PSIHOLOGIJA, 2009, Vol. 42 (3), str. 307-327 UDC 159.931.072

DOI: 10.2298/PSI0903307K

EYE-TRACKING STUDY OF ANIMATE OBJECTS[1]

Vanja Kovic[2]
Department of Psychology, University of Novi Sad, Serbia

Kim Plunkett
Department of Experimental Psychology, University of Oxford, UK

Gert Westermann
Department of Psychology, Oxford Brookes University, UK

This study involved presentation of animate objects under labelling and non-labelling conditions and examination of participants' looking pattern across these conditions. Results revealed a surprisingly consistent way in which adults look at the pictures of animate objects. The head/eyes of the animals were a typical region attracting a number of fixations, but also some other parts of animals (e.g. the tail in cats, or the udder in cows and the body in snakes). Furthermore, not only did participants tend to look at similar regions of the pictures of animate objects, but also the looking order to these regions was consistent across participants. However, contrary to the original predictions, these patterns of fixations were similar across the naming and non-naming conditions ('Look at the <target>!', 'Look at the picture!' and 'What's this?', respectively), which led to the conclusion that participants' consistency in processing animate objects was not reflecting underlying mental representation evoked by labels, but was rather driven by the structural similarity of animate objects, in particular the presence of a head.

Key words: *animate objects, eye-tracking, mental representations*

[1] This work was funded by ESRC (RES-000-23-1322) and Leverhulme Trust grants awarded to Prof. Kim Plunkett and a Clarendon grant awarded to Vanja Kovic.
[2] ✉: vkovic@ff.uns.ac.rs

Imagine you were shown a picture of a cat. Are there some features that you would be more likely to attend to? Perhaps the head? The tail? Would other people do the same? And, if so, what does this tell us about the nature of the underlying representation in your mind?

Now imagine hearing the word "cat" before an image is presented. Would invoking of a mental representation prior to seeing the picture change your response to the picture? Would your picture processing differ if you were told that you were going to see a picture of a dog, or a table, when you were in fact presented with a picture of a cat?

The study of saccadic exploration of complex images was pioneered by the Russian psychologist Yarbus (1967). Using relatively crude equipment he was able to record the fixations and saccades observers made while viewing natural objects and scenes. In one example he instructed an observer to explore a face during a three minute interval. He found that most fixations were allocated around the eyes, nose, mouth and the general outline of the face. More interestingly, he found that eye movements were not only dependent on the structure of the picture, but also on the task itself. Saccadic movements were rather different when a picture was examined freely without any specific objective, compared to when performing a task like estimating the age or material circumstances of the people in the picture.

Ever since, the eye-tracking methodology has been widely used in investigating various aspects of language and vision, such as: scene perception (Henderson, 1999; Loftus & Mackworth, 1978), face perception (Althoff & Cohen, 1999), reading (Ferreira & Clifton, 1986; Just & Carpenter, 1987; Mak et al., 2002; Rayner, 1998; Tanenhaus et al., 1995), language production (Griffin & Bock, 2000; Meyer et al., 1998) and comprehension (Frisson & Pickering, 1999), problem solving (Grant & Spivey, 2003; Hegarty & Just, 1993) etc.

The underlying hypothesis in eye-tracking research is that eye movements are at least partially correlated with attentional processes. This means that eye movements may provide a tangible trace of the attentional processes underlying performance in such tasks. Recording eye movements is then expected to allow a fine-grained analysis of the spatial and temporal aspects of performance, that is to say the presence of eye fixations to object dimensions is taken as a proximal measure of attention to those dimensions. Nowadays, we know that it takes about 200ms to initiate a *saccade* (Altmann & Kamide, 2004; Dahan et al., 2001; Huettig & Altmann, 2005), i.e. from the time a stimulus is presented until the eye starts moving, and another 30-120ms to complete the saccade. During saccades processing of the visual image is suppressed (but not entirely inhibited)[3]. Thus, processing of the retinal image takes place mainly between the saccades, during the so-called *fixations*, which last for about 100-600ms.

Given this interest, it is important to consider the nature of the link between attention and eye movements. At present, this is a controversial issue. It is generally

[3] Actually, saccadic suppression begins about 50ms before saccades are initiated (Barber & Legge, 1976)

accepted that eye movements and shifts of attention are at least partially dissociable; this is indicated by the large body of research showing that attention can be shifted covertly in the absence of concomitant eye movements (e.g., Downing, 1988; Posner, 1980; Posner et al., 1978). Using the non-predictive, peripheral, double cue paradigm that calls attention without eye movements, Posner (1980) found that shifts in attention normally occur more quickly than saccadic eye-movements, actually in less then 100ms, long before saccadic eye movement can be initiated, thus providing support for the dissociation between eye-movements and shifts in the attention.

Cooper (1974) and Tanenhaus et al. (1995) were first to demonstrate how visual attention towards the external world can be mediated by the unfolding spoken language. This paradigm known as 'visual world paradigm' was extensively used subsequently by Altmann (1999); Altmann & Kamide (2004); Huettig & Altmann (2005); Kamide & Altman (2003) and Altmann & Kamide (2007) to demonstrate that language mediated eye movements permit "investigation of the interplay between the mental world and the external visual world" (Altmann & Kamide, 2007).

These studies of anticipatory eye-movements typically involved presentation of few object on the visual scene while participants listened to the description of the visual scene. For example, Kamide et al. (2003) demonstrated language-guided visual exploration for scenes such as one showing a man, a girl, beer and sweets whereby, participants when hearing 'the man will taste the beer' or 'the little girl will taste the sweets' tended to look more during the verb 'taste' at whatever object was most plausibly tasted (i.e., 'beer' when the subject was 'the man', and 'sweets' when the subject was 'the little girl'). Following on from this idea, the aim of the current study was to present participants with a single object at the time, which would either be named or not named prior to visual-object presentation and examine which features of the objects participants tent to pay attention to in the naming and non-naming conditions. Evoking the mental representation prior to visual presentation of the object achieved through labelling of the objects was expected to reveal different pattern of eye-movements in comparison to the non-naming condition.

The questions of particular interest here are the following: What visual features do people attend to in the early stages of visual object processing? Which parts of object tend to attract participants' attention? Is processing of familiar objects category specific? And is participants' looking behaviour language-driven, that is to say, does the labelling of the object result in different looking patterns when the objects they are looking at have been named? Are these looking patterns stable across different exemplars and irrespective of picture orientation?

Adult participants were expected to be faster at initiating and programming their eye-gaze towards the visual features of animate objects in a naming condition, where image presentation occurred directly after the animal was named compared to a non-naming condition. A different, less diffuse pattern of eye-movements was expected when the objects were named compared to non-naming conditions.

Hearing the label was expected to evoke a mental representation of a named animal which was then expected to focus or direct participants' attention to the more relevant dimensions of the animal (such as the tail in a cat or ears in a rabbit).

METHOD

Participants

Thirty-six healthy, right-handed, native speakers of English were recruited for the first eye-tracking study. They were all first year undergraduate students from the University of Oxford with normal hearing and normal or corrected to normal vision. They received course credits for their participation. Four of the recruited participants were excluded from the analysis due to the failure of calibration or experimenter error.

Stimuli

Visual stimuli: The visual stimuli were photographs of real animals which were chosen from the CD-ROM Graphic Interchange Format Data (Hemera, 2000) and edited using Adobe Photoshop CS software. For each of the chosen animal labels, three versions of the corresponding static computer images were chosen, so that the whole sample consisted of 72 (24x3) images in total. Most of the chosen images were presented in the profile view, to facilitate ease of recognition. The exceptions were the images of the snake, scorpion, spider and butterfly which we presented in the alternative, top view perspective which was assumed to be clearer and easier for participants to recognize. All the presented images were of the same size (400 x 400 pixels) and all of them had ten percent grey background to avoid brightness on the screen.

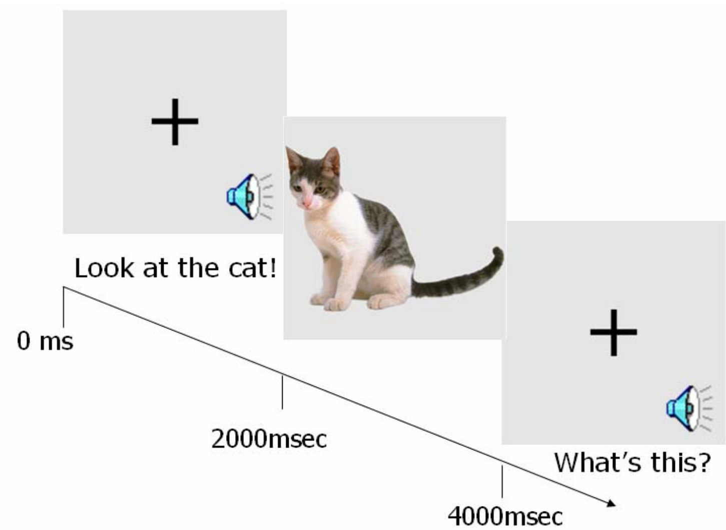
The original stimuli were 2-D jpeg images, which were later converted to AVI files that is, short films for each of the experiments, using Flash software. This was done in order to avoid short intervals of dark gaps which tend to appear just before picture presentation due to the time necessary for uploading pictures. These sudden shifts between bright and dark contrast might cause changes in pupil size, and consequently introduce error and less precise data from the eye-tracker. Visual and audio stimuli in this study were presented using the Preferential Looking Program, 'Look', developed at the Oxford BabyLab, implemented in Visual Basic by Woodford, Bellis, Baker, Plunkett and Schafer (2000).

Auditory stimuli: The twenty-six sentences were digitally recorded in stereo on the same session at 44.11 kHz sampling rate into signed 16-bit files. Audio stimuli were edited to remove background noise, head and tail clicks and to match for peak-to-peak amplitude by using the GoldWave 5.10 software. For the naming condition, twenty four animal labels were selected (bear, bee, bird, bison, butterfly, cat, cow, deer, dog, fish, frog, giraffe, goat, hamster, hedgehog, mouse, rabbit, rhinoceros, scorpion, sheep, snake, spider, turtle and zebra) and recorded within the carrier phrase: 'Look at the <target>'. For the other two, non-naming conditions the following sentences were recorded: 'Look at the picture' and 'What's this?', respectively.

Experimental design

The experiment consisted of three experimental conditions and 72 trials in total, that is, 24 trials per condition. The first experimental condition was the "naming condition". A typical trial in the naming condition involved presentation of the fixation cross for 2000 ms, during which a sentence containing the name of the animal would be uttered (i.e. 'Look at the <cat>').

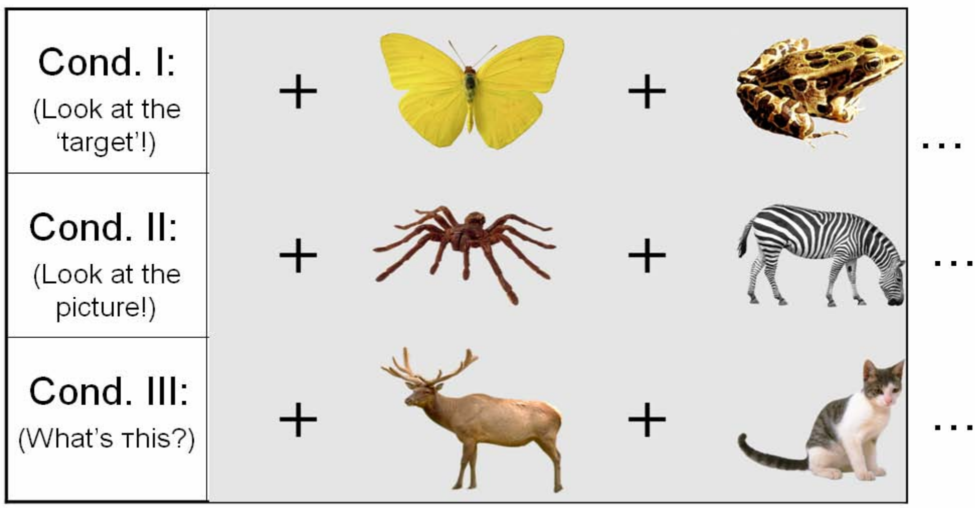
Figure 1. The time sequence of the auditory/visual stimuli presentation



The offset of the linguistic phrase was right-aligned with the offset of the fixation cross. The visual stimulus was presented at the offset of the auditory stimulus and remained on the screen for 2000 ms precisely. The next trial was presented immediately after the offset of the visual stimulus. (see Figure 1).

The other two non-naming condition had exactly the same timing, except that the uttered sentences were 'Look at the picture' and 'What's this?', respectively. Three different images of each animal category in two side profiles pointing either to the left or to the right were presented under the three auditory conditions, so that participants would not see any of the pictures more than once. The presentation of the auditory conditions, visual conditions as well as the profile views was counterbalanced across participants using a Latin Square design and the presentation order was randomized for each subject (see Figure 2).

Figure 2. The three experimental conditions



Procedure

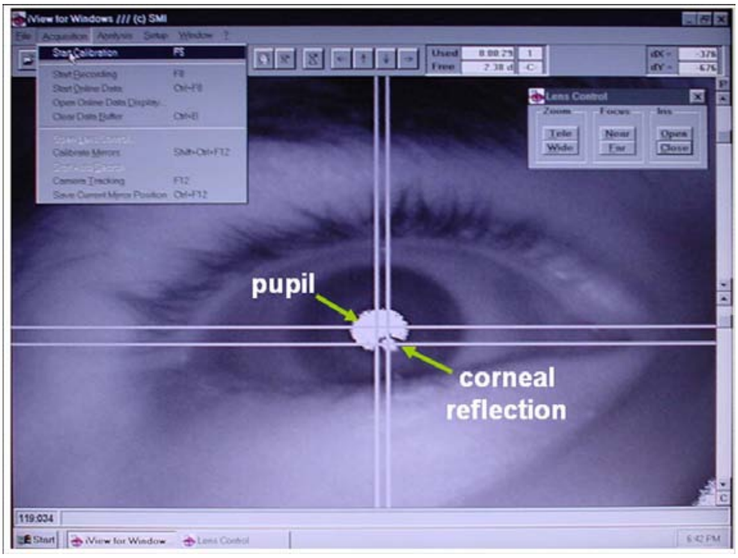
Participants were seated in a dark room approximately a meter away from a monitor displaying centrally presented visual stimuli (~6° of visual angle). They were given brief instructions at the beginning of the experiment explaining the task. The instructions were to focus on the fixation cross while it is displayed and look freely when the visual stimuli are presented on the screen, as well as to pay attention to the auditory stimuli presented through loudspeakers. Gazing at the fixation cross in the centre of the screen meant that all participants had the same starting point for all objects presented to them. Furthermore, the participants were also asked to try to find the most comfortable position before the start of the experiment and to sit as still as possible during the experiment given that the accuracy of the recorded data depended on their minimizing head-movements. At the beginning of the experiment they were asked to sign a consent form. At the end of the experiment, they were given a full explanation of the experimental procedure as well as a demonstration of the eye-fixations they made during the study.

Once the participant had settled down, the experiment started. The experimental room remained dark and silent throughout the experiment. The participant sat centrally in front of a screen on a chair with an adjustable chin-rest attached to it to help prevent head-movements. The monitor displayed a single (centrally presented) picture in each trial. The experimental area consisted of a rectangular booth measuring 1.55 x 2.45 m, made up of grey wooden panels. The entrance to the booth was covered with a suspended black curtain during testing. On the front panel, there was a wide screen of 1.30 x 0.39 m facing the participant. Two loudspeakers presenting the auditory message were mounted in the front panel, centrally above the screen. All of the images were shown at the participants' eye-level, at a distance of approximately 100cm. The eye-tracking camera permitted recording of participants' eye-paths for each image. The experimenter remained out of the participants' sight during the task and managed the experimental computer in a separate control room. After finding and adjusting a clear image of the participant's eye on the eye-tracking computer and setting the two threshold values for the pupil and corneal reflection, the system was calibrated using nine calibration points on the screen (centre, top left, top right, bottom left, bottom right, mid left, mid top, mid right and mid bottom). During the calibration process, the system 'learns' the relationship between the subject's eye-movement and gaze position. The measuring system has to be calibrated before every experimental session. After calibrating the system, gaze position could be measured within the boundary of the calibration area. For the majority of the participants the calibration procedure was fairly simple, but it could take longer for those wearing glasses or contact lenses. Typically, the picture of the eye would be slightly blurred if participants wore contact lenses or the eye-image was increased to avoid the dark frame for participants wearing glasses. Once the calibration was successfully completed, recording would start, immediately followed by experiment onset.

Eye-tracking methodology

An iView 3 RED II remote eye-tracking device from SensoMotoric Instruments GmbH (SMI) was used for tracking the participants' eye-movements. The fixed camera with the source of infra-red light was positioned below the monitor displaying the visual stimuli. The fixed camera with the source of infra-red light was directed onto the participants' pupil, so that the reflection of the infra-red light on the eye was captured by the camera. The pupil itself had to be the darkest object in the camera picture, so that it could be easily selected using a threshold value. At the same time a small reflection of light was visible on the cornea of the eye, as the brightest spot in the video set selected by a second threshold value, see Figure 3.

Figure 3. Setting calibration parameters: pupil and corneal reflexion thresholds



The displacement of the corneal reflection is mathematically related to the eye movement. Horizontal and vertical gaze position (x and y coordinates on the screen) were sampled every 20 ms with the accuracy of 0.5-1 degrees visual angle and the minimum duration for the fixation was defined to be 100 ms. The technique is non-invasive, but the system is very sensitive to the participants' head-movements. The data were always recorded from participants' left eye.

Measurements

In order to assess participants' looking behaviour across the three different experimental conditions, the naming and two non-naming conditions, a number of different measurements were used:

1. A *first look* measurement was used to asses how much time from the presentation of the picture participants would take to initiate their eye-gaze towards any part of the picture. In other words, the first look was an index of the amount of time from the offset of the fixation cross (when participants were expected to be focusing at the centre of the picture), until the start of the fist saccade when the object was presented. In more complex visual displays the minimal initiation of the

first look takes around 200 ms (Altmann & Kamide, 2004; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Huettig & Altmann, 2005).

2. A *longest look* measurement was calculated as a single longest fixation made during the interval of the picture presentation. The minimal fixation duration was defined as 100 ms. It is generally excepted view that the processing of the retinal image takes place mainly between the saccades, during *fixations* that last for about 100-600 ms (Barber & Legge 1976).

3. *Total looking time* (TLT) was calculated by summing all the fixations that participants made during the picture presentation. The rational for using TLT measure was similar to the previous one. Given that it is generally assumed that the processing of the pictures is suppressed if not even inhibited during eye movements (saccades), the total amount of time participants spend looking at the objects was another important measure to take into account.

4. The total *number of fixations* that the participants made during the presentation of each of the pictures was also calculated. Given that the fixation duration could vary substantially and given the limited amount of time participants were given to process visual objects in this paradigm it was important to take into account how many fixations they made.

5. In addition to these measurements, a spatial analysis of the *positions of the fixations* was also performed based on the x and y pixel coordinates on the screen. This type of analysis was useful for understanding which areas of animate objects tended to attract participants' attention. At the start of each trial, while focusing on the fixation cross, participants' eye-gaze is expected to fall at around 512x384 pixels. When the pictures are presented on the screen, the fixations are expected to fall onto a specific set of animal features. All of the fixations participants made while looking at the animate objects were firstly extracted for each of the participants for each single picture. At the later stage, clusters of fixations were plotted on top of the pictures.

RESULTS

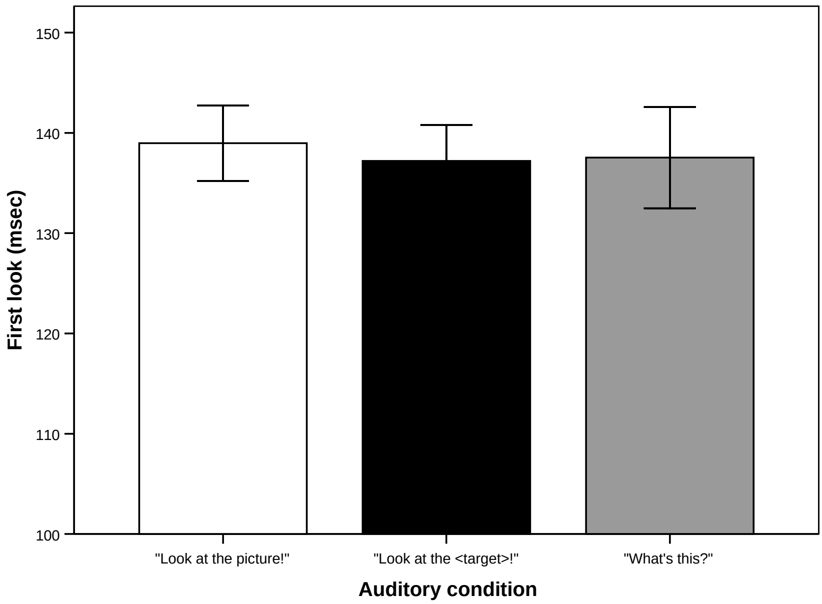
Separate analyses were performed for each of the previously described temporal and spatial aspects of performance across the labelling and no-labelling conditions.

Analysis of the first look

In order to asses the amount of time participants took to initiate their first look from the centre of the screen towards any part of the picture across the experimental conditions a 3x2 way ANOVA with factors Auditory Condition ('Look at the

<target>', 'Look at the picture' and 'What's this?') and Profile (left and right picture orientation) was run.

Figure 4. Average initiation time of the first look across the conditions



No significant main effects of Auditory Condition (F(2,414)=0.58, p=0.944), or picture Profile (F(1,414)=1.5, p=0.221) were found and no interaction either. Regarding of the auditory condition, the average amount of time participants took to initiate the first gaze upon hearing the name of the visual stimuli was M=137.20ms (s.e.m.= 3.57), whereas in the 'Look at the picture' condition M=138.97ms (s.e.m.=3.76) and in the 'What's this?' condition it was M=137.53ms (s.e.m.=5.05). See Figure4[4]. Thus, participants were very quick to initiate their first looks in all three conditions, but there was no systematic difference between participants across the labelling and no-labelling conditions.

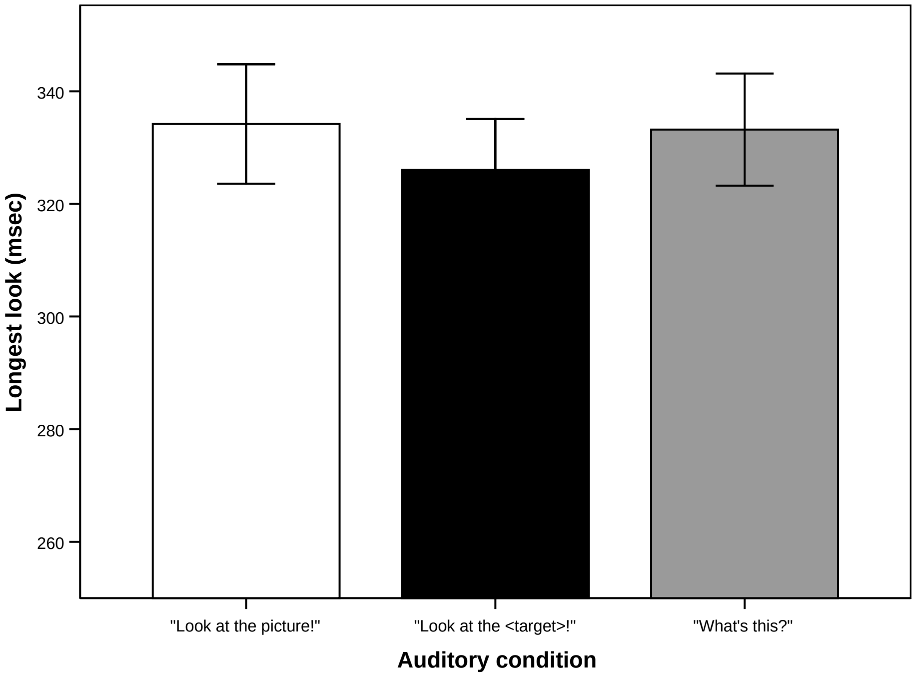
Analysis of the longest look

A 3x2 ANOVA with factors Auditory Condition and picture Profile revealed a similar pattern of results for the longest look measure. There was no main effect of

[4] S.e.m. in all the graphs is ±1. All of the tests are two-tailed and the significance level for all the tests was 0.05.

Auditory condition (F(2,414)=0.252, p=0.777) and no main effect of picture Profile (F(1,414)=0.031, p=0.861). There was no significant interaction effect either.

Figure 5. Average longest look across the three experimental conditions



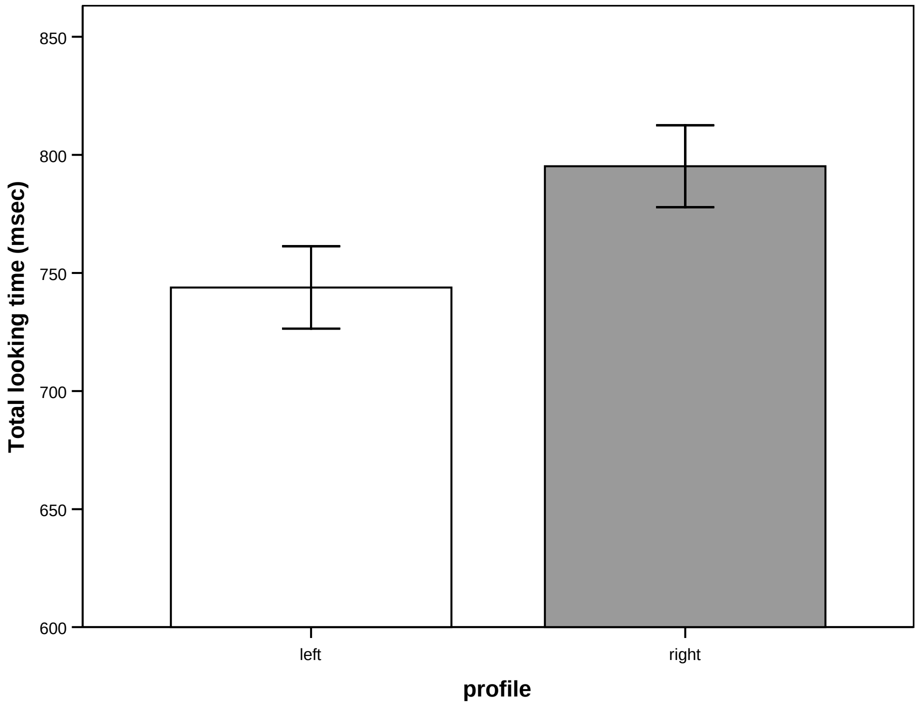
The single longest look in the naming (‘Look at the <target>’) condition lasted M=326.03 ms (s.e.m.= 9.03) and did not differ significantly from the other two conditions. The average duration of the longest look in the ‘Look at the picture’ and ‘What’s this?’ conditions was M=334.19 ms (s.e.m.=10.60) and M=333.25 ms (s.e.m.=9.95), respectively (see Figure5).

Analysis of total looking time (TLT)

A 3x2 ANOVA with factors Auditory condition and picture Profile revealed no main effect of the Auditory Condition (F(2, 414)=0.005), p=0.995), but there was a significant main effect of picture Profile (F(1,414)=4.28, p<.039). The left oriented pictures received less of the total looking time (M=743,84ms, s.e.m.=17.46) in comparison to the pictures presented in the right profile (M=795.16ms, s.e.m.=17.35), see Figure6. The pictures presented in the ‘Look at the <target>!’

condition received M=769.88ms (s.e.m.=21.06) of total looking time and did not differ significantly from the other two conditions. The TLT in the 'Look at the picture!' condition was M=764.06ms (s.e.m.=22.13) and in the 'What's this?' condition it was M=770.83ms (s.e.m.=21.25). The interaction effect was found not to be significant.

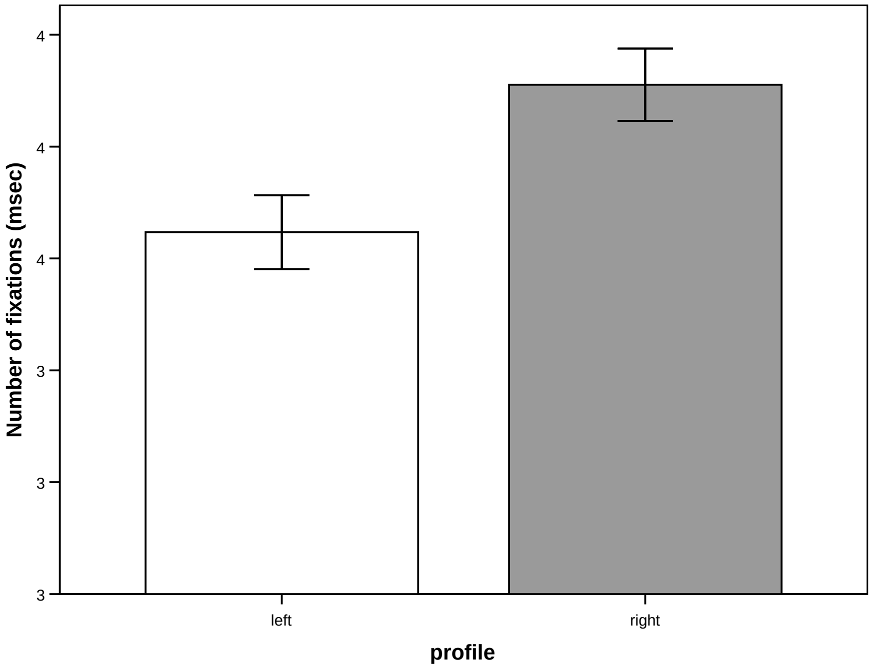
Figure 6. Average total looking time for the left and right picture profiles



Analysis of number of fixations

A 3x2 ANOVA regarding the average number of fixations revealed no significant effect of the Auditory condition (F(2,414)=0.054, p=0.945), but a significant effect of picture Profile (F(1,414)=8.073, p<0.005). There was no significant interaction. Planned comparisons revealed that the left oriented pictures received fewer fixations M=3.65 (s.e.m.= 0.066) than the pictures presented in the right profile M=3.91 (s.e.m.=0.064) (see Figure7). The average number of fixations participants made while looking at the pictures presented in the 'Look at the <target>!' condition was M=3.77 (s.e.m.=0.081) and in the <Look at the picture!> and <'What's this?'> conditions it was M=3.75 (s.e.m.=0.079) and M=3.80 (s.m.e=0.081), respectively.

Figure 7. Number of fixations for the left and right picture profiles



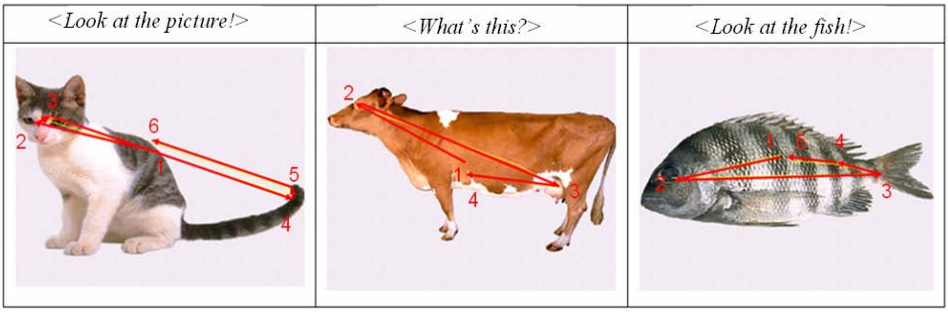
Cluster analysis

The participants' scan-paths and eye-fixations were examined across the three experimental conditions.

The visual examination of the participants' eye-gazes revealed surprisingly consistent patterns of looking across participants. Figure8 demonstrates participants' scan paths. Participants typically started the examination of the pictures by looking at the eyes/head of the animals, before focusing on other parts, such as a cat's tail or a cow's udder. The order of eye-movements was marked for each of the three examples presented in the Figure 8.

In order to analyze the patterns of eye-movements and to compare them across the three experimental conditions, a cluster analysis was performed. The fixations for each of the 24 animals across all of experimental conditions (3 auditory conditions x 3 visual conditions x 2 side profiles) were extracted for each of the participants. Fixations were clustered using Ward's method (Ward, 1963), which uses an analysis of variance approach to evaluate the distances between clusters. In short, this method attempts to minimize the Sum of Squares of any two (hypothetical) clusters that can be formed at each step. This method is regarded as very efficient; however, it tends to create clusters of small size.

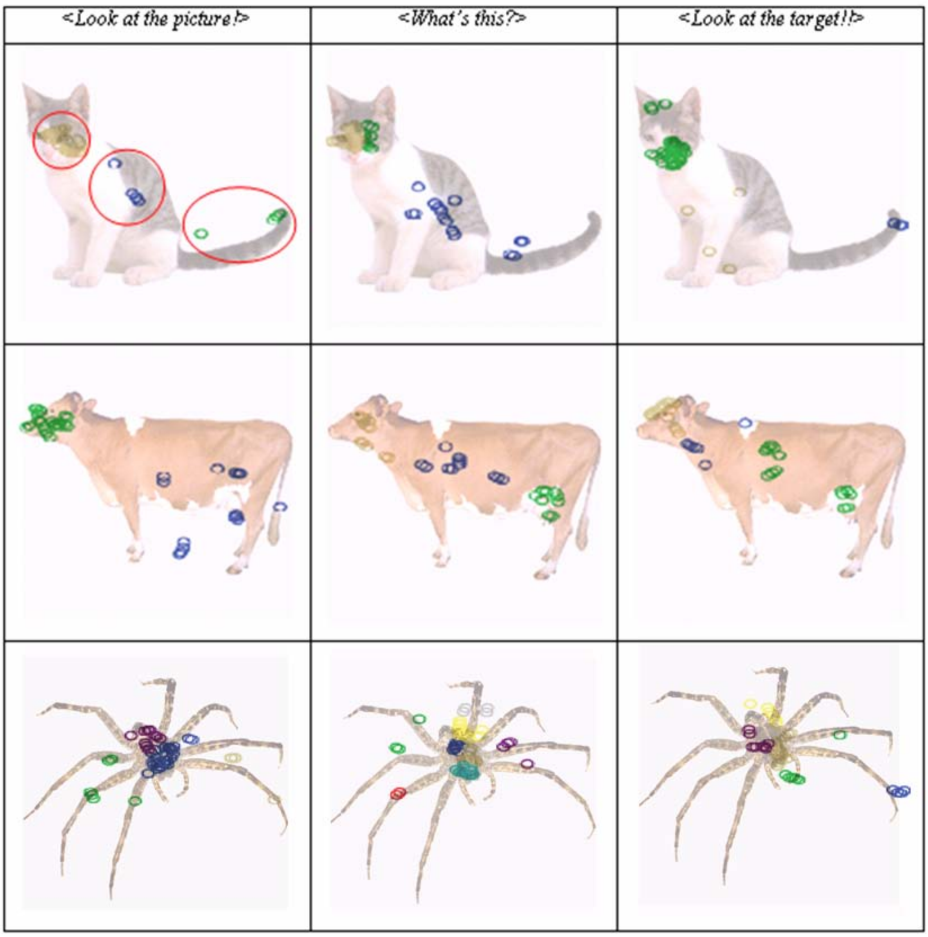
Figure 8. Typical eye-gazes at pictures of a cat, cow and fish



The Clastan software (Wishart, 2004) was used to identify clusters of fixations for each animal under each of 18 conditions. The cluster identification in this software was equivalent to k-means clustering and was based on Ward's method, except that it does not require initial specification of the number of clusters to be created. It actually provides all possible solutions for cutting the cluster tree (that is from minimal to maximal number of clusters) and gives significance levels for each of the models visualising the perfect cut on the clustering tree. For purposes of the current study, the model with a maximum number of significant clusters of fixations was consistently chosen for each of the individual objects. After this part of the analyses was completed, the clusters of fixations were plotted on top of images and each cluster of fixations was presented in a different colour (see Figure 9).

The cluster analysis for a picture of the cat revealed three clusters of fixations in each of the three auditory conditions: $F(2,37)= 341.002$, $p<.001$ in the "Look at the picture!", $F(2,45)= 745.552$, $p<.001$ in the "What's this?" and $F(2,45)= 89.476$, $p<.001$ in the "Look at the cat!" condition. In the 'Look at the picture condition' the first cluster of fixation presented in brown was around the head, the second cluster (fixations in blue) was around the central part of the body and the third cluster (fixations in green) around the tail. A similar distribution of clusters and fixations was found for the "Look at the cat!" condition, whereas in the "What's this?" condition there were two clusters around the head region (fixations in brown and green) and only one cluster grouping all of the other fixations around the central part of the body and tail (fixations in blue). For a picture of the cow in the "Look at the picture!" condition, all of the fixations clustered in two groups ($F(1,48)= 314.509$, $p<.001$), with one cluster around the head and another one around the central part of the body, tail and udder. In the other two conditions, cluster analysis demonstrated three clusters of fixations ($F(2,34)= 540.47$, $p<.001$ in the "What's this?" and $F(2,29)= 295.695$, $p<.001$ in the "Look at the cow!" conditions, respectively). In the "What's this?" condition fixations grouped around the head (fixations in brown), the central part of the body (fixations in blue) and the udder (fixations in green), respectively, whereas in the "Look at the cow!" condition there were two clusters of fixations (in blue and brown) around the head and one broader cluster around the central part of the body and the udder (fixations in green).

Figure 9. Plotting clusters of fixations on top of images



The cluster analysis for a picture of the cat revealed three clusters of fixations in each of the three auditory conditions: $F(2,37)= 341.002$, $p<.001$ in the "Look at the picture!", $F(2,45)= 745.552$, $p<.001$ in the "What's this?" and $F(2,45)= 89.476$, $p<.001$ in the "Look at the cat!" condition. In the 'Look at the picture condition' the first cluster of fixation presented in brown was around the head, the second cluster (fixations in blue) was around the central part of the body and the third cluster (fixations in green) around the tail. A similar distribution of clusters and fixations was found for the "Look at the cat!" condition, whereas in the "What's this?" condition there were two clusters around the head region (fixations in brown and green) and only one cluster grouping all of the other fixations around the central part of the body and tail (fixations in blue). For a picture of the cow in the “Look at the picture!” condition, all of the fixations clustered in two groups ($F(1,48)= 314.509$, $p<.001$), with one cluster around the head and another one around the central part of the body, tail and udder. In the other two conditions, cluster analysis demonstrated three clusters of fixations ($F(2,34)= 540.47$, $p<.001$ in the “What’s this?” and $F(2,29)= 295.695$, $p<.001$ in the “Look at the cow!” conditions, respectively). In the “What’s this?” condition fixations grouped around the head (fixations in brown), the central part of the body (fixations in blue) and the udder (fixations in green), respectively, whereas in the “Look at the cow!” condition there were two clusters of fixations (in blue and brown) around the head and one broader cluster around the central part of the body and the udder (fixations in green).

Generally, fixations for the vast majority of the animals typically clustered in three groups (clusters), occasionally in two or four and very rarely in more than four. An example of such an animal was a spider. All fixations in the naming conditions clustered in four clusters for a picture of the spider (“Look at the picture!”: $F(3,43)= 88.49$, $p<.001$); in the “What’s this?” condition fixations clustered in five clusters: $F(4,50)= 212.85$, $p<.001$ and in eight clusters in the “Look at the spider!”: $F(7,59)= 110.198$, $p<.001$, respectively).

Notice that the F-tests reported here can be used only for descriptive purposes. Since clusters of fixations are chosen to maximize the differences among fixations in different clusters and that the maximum number of significantly different clusters was consistently chosen for each of the animals, observed significance levels are not corrected for this and thus cannot be interpreted as a test of the hypothesis that the cluster means are equal. Therefore, further quantification of the spatial distribution across the naming and non-naming conditions was calculated by extracting fixations’ mean distance from their cluster centroid. A one-way ANOVA demonstrated no difference between “Look at the picture!”, “Look at the <target>!” and “What’s this?” conditions regarding fixations’ distance from cluster centroid ($F(1,404)=0.101$, $p=0.904$; M(“Look at the picture!”)=21.88, s.e.m.=0.395, M(“Look at the <target>!”)=21.07, s.e.m.=0.405, M(“What’s this?”)=21.83, s.e.m.=0.391).

Typically, one would want to further examine how much time participants tend to spend on certain parts of the images, or how many fixations they made within the cluster region. But, very often fixations were found to be in between the two regions of interest or there were two clusters within the (what looks like) same region of interest (see Figure9 – picture of a cat in ‘What’s this?’ condition, head region or picture of a cow in ‘Look at the cow!’ condition). Thus, further quantification of the spatial distribution of the fixation clusters was very hard due to the difficulty of making a clear cut-off between areas of interest.

Another interesting finding was that the first cluster of fixations, irrespective of the auditory conditions in which the pictures were presented, was typically found to be around eyes (head) of the animal. The second cluster of fixations grouped around the tail in the cat, the udder in cow, shape of the body in snakes and so on (see

Figure9). Furthermore, there was usually one cluster of fixations around the central part of the body, and fixations in this region usually occurred at the beginning of the trial, before initiation of the first look, or at the end of the trial, when participants were moving to the position of the fixation cross, preparing for the next trial. In order to further quantify seemingly reliable order effects of participants' fixations, the order of fixations was correlated with cluster membership. If participants were consistently attending to the animals' head first, followed by fixation to a certain area of interest which varied from picture to picture, one would expect a significant correlation between order of participants' looks and cluster membership of these looks. Indeed, Spearman's rank ordering correlation revealed significant overall correlation between order of participants' looks and cluster membership of these looks ($r=.107$, $p<.005$). This correlation was significant in the naming condition ("Look at the <target>!": $r=.160$, $p<.005$) and in one of the non-naming conditions ("Look at the picture!": $r=.209$, $p<.005$), but not in the other ("What's this?": $r=.002$, $p>.05$).

DISCUSSION

Using the initiation and programming of the first look as an index of visual object processing it was found that the participants were taking approximately the same amount of time to initiate their eye-gaze irrespective of the auditory condition under which the pictures of the animals were presented to them. Furthermore, the single longest look towards a picture, total looking time (that is, summation of all the fixations) and number of fixations participants made while looking at the pictures of animals revealed no systematic difference across the naming ('Look at the <target>!') and two non-naming conditions ('Look at the picture!' and 'What's this?', respectively). These results do not support the first hypothesis, according to which it was expected that labelling an object prior to visual presentation would lead to different looking behaviour and in particular, faster initiation of the first look towards pictures of the animate objects when the objects were named. A possible explanation for this failure may have to do with the lack of inter-stimulus-interval (ISI) between the offset of the auditory stimuli and the onset of the visual stimuli, that is to say lack of time for evoking a mental image prior to visual presentation of the images.

The analysis of total looking time revealed an interesting difference between processing of the left vs. right oriented pictures, whereby the pictures presented in a left profile-view received less of total looking time than did pictures presented in the right-profile view. Consistent with this result was a finding that left-oriented pictures received fewer fixations in comparison to right-oriented pictures. These results may be explained by a left-right visual processing advantage or participants' handedness, given that all of the tested subjects were right-handed. However, these interpretations would require further experimental exploration, which was not the main purpose of the present research. Also, notice that total looking time and number of fixations were not inversely correlated, as one would expect, but showed similar effects. The possible explanation for this finding is that the participants were making longer saccades between fixations in this task.

Examination of spatial location of participants' fixations showed that the participants were very consistent in the way they looked at the pictures of animate objects. The eyes or head of the animals were a typical region attracting a number of fixations, but also some other parts of animals (e.g. tail in a cat, or udder in a cow, body in a snake). Also, for most pictures of animate objects there was a cluster of fixations around the center of the screen. Given that participants were asked to focus on the fixation cross prior to presentation of the objects, these fixations mainly occurred at the start of the trial that is, prior to the initiation of the first look. Also, fixations in this region occurred at the end of the trial, when participants were possibly preparing for the start of the next trial, given the brief presentation of pictures (see Figure8).

Furthermore, not only did participants tend to look at similar regions of the pictures, but also the looking order to these regions was consistent across participants. However, contrary to the original predictions, these patterns of fixations were similar across the naming and non-naming conditions ('Look at the <target>!', 'Look at the picture!' and 'What's this?', respectively).

Besides an insufficient ISI and possible time-shortage for evoking a mental representation in the current experimental design, another explanation for the similarity of the eye-movement behaviour across the naming and non-naming conditions may have to do with a high consistency of visual structure for the majority of objects, and in particular the presence of a head which was attended first and which attracted many fixations in all of the animate pictures. It could be that less structurally similar objects would reveal different patterns of eye movements across the variety of objects and across the naming and non-naming conditions. In order to test this hypothesis, the following study involved a series of inanimate objects and contrasted eye-movements for pictures of inanimate objects that were either named or not named prior to visual presentation.

REFERENCES

Althoff, R. R, & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 997–1010.

Altmann, G. T. M, (1999). Thematic role assignment in context. *Journal of Memory and Language, 41*, 124-145.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm', *Cognition, 93*, 79–87.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518.

Barber, P. J., & Legge, D. (1976). *Perception and Information*. London: Methuen.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*, 507-534.

Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 188-202.

Ferreira, F., & Clifton J. C. (1986). The independence of syntactic processing, *Journal of Memory and Language, 25*, 348–368.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 1347–1365.

Grant, E., & Spivey, M. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*(5), 462-466.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274-279.

Hegarty, M. & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717-742.

Henderson, J. M. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 210–228.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition, 96*, 23-32.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.

Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye-movements. *Journal of Memory and Language, 49*, 133-159.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 565–572.

Mak, W. M., Vonk, W., & Schriefers, H. J. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*(1), 50-68.

Meyer, A. S., Sleiderink A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), 25-33.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-25.

Posner, M., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. L. Pick & I. J. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137-157). Hillsdale, N.J.: Erlbaum.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Tanenhaus, M. K., Spivey, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic infor-mation in spoken language comprehension. *Science, 268,* 1632-1634.

Ward, J. H. (1963). Hierachical grouping to optimize an objective function. *Journal of the American Statistical Association, 58,* 236-244.

Wishart, D. (2004). *ClustanGraphics Primer: A guide to Cluster Analysis, third edition.* Edinburgh: Clustan Ltd.

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

ABSTRAKT

VIZUELNO PROCESIRANJE ŽIVIH OBJEKATA – STUDIJA OČNIH POKRETA

Vanja Ković
Odsek za psihologiju, Univerzitet u Novom Sadu

Kim Plunkett
Odsek za ekeprimentalnu psihologiju, Univerzitet u Oksfordu

Gert Westermann
Odsek za psihologiju, Oksford Bruks Univerzitet

U ovoj studiji ispitanicima su prikazivane imenovane ili neimenovane slike živih objekata, a proučavane su šeme očnih pokreta tokom razgledanja slika. Rezultati su pokazali da ispitanici na iznenađujuće sličan nacin razgledaju slike živih objekata. Glava ili oči živih objekata su po pravilu djelovi slike na koje ispitanici prvo obrate pažnju, o čemu svjedoče brojne fiksacije u tom regionu, a potom razgledaju i druge djelove slike (poput repa kod mačke, vimena na slici krave, tijelo zmije). Takođe, ne samo da su ispitanici imali tendenciju da razledaju iste djelove prikazanih im slika, već je i redosled posmatranja djelova koji su im privukli pažnju bio konzistentan. Međutim, suprotno početnim pretpostavkama, šeme očnih pokreta bile su veoma slične, nezavisno od toga da li je objekat imenovan ili ne, tj. nezavisno od tri eksperimentalna uslova (‘Pogledaj <imenovanu sliku>!’, ‘Pogledaj sliku!’ i ‘Šta je ovo?’), što navodi na zaključak da konzistentnost u načinu procesiranja vizuelnih objekata u ovoj studiji zapravo ne odražava sadržaj mentalnih reprezentacija (pobuđenih u uslovu imenovanja), već je posledica strukturalne sličnosti živih objekata, prije svega lica i očiju životinja koje prve privuku pogled ispitanika.

Ključne reči: *živi objekti, praćenje očnih pokreta, mentalne reprezentacije*

RAD PRIMLJEN: 11.05.2009.