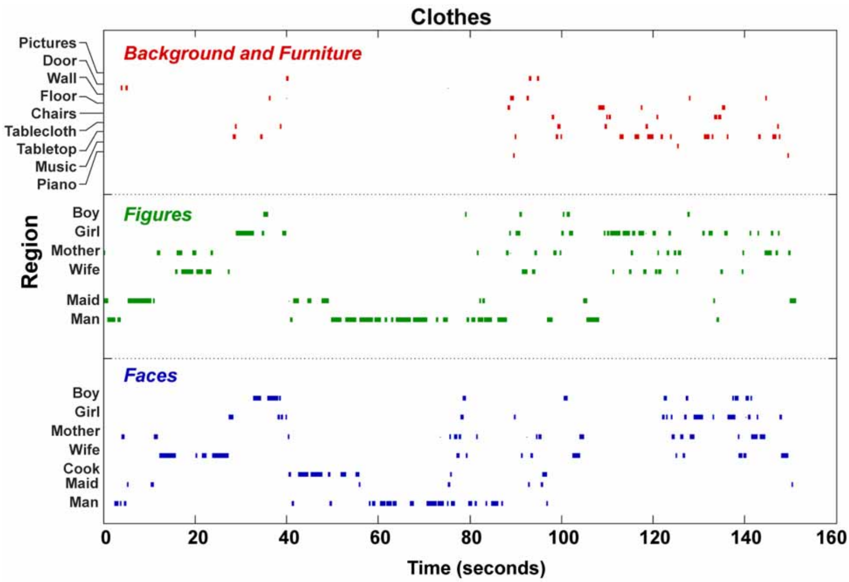

Figure 12. Temporal order of fixations during "Remember the clothes" task.

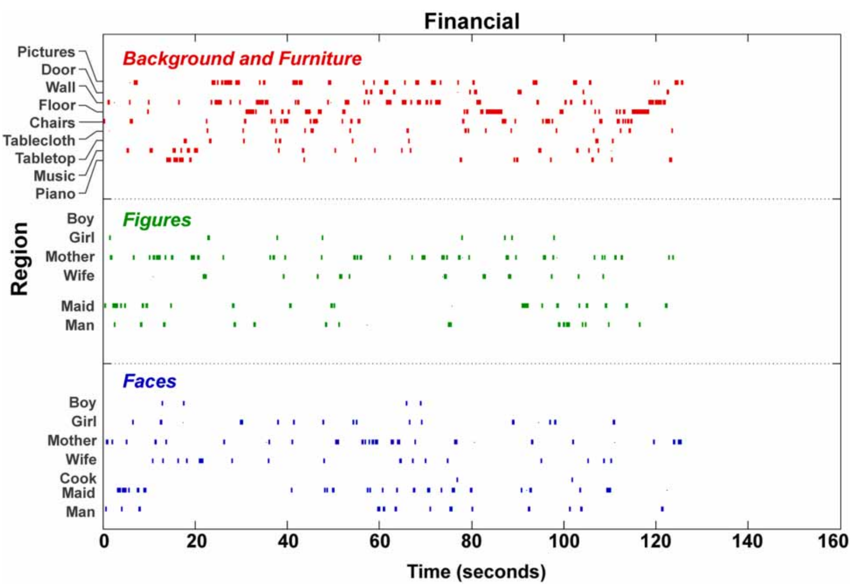

Figure 13. Temporal order of fixations during "Financial" task.

From these visualizations we see that our observer did indeed fixate on primary regions (faces) throughout the viewing time. However, Yarbus' notion that an observer would repeat the same scanpath over and over, without moving on to secondary objects is too severe a generalization.

CONCLUSIONS

Alfred Yarbus' classical experiment showed the influence of task on oculomotor behaviour. In his experiment, Yarbus had one observer repeatedly view a painting seven times for 3 minutes each time; before each viewing the observer was given a different instruction. From records of the spatial pattern of eye movements, Yarbus' concluded that the observer's fixations repeatedly fell on regions that were most important or "informative" for the particular task, forming a cyclic pattern throughout the 3-minute viewing time.

We replicated the experiment with 17 naïve observers using a free-head, video-based eyetracker. Viewing times were self-terminated. Observers performed each of the seven tasks that Yarbus used in his experiment while viewing a large-field reproduction of I. E. Repin's painting, "They Did Not Expect Him". The tasks were: (1) Freeview (no instruction); (2) "Estimate the financial circumstances of the family in the picture"; (3) "Give the ages of the people"; (4) "Surmise what the family had been doing before the arrival of the 'unexpected visitor'"; (5) "Remember the clothes worn by the people"; (6) "Remember the position of people and objects in the room"; and (7) "Estimate how long the 'unexpected visitor' had been away from the family".

The average view time for each task was 9, 19, 50, 25, 24, 29, and 15 s, respectively; these times are significantly less than the enforced 3-minute view time of Yarbus' observer. The task, "Give the ages of the people", elicited both the shortest and longest fixation durations; 22% of fixations were 100 ms or shorter, and 7% were longer than 1 s. All other tasks had similar distributions of gaze durations with median durations near 200 ms.

Most observers' eye movement patterns were subjectively similar to those reported by Yarbus, with faces invariably fixated, and the overall viewing pattern varying with instruction. A few observers, especially those with shorter viewing times, did not show dramatic shifts with instruction.

A Euclidean-distance metric between the feature vectors describing the distribution of gaze within regions of interest in the painting was used as a measure of both between-task and between-observer variability. All tasks except for "How long away" resulted in between-observer average distances that were less than the between-task distance. The "Give the ages" task resulted in the smallest distance of all tasks, indicating that for this task the eye movement patterns were most similar among the observers. The "How long away" task showed the most variability among observers, suggesting that observers used different viewing strategies to complete this task.

Yarbus' observation of a "cyclic" behaviour of eye movements was also investigated by examining the temporal sequences of fixations of one observer who performed each of the seven tasks with an enforced 3-minute view time. The viewing behaviour for the "Financial" task was consistent throughout the viewing time in that the observer frequently shifted his gaze among faces, figures, and background objects. Other tasks, including "Remember the clothes", show a distinct change in behaviour. For this task, the observer began the task by examining the clothes and faces of the people in the scene, with some fairly long gaze durations. Halfway through the viewing time, the observer began to examine furniture and background elements. Yarbus' observer did not show this behaviour of moving attention to secondary elements, but instead spent the entire 3 minutes refixating on the "informative" regions. This cyclic behaviour may in part be a result of the uncomfortable setup of his experiment, which may have made his observer conscious of where he was looking. In doing so, the observer may have changed his viewing behaviour to adhere strictly to the given instructions. It is also possible that an observer, even a naïve observer, having looked at the same painting for over 20 minutes, would begin to consciously narrow his or her gaze based on the instructions.

Yarbus' basic result, that eye movement patterns are tied to the immediate task, is fundamental to the so-called "bottom-up/top-down debate". It is in this context that many of the references to his work are made (e.g., Baars, 1988, Ballard, Hayhoe, Pook, & Rao, 1997; Gould, 1976; Hoffman & Subramanium, 1995), and recent research continues to show the importance of top-down influence on eye movement patterns (e.g., Einhäuser, Rutishauser, & Koch, 2008, Rothkopf, Ballard, & Hayhoe, 2007). Yet a significant literature exists supporting bottom-up models that predict fixation density or scanpath sequences based on low-level image features. In an early instantiation, Koch and Ullman (1985) proposed a network that calculated the relative conspicuity of each location in the visual scene based on the low-level features. Work on such "saliency-map" methods continues, with some researchers attempting to merge the bottom-up and top-down approaches (Canosa, 2005; Itti & Koch, 2001; Parkhurst, Law, & Niebur, 2002; Peters & Itti, 2008). The importance and visibility of Yarbus' work is likely to grow in the foreseeable future as the community continues to work toward models of that successfully merge the two approaches.

REFERENCES

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Babcock, J. B. (2002). *Eye tracking observers during color image evaluation tasks*. Unpublished MS thesis. Rochester Institute of Technology, Rochester, NY.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–742.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.

Canosa, R. L. (2005). Modeling selective perception of complex, natural scenes. *International Journal on Artificial Intelligence*, *14*(1–2), 233–260.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, 8(2):2, 1–19.

Gould, J. D. (1976). Looking at pictures. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes* (pp. 323–346). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hoffman, J. E., & Subramanium, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics*, *57*, 787–795.

Iarbus, A. L. (1961). Eye movements during examination of complex objects (in Russian). *Biofizika*, *6*, 207–212.

Iarbus, A. L. (1965). *Rol' dvizheniya glaz v protsesse zreniya* [The role of eye motion in vision processes] (in Russian). Moscow: Nauka Press.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203.

Judd, C. H. (1905). The Müller-Lyer illusion. *Psychological Review Monographs*, *7*(Suppl. 29), 55–82.

Judd, C. H., & Courten, H. C. (1905). The Zöllner illusion. *Psychological Monographs*, *7*, 112–139.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of visual selective attention. *Vision Research*, *42*(1), 107–123.

Peters, R. J., & Itti, L. (2008). Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions in Applied Perception*, *5*(2), 1–19.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, 7(14):16, 1–20.

Rothkopf, C. A., & Pelz, J. B. (2004). Head movement estimation for wearable eye tracker. In *Proceedings of the Eye Tracking Research and Application symposium* (ETRA 2004) (pp. 123–130). San Antonio, TX: ACM.

Yarbus, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York: Plenum Press.