Attention and Information Acquisition: Comparison of Mouse-Click with Eye-Movement Attention Tracking

Steffen Egner
MediaAnalyzer GmbH,
Hamburg, Germany

Stefanie Reimann
University of Luneburg,
Germany

Rainer Hoeger
University of Luneburg,
Germany

Wolfgang H. Zangemeister
University of Hamburg,
Germany

Attention is crucial as a fundamental prerequisite for perception. The measurement of attention in viewing and recognizing the images that surround us constitutes an important part of eye movement research, particularly in advertising-effectiveness research. Recording eye and gaze (i.e. eye and head) movements is considered the standard procedure for measuring attention. However, alternative measurement methods have been developed in recent years, one of which is mouse-click attention tracking (mcAT) by means of an online based procedure that measures gaze motion via a mouse-click (i.e. a hand and finger positioning maneuver) on a computer screen.

Here we compared the validity of mcAT with eye movement attention tracking (emAT). We recorded data in a between subject design via emAT and mcAT and analyzed and compared 20 subjects for correlations. The test stimuli consisted of 64 images that were assigned to eight categories. Our main results demonstrated a highly significant correlation $(p<0.001)$ between mcAT and emAT data. We also found significant differences in correlations between different image categories. For simply structured pictures of humans or animals in particular, mcAT provided highly valid and more consistent results compared to emAT. We concluded that mcAT is a suitable method for measuring the attention we give to the images that surround us, such as photographs, graphics, art or digital and print advertisements.

Introduction

Visual Attention is a research topic of increasing impact. Not only is attention an interesting topic in itself, it also plays a crucial role in perception and motor control. Moreover, measuring attention also yields valuable data for studying higher cognitive functions such as interest, understanding and reading. Measurement of attention is traditionally being seen as parallel to eye-tracking (Yarbus, 1961, 1967; Noton & Stark, 1971; Groner, 1988; Groner & Groner, 1989; Groner & Groner, 2000). The underlying rationale is that humans direct the region of their retina with the highest resolution (fovea) to aspects of the optical scene which are of high relevance for the organism. However, experiments with response latency tasks (Posner, Snyder, & Davidson, 1980) clearly indicate that there are attention shifts that are not measurable as eye movements (covert attention). Moreover, the visual

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

modalities do not seem to be specifically linked to attention. In fact, attention seems to be modality-unspecific (Allport, 1987; Neumann, 1987). The two classical ways to measure attention have specific advantages and disadvantages.

Eye-Tracking and attention

As indicated by its name, eye tracking measures the position and orientation of the eye(s). Based on these raw data, gaze position in the environment can be determined. Eye-tracking is a technique to measure an individual's visual attention, focus, and eye movements. This experimental methodology has proven useful both for human-computer interaction research and for studying the cognitive processes involved in visual information processing, including which visual elements people look at first and spend the most time on (Jacob & Karn, 2003).

Fixation criteria are often unclear. Blinks, correction saccades, physiological and technical noise contribute to difficulty of measurement: This is attention tracking by eye movement – emAT.

Response latency tasks assume that the reaction upon an event that happens at a specific location – for instance the onset of a stimulus – will be quicker when the position of the emerging stimulus is expected at that particular moment. This method allows the measurement of covert attention (Posner et al., 1980), which precedes eye-movements in some cases. Each trial of a respondent only reveals one attended location at the most. Therefore, we cannot measure a full path of attentional shifts, but only individual locations. Some authors propose that attention can be measured in other ways than the two methods mentioned above.

Some of these other ways employ the computer mouse to indicate attention locations. One of these methods, mouse-click based attention tracking (mcAT) will be examined in more detail in this paper. It seems obvious that a method that relies on the computer and a mouse as the only necessary devices would have many practical advantages over other methods. But the question is: Is it a method that generates results with validity comparable to eye tracking?

Both, *salience and conspicuousness* of a stimulus in terms of its environment (Borji & Itti, 2013) as well as *relevance of a stimulus* are decisive criteria for the allocation of attention. They seem to be based upon two inde-pendent systems (Schütz, Braun, & Gegenfurtner, 2011); however, to quantify the impact of bottom-up and top-down mechanisms within a certain setup may be highly difficult to differentiate with respect to their individual importance to the actual perception (Fischer, 1999; Parkhurst, Law, & Niebur, 2002). Various studies have shown that the degree of exogenous and endogenous direction of attention depends upon a number of factors. Novel or unfamiliar stimuli situations in free-viewing tasks, or the viewing of images in advertising are usually thought to be dominated by bottom-up processes, especially at the beginning of the viewing time (Anderson, Ort, Kruijne, Meeter, & Donk, 2015; Berger, 2009; Duchowski, 2007). However, with increased viewing time, and with known visual performances and situations as well as in the search for certain stimuli, the situation is dominated by top-down processes.

Viewing time: The total amount of time within an AOI approximately complies with the fixation duration – the time between two successive clicks, generally half the fixation before (max. 500ms) and half of the fixation attributed thereafter (max. 500 ms).

Selective attention and eye movements – the classical relationship to study.

Selective attention is the gateway to conscious experience, affecting our ability to perceive, distinguish and remember the various stimuli that come our way (James, 1890). Selective attention denotes the allocation of limited processing resources to some stimuli or tasks at the expense of others (Dosher, Sperling, & Wurst, 1986; Kowler, Anderson, Dosher, & Blaser, 1995; Norman & Bobrow, 1975; Reeves & Sperling, 1986; Shaw, 1982, 1984). Apart from its effects on perception or memory, selective attention is a significant contributor to motor control, determining which of the various objects in the visual field is to be the target used to plan and guide movement. As selective visual attention allows us to concentrate on one aspect of the visual field while ignoring other things, it is modulated by both involuntarily bottom-up and voluntary top-down mechanisms (Kowler et al., 1995), within a brainstem-parietotemporal and basal ganglia-frontal neuronal network (Kastner & Ungerleider, 2000).

Selective visual attention for spatial locations is under the control of the same neural circuits as those in charge of motor programming of saccades (Dubois & Pillon,

1997; Mink, 1996; Rizzolatti, 1983; Rizzolatti, Riggio, Dascola, & Umiltá, 1987; Sheliga, Craighero, Riggio, & Rizzolatti, 1997).

Directing visual attention to a certain location as well as ocular saccades in visual attention tasks depend upon accurate saccade programming. Programming the eye saccade is thought to lead to an obligatory shift of attention to the saccade target before the voluntary eye movement is executed, which is due to two parameters: correct programming of the saccade and correct saccade dynamics. (Deubel & Schneider, 1996; Henderson & Hollingworth, 1999; Sheliga, Riggio, & Rizzolatti, 1994; Umiltà, Riggio, Dascola, & Rizzolatti, 1991).

Therefore, the alertness of central, top-down programming influences oculomotor function and, conversely, a resulting oculomotor dysfunction could have a direct, bottom up impact on results of visual attention tasks.

Visual selective attention can be investigated by visual search tasks.

Visual search means to look for something in a cluttered visual environment. The item that the observer is searching for is termed the target, while non-target items are termed distractors. Many visual scenes contain more information than we can fully process all at once. Accordingly, mechanisms like those subserving object recognition might process only a selected/restricted part of the visual scene at any one time. Visual attention is used to control the selection of the subset of the scene, and most visual searches consist of a *series of attentional deployments*, which ends either when the target is found, or when the search is abandoned. Overt search refers to a series of eye movements around the scene made to bring difficult-to-resolve items onto the fovea. Only if the relevant items in the visual scene are large enough to be identified without fixation can the search be successfully performed while the eyes are focused upon a single point. In this case, attentional shifts made during a single fixation are termed *covert*, because they are *inferred rather than directly observed*.

While under laboratory conditions, many search tasks can be performed entirely with covert attention, under real world conditions a new point of fixation is selected 3 to 4 times per second. *Overt fast movements of the eye, saccades, and covert deployments of attention are closely related* (Kowler et al., 1995), as the sample rate of saccades is 4/sec. With stimuli that do not require direct foveation, 4–8 objects can be searched during each fixation. As estimates of the minimum time required to recognize a single object are almost always greater than 100 ms, multiple items may be processed in parallel (Palmer, 1995). *Volitional* deployments of attention are much slower than *automatic* deployments (Wolfe, Alvarez, & Horowitz, 2000), and occur at a rate similar to saccadic eye movements, i.e. a sample rate of 4/sec (Gilchrist & Harvey, 2006). *Search termination* happens after finding the target, or one could declare the target to be absent after rejection of every distractor object, although it may be difficult to determine when this point has been reached.

Mouse-click attention tracking – Background

The mouse-click Attention Tracking (mcAT) method measures attention by mouse clicks. They can be counted and concatenated to a time sequence that is analogous to the eye movement scanpath. Egner and Scheier developed this method in collaboration with Laurent Itti (Egner, Itti, & Scheier, 2000) at the California Institute of Technology (USA). They assumed the predictive power of a computerized attention model with three categories of visual stimuli (photographs of natural scenes, artificial laboratory stimuli and sites).

Based upon empirical evidence of a close link between attention, eye movements/fixations and pointing movements (Scheier, Reigber, & Egner, 2003) the eye tracking data (emAT), touch screen and click data with a computer mouse (mcAT) were highly correlated (an overview can be found in Berger, 2009; Egner et al., 2000; Scheier & Egner, 2005). The mcAT method was patented in the US and Europe as a mouse-click based AT procedure for measuring visual attention (Scheier & Egner, 2003a; Scheier & Egner, 2003b).

The central idea of the mcAT method is the natural coupling of the use of mouse clicks with eye movement measures, which in turn represent a valid indicator of the attention.

In the following years, also other researchers have explored the relationship between users' mouse movements and eye movements on web pages (Granka, Joachims, & Gay, 2004).

Deng, Krause and Fei-Fei (2013) used a bubble paradigm of Gosselin and Schyns (2001) that was used to

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

discover the object/image regions people explicitly choose to use when performing.

Chen, Anderson and Sohn (2001) described in their paper a study on the relationship between gaze position and cursor position on a computer screen during web browsing. Users were asked to browse several web sites while their eye/mouse movements were recorded. The data suggested that there was a strong relationship between gaze position and cursor position. The data also showed that there were regular patterns of eye/mouse movements. Based on these findings, they argued that a mouse could provide more information than just the x, y position where a user was pointing. They speculated that by understanding the intent of every mouse movement, one should be able to achieve a better interface for human computer interaction.

Using eye and mouse data, Navalpakkam et al. (2013) demonstrated that the mouse, like the eye, is sensitive to two key attributes of page elements: their position (layout), and their relevance to the user's task. They identified mouse measures that were strongly correlated with eye movements and developed models to predict user attention (eye gaze) from mouse activity.

Our approach is different from the viewing window approach of (Deng et al., 2013) in that we explicitly collect the path of discretized click data, as each click represents a conscious choice made by the user to reveal a portion of the image. Since the clicks correspond to individual locations of attention, we can directly compare them to eye fixations.

Kim et al. (2015) investigated the utility of using mouse clicks as an alternative for eye fixations in the context of understanding data visualizations. They developed a crowdsourced study online in which participants were presented with a series of images containing graphs and diagrams and asked to describe them. They compared the mouse click data with the fixation data from a complementary eye-tracking experiment by calculating the similarity between resulting heatmaps and got a high similarity score and suggested that this methodology could also be used to complement eye-tracking studies with an additional behavioral measurement, since it is specifically designed to measure which information people consciously choose to examine for understanding visualizations.

Aim of our study

The question is, can mouse clicks approximate human fixations in the context of data visualization understanding? When we compare eye movement/fixations and hand mouse movement/clicks, we assume that the sensory-attentional and the cognitive part of these actions are highly similar, whereas the motor part is obviously different. From this reasoning we can infer three questions:

1. What *are* the differences between eye and hand movements that have been described by many researchers (Stark, 1968; Lacquaniti & Soechting, 1982) and how do they relate to our findings?

2. What are the similarities of the two responses?

3. Are there non-motor differences related to attention/cognition and how do they relate to our findings?

The present study is of interest to the eye movement research community for the following reasons. While eye-tracking is the well-established method for measuring visual attention, the eye movement data does not allow to make a distinction between eye-movement-specific and attention-specific effects. The alternative measurement described and used in the present article uses the hand (computer mouse) to measure attention. The resulting data is highly similar to eye-tracking data, and it is not affected by eye-movement-specific processes. Thus, it allows to separate eye-movement-specific and attention-specific effects. Additionally, the alternative measurement enables a method comparison, which enriches our knowledge about eye-tracking methodology. Last, the new method can help to gain a better understanding of attention, which is also the goal of much eye and mouse tracking research on web page viewing. So, both methods contribute to the same goal.

Therefore, two practical issues were addressed:

i. Does the spatial dimension of fixations and clicks correlate highly positively?

ii. Does this putative correlation depend upon the stimulus material?

In view of our experiences, we hypothesized that:

i. The overall pattern of fixations/clicks correlate highly.

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

ii. The amount of the agreement between both recording methods depends upon the nature of the stimuli.

Methods

We used an independent experimental design i.e. a between-subject design selected to prevent both carry-over effects (affecting a later experimental condition by a previous condition) as well as to prevent position effects like fatigue and exercise.

Subjects

The data of twenty participants were used. One group of subjects underwent one experimental condition only: In ten subjects emAT was measured via eye movement recordings while viewing the stimulus material.

The remaining ten subjects were subject to mcAT, measured via mouse-click recordings while viewing the stimulus material. All other experimental conditions (stimulus material, type of presentation, site of examination, demographics of the subjects, experimenter) were kept constant. Before the actual test took place, multiple testing of subjects to clarify, check and optimize instructions and operation of the equipment took place.

The study has been approved by the local ethics committee. It complies with the ethical practices and follows the Code of Conduct and Best Practice Guidelines outlined by the Committee on Publication Ethics. Informed consent for the research was obtained by an oral introduction and overview at the LEUPHANA University of Luneburg, Engineering Psychology research lab.

Sample

The whole study consisted of three phases: (1) compilation of the stimulus material, (2) pilot run, and (3) final experiment.

The participants of all three trials – of the preliminary selection of images and classification in the classification scheme (n = 12), the trial run (n = 6) and the final test with emAT and mcAT measurements (n = 29) – were students recruited from the University of Lueneburg. All participants gave their written informed consent following the rules of the Helsinki Declaration. None of the subjects (Ss) had participated in more than one of the tests or knew about the exact purpose of the investigation.

The comparison of various studies shows a wide range in the number of required study participants. An overview of Borji and Itti (2013) lists over 19 trials in which attention stimuli computer programs were used based on emAT; the subject numbers vary between 5 and 40, but more than half of the listed studies used less than 15 subjects. To keep the failure rate of the emAT test low, only subjects that did not rely on visual aids were invited.

Four out of 14 emAT respondents were removed from the evaluation. The criterion for removing such recordings was the calibration quality. The calibration, which was performed in the beginning of the recording, was checked at the end of each recording. The respondent had to redo the calibration procedure. If the results were significantly different from the initial calibration, we removed the recording. This was decided upon face validity. The reasoning behind this is that, if the calibration parameters have changed throughout the recording process, this is due to a distortion that happened during the recording. The remaining ten were between 18 and 22 years old (average age: 20). Four were female. Also, five out of 15 mcAT respondents were removed from the evaluation. This was decided upon the mouse behavior during the recording. The recording was excluded if:

- the click rate went below 1.5 clicks per second,

- the respondent stopped moving the mouse,

- the click pattern revealed that the respondent did not understand the instructions. The last point was decided upon face validity.

Subjects for mcAT measurements had no restrictions concerning visual aids. Of the 15 published subjects, ten recordings could be used for further analysis. From the ten remaining subjects, the age range was 20–26 years (average age: 23) and the sex ratio was even.

Stimuli

Until now, no generally accepted classification schemes, neither number nor type of classes have been suggested in attention research. Examples of the classification of the photos are: Natural landscapes and portraits (Judd, Ehinger, Durand, & Torralba, 2009); animals in natural environments, street scenes, buildings, flowers and natural landscapes (Kootstra, de Boer, & Schomaker, 2011); nature/landscape scenes, urban environments and artificial laboratory stimuli (Wilming, Betz, Kietzmann,

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

& König, 2011) as well as images with obvious and unclear/non-existent AOI (Oyekoya & Stentiford, 2004). None of the authors stated the reasons/justification for the particular classification that was selected. Therefore, we developed our own classification scheme that reflects the suggestions in the literature but has also been derived from features that mirror the mechanisms of control of attention.

Figure 1 shows the distribution of images at their different levels. At the highest level were photographs that represented animate and inanimate matter. At the second level, the class animate represented pictures of people/animals-plants; the inanimate class included artificial/natural environments. Each of these four classes is

tos from the pixabay.com website, a database for a Creative Commons CCO.

First, we chose a preselection of images per category (120). To ensure that the classification of images was done objectively despite different possible interpretations of schema classes, we reduced the dataset to eight photos per category using 12 subjects for classification. The subjects were given the task of classifying each of the 120 sequentially presented images in the classification scheme shown above without further explanation of the schema classes. In this way, an impartial classification could be performed.

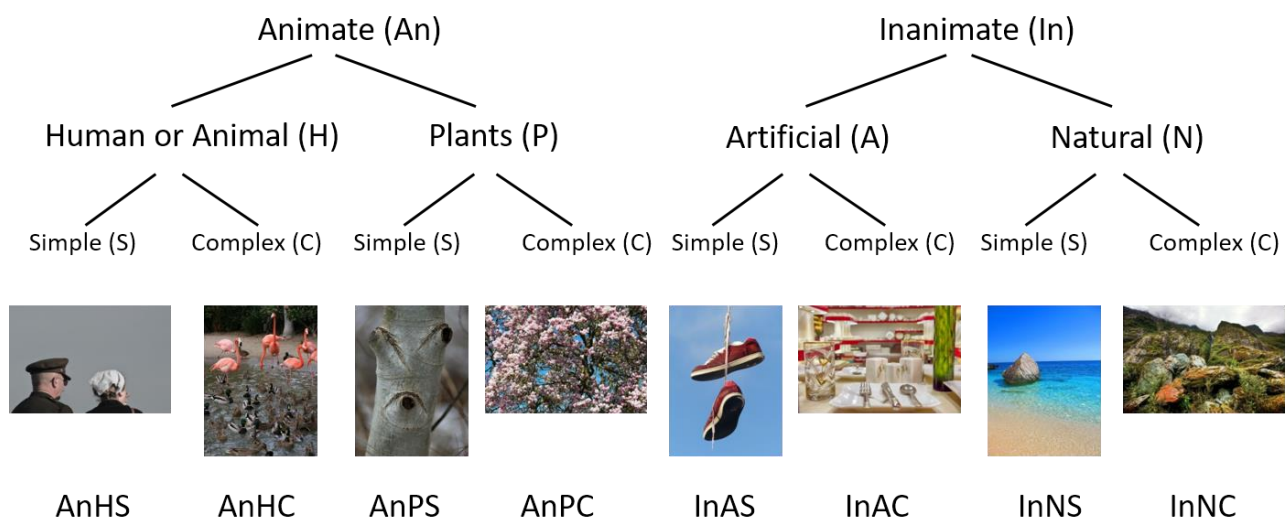Only those images were chosen for the experiment for

Figure 1. Images and their category levels. Utilized stimulus material: human/animal, easy, complex, plants, inanimate, artificial, natural

independent of content aspects and divided into simple and complex designs.

Although complexity in different contexts may be an important determinant of attention processes (Pieters, Wedel, & Batra, 2010; Schütz et al., 2011), different authors define the term complexity in many different ways.

Image selection

The photographs originate from two image databases: i. (Borji & Itti, 2013) which is currently the largest, freely available and commonly used image data set that was also used by Judd et al. (2009). Another source was pho-

which a minimum of two thirds of the subjects chose a particular category.

During the evaluation, it was found that some subjects had difficulties differentiating between the categories "natural" and "plants"; the terms "complex" and "simple" were also interpreted quite differently. Subsequently, the selected 64 images were cut to three uniform sizes to reflect the screen used in the following test: 8 photos are dimensioned in portrait format (690 x 920 pixels); 30 photos are in landscape format (1226 x 920 pixels); 26 photos are in landscape format with 1250 x 833 pixels.

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

Distribution of image material at different levels.

Photographs with representation of animate and inanimate matter were at the highest level. The second level, the "animate" class, represented images of people/animals, and plants, and the "inanimate" class in artificial and natural environments. Each of these four classes was again independent of content aspects and divided into simple and complex designs. Although complexity was in different contexts an important determinant of attention processes (Pieters et al., 2010; Schütz et al., 2011), many authors have used the term differently.

A priori AOIs – grid application

To compare the spatial distribution of the viewing and click data in a meaningful way, regions (ROI) or areas of interest (AOI) had to be defined. For this, semantically-based AOIs have been frequently used, in particular in the analyses of advertisements or sites. As we were less interested in gaze behavior with respect to specific image regions and were more interested in the global eye movements over the entire image, we used a grid laid over the image that divided the image into a certain number of fields. In this way, we excluded subjective preferences that might confound the results of our analysis. Of course, this procedure had its disadvantages: The choice of the grid field's size sometimes played a particularly important role by defining objects too inaccurately, e.g. a face might be divided into several fields and, therefore, subdivided by single fixations. As the image data set used here contained many complex stimuli for which AOIs were difficult to define, gridding was the only meaningful way to analyze the data. As a compromise, we selected a 5 x 7 grid (35 fields), so that the fields had an average square size. The attention parameters were steadily distributed features and could assume values between 0.0 and 1.0. For example, a value of 0.24 in a grid meant that 24% of all clicks or fixations were made in this particular field. The contact parameters, however, were discretely distributed. 24% of all clicks or fixations were made in this field. The contact parameters, however, were discretely distributed features, with eleven possible specifications between 0 and 1 (in increments of 10) for ten subjects; for example, a value of 0.3 means that 30% of the subjects had looked or clicked in a field.

An arbitrarily chosen grid definitely has disadvantages in the evaluation. Among other things, adjacent fixations (or clicks) may fall into separate grid cells, even though both fixations belong to the same object. It would be desirable to evaluate fixations on the same object together. As an alternative to the grid approach, one can also define regions. Ideally, the regions are set to correspond to fixation goals (objects). This bypasses the above cited disadvantage. However, this approach also has a significant disadvantage. The manually selected regions can strongly distort the results if chosen unfavourably. They could also be used to deliberately distort results. We will not solve this general problem of the eye tracking community with our article (see Riche, Duvinage, Mancas, Gosselin, & Dutoit (2013) for a methodological overview). That is beyond the scope of our paper. The JEMR paper by Oliver Hein and Wolfgang H. Zangemeister (2017) offers one possible solution.

In summary, to calculate the fixation and click data, the *contact value* was calculated, i.e. the proportion of subjects that viewed or clicked in a particular field. Also, the *attention value* of the subjects was calculated by averaging the *single grid* percentage clicks or fixations with respect to the total clicks or fixations per image.

For automatic algorithmic generation of particular grid sizes and/or content specific AOIs see: Privitera and Stark (1998) and Hein and Zangemeister (2017).

Experimental Setup

In order to keep the experimental conditions for both measuring methods as constant as possible, data collection was carried out in both emAT and mcAT in the eye movement laboratory of the University of Lüneburg between November 1 and 16, 2015. The stimulus material was presented on a 21.5-inch monitor (Acer) with a resolution of 1920 x 1080. To avoid sequence effects, the 64 images in both experiments were presented in randomized order each for the duration of 5 s. This timing was chosen in accordance with many other related studies (Jiang, Huang, Duan, & Zhao, 2015; Kootstra et al., 2011; Oyekoya & Stentiford, 2004). We separated the individual images by means of a blank screen (here for a duration of 2 s) on which a commonly used fixation cross was shown in the middle (Holmqvist et al., 2011). A fixation cross was used for both measurement methods for all pictures to ensure a common starting position for the eyes and the computer mouse. Following both tests, the subjects answered a short questionnaire on their demographic data. They also completed a recognition test

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

and had to judge whether a series of images in the previous experiment was shown or not. The mcAT-subjects were also requested to answer three qualitative questions with click behavior.

Eye movement Attention Tracking (emAT)

The eye movement measurement was carried out with the SMI iView X™ Hi-Speed 1250 eye tracker. This is a tower-mounted dark-pupil system recording movements of one eye with a sampling rate of 500 Hz (SMI SensoMotoric Instruments GmbH; SMI SensoMotoric Instruments GmbH). The distance between the chin rest of the iView X™ and the screen was 60cm. Programming, evaluation and control of the experiment was carried out using SMI's BeGaze Analysis (version 3.5). The iViewX program was used to control the recording of eye movement.

Subjects received standardized verbal instructions, during which they were informed of the calibration and test procedure. Automatic calibration then followed, with a spatial accuracy of at least 0.5°. We used additional manual calibration in case accuracy was insufficient. This took place before and after the presentation of 64 images for 10 seconds for each image. Thus, both the measurement accuracy and precision were validated to provide assessment of data quality. After the first calibration, further instructions were carried out on the screen. Subjects were asked to view the following images as they chose in a free-viewing task to create almost natural viewing conditions without any viewing strategies (see as refs.: Borji & Itti, 2013; Parkhurst et al., 2002). Overall, the presentation of images took about ten minutes.

Mouse click Attention Tracking (mcAT)

For the mcAT test, the instructions (Appendix 1), click training, click test with 64 images and the demographic data collection was programmed using MALight software from MediaAnalyzer in an online questionnaire. While the subjects initially completed the click training, the experimenter looked to answer questions and provide guidance for clicking behavior in case the instructions were not understood. The subsequent viewing of images was done similarly to the emAT without any task, with the supplementary advice: "You can click everywhere you are looking at". Completion of the click training, and the click test took about 12 minutes.

Data Analysis

Default settings algorithm parameters of "BeGaze" (SMI SensoMotoric Instruments GmbH) were: Saccade detection parameters: min. duration 22ms, peak velocity threshold 40°/s, min. fixation duration 50ms. Peak velocity start: 20% of saccade length; end: 80% of saccade length.

First the emAT fixations were calculated. BeGaze contained both a dispersion-based and a speed-based algorithm, which was used here because the SMI machine is a high-speed device. The minimum fixation duration was based upon inspection of selected images and subjects for durations of 50 ms, 100 ms and 200 ms. Based upon this visual analysis and the information in the literature of Holmqvist et al. (2011), we used a minimum fixation duration of 100 ms (instead of the default 50 ms) as the parameter setting.

Data Cleansing

Next, the data quality of each subject was checked by their fixations at the beginning and at the end of the experiment. In case deviation between initial and final calibration fixations (precision and accuracy) was too high, we had to exclude four subjects (s. Annex 4). It can be assumed that the first specific fixation does not necessarily start with the very first fixation, but after a certain period of time that we defined to be 500ms: Therefore, the first 500ms were excluded from further evaluation. Compared to the first saccadic eye fixations, the manual start of the mouse clicks i.e. the "mouse-fixation" was slightly slower than the sequence of eye movement fixations, due to the inertial load difference between eye and hand. Therefore, for mouse clicks we excluded the first 800ms from further evaluation. Decisive for the quality of click data was a minimum click speed that can be controlled. The demanded click rate was 1.5 per second or higher. This was attained by all subjects. The click data in the attention and contact parameters were transformed per grid and averaged across all subjects by means of Microsoft Excel. At 35 fields per frame and a total of 64 images, 2240 values per sample were observed for each method (mcAT and emAT).

Click test. Subjects view a series of stimuli on a screen for 5–7 seconds – usually advertising materials, website or shelf view, mostly in combination with distractors that are shown before and after the test material.

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

They are prompted to click quickly and without thinking on those places that they consider to be attractive. This measurement is carried out as a "click test" during an online survey, which they can have performed by an online panel of recruited volunteers with their own computers at home. As the subjects are required to perceive the mouse as an "extension" of the eye, a short time of training for the exercise is required. This click training is an interactive and playful method, based upon five tasks during which the subjects get accustomed to clicking continuously fast – at least 1–2 times per second – while they control certain image regions with the mouse. Meanwhile, they receive real-time evaluation feedback on their click behavior. Only subjects that pass all tasks, i.e. also after several attempts have been successfully completed, can they take part in the next click test. The aim of the training is to teach subjects to click as spontaneously and unconsciously as they direct their gaze, so that a fixation and a mouse-click become equivalent.

The data collected from the click test and the survey are stored on a server, and MediaAnalyzer uses special software to statistically analyze and interpret the data. Firstly, there is a verification of data quality and possibly a data cleansing. Despite click training, few subjects fail to maintain the clicks throughout the test or show a lack of motivation by clicking only on the same spot. Data from these subjects can be detected and filtered by algorithms (Egner & Scheier, 2002; Scheier & Egner, 2005; Scheier et al., 2003). During the evaluation, a click is taken as a fixation. It is analyzed similarly to an etAT based upon semantically-derived, predefined areas of interest (AOIs) – at an ad e.g. based on the logo or the name – and on the average results of all subjects.

Typical parameters are:

i. Time to contact: The time to first click in an AOI

ii. Percent attention: Share of clicks in an AOI relative to the total number of click-stimulus corresponds approximately to the relative fixation frequency; thereafter referred to as attention value.

iii. Percent contact: Relative proportion of subjects that clicked at least once in an AOI; thereafter referred to as contact value.

Statistical analysis

Using IBM SPSS, various summary measures were calculated to determine the relationship between the con-
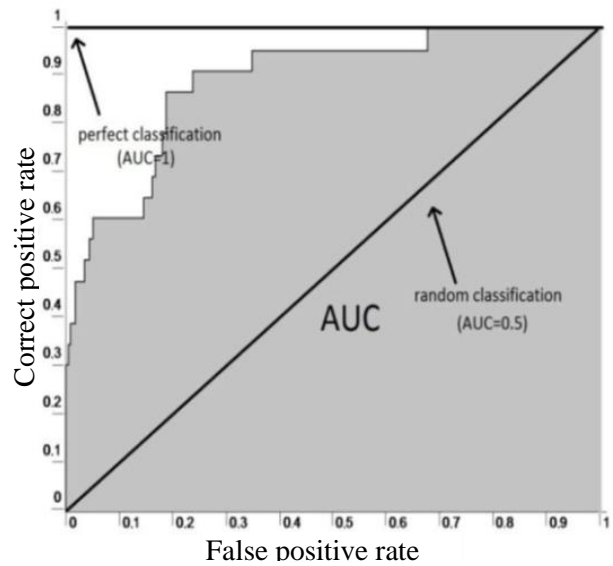

Figure 2. ROC curve example (Cadek, 2015)

verted data from mcAT and emAT: Pearson product-moment correlation coefficient, the area under curve (AUC) and the receiver operating characteristic (ROC) curve.

The correlation analysis and the calculation of the ROC curve are the most commonly used methods for analyses of these data (Borji, Sihite, & Itti, 2013; Wilming et al., 2011; Zhao & Koch, 2011). The use of two or more evaluation methods is recommended to ensure that the observed effects are independent of the summary measure (Borji et al., 2013.).

The *Pearson product-moment correlation coefficient* provides information about the amount and direction of the linear relationship between two interval-scaled variables, in this case between the pairs of values of the two samples (emAT and mcAT parameters per grid). The correlation coefficient r can take values between -1 and +1 that specify the strength and the sign of the direction of the relationship. If r = 0, no linear relationship between the variables is evident.

Correlations with r ≥ 0.5 are considered as high, and r ≥ 0.7 as very high (Cohen, 1988).

Receiver operating characteristic - ROC curve

The ROC curve originated from the signal detection theory and is used in medicine as a tool for the evaluation of known diagnostic tests. Transmitted to the two methods of attention measurement the ROC curve or AUC

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

(area under the curve) measures the goodness of this measure (mcAT parameters) to predict the occurrence or absence of the variable of the other method (emAT parameters). There are four possibilities of prediction: right positive, false positive, right negative and false negative.

The ROC curve is created by a diagram of the *correct positive rate* (known as the *"hit ratio" or "sensitivity"*) and is deducted from the *false positive rate* (also known as *"one minus specificity"*) (Fig. 2), wherein the threshold of the classifier (the AT parameter) is continuously varied. The closer to the diagonal, the more the right-positive rate corresponds to false positive rate – which is expected of the right-positive frequency of a random process equivalent. Thus, the greater the area under the curve (AUC), the better is the prediction; and thus, the agreement between the two variables (Fawcett, 2006; Janssen & Laatz, 2013).

Sensitivity i.e. probability of detection (see 2 refs. above) – measures the *number* of positives that are correctly identified as such (e.g. the percentage of mouse clicks that resemble *true eye fixations*). *Specificity* (also called the true negative rate) measures the number of negatives that are correctly identified as such (e.g., the *number* of mouse clicks not resembling eye fixations, false alarms). Thus, *sensitivity* quantifies the avoiding of false negatives, as specificity does for false positives. For any test, there is usually a trade-off between these measures. This trade-off can be represented graphically as a receiver operating characteristic, *ROC curve* (Fig. 2).

A perfect predictor would be described as 100% sensitive and 100% specific; but any predictor will possess a minimum error bound (Bayes error rate). The ROC curve is the *sensitivity as a function of fall-out*, i.e. the proportion of non-relevant measures that are retrieved, out of all non-relevant measures available: In general, if the probability distributions for both detection and false alarm are known, the ROC curve can be generated by plotting the cumulative distribution function, i.e. the area under the probability distribution for the discrimination threshold of the detection probability in the y-axis versus the cumulative distribution function of the false-alarm probability in x-axis.

In the medical field, the divisions to assess the test accuracy are: An AUC value ≥ 0.7 is considered acceptable,
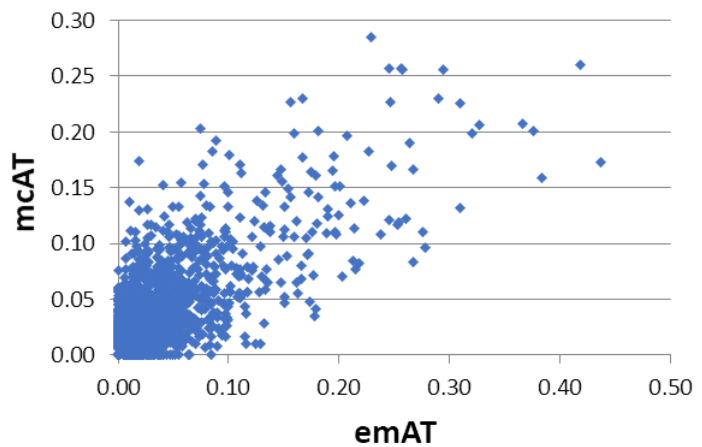

Figure 3. Attention values mcAT as a function of emAT for all pictures.

≥ 0.8 good acceptable, and ≥ 0.9 as *excellent* (Janssen & Laatz, 2013).

Another method used is the assessment of the predictive power of a computer model as a reference value, where the inter-subject variance or inter-subject homogeneity to validate emAT is employed. (Wilming et al., 2011; Stankiewicz, Anderson, & Moore, 2011). For this purpose, the AUC for the prediction of the emAT data for one half of the subjects is determined by the other half of the subjects. The higher the value, the lower the variance – or the higher the homogeneity among emAT subjects. This value is considered to be the theoretically achievable AUC or the upper limit of a computer model for predicting fixations (Stankiewicz et al., 2011; Zhao & Koch, 2011).

To address the *first hypothesis*, i.e. to determine the relationship between the two samples, the correlation coefficient (correlation of measurement of pairs per grid) and the AUC are calculated based on the total stimulus material. A high positive correlation between mcAT and emAT values is observed for both parameters – *contact, attention* – if AUC is significantly above the chance level of 0.5.

To determine the ROC curve using the present emAT and mcAT data it was necessary to clarify which contact or attention value of an emAT sample was interpreted as "seen"; as only the "not seen" and "seen" classes were used for the "seen" calculation.

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

Therefore, *three possible limits* were initially set per parameter and the curves for all three were calculated and compared.

As no reference values were found in the literature for this problem, the limits were set primarily by theoretical considerations on the respective central test limits (at contact: 0.3; in attention: 0.05) and set for calculation of the AUC values of the individual categories used. As the measured values were not normally distributed and thus a prerequisite for calculating the significance tests for correlations was not met (Eid, Gollwitzer, & Schmitt, 2011), the confidence intervals of the correlation coefficients were determined using bootstrapping (for the bootstrapping method see IBM, 2011). Bootstrapping is a resampling technique used to obtain estimates of summary statistics. It can refer to any test or metric that relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy to sample estimates. It allows estimation of the sampling distribution of almost any statistic using random sampling methods.

Fixation/Click rates, scatter plots and difference histograms

Figure 3 shows the attention values of mcAT as a function of emAT for all pictures. Graphically, it demonstrates a close relationship between the two methods

Results

To address the second hypothesis, the two summary measures were calculated and compared for each category of images. In the main "Simple" category in particular, higher compliance – i.e. a higher correlation coefficient and a higher AUC – between mcAT and emAT was postulated than in the adjacent category "complex", as well as in the "human/animal" category

compared to the other categories of the same level. In order to check whether the detected correlation coefficients of the various categories differ significantly from each other, the online calculator used significance testing with correlations suggested by (Wolfgang Lenhard & Alexandra Lenhard, 2014), specifically the test for comparison of two correlation coefficients of independent samples. The average attention value (n = 10) in the emAT test amounted to between 0.00 and 0.44. This means that one single grid received up to 44% of all fixations while an image was being viewed. The highest number of attention value in the mcAT test was 0.28, i.e. one single grid received up to 28% of all clicks. With respect to the contact values, in both experiments all values were between 0, i.e. fields that nobody paid any attention to, and 1, i.e. fields all subjects did notice. The distribution of emAT and mcAT value pairs per grid-box is graphically depicted in Figure 3 by means of a scatter plot.

First Research Question - The ROC curve results

In nearly all image categories the correlation amounted to r = 0.76 (*attention*) and r = 0.71 (*contact*). Both correlations are highly significant and greater than zero (P <0.001). The confidence intervals determined by bootstrapping (.72 to 0.78 (*attention*) and 0.68 to 0.74 (*contact*)), also indicate that, in our sample, the correlation coefficients can be classified as high or very high.

Figure 4 shows the ROC curves for the *entire stimulus material* with three different thresholds. In the attention value for the selected limit of 0.05, the AUC size is 0.88 with a confidence interval of 0.87 to 0.90. In the contact value with the limit 0.3, the AUC size is 0.85 with a confidence interval of 0.84 to 0.87. Both values are thus significantly different from 0.5 (p <0.001).

Table 1: Correlation coefficients of the between-subjects (1st row) compared to the within-subject correlations (2nd and 3rd rows).

| | attention | | contact | | |
	r	AUC	r	AUC	N
emAT-mcAT	0.76	0.88	0.71	0.85	10
emAT1-emAT2	0.68	0.87	0.66	0.82	5
mcAT1-mcAT2	0.66	0.84	0.63	0.81	5

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
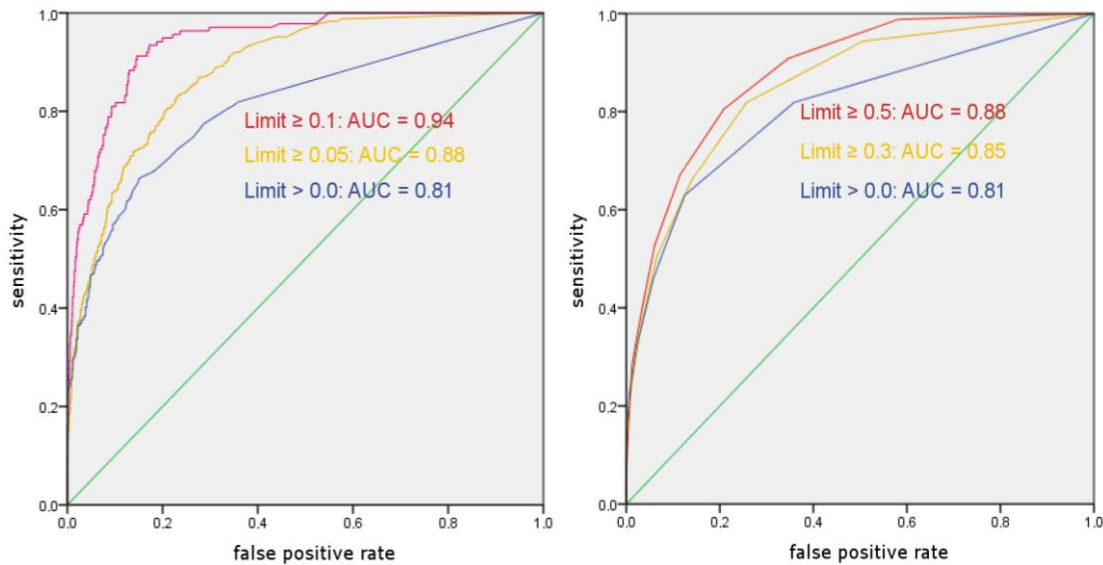Comparison of Mouse-Click with Eye-Movement Attention Tracking

Figure 4. ROC curves with different limits. Left: attention values; right: contact values.

The above described correlation coefficient and the AUC for the subject's internal prediction of fixations is 0.88 (*attention*) and 0.85 (*contact*); the correlation within the ET-sample amounts to r = 0.68 (*attention*) and r = 0.66 (*contact*). This means there is a closer link between the mcAT and emAT data (n = 10) than between the emAT data of the one with the other sub-group (n = 5) (see Tab1).

Inter-subject variance of the emAT data was determined through a within-subject analysis. Correlation coefficients and AUC values were lower than the results

from the between-subject design.

The finding that mouse clicks were more similar to eye fixations than eye fixations to themselves seems hard to understand at first glance. It may be a consequence of the way we generate eye movements. Eye movements are very fast in three respects: We perform many movements per second, eye movements are generated with a short response latency, and saccades are the fastest movements we can generate. This may lead to the effect that eye movements are somewhat inexact, more often than Mouse clicks. We can observe the inaccuracy of eye

Figure 5: Comparison of the viewing (left) and click data (right) image InAS #5 (inanimate, artificial, simple)

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
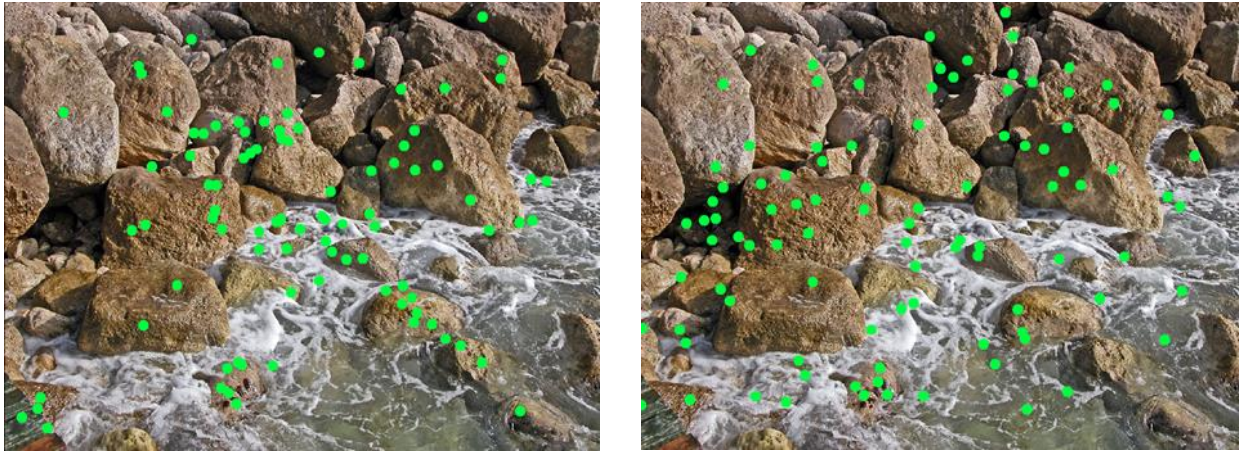Comparison of Mouse-Click with Eye-Movement Attention Tracking

Figure 6. Comparison of the viewing (left) and click data (right) image InNC #8 (inanimate, natural, complex).

movements in any eye-tracking recording: Fixations on a given target are located in an area around the target that is about one degree of visual angle. In comparison, Mouse clicks seem to have a higher accuracy than eye fixations.

Statistically speaking, the inaccuracy of eye movements leads to noise in the recorded data. When we compare eye fixation data with eye fixation data, we compare two noisy sources. In contrast, when we compare gaze data with click data, we compare one more and one less noisy source. This explains, why the comparison of two eye fixation data yields a higher difference than the comparison of eye fixation data with click data.

Second Research Question – the picture categories and emAT vs. mcAT

The correlation coefficients of the individual images demonstrated a large scattering width depending upon the different picture categories (see Fig.1 for reference). The picture with the highest obtained correlation (r = 0.95 (*attention*) and r = 0.94 (*contact*)) was within the category inanimate artificial simple (InAS) (Fig.5).

The picture for both parameters with the lowest correlation (r = 0.22 (*attention*) and r = 0.17 (*contact*)) was one of eight images from the category *inanimate natural complex* (InNC, Fig. 6).

Of the total of 64 images obtained, seven (*attention*) and four (*contact*) images showed a correlation of only r
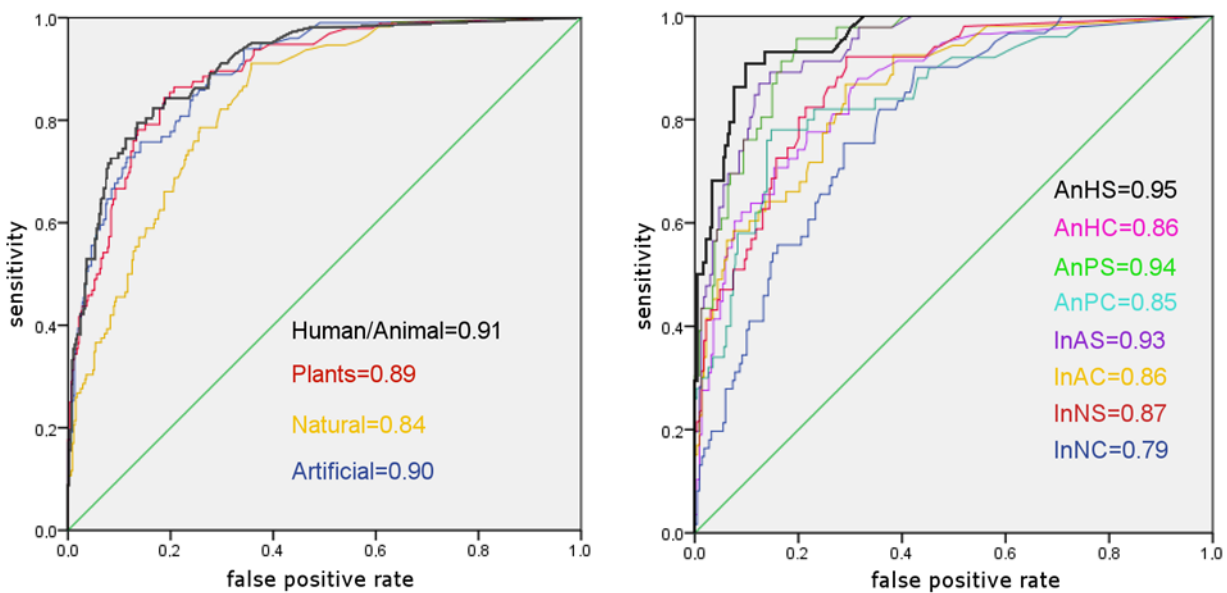

Figure 7. Comparison of the AUC values: Left: Basic categories, attention; right: Super categories, 2nd level, attention

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

<0.5, i.e. a low effect size. On the other hand, six (*attention*) and two images (*contact*) showed a correlation of r > 0.9.

Similar results were obtained in the evaluation of the AUC of the individual super categories represented in Figure 7.

All AUC values differ significantly from chance level with an AUC of 0.5, and all categories are at least in an acceptable, almost good range (Janssen & Laatz, 2013).

But in this statistical analytical method there are also large, significant differences between the super categories. The image category with the highest values was AHS (*animate human simple*) with an AUC of 0.95 (*attention*) or 0.94 (*contact*).

With an AUC of 0.79, INC (*inanimate natural complex*) is the category with the lowest correlation between emAT and mcAT data, and *attention* values are medium (0.5). For *contact*, this applies to the category inanimate artificial complex (IAC) with an AUC of 0.771, similar to the AUC of category INC (0.774).

Generally, attention showed slightly higher values than contact. The "human/animal" category consistently showed higher values than neighboring categories. Most of the test results regarding significance between the different categories were similar for both attention and contact. Both the attention and contact values clearly show the difference between the correlations for the AHS (animated human simple) category and almost all other categories. Both also very clearly differ regarding the "easy" and "complex" image categories. Furthermore, in the "human/animal" super category, the differences with respect to all three categories on the same level were highly significant. The "human/animal" super category also shows a highly significant difference between the "animate" and "inanimate" categories.

We conclude with the observation that the comparable significant similarity of this data demonstrates convincingly the close link and resemblance of the mcAT and emAT methods for searching, recognizing and perceiving the images shown.

Discussion

We investigated the conformity of the mcAT-measurement data (clicks) (n = 10) and the emAT measurement data (fixations) (n = 10). This was based on 64 photographs that were viewed by our subjects. These images were divided into eight categories of our classification scheme. The comparison of click and fixation rates demonstrated that clicks yielded highly similar results to eye movements within our paradigm. In accordance with suggestions in the literature, clicks were on average slightly slower, and occurred with smaller numbers than fixations.

To what extent fixations and clicks match each other, and whether there are differences depending upon the stimulus material?

We found a highly positive and significant correlation between mcAT and emAT data. The AUC values were significantly (p<0.01) above chance level of 0.5 with correlations of r = 0.76 for attention values, and r = 0.71 for contact values. Inter-subject variance of the emAT data was determined through a within-subject analysis. Correlation coefficients as well as AUC-values were below results from the between-subject design. Due to the small number of participants (5 vs. 5 of the within-subject designs compared to 10 vs. 10 of between-subject designs) the variance was high.

This was comparable to other studies with a higher number of subjects: Rajashekar, van der Linde, Bovik and Cormack (2008) reported a larger inter-subject variance with r = 0.75, compared to r = 0.68 (attention) and 0.66 (contact). Their study was based on a sample of emAT with n = 29, and on stimuli excluding images with top-down features.

Interestingly, using a calibration function built by an algorithm (Kasprowski & Harezlak, 2016) it was possible to predict where a user will click with a mouse: The accuracy of the prediction was about 75% − which points to a high correlation as also shown here.

Our second hypothesis was also confirmed: The correlation coefficient and AUC values of individual categories differed from each other. Both the basic eight categories as well as the super-categories showed significant differences for many variations between categories of the same level. As expected, we found that the main "simple" category showed significantly higher correlation values (r = 0.81) for attention when compared to the "complex" category (r = 0.66).

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

The same was true for the "human/animal" super-category (r = 0.82) with respect to the other three categories of the second level "plants" (r = 0.74), "natural" (r=0.74) and "artificial" (r=0.72). This is important evidence for the validity of the mcAT procedure. It has been demonstrated by many researchers that viewers prefer images of humans and animals. The preference for pictures of humans and animals, especially of faces that are simply structured, is also reflected in the eight fundamental categories. These show by far the highest correlations of 0.86 for attention and an excellent AUC of 0.95. Thus, this category is significantly superior to all others. The lowest correlation of mcAT and emAT data is demonstrated by images that represent complex natural structures, but still with a high linear correlation. As print ads and websites often depict people with clearly defined AOIs such as title, motif, slogan, and brand logo, we conclude that the mcAT procedure is highly suitable for measuring attention in this type of stimuli.

It is interesting to note that our results are in line with most previously published reports on our categories: Animate (human/animal plants), Inanimate *(artificial natural)*, as we demonstrate in the following descriptions.

The face recognition system is capable of extremely fine within-category judgments to recognize and discriminate between faces and different facial expressions displayed by the same face (Haxby, Hoffman, & Gobbini, 2000; Kanwisher, 2000). To support this ability, it has been proposed that a separate system evolved to mediate face recognition.

These results indicated the existence of an experience-independent ability for face processing as well as an apparent sensitive period during which a broad but flexible face prototype develops into a concrete one for efficient processing of familiar faces. (Sugita, 2008).

A cortical area selective for visual processing of the human body was described by Downing, Jiang, Shuman and Kanwisher (2001). Despite extensive evidence for regions of human visual cortex that respond selectively to faces, few studies have considered the cortical representa-
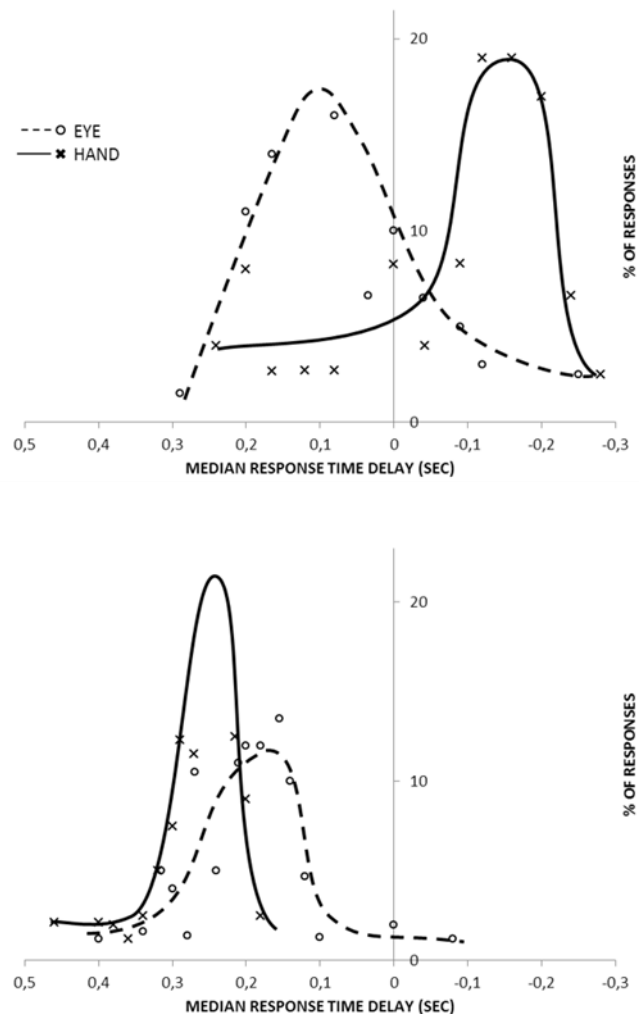

Figure 8: The dependence of median response time (i.e. latency) to frequency of repetitive square-wave patterns (Stark, 1968). Histograms of response time delays for hand and eye. Top: predictive square waves at 1.2 cps. Bottom: random target.

tion of the appearance of the rest of the human body. They presented a series of functional magnetic resonance imaging (fMRI) studies revealing substantial evidence for a distinct cortical region in humans that responds selectively to images of the human body, as compared with a wide range of control stimuli. This region was found in the lateral occipitotemporal cortex in all subjects tested (n = 19) and apparently reflected a specialized neural system for the visual perception of the human body.

Several lines of evidence suggest that animate entities have a privileged processing status over inanimate objects – in other words, that animates have priority over

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

inanimates. The animate/inanimate distinction parallels the distinction between "living" and "nonliving" things that has been postulated to account for selective deficits in patients (for a review, see Capitani, Laiacona, Mahon, & Caramazza, 2003). Animates belong to the general category of living things.

Their studies revealed better recall for words denoting animate than inanimate items, which was also true with the use of pictures. The findings provided further evidence for the functionalist view of memory championed by Nairne and co-workers (Nairne & Pandeirada, 2010; Nairne, 2010).

Evidence from neuropsychology suggests that the distinction between animate and inanimate kinds is fundamental to human cognition. Previous neuroimaging studies have reported that viewing animate objects activates ventrolateral visual brain regions, whereas inanimate objects activate ventromedial regions.

Comparison of eye and hand movements from a neuro-bioengineering perspective

A saccade made to a target that appears eccentric to the point of fixation is sometimes called a 'reflexive' (or 'stimulus-elicited') saccade in contrast to those made in situations that depend more heavily upon voluntary (or 'endogenous') cognitive control processes (for example when directed by a simple instruction "look to the left"). Most saccades are essentially voluntary in nature, as an observer can always decide not to move the eyes. Also, if the time and place of a target's appearance can be predicted, an anticipatory saccade often occurs before the target itself appears, or too briefly subsequently for visual guidance to have occurred.

When reaching for targets presented in peripheral vision, the eyes generally begin moving before the hand (Bekkering, Adam, van den Aarssen, Kingma, & Whiting, 1995; Biguer, Jeannerod, & Prablanc, 1982; Jeannerod, 1988; Johansson, Westling, Bäckström, & Flanagan, 2001; Land, Mennie, & Rusted, 1999; Prablanc, Echallier, Komilis, & Jeannerod, 1979). This is the case because much of the delay in hand movement onset, relative to eye movement onset, can be attributed to the greater inertia of the arm. Recent studies demonstrate that the motor commands underlying coordinated eye and hand movements appear to be issued in close temporal proximity and that commands for hand movement may even precede those for eye movement (Biguer et al., 1982; Gribble, Everling, Ford, & Mattar, 2002; Sailer, Flanagan, & Johansson, 2005).

Furthermore, hand movement can influence saccadic initiation. Saccadic reaction time (SRT) is greater when eye movement is accompanied by hand movement compared to when the eyes move alone (Mather & Fisk, 1985; Navas & Stark, 1968), and SRT and hand reaction time (HRT) both increase when reaching for targets in contralateral versus ipsilateral space (Fisk & Goodale, 1985). In addition, in eye-hand coordination saccades are faster when accompanied by a coordinated arm movement (Snyder, Calton, Dickinson, & Lawrence, 2002). Because of the large variation in response characteristics of both, hand and eye systems with different subjects, both systems have been compared by measuring the systems´ responses simultaneously (Stark, 1968) (Navas & Stark, 1968). When eye and hand responses were recorded simultaneously to random steps and predictive regular steps, the eye shows shorter response times than those of the hand (Fig.8).

The eye muscles have considerable power with respect to their constant load, the eyeball, and show faster rise times than the hand, especially when tracking rapidly alternating signals. With random targets, the hand response lags behind the eye response due to the eye's considerably smaller load. At moderate frequencies of 0.7 to 1.0 cps the hand develops prediction faster and to a greater extent than the eye. At higher frequencies of >1.1 cps the hand shows considerable prediction, while the median eye response time starts to lag despite the higher frequency characteristics of the actual movement dynamics of the eye. In general, at low frequencies there is some correlation evident between eye and hand response.

Obviously, eye movement is not necessary for hand movement. Conversely, the physical movement of the hand appears to help the eye movement system: Adequate eye tracking may occur with comparatively high frequencies, if the hand is tracking and may not occur if the hand is still. When hand tracking improved eye performance and then stopped, the eye performance deteriorated significantly (Stark, 1968; Navas & Stark, 1968).

With respect to our paradigm, this means that the mcAT method might help the sequence of eye fixations that must go along with the mouse clicks when viewing and perceiving the test images. This is particularly true in

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

test settings where sensory-motor actions tend to be time-optimal due to limited time: This was the case in our test set with the time limit of 5 sec.

Interestingly, Bednarik, Gowases, & Tukiainen (2009) showed that users with gaze-augmented interaction outperformed two other groups – using mouse or dwell time interaction – on several problem-solving measures: they committed fewer errors, were more immersed, and had a better user experience. As mentioned, the slower manual action of the hand/finger-movement when activating the mouse may well be due to the physiological properties rather than solely on to the hidden cognitive processes of augmented gaze in problem solving as the authors speculate. A combined eye and mouse interaction may show an even more successful result, since hand and eye movements in coordination could act faster and more precisely in many situations (Navas & Stark, 1968).

In continuation of the early results by Stark (1968), Helsen, Elliott, Starkes, and Ricker (2000) studied eye and hand movement coordination in goal directed movements. They found a remarkable temporal relationship between the arrival of the eye and the hand on the target position. Because point of gaze always arrives on the target area first, at roughly 50% of the hand response time, there is ample opportunity for the eye to provide extra-retinal information on target position either through oculomotor efference or visual and proprioceptive feedback resulting from the large primary saccade. They demonstrated an invariant relationship between the spatial-temporal characteristics of eye movements and hand kinematics in goal-directed movement that is optimal for the pickup of visual feedback. Interestingly, the natural coupling between eye and hand movements remains intact even when hand movements are merely imagined as opposed to being physically executed. So, it appears that the bioengineering and neurophysiological literature shows indeed a solid background for the highly correlated relationship of eye and mouse movements firstly described by Egner and Scheier in 2002.

Bieg, Chuang, Fleming, Reiterer and Bülthoff (2010) showed, when target location was unknown (quasi random), the eyes lead the mouse by 300 ms on average. When the approximate location of the target was known (i.e. predictive), the cursor often led gaze in acquiring the target, and fixations on the target occurred later in the pointing process. This again corresponds to the early results of Stark and Navas inasmuch the degree of prediction of a target influences the result.

Knowledge about the target location is likely to be very important especially in non-laboratory settings. This was shown by Liebling and Dumais (2014) presented an in-situ study of gaze and mouse coordination as participants went about their normal activities before and after clicks. They analyzed the coordination between gaze and mouse, showing that gaze often leads the mouse, about two thirds of the time, and that this depends on type of target and familiarity with the application; but not as much as previously reported, and in ways that depend on the type of target.

Rodden, Fu, Aula and Spiro (2008) tracked subjects´ eye and mouse movements and described three different types of eye-mouse coordination patterns. However, they found that the users were not easy to classify: each one seemed to exhibit several patterns of the three types to varying degrees. There was also substantial variation between users in high-level measures: The mean distance between eye and mouse ranged from 144 to 456 pixels, and the proportion of mouse data points at which the eye and mouse were in the same region ranged from 25.8% to 59.9%. This corresponds with Huang, White and Dumais´ (2011) results. During Web search, Huang et al. found that eye and mouse most often were correlated. The average distance between the eye and mouse was 178 px, with the differences in the x-direction being much larger (50 px) than in the y-direction (7 px).

Later, Chen et al. (2001) also reported that in web browsing during certain subtasks, mouse and gaze movements were very often correlated. They found that the average distance between mouse and gaze was 90 pixels during transitions from one area of interest (AOI) to another, and that 40% of the distances were closer than 35 pixels.

In summary, as gaze provides a measure of attention, knowing when the mouse and gaze are aligned i.e. highly correlated, this confirms the usability of the mouse as indicator of attention as shown by Egner and Scheier (2002).

Limitations of emAT and mcAT

Both methods used in this study, emAT and mcAT, were geared towards determining the respondents' current attention location. However, as attention is an inter-

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

Table 2: Comparison between the two attention measurements used in the present article. See also Appendix 3.

SOURCE OF NOISE	emAT	mcAT
— PHYSIOLOGICAL NOISE SOURCES —		
Motor response lags behind attention	Leads to a consistent lag in measured data	Ditto. Lag is even bigger.
Inaccuracy of motor responses	Generates noise through misplaced fixations and correction saccades	Analogously
Blinks	Ca. 1 blink per second; noise can often be removed, automatically	-
Mis-clicks	-	Rapid clicking behavior leads to erroneous clicks; noise
— TECHNICAL NOISE SOURCES —		
Fixation detection	Arbitrary parameter settings; noise through misclassified fixations	-
Unreliable position measurement	Eye position cannot be detected reliably; small deviations in eye position create big deviations in gaze position	-
Sophisticated hardware	Eye-tracking hardware susceptible to external noise	-
— BEHAVIORAL NOISE SOURCES —		
Clicking enforcement	-	Respondents are trained to rapidly click where they look. Since targeted clicking involves attention, the natural scan path may be altered by the clicking enforcement.
Voluntary control of motor responses	Eye movements can be controlled voluntarily, but this can probably be neglected.	Mouse behavior is under voluntary control. This may affect the natural scan path.

nal process in the brain, we are measuring external responses (gaze direction, mouse location) and use the measured data to infer the actual attention location. These external responses as well as their measurements are subject to noise (Tab.2) (Appendix 2).

Also, past research has found a correlation between gaze and cursor positions (Chen et al., 2001; Cooke, 2006; Guo & Agichtein, 2008; Huang et al., 2011; Rodden et al., 2008) and that cursor movements can be useful for determining relevant parts of the Web page with vary-ing degrees of success (Goecks & Shavlik, 2000; Hijika-ta, 2004; Huang et al., 2011; Shapira, Taieb-Maimon, & Moskowitz, 2006). Cursor interaction spans a variety of behaviors including reading, hesitating, highlighting, marking, and actions such as scrolling and clicking. Huang et al. (2011) showed that user, time, and search task (to a lesser extent) each contribute to the variation in gaze-cursor alignment. The gaze and cursor positions are also better aligned when the gaze position is compared to a future cursor position. Furthermore, by distinguishing

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

between five different cursor behaviors—inactive, examining, reading, action, and click—one might get a better idea of the strength of alignment. Identifying these behaviors was beyond the scope of this paper. For further discussion of these problems see Huang et al. (2011).

Demšar and Çöltekin (2017) investigated the actual connection between the eye and the mouse when searching web pages. They found that there seemed to be natural coupling when eyes were not under conscious control, but that this coupling breaks down when instructed to move them intentionally. Therefore, they suggested that for natural tracing tasks, mouse tracking could potentially provide similar information as eye-tracking and so be used as a proxy for attention. Since our paradigm used a clear-cut task that asked our subjects to consciously coordinate eye and mouse movements, this aspect of Demšar and Çöltekin appears not to be relevant in our context.

Starting points for further evaluations and investigations

Based on the results and limitations of this study, and the explanations of the theoretical part of this work, recommendations for further research can be derived that could complete the results of this work and lead to a broader assessment of the contribution and validity of mcAT. Firstly, it describes aspects that are already present in the data that can be evaluated in secondary analyzes. A further method would be *the analysis of spatial correspondence of fixations and clicks*, which would provide an extension to the viewing or clicking motion paths. The question of whether clicking may be influenced by short-term memory in relation to the presented stimuli cannot be answered here in the context of the evaluated tests but should be studied. The evaluation of the qualitative survey could also shed some light on possible gaze and click strategies with mcAT. Furthermore, the influence of the number and size of the grids on the images varies and should be checked systematically to determine whether the effects found in this study can be confirmed.

Another possibility would be to select from the present stimulus material only those images that are semantically oriented instead of using grids. AOIs could be defined in a next step and evaluation methods for these new AOIs then repeated. For checking the coincidence of time of fixation and clicks, other parameters such as the viewing time and time to contact for both could be calculated and compared. Another question is how precise are spatial mouse clicks set? The variation of the presentation duration for certain looking/click strategies of the subjects would enable a time-based evaluation in which the extent to which the match of gaze and click data is also dependent upon the presentation duration could be checked.

Conclusion

The aim of this work was a fundamental review of the suitability of mcAT to measure the attention of participants by registering clicks on the computer screen. The validation was carried out under a between-subject design based on eye tracking (emAT), the standard method of attention measurement. The focus of the investigation was the analysis of the spatial relationship of the measured data of both methods. With respect to findings reported in the literature, it was assumed that, in general, a close relationship between gaze and click data exists. However, the extent of the relationship varies for different image categories. Both hypotheses were confirmed. As eye tracking (emAT) is predominantly accepted as the valid method for measuring attention, we can conclude that mouse-click tracking (mcAT) is similarly highly valid.

A further finding of our research was that this innovative method obtains particularly valid results with stimuli that are simply structured and where humans or animals are shown. This suggests that the use of mcAT for attention measurement is well suited for print ads, and thus for advertising research a valid alternative to eye tracking, with benefits regarding practicability. Due to some emAT's restrictions, we suggest that, in other fields, mcAT could replace the registration of eye movements in cases where eye tracking may be inaccurate or technically unfeasible and provides a promising additional method for usability research (Groner, Raess, & Sury, 2008)

Ethics and Conflict of Interest

The author(s) declare(s) that the contents of the article are in agreement with the ethics described in http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.ht

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

ml and that there is no conflict of interest regarding the publication of this paper.

References

Allport, A. (1987). Selection-for-action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer (Ed.), *Perspectives on perception and action* (pp. 395–420). Hillsdale NJ u.a.: Erlbaum.

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, *15*(5), 9. https://doi.org/10.1167/15.5.9

Bednarik, R., Gowases, T., & Tukiainen, M. (2009). Gaze interaction enhances problem solving: Effects of dwell-time based, gaze-augmented, and mouse interaction on problem-solving strategies and user experience. *Journal of Eye Movement Research, 3(1)*. http://dx.doi.org/10.16910/jemr.3.1.3

Bekkering, H., Adam, J. J., van den Aarssen, A., Kingma, H., & Whiting, H. T. (1995). Interference between saccadic eye and goal-directed hand movements. *Experimental Brain Research*, *106*(3), 475–484.

Berger, S. (2009). *Spotlight-Viewer. Das Aufzeichnen von manuellen Zeigebewegungen als neues Verfahren zur impliziten Messung der Effektivität von Werbeanzeigen* (Dissertation). Universität Wien, Wien.

Bieg, H.-J., Chuang, L. L., Fleming, R. W., Reiterer, H., Bülthoff, H. H. (2010). Eye and Pointer Coordination in Search and Selection Tasks. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*, 89–92. http://dx.doi.org/10.1145/1743666.1743688.

Biguer, B., Jeannerod, M., & Prablanc, C. (1982). The coordination of eye, head, and arm movements during reaching at a single visual target. *Experimental Brain Research*, *46*(2), 301–304.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 185–207. https://doi.org/10.1109/TPAMI.2012.89

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, *22*(1), 55–69. https://doi.org/10.1109/TIP.2012.2210727

Cadek, V. (2015). Why are we predicting a probability rather than a binary response? Retrieved from https://www.kaggle.com/c/GiveMeSomeCredit/forums/t/14823/why-are-we-predicting-a-probability-rather-than-a-binary-response/100798

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive neuropsychology, 20(3/4/5/6),* 213–261.

Chen, M. C., Anderson, J. R., & Sohn, M. H. (2001). What can a mouse cursor tell us more? correlation of eye/mouse movements on web browsing. In *CHI EA, ACM*, 281-282.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Hillsdale: Erlbaum*.

Cooke, L. (2006). Is the mouse a poor man's eye tracker? In: *Proceedings of STC*.

Demšar, U., & Çöltekin, A. (2017) Quantifying gaze and mouse interactions on spatial visual interfaces with a new movement analytics methodology. *PLoS ONE 12(8): e0181818*. https://doi.org/10.1371/journal.pone.0181818

Deng, J., Krause, J., & Fei-Fei, L. (2013). Fine-grained crowdsourcing for fine-grained recognition. In *CVPR*, 580-587.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, *26*(6), 973–990.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science (New York, N.Y.),*

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

293(5539), 2470–2473.
https://doi.org/10.1126/science.1063414

Dubois, B., & Pillon, B. (1997). Cognitive deficits in Parkinson's disease. *Journal of neurology*, *244*(1), 2–8.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice* (2. ed.). London: Springer. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10202868

Egner, S., Itti, L., & Scheier, C. R. (2000). Comparing attention models with different types of behavior data, In: *Investigative Ophthalmology and Visual Science (Proc. ARVO 2000), Vol. 41, No. 4*, p. S39.

Egner, S., & Scheier, C. R. (2002). Erfassung der Kundenwirkung von Bildmaterial. *Künstliche Intelligenz.* (1), 84–86. German.

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden* (1. Ed). Weinheim u.a., Beltz.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fischer, B. (1999). *Blick-Punkte: Neurobiologische Grundlagen des Sehens und der Blicksteuerung*. Bern u.a.: Huber.

Fisk, J. D., & Goodale, M. A. (1985). The organization of eye and limb movements during unrestricted reaching to targets in contralateral and ipsilateral visual space. *Experimental brain research*, *60*(1), 159–178.

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual cognition*, *14*(4-8), 704–715. https://doi.org/10.1080/13506280500193719

Goecks, J., & Shavlik, J. (2000). Learning users' interests by unobtrusively observing their normal behavior. In *Proceedings of IUI*, 129–132.

Gosselin, F., & Schyns, P. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *VR 41, 15,* 2261-2271.

Granka, L., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in www search. In *Proceedings of the 27th annual international ACM SIGIR*

conference on Research and development in information retrieval, pages 478-479.

Gribble, P. L., Everling, S., Ford, K., & Mattar, A. (2002). Hand-eye coordination for rapid pointing movements: Arm movement direction and distance are specified prior to saccade onset. *Experimental brain research*, *145*(3), 372–382. https://doi.org/10.1007/s00221-002-1122-9

Groner, R. (1988). Eye movements, attention and visual information processing: some experimental results and methodological consideration. In G. Lüer, U. Lass & J. Shallo-Hoffmann (Eds.), *Eye movement research, physiological and psychological aspects*. Göttingen: Hogrefe.

Groner, R., & Groner, M. (1989). Attention and eye movement control: an overview. *European Archives of Psychiatry and Neurological Sciences*, 239, 9-16.

Groner, R., & Groner, M. T. (2000). The issue of control in sensory and perceptual processes: attention selects and modulates the visual input. In W.J. Perrig & A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness*, pp. 125-135. Mahwah, N.J.: Lawrence. Erlbaum.

Groner, R., Raess, R., & Sury, P. (2008). Usability: Gestaltung und Optimierung von Benutzerschnittstellen. In B. Batinic (Hrsg.), *Lehrbuch Medienpsychologie, Kapitel 18*. Berlin: Springer.

Guo, Q., & Agichtein, E. (2008). Exploring mouse movements for inferring query intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval,SIGIR '08*, pages 707-708, New York, NY, USA.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223-233.

Hein, O., & Zangemeister, W. H. (2017). Topology for gaze analyses - Raw data segmentation. *Journal of Eye Movement Research, 10*(1). http://dx.doi.org/10.16910/jemr.10.1.1.

Helsen, W. F., Elliott, D., Starkes, J. L., & Ricker, K. L. (2000). Coupling of eye, finger, elbow, and shoulder movements during manual aiming. *Journal of Motor Behavior, 32*, 241-248

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271. https://doi.org/10.1146/annurev.psych.50.1.243

Hijikata, Y. (2004). Implicit user profiling for on demand relevance feedback. In Proceedings of IUI,198–205.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures* (First edition). Oxford, New York, Auckland: Oxford University Press.

Huang, J., White, R. W., & Dumais, S. (2011). No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of CHI*, 1225–1234.

IBM Corporation. (2011). IBM SPSS Bootstrapping 20. Retrieved from ftp://public.dhe.ibm.com/software/analytics/spss/documenttion/statistics/20.0/de/client/Manuals/IBM_SPSS _Bootstrapping.pdf

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research. In: *The Mind's Eye. Elsevier*, pp. 573-605.

James, W. (1890). *The principles of psychology* (1. Ed). New York: Dover Publications.

Janssen, J., & Laatz, W. (2013). *Statistische Datenanalyse mit SPSS: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests* (8. Aufl.). Berlin u.a.: Springer Gabler.

Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements. Oxford psychology series: Vol. 15*. Oxford: Oxford Univ. Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0638/8702 4669-d.html

Jiang, M., Huang, S., Duan, J., & Zhao, Q. (2015). SAL-ICON: Saliency in Context. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1072–1080). https://doi.org/10.1109/CVPR.2015.7298710

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *Journal of neuroscience*, *21*(17), 6917-6932.

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *IEEE 12th International Conference on Computer Vision, 2009. Kyoto, Japan, 29 September - 2 October 2009* (pp.2106–2113). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICCV.2009.5459462

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763.

Kasprowski, P., & Harezlak, K. (2016). Implicit Calibration Using Predicted Gaze Targets. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*. ACM, New York, NY, USA, 245-248.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual review of neuroscience*, *23*, 315–341. https://doi.org/10.1146/annurev.neuro.23.1.315

Kim, N. W., Bylinskii, Z., Borkin, M., Oliva, A., Gajos, K. Z., & Pfister, H. (2015). A Crowdsourced Alternative to Eye-tracking for Visualization Understanding. *HI'15 Extended Abstracts, April 18-23, 2015, Seoul, Republic of Korea. ACM 978-1-4503-3146-3/15/04*. http://dx.doi.org/10.1145/2702613.2732934

Kootstra, G., de Boer, B., & Schomaker, L. R. B. (2011). Predicting Eye Fixations on Complex Visual Stimuli Using Local Symmetry. *Cognitive computation*, *3*(1), 223–240. https://doi.org/10.1007/s12559-010-9089-5

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916.

Lacquaniti, F., & Soechting, J. F. (1982). Coordination of arm and wrist motion during a reaching task. *Journal of Neuroscience 2 (4),* 399-408.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. https://doi.org/10.1068/p2935

Lenhard, W., & Lenhard, A. (2014). *Testing the Significance of Correlations*. Retrieved from http://www.psychometrica.de/korrelation.html

Liebling, D. J., & Dumais, S. T. (2014). Microsoft Research, One Microsoft Way, Redmond, WA 98052 USA. Gaze and Mouse Coordination in Everyday Work, UbiComp '14, September 13 - 17 2014, Seattle,

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

WA, USA,
http://dx.doi.org/10.1145/2638728.2641692)

Mather, J. A., & Fisk, J. D. (1985). Orienting to targets by looking and pointing: Parallels and interactions in ocular and manual performance. *The Quarterly Journal of Experimental Psychology Section A*, *37*(3), 315–338. https://doi.org/10.1080/14640748508400938.

Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology*, *50*(4), 381–425.

Nairne, J. S., & Pandeirada, J. N. S. (2010). Adaptive memory: ancestral priorities and the mnemonic value of survival processing. *Cognitive psychology*, *61*(1), 1–22. https://doi.org/10.1016/j.cogpsych.2010.01.005

Nairne, J. S. (2010). Adaptive memory: Evolutionary constraints on remembering. In B. H. Ross (Ed.), *The psychology of learning and motivation. Volume 53* (1st ed., pp. 1–32). London: Academic Press.

Navalpakkam, V., Jentzsch, L., Sayres, R., Ravi, S., Ahmed, A., & Smola, A. (2013). Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. *Www 2013 - Proceedings of the 22nd International Conference On World Wide Web*. 953-963.

Navas, F., & Stark, L. (1968). Sampling or Intermittency in Hand Control System Dynamics. *Biophysical Journal*, *8*(2), 252–302. https://doi.org/10.1016/S0006-3495(68)86488-4

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer (Ed.), *Perspectives on perception and action* (pp.361–394). Hillsdale NJ, Erlbaum.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive psychology*, *7*(1), 44–64. https://doi.org/10.1016/0010-0285(75)90004-3

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science (New York, N.Y.)*, *171*(3968), 308–311.

Oyekoya, O., & Stentiford, F. (2004). Exploring human eye behaviour using a model of visual attention. In J. Kittler (Ed.), *Proceedings of the 17th International Conference on Pattern Recognition. August 23 - 26,* *2004, Cambridge, UK* (945-948 Vol.4). Los Alamitos, Calif.: IEEE Computer Soc. https://doi.org/10.1109/ICPR.2004.1333929

Palmer, J. (1995). Attention in Visual Search: Distinguishing Four Causes of a Set-Size Effect. *Current Directions in Psychological Science*, *4*(4), 118–123. https://doi.org/10.1111/1467-8721.ep10772534

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

Pieters, R., Wedel, M., & Batra, R. (2010). The Stopping Power of Advertising: Measures and Effects of Visual Complexity. *Journal of Marketing*, *74*(5), 48–60. https://doi.org/10.1509/jmkg.74.5.48

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology*, *109*(2), 160–174.

Prablanc, C., Echallier, J. F., Komilis, E., & Jeannerod, M. (1979). Optimal response of eye and hand motor systems in pointing at a visual target.: I. Spatio-temporal characteristics of eye and hand movements and their relationships when varying the amount of visual information. *Biological cybernetics*, *35*(2), 113–124.

Privitera, C. M., & Stark, L. W. (1998). Evaluating image processing algorithms that predict regions of interest. *Pattern Recognition Letters*, *19*(11), 1037–1043. https://doi.org/10.1016/S0167-8655(98)00077-4

Rajashekar, U., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2008). GAFFE: a gaze-attentive fixation finding engine. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, *17*(4), 564–573. https://doi.org/10.1109/TIP.2008.917218

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological review*, *93*(2), 180–206.

Riche, N., Duvinage, M., Mancas, M., Gosselin, B., & Dutoit, T. (2013). Saliency and human fixations: State of the art and study of comparison metrics. *Computer Vision (ICCV), IEEE, Internatl. Conf. on IEEE*.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premo-

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

tor theory of attention. *Neuropsychologia*, *25*(1A), 31–40.

Rizzolatti, G. (1983). Mechanisms of Selective Attention in Mammals. In J.-P. Ewert, R. R. Capranica, & D. J. Ingle (Eds.), *Advances in Vertebrate Neuroethology* (pp. 261–297). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-4412-4_12

Rodden, K., Fu, X., Aula, A., & Spiro, I. (2008). Eye-mouse coordination patterns on web search results pages. In *CHI EA, ACM*, 2997-3002.

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *The Journal of neuroscience*, *25*(39), 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.

Scheier, C. R., & Egner, S. (2003a). Apparatus and method for examination of images. *U.S. Patent No. 7,512,289. Washington, DC: U.S. Patent and Trademark Office.*

Scheier, C. R., & Egner, S. (2003b), Vorrichtung und Verfahren zum Untersuchen von Bildern. *EU-Patent No. 3012738, Munich, Germany: European Patent Office.*

Scheier, C. R., & Egner, S. (2005). Beobachten statt Fragen- Internetgestützte Verhaltensmessung mit Tracking. *Planung & Analyse*. (1), 1–7.

Scheier, C. R., Reigber, D., & Egner, S. (2003). Messen der Aufmerksamkeit bei Internet-Nutzern — Ansatz und Einsatz eines neuen Verfahrens zur Online-Messung von Aufmerksamkeit. In A. Theobald, M. Dreyer, & T. Starsetzki (Eds.), *Online-Marktforschung. Theoretische Grundlagen und praktische Erfahrungen* (2nd ed., pp.309–324). Wiesbaden, Gabler Verlag. https://doi.org/10.1007/978-3-663-10948-8_21

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: a selective review. *Journal of Vision*, *11*(5), 1–30. https://doi.org/10.1167/11.5.9

Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive psychology*, *14*(3), 353–409. https://doi.org/10.1016/0010-0285(82)90014-7

Shaw, M. L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 109–121). London u.a.: Erlbaum.

Sheliga, B. M., Craighero, L., Riggio, L., & Rizzolatti, G. (1997). Effects of spatial attention on directional manual and ocular responses. *Experimental brain research*, *114*(2), 339–351.

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental brain research*, *98*(3), 507–522.

Shapira, B., Taieb-Maimon, M., & Moskowitz, A. (2006). Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests. In *Proceedings of SAC,* 1118–1119.

Snyder, L. H., Calton, J. L., Dickinson, A. R., & Lawrence, B. M. (2002). Eye-hand coordination: Saccades are faster when accompanied by a coordinated arm movement. *Journal of neurophysiology*, *87*(5), 2279–2286. https://doi.org/10.1152/jn.00854.2001

Stankiewicz, B. J., Anderson, N. J., & Moore, R. J. (2011). Using performance efficiency for testing and optimization of visual attention models. In S. P. Farnand & F. Gaykema (Eds.): *SPIE Proceedings.* International Society for Optics and Photonics. https://doi.org/10.1117/12.872388

Stark, L. L. (1968). *Neurological Control Systems: Studies in Bioengineering* (1. ed.): Plenum Press.

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(1), 394–398. https://doi.org/10.1073/pnas.0706079105

Umiltà, C., Riggio, L., Dascola, I., & Rizzolatti, G. (1991). Differential effects of central and peripheral cues on the reorienting of spatial attention. *European Journal of Cognitive Psychology*, *3*(2), 247–267. https://doi.org/10.1080/09541449108406228

Wilming, N., Betz, T., Kietzmann, T. C., & König, P. (2011). Measures and limits of models of fixation selection. *PloS one*, *6*(9), 1-19. https://doi.org/10.1371/journal.pone.0024038

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature*, *406*(6797), 691. https://doi.org/10.1038/35021132

Yarbus, A. L. (1961). Eye movements during the examination of complicated objects. *Biofizika, 6*(2), 52–56.

Yarbus, A. L. (1967). Eye Movements and Vision. Plenum Press, New York. http://dx.doi.org/10.1007/978-1-4899-5379-7

Zhao, Q., & Koch, C. (2011). Learning a saliency map using fixated locations in natural scenes. *Journal of Vision*, *11*(3), 1–15. https://doi.org/10.1167/11.3.9

Appendix 1

Utilized stimulus material

ANIMATE
    human/animal
        simple
        complex
    plants
        simple
        complex
INANIMATE
    artificial
        simple
        complex
    natural
        simple
        complex

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

Appendix 2

Instructions for mcAT trial

Follow your viewing of the image with the computer mouse, so that the fixation points are transferred into mouse clicks. In order to optimize your eye-hand coordination, get a feel for the necessary click speed; initially you will have to complete a short click training. The first step is done when you have completed the training successfully. Now the real test starts, in which you have to follow your viewing process for 5 seconds for 64 images, each preceded and followed by a separating fixation cross. Throughout the trial period try to keep your click speed constant with a minimum frequency of at least 1-2 clicks per second. Note that it is not possible to stop the click test in between. Also, try to focus and concentrate fully until to the end of the test (duration: 7.5 minutes). At the end of the click test there will follow a brief survey of the pictures that you have just viewed. Tell us if you have seen the following 16 images mentioned in the test or not: 16 images (1 per page; 8 true, 8 wrong; 1 from each category). This will allow us to evaluate the technical quality of your click performance.

Appendix 3

Sources of noise

Sources of noise in classical eye-tracking (emAT):

- Gaze direction is not always identical to attention direction (Anderson et al., 2015; Chun & Wolfe, 2005). It can be assumed that, under normal viewing conditions, the gaze is preceded by attention. This introduced a delay in our measured data.

- When the gaze follows the attention direction, it typically performs saccades that target the current attention direction. However, saccades are not always precise. When a saccade misses the actual attention direction, an additional saccade (correction saccade) is performed. In such a case, emAT measures multiple saccades and fixations while only one attentional shift has been performed. The additional saccades and fixations are not corresponding to attention; they are noise. Of course, saccades are possible that represent a strategic under-shoot or over-shoot.

- Blinks are an additional source of noise. However, we attempted to remove all blinks from the emAT data in this study.

- An underlying assumption of emAT is that fixations directly correspond to attention directions. However, it is often hard to classify data from the eye-tracker into fixations. Depending upon the choice of parameters of the fixation detection algorithm (temporal, spatial), it classifies differing portions of the trajectory as fixation. There is no optimal parameter regime; therefore, there will always be some mis-detected fixations, which can be seen as noise.

- The measurement of gaze direction is technically demanding, because small differences in the eye position correspond to large differences in the gaze position. Therefore, the measurement has to have a high accuracy on the raw data level to avoid big mistakes on the level of gaze positions. Head movements in all directions create additional difficulty. Altogether, the technical difficulties lead to more or less noisy data.

Analogously to the physiological and technical sources of noise in emAT, mcAT must deal with both kinds of noise sources:

- The hand is much slower than the eye. It can, therefore, be assumed that the lag between the attention location and the measured position is even bigger than in emAT.

- Mouse movements are also not precise.

Journal of Eye Movement Research
11.6.4

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W.H. (2018)
Comparison of Mouse-Click with Eye-Movement Attention Tracking

Stimulus material: There remains a certain arbitrariness in the derivation and definition of image categories. In addition, in reality mixed forms of categories such as simultaneous displays of plants and animals usually occur.

Sample: Although both samples are similar in their demographic structure, subjects consisted exclusively of students aged between 18 and 26 years. This accounts for a small population segment only. One of the advantages of the mcAT procedure is that it allows for a large population and many different audiences to perform the mcAT.

Statistical analysis: As determination of semantically-oriented AOIs involves certain disadvantages in terms of the stimulus material used, the division of the images in grids was the better alternative. However, the optimal number of fields per frame could not be determined. As each emAT and mcAT data pair per AOI counted as one case in the analysis, a large or too large field or case number might mean small, essentially insignificant effects get rated as significant. For this reason, we calculated and assessed effects across both linked measures (correlation and AUC) and parameters (attention and contact). To calculate the significance between correlation values of the individual picture category, the online calculator (ref: http://www.socscistatistics.com/tests/pearson/) for comparison of two correlation coefficients was used. The emAT and mcAT metrics are independent in terms of subjects and methods of measurement, but they are not independent with respect to the stimulus material: i.e. the stimuli are the same for both maneuvers, but different in time and dynamics.