Does Cognitive Load Affect Eye Movements/Oculomotor Behavior in Natural Scenes?

Kerri Walter (✉ walter.ker@northeastern.edu)

Northeastern University

Peter Bex

Northeastern University

Research Article

Keywords: Cognitive, Oculomotor, Natural Scenes

Posted Date: February 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-199698/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Cognitive neuroscience researchers have identified relationships between cognitive load and eye movement behavior that are consistent with oculomotor biomarkers for neurological disorders. We develop an adaptive visual search paradigm that manipulates task difficulty and examine the effect of cognitive load on oculomotor behavior in healthy young adults. Participants (N=30) free-viewed a sequence of 100 natural scenes for 10 seconds each, while their eye movements were recorded. After each image, participants completed a 4 alternative forced choice task in which they selected a target object from the previously viewed scene, among 3 distracters of the same object type but from alternate scenes. Following two correct responses, the target object was selected from an image increasingly farther back (N-back) in the image stream; following an incorrect response, N decreased by 1. N-back thus quantifies and individualizes cognitive load. The results show that response latencies increased as N-back increased, and pupil diameter increased with N-back, before decreasing at very high N-back. These findings are consistent with previous studies and confirm that this paradigm was successful in actively engaging working memory, and successfully adapts task difficulty to individual subject's skill levels. We hypothesized that oculomotor behavior would covary with cognitive load. However, there were no significant differences between the number or duration of fixations and saccades for high/low performing subjects, or between high/low performing trials for a given subject. Similarly, oculomotor behavior did not act as a predictor of correct/incorrect responses with increasing demand from the N-back task. Similarly, the proportion of each scene viewed was not related to N-back and was not a significant predictor of accuracy. These results suggest that cognitive load can be tracked with an adaptive visual search task, but that oculomotor strategies generally do not change as a result of greater cognitive demand in healthy adults.

Introduction

Where do we look?

The human visual system only allows for high-resolution visual information to be encoded from the fovea (the central $\sim 2^\circ$ of vision). As a result, to estimate the contents of a scene, we move our eyes rapidly around a scene (*saccades*) in order to focus our central vision on multiple discrete areas (*fixations*) (for review, see ^{1,2}).

Human vision is reliant on eye movements, however there is still relative debate about what determines where observers will look when told to view a scene. It is well documented that subjects adopt different viewing strategies when performing different tasks, as the way someone looks around a scene is dependent on the current task they are trying to accomplish ^{3,4}. However, it is still unclear how subjects decide where to look when they are given no task or instruction, otherwise known as “free-view”. During free-view, fixation locations may vary significantly from subject to subject ⁵. Because the ways different individuals view a scene are idiosyncratic, it is unclear what exactly guides eye movements during free-view.

Two main approaches have attempted to explain what guides eye movements during free-view: salience and meaning. Evidence suggests that fixation locations may be driven by areas of higher salience⁶⁻⁹, while opposing evidence suggests fixation locations are driven by areas of higher semantic meaning¹⁰⁻¹⁶. The salience approach is based on bottom-up processes, stating that fixations are guided by image features that contrast their surroundings, while the meaning approach is based on top-down processes, stating that fixations are guided by prior experience. Additionally, some evidence suggests that fixation durations may be guided by peripheral content and image features^{17,18}. Yarbus' original study demonstrates that participants will have different scan-paths for the same image, even while performing the same task, suggesting that low level information is not sufficient to predict human gaze⁴. Recently, deep learning models of gaze-guidance have trained convolution neural networks on the gaze patterns of human subjects (REFs), and have demonstrated greater performance than salience or meaning models alone¹⁹. These approaches therefore indirectly incorporate both feed-forward scene statistics with the use of high-level image meaning that guided the fixations of observers who supplied the training eye movements.

Eye movements have proven to be useful diagnostic tools and biomarkers for cognitive functioning. For example, children with reading difficulties exhibit atypical oculomotor behaviors while reading²⁰, and children with autism spectrum disorder exhibit subtle atypical oculomotor behaviors when processing language and social information^{21,22}, as well as exhibiting a center bias on images, demonstrating reduced saliency for social-gaze related locations, and prioritizing saliency for pixel-specific locations rather than saliency for overall semantic knowledge²². Eye movements can also serve as screening methods for degenerative diseases, such as Alzheimer's, as saccades and smooth pursuit become slowed and less accurate, and viewing strategies become erratic and seemingly random²³.

Cognitive load:

Cognitive load refers to the amount of active effort being invoked by working memory²⁴. N-back tasks have been widely utilized to measure working memory function, therefore cognitive load can be manipulated with the use of an N-back task. An N-back task presents participants with visual or auditory information and asks the participant to remember that information a specified number (N) of trials later²⁵. Generally, as N-back increases, response latencies increase and response accuracies decrease²⁶⁻²⁸. Increasing the demands of an N-back task has also been shown to activate various areas of the brain associated with working memory²⁶⁻³⁰.

Some studies have demonstrated that eyetracking technology can be used to measure cognitive load, with features such as pupilometry: pupil diameter has been shown to increase in response to increasing levels of cognitive load³¹⁻³⁴. Different aspects of oculomotor properties (fixation number and duration; saccade length, angle, and velocity; pupil dilation; blink rate and velocity) have been linked to cognitive load³⁵⁻³⁷, and combinations of these features have been proposed as a model for measuring cognitive load³⁶.

Cognitive load vs Perceptual load:

Top-down processing can be affected by an increase in cognitive load, but not perceptual load³⁸. Perceptual load refers to the amount of visual information being presented, and is related to the levels of clutter, distractors, or edges within a scene. Perceptual load is therefore distinct from cognitive load, which refers to the amount of information being processed in the brain, and is related working memory³⁸. Belke et al. (2008) demonstrated that tasks which required semantic knowledge, such as matching a written word with its line drawing, were influenced by the presence of a competitor object, (an object similar in semantic meaning), when assigned an additional working memory task (increased cognitive load), but were not influenced when the number of objects on screen increased (increased perceptual load).

To summarize: if visual search strategies are guided by top-down processing, and increasing cognitive load disrupts top-down processes, then increasing working-memory demands (which we will do with an N-back task) should alter a participant's visual search strategy. We are interested in seeing if this increased cognitive demand affects a subject's oculomotor strategies, such as the number and duration of fixations and saccades. Similarly, do subjects who excel with this task, (subjects who can hold a higher number of scenes in working memory, or subjects with higher cognitive load capacities), utilize different oculomotor strategies than subjects who struggle with this task (subjects with low cognitive load capacities)? Do certain oculomotor strategies predict accuracy on this task? We modify an N-back paradigm for visual search in natural scenes and implement an adaptive procedure to maintain constant cognitive load, given large individual differences in visual search performance. The task allows for the analysis of oculomotor behaviors under varying levels of cognitive load.

Similar studies demonstrate a close relationship between attention, cognitive function, and the deployment of eye movements. We therefore hypothesize that changes in attention demand and cognitive load should lead to reliable changes in oculomotor behavior. We also hypothesize that individual differences in performance on a demanding cognitive task should be associated with differences in patterns of oculomotor behavior. In this study, we manipulate cognitive load in a healthy population of young adults and measure eye movement behavior as they perform a demanding visual search task in natural scenes. In this study, we examine if oculomotor behavior, regardless of scene context, can explain some of the differences between how individual subjects view a scene. In a companion paper (Walter et al, 2021), we examine how semantic information in natural images affects oculomotor behavior. We propose an Adaptive N-back task that allows for the comparison of oculomotor behaviors under varying levels of cognitive load.

Methods

Apparatus

Stimuli were presented on a 60cm x 34cm BenQ XL2720Z LCD monitor (BenQ Corporation, Taipei, Taiwan) set to a screen resolution of 1,920 × 1,080 pixels at 120 Hz and run using a Dell Optiplex 9020 desktop computer (Dell Inc. Round Rock, TX) with a Quadro K420 graphics card. The experiment was programmed and run using MATLAB (The MathWorks, Inc., Natick, MA) and the Psychophysics Toolbox Version 3³⁹. Observers were seated 63 cm from the monitor with head stabilization secured via chinrest. Eye movements were recorded using an SR Research Eyelink 1000 (SR Research Ltd. Mississauga, Ontario, Canada) and the MATLAB Eyelink Toolbox⁴⁰. The sampling rate was set to 1,000 Hz (note that sampling rate was set to 250 for one subject due to experimenter error, however this did not impede data collection or analysis).

Participants

In total, 33 naïve subjects (7 male, 26 female) with self-reported normal or corrected vision from the Northeastern undergraduate population participated in this study. 3 subjects were excluded due to program crashes (N = 2) or Eyelink calibration issues (N = 1). Subjects were excluded as soon as issues arose, and data collection continued until 30 subjects with usable data were collected (7 male, 23 female). Subjects received course credit as compensation for their time. All subjects read and signed an informed consent form approved by the University Ethics Board before the experiment began, the experimental procedure was approved by the institutional review board at Northeastern University, and the experiment was performed in accordance with the tenets of the Declaration of Helsinki.

Images

In total, 100 images (50 indoor, 50 outdoor), were selected from the LabelMe database⁴¹. The database, comprised 75,353 total images at the time of selection, was filtered down as a result of the steps listed in Table 1. All images were landscape oriented and were in color.

Table 1

Steps taken to filter through the LabelMe database. List of filters applied, and number of images remaining after filtering, that lead to the unbiased selection of 100 experimental images.

Images Removed	Images Remaining
< 75% of image surface labeled	11,822
< 15 unique objects	2,186
< 1000 x 1000 pixel resolution	1,629
Portrait images	1,523
$\geq 25\%$ of unique objects are parts of larger objects	1,364
$\geq 50\%$ of image is taken up by a single object	1,311
< 15 unique objects (excluding broad scenery objects)	975

From these remaining 975 images there were 76 indoor and 899 outdoor scenes, from which we hand selected 50 indoor and 50 outdoor images. Images were manually removed based on criteria similar to above: we removed images with objects taking up a large portion of the frame, blurry images, images with few distinct objects, etc. We also avoided including images that were taken of the same setting at different angles, to ensure no identical objects were overlapping in the database. We sought to ensure that the image database used for this experiment was varied, but also that each image had enough common, unique objects to satisfy the decision task.

Procedure

Participants were shown a short schematic of the instructions (in the form of a PowerPoint presentation) before the experiment began. Subjects were asked if they understood the task before the start of the experiment. All subjects reported yes, and none reported that they struggled with the task due to misunderstanding the instructions. Participants were shown an image for 10 seconds and were instructed to view the scene freely. After 10 seconds, the image was removed and replaced with four small snapshots from different scenes, each centered on objects with the same label from the LabelMe database. One of these snapshots was from the image the participant had previously viewed, and the goal was to identify the corresponding object by clicking a mouse cursor on it. For example, a forced choice task could be of four different lamps, with one of the lamps from the target scene, and the other three from other scenes within the experiment, without replacement. Participants received immediate feedback on their answer. Whenever a subject answered two trials correctly in a row, they received a prompt that read “Now look for objects from the image (N) back”. N would change depending on subject’s performance. N started at zero, meaning the choice task was referring to the image immediately

preceding it. Every time a subject answered two trials in a row correctly, N was increased by one. If at any point a subject answered incorrectly, N was decreased by one (Fig. 1).

The experiment was composed of 100 trials across 4 blocks (25 trials per block). A standard Eyelink 9-point eye tracker calibration task was completed before the start of each block. Images were presented in random order for each participant. There was a mandatory break between blocks, and participants were instructed to tell the experimenter when they were ready to continue. Participants were told that they did not have to remember the previous image stream during a break, as N was reset to zero at the start of each new block.

All images were scaled to be approximately the same size (1,280 x 960 pixels) when presented in the experiment. Images were rescaled according to their largest dimension in order to maintain their original aspect ratio. The forced choice task was comprised of objects taken from the 100 images used in the dataset. For each trial, one object was randomly chosen from the list of labeled objects in the LabelMe file for each image. The full database was scanned for matches of that object label. If the object did not reoccur at least 3 times within the dataset, a different object was chosen. Three objects with the same label were chosen at random and used as distracters alongside the target object in the forced choice task. Only one object was sampled from each image at a time. Only objects larger than 100 x 100 pixels were used to prevent excessive magnification in the alternative choice display. Objects were taken from a rectangular section of the original image, with a surrounding 10% of the object's dimensions included. This was done to provide a small amount of image context for each object. In pilot studies, we grabbed only the object with no background context for the alternative choice display, however, this proved to be too difficult for subjects to complete. The objects were scaled to be approximately the same size as each other (maximum dimension of 300 pixels), while maintaining their original aspect ratios, but different from their size in the original scene.

Results

In total, 1% of trials were missing due to Eyelink error (30 out of 3000 total trials). There were high levels of individual differences in performance on this task: the highest maximum N-back reached was 10 (1 participant), and the lowest maximum N-back reached was 2 (1 participant). The median N-back reached was 5, and the mode was 4 (Fig. 2). This wide distribution of maximum N-back achieved demonstrates the variability of subjects on this task, while simultaneously exemplifying the notion that this adaptive task can be suited to a number of participants, regardless of overall ability on the cognitive load task.

There were both learning and fatigue effects throughout the experiment, providing evidence that our task was successful in increasing cognitive load. We compared the rate of learning across each block by performing individual t-tests on the b value of our fit equation ($y = a \cdot (x-1)^b$). We fit all 4 curves individually, found their average a value (0.5096), and set this as the constant a . By fitting all 4 blocks with an average constant, we were able to compare strictly the b value of each curve, or the rate of learning. Throughout each block there was a steady learning effect, and as the blocks continued, the rate

of learning generally increased (Fig. 3). Compared to block 1, the rate of learning was faster in block 2 ($t(29) = -6.824, p < .001$) and in block 4 ($t(29) = -4.276, p < .001$), but not in block 3 ($t(29) = -1.140, p = .1318$), demonstrating a possible fatigue effect that occurs just after the halfway point in the experiment. Learning is recovered in block 4, where the rate is significantly higher than in block 3 ($t(29) = -3.386, p = .001$). The rate of learning was highest in block 2, where it was significantly faster than block 1 ($t(29) = -6.824, p < .001$), block 3 ($t(29) = -6.082, p < .001$), and block 4 ($t(29) = -2.094, p = .023$).

Response Latency:

We used a Pearson's correlation to examine the relationship between N-back and response latency. Replicating previous studies^{26–28}, there was a significant correlation between increasing response latency and N-back ($r(2998) = 0.292, p < .001$). There was also a significant correlation between mean response time and N-back for each subject ($r(328) = 0.374, p < .001$) (Fig. 4). This suggests that our paradigm was successful in actively engaging working memory, as subjects demonstrated more difficulty in recalling the correct response as the N-back increased. This increase in response time is indicative of subjects having to work harder to search short term memory as difficulty of the task increases. Furthermore, when analyzing each subject individually, 26/30 subjects (86.7%) showed significant correlations ($p < .05$) between response latency and N-back. These results suggest that our paradigm successfully increases cognitive load and also adapts to individual differences in skill level on the task, and thus can easily accommodate the ability of different subjects.

Pupilometry:

Evidence suggests cognitive load can be measured through pupil diameter, where an increase in cognitive demand is associated with an increase in pupil size^{31–34}. Our results replicate this finding, with a univariate ANOVA reporting a significant interaction of pupil size and N-back ($F(10) = 1.925, p = 0.038$). Pupil size slightly increases as N-back increases, and then sharply drops off at an N-back of 9 or 10 (Fig. 5). This is consistent with previous reports, which have shown that pupils dilate with the increasing demands of a working memory task, and then constrict again when cognitive load capacity has been surpassed³¹.

Fixations and Saccades:

We used the threshold criteria of the Eyelink 1000 to analyze the number of fixations and saccades, and durations of fixations and saccades. Standard settings on the Eyelink use a velocity threshold of $30^\circ/\text{s}$ and an acceleration threshold of $8000^\circ/\text{s}^2$ to determine the onset and offset of saccades (samples below these thresholds are considered to be fixational/microsaccadic eye movements). We only counted fixations or saccades occurring within the scene region, any events falling outside the image presented were discarded (amounting to a total of 1.59% data removal). Events for each trial were taken from one eye only: the eye used was determined by smoothing the position data of each eye and comparing the smoothed data to the original binocular data, and the eye with a smaller error was used. The total number of fixations and saccades that the Eyelink recorded during a trial were recorded, and the duration of

fixations and saccades were the total cumulative time spent performing each type of event. An example of a subject's scan-path is presented in Fig. 6.

Averages across subjects for maximum N-back:

We calculated the mean number and duration of fixations and saccades made by each subject, and compared those means across the maximum N-back achieved by each subject. We hypothesized that subjects who could achieve a higher N-back were generally better at this task than subjects who maintained a lower N-back and may use different oculomotor strategies than subjects who struggle with this task. However, there was no significant differences in oculomotor parameters between any of the N-back groups. We ran a one way ANOVA with unequal sample sizes, as each maximum N-back had a different number of subjects who had achieved it. There were no significant differences between the number of fixations ($F(8,21) = 0.848$, $p = 0.572$), duration of fixations ($F(8,21) = 0.693$, $p = 0.694$) (Fig. 7A), number of saccades ($F(8,21) = 0.709$, $p = 0.681$), or duration of saccades ($F(8,21) = 0.279$, $p = 0.966$) (Fig. 7B), across all groups (post hoc analysis showed no significant differences between any two maximum groups of N-back). This suggests that subjects who performed well in this task did not use different oculomotor strategies (e.g. looking more frantically around an image with a large number of brief fixations or making fewer, longer fixations), to achieve success.

Correct vs incorrect responses:

We also analyzed oculomotor events according to subject's responses, in order to test if there are differences in oculomotor strategies that lead to more success in this task. We performed a univariate analysis of variance (two way ANOVA) to analyze the interaction of N-back and response accuracy. There was no significant interaction between N-back and response accuracy for the number of fixations made ($F(9) = 0.128$, $p = 0.999$) (Fig. 8A), or for the duration of fixations made ($F(9) = 0.385$, $p = 0.943$) (Fig. 8B). This suggests that the ability to perform this task successfully across various N-backs is not related to the number or duration of fixations made. There were also no significant interactions between N-back and response accuracy for the number of saccades made ($F(9) = 0.309$, $p = 0.972$) (Fig. 8C), or for the duration of saccades made ($F(9) = 1.350$, $p = 0.205$) (Fig. 8D). This suggests that the ability to perform this task successfully across various N-backs is not affected by the amount or duration of saccades made.

Proportion of image looked at:

Our analysis of the number and duration of fixations and saccades showed no relationships between task performance and high or low scoring subjects. We therefore looked at the proportion of each image viewed by each subject for each trial to examine whether there were any effects of the efficiency of eye movements and fixations. We used the `convhull()` function in Matlab to estimate the image area falling within the polygon defined by the farthest reaching positions recorded by the Eyelink (positions that fell outside of the image region were ignored). We used this as a measure of the approximate area of the image that was viewed by the subject. Values are represented as percentages, where the area of the image viewed was divided by the total size of the image (Fig. 9).

Averages across subjects for max N-back:

A one-way ANOVA with unequal sample sizes found no significant differences across groups of maximum N-back reached ($F(8) = 0.448$, $p = .878$) (Fig. 10). This suggests that subjects who were better at this task, (those who were able to reach a higher N-back), on average, did not look at a greater proportion of the image than subjects who performed poorly at this task.

Correct vs incorrect responses:

We performed a univariate analysis of variance (two way ANOVA) to analyze the interaction of N-back and response accuracy. Once again, there was no significant interaction between N-back and response accuracy ($F(9) = 0.803$, $p = 0.613$) (Fig. 11). When looking at Fig. 11, there is a slight trend: as N-back increases, for correct responses there is a small increase in the proportion of the image viewed, whereas for incorrect responses there is a small decrease in the proportion of the image viewed. This suggests that maybe subjects are more successful when viewing more of the image, however there was no significant difference between correct and incorrect responses ($F(1) = 2.946$, $p = 0.086$).

Discussion

We developed novel adaptive paradigm to study of how subjects view scenes under varying levels of cognitive demand. We found that our paradigm was successful in engaging working memory across various difficulties for individual subjects, as reflected in response latency and pupilometry analyses. Our paradigm demonstrates flexibility between subjects: the difficulty of the task is determined entirely by a subject's ability to perform it. This allows the model to fit a variety of different participants with varying cognitive load capacities, while still being able to compare performance between and within subjects at different performance levels. A subject who can only reach an N-back of 2 still has a personalized low-load and high-load range that can be measured: $N = 0$ being low cognitive demand and $N = 2$ being high cognitive demand for this subject. Comparatively, a subject who can reach up to an N-back of 10 is also studied across their performance range, they still complete trials at low and high levels of cognitive load. In this way, the paradigm easily adapts to the subjective ability of individual participants. This feature potentially allows the paradigm to be deployed in special populations, an avenue we are currently investigating.

In our task, observers are required to free-view a sequence of natural images, and identify objects from those images at a later stage. Belke et al's results demonstrate that a variety of different natural images can be presented in our task without the fear of perceptual load influencing oculomotor strategies, and provides assurance that any differences that occur in oculomotor behaviors are due to the manipulation of cognitive load, rather than perceptual load.

When looking at the number and duration of fixations and saccades, we hypothesized that as N-back increased, the number of fixations and saccades would increase as subjects looked more exhaustively around the scene. An alternative hypothesis might state the number of fixations and saccades would

decrease as subjects focused more steadily on significant portions of the scene. However, neither of these hypotheses were supported by our results: there was no significant difference in the number or duration of fixations or saccades. Different subjects who struggled or excelled at this task did not show differences in these general oculomotor behaviors. Similarly, for a given subject there were no differences in oculomotor behaviors on trials where the subject correctly identified the target or incorrectly identified a distractor. These results suggest that increasing the demands of a cognitive load task does not affect oculomotor strategies, and different oculomotor behaviors do not predict better performance on this task. These results challenge the assumption that oculomotor behavior differences between different neurological populations directly relate to attention and cognitive load.

Furthermore, there was no simple relationship between the proportion of the image viewed on average and performance in this task. Subjects who were more successful at this task overall did not fixate a higher overall area of each image. Viewing more of each image did slightly increase the probability of correct responses across N-back, however this result was not significant. These results together suggest that simply viewing “more” of an image does not necessarily improve performance. Searching out to the corners of each image does not predict better performance than focusing on a smaller, central area.

Higher demands of cognitive load did not affect the oculomotor behaviors between participants. Because performance in this task is not correlated with differences in oculomotor behavior, we hypothesize that the variability in success may be reliant on scene context. Perhaps it isn't the potential *amount* of information gathered during free-view, but rather the *context* of what was viewed. We are currently using semantic information of the fixated locations¹⁵ to examine whether success in this task correlates with salience-based viewing methods, or meaning-based ones.

Overall, this paradigm has great potential in measuring eye-movement data while controlling individualized cognitive load. Our pupilometry and performance data demonstrates that this task is successful in manipulating cognitive load while tailoring difficulty to the individual. Concurrently, our eyetracking data is contradictory to the emerging idea that oculomotor behavior is a covert metric for cognitive load.

Declarations

Acknowledgements

Supported by NIH R01 EY029713.

Authors Contribution Statement

K.W. and P.B. conceived the experiment, K.W. conducted the experiment and analyzed the results. Both authors reviewed the manuscript.

Additional Information

The authors declare no competing interests.

References

1. Henderson, J. Human gaze control during real-world scene perception. *Trends Cogn. Sci.* **7**, 498–504 (2003).
2. Rayner, K. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **62**, 1457–1506 (2009).
3. Buswell, G. T. How people look at pictures: a study of the psychology and perception in art (University of Chicago Press, 1935).
4. Yarbus, A. L. Eye movements during perception of complex objects. (Springer, 1967).
5. Andrews, T. J. & Coppola, D. M. Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Res.* **39**, 2947–2953 (1999).
6. Borji, A., Sihite, D. N. & Itti, L. Objects do not predict fixations better than early saliency: A re-analysis of Einhäuser et al.'s data. *J. Vis.* **13**, 18–18 (2013).
7. Harel, J., Koch, C. & Perona, P. Graph-Based Visual Saliency. *Adv. Neural Inf. Process. Syst.* **19**, 545–552 (2007).
8. Itti, L. & Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2**, 194–203 (2001).
9. Parkhurst, D., Law, K. & Niebur, E. Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* **42**, 107–123 (2002).
10. Henderson, J. M., Hayes, T. R., Peacock, C. E. & Rehrig, G. Meaning and attentional guidance in scenes: A review of the meaning map approach. *Vis. Switz.* **3**, (2019).
11. Hwang, A. D., Wang, H. C. & Pomplun, M. Semantic guidance of eye movements in real-world scenes. *Vision Res.* **51**, 1192–1205 (2011).
12. Nyström, M. & Holmqvist, K. Semantic Override of Low-level Features in Image Viewing – Both Initially and Overall. *J. Eye Mov. Res.* **2**, 11 (2008).
13. Onat, S., Açıık, A., Schumann, F. & König, P. The Contributions of Image Content and Behavioral Relevancy to Overt Attention. *PLoS ONE*. **9**, e93254 (2014).
14. Rider, A. T., Coutrot, A., Pellicano, E., Dakin, S. C. & Mareschal, I. Semantic content outweighs low-level saliency in determining children's and adults' fixation of movies. *J. Exp. Child Psychol.* **166**, 293–309 (2018).
15. Rose, D. & Bex, P. The Linguistic Analysis of Scene Semantics: LASS. *Behav. Res. Methods*. <https://doi.org/10.3758/s13428-020-01390-8> (2020).
16. Stoll, J., Thrun, M., Nuthmann, A. & Einhäuser, W. Overt attention in natural scenes: Objects dominate features. *Vision Res.* **107**, 36–48 (2015).
17. Einhäuser, W., Atzert, C. & Nuthmann, A. Fixation durations in natural scene viewing are guided by peripheral scene content. *J. Vis.* **20**, 15 (2020).

18. Nuthmann, A. Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychon. Bull. Rev.* **24**, 370–392 (2017).
19. Pedziwiatr, M. A., Kümmerer, M., Wallis, T. S. A., Bethge, M. & Teufel, C. Meaning maps and saliency models based on deep convolutional neural networks are insensitive to image meaning when predicting human fixations. *Cognition*. **206**, 104465 (2021).
20. Ozeri-Rotstain, A., Shachaf, I., Farah, R. & Horowitz-Kraus, T. Relationship Between Eye-Movement Patterns, Cognitive Load, and Reading Ability in Children with Reading Difficulties. *J. Psycholinguist. Res.* **49**, 491–507 (2020).
21. Howard, P. L., Zhang, L. & Benson, V. What Can Eye Movements Tell Us about Subtle Cognitive Processing Differences in Autism? *Vision*. **3**, 22 (2019).
22. Wang, S. *et al.* Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron*. **88**, 604–616 (2015).
23. Molitor, R. J., Ko, P. C. & Ally, B. A. Eye Movements in Alzheimer's Disease. *J. Alzheimers Dis. JAD.* **44**, 1–12 (2015).
24. Sweller, J. CognitiveLoadDuring ProblemSolving: Effects on Learning. *Cogn. Sci.* **12**, 29 (1988).
25. Kirchner, W. K. Age differences in short-term retention of rapidly changing information. *J Exp Psychol.* **55**, 352–358 (1958).
26. Carlson, S. Distribution of cortical activation during visuospatial n-back tasks as revealed by functional magnetic resonance imaging. *Cereb. Cortex.* **8**, 743–752 (1998).
27. Jonides, J. *et al.* Verbal Working Memory Load Affects Regional Brain Activation as Measured by PET. *J. Cogn. Neurosci.* **9**, 462–475 (1997).
28. Perlstein, W. M., Dixit, N. K., Carter, C. S., Noll, D. C. & Cohen, J. D. Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biol. Psychiatry.* **53**, 25–38 (2003).
29. Braver, T. S. *et al.* A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory. *NeuroImage* **5**, 49–62(1996).
30. Manoach, D. S. *et al.* Prefrontal cortex fMRI signal changes are correlated with working memory load. *NeuroReport.* **8**, 545–549 (1997).
31. Granholm, E., Asarnow, R., Sarkin, A. & Dykes, K. Pupillary responses index cognitive resource limitations. *Psychophysiology.* **33**, 457–461 (1996).
32. Kahneman, D. Attention and effort. (Prentice-Hall 1973).
33. Klingner, J., Kumar, R. & Hanrahan, P. Measuring the task-evoked pupillary response with a remote eye tracker. in *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08* 69 (ACM Press, 2008). doi:10.1145/1344471.1344489.
34. Rafiqi, S. *et al.* PupilWare: towards pervasive cognitive load measurement using commodity devices. in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '15* 1–8 (ACM Press, 2015). doi:10.1145/2769493.2769506.

35. Stuyven, E., Claeys, K. & Crevits, L. The effect of cognitive load on saccadic eye movements. *Acta Psychol. (Amst.)*. **104**, 69–85 (2000).
36. Zagermann, J., Pfeil, U. & Reiterer, H. Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. in Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV '16 78–85(ACM Press, 2016). doi:10.1145/2993901.2993908.
37. Zagermann, J., Pfeil, U. & Reiterer, H. Studying Eye Movements as a Basis for Measuring Cognitive Load. in Extended Abstracts of the 2018 *CHI Conference on Human Factors in Computing Systems* 1–6 (ACM, 2018). doi:10.1145/3170427.3188628.
38. Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S. & Telling, A. L. Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Percept. Psychophys.* **70**, 1444–1458 (2008).
39. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
40. Cornelissen, F. W., Peters, E. M. & Palmer, J. The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behav. Res. Methods Instrum. Comput.* **34**, 613–617 (2002).
41. Russell, B. C., Torralba, A., Murphy, K. P., Freeman, W. T. & LabelMe A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **77**, 157–173 (2008).

Figures

Figure 1

Illustrations of N-back Visual Search. A) Example stimuli selected from the LabelMe database, here illustrated as three images presented sequentially in the stream. Illustrations of the experimental display for B) N-back increase or C) N-back decrease. In both examples B and C, assume the subject was previously at an N back of 1. B) Example of a subject getting 2 correct responses in a row and moving up to an N back of 2. The image presented is last in the photo stream (A), but the forced choice task will be based on the scene presented 2 images prior (first in the photo stream (A)). C) Example of a subject getting a response incorrect and moving down to an N back of 0. The image presented is last in the photo stream (A), and the forced choice task is to select a target object from that same scene.

Distribution of Maximum N-Back Reached Across Subjects

Figure 2

Distribution of maximum N-back achieved by each subject. Top plot represents a histogram of maximum N-back achieved by each participant. Bottom plot represents a box and whisker summary of the same data.

Figure 3

Mean N-Back across Experiment. Each sub-plot represents a block from the experiment, note N-Back resets to 0 at the start of each block. Means are computed as the average N-back from every subject at a given trial. Performance curves are fitted with a quadratic polynomial ($y=a*(x-1)^b$).

Figure 4

Mean response latency at each N-back level for each subject. Note N-back=0 has 30 points as all 30 subjects experiences an N-back of 0, while N-back=10 only has 1 point because only 1 subject reached an N-back of 10. Error bars show ± 1 SEM.

Figure 5

Mean pupil size as a function of N-back. Separated into correct (red) and incorrect (blue) responses. Pupil size increases gradually with increasing N-back, then constricts abruptly once N-back exceeds 8. Error bars show ± 1 SEM.

Figure 6

Example of a representative gaze-pattern from one subject. Blue points represent fixations, red dotted lines represent saccades.

Figure 7

Mean number (A) and duration (B) of fixations (blue) and saccades (red) made by each subject at their maximum N-back reached. Duration is represented as the total cumulative sum of all fixation or saccade events during a trial. Error bars represent 95% confidence intervals.

Figure 8

Relationship between maximum N-back achieved for correct (blue) and incorrect (red) responses for A). Mean number of fixations; B) mean cumulative duration of fixations; C) mean number of saccades; and D) mean cumulative duration of saccades. Error bars represent ± 1 SEM.

Figure 9

Percentage of image viewed analysis. Example of a subject's raw gaze-data (blue x's), overlaid with the `convhull()` polygon (teal lines). Area of the polygon is divided by the area of the image (not including gray stimulus background), to produce an approximate estimate of the percentage of the image viewed by the subject.

Average Proportion of Image Viewed by Each Subject

Figure 10

Mean proportion of each image viewed by each subject as a function of their maximum N-back reached. Error bars represent 95% confidence intervals.

Figure 11

Mean proportion of each image viewed as a function of N-back. Separated into correct (blue) and incorrect (red) response. Error bars represent ± 1 SEM.