Philipps Universität Marburg

Measuring gaze and pupil in the real world: object-based attention, 3D eye tracking and applications

Dissertation
zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Physik
der Philipps-Universität Marburg

vorgelegt von

Josef Stoll

aus
Waldshut-Tiengen, Deutschland

Marburg/Lahn, 2015

.

Vom Fachbereich Physik der Philipps-Universität als Dissertation

angenommen am 6. Februar 2015

Erstgutachter: Wolfgang Einhäuser

Zweitgutachter: Wolfgang Parak

Tag der mündlichen Prüfung: 23. April 2015

Hochschulkennziffer 1180

Supervisor of the Dissertation: Prof. Dr. Wolfgang Einhäuser

Processed at AG Neurophysik, FB Physik, Karl-von-Frisch-Str. 8a, Lahnberge.

Part of this cumulative thesis is published. For detailed information the publications' author contributions, see 1.3.6.

Vorwort

Diese kumulative Dissertation setzt sich durch eine Mischung von Problemlösungen in unterschiedlichen Fachgebieten zusammen. Die wissenschaftlichen Fragestellungen sind grundlegend durch Kenntnisse und Ideen aus der Kognitions-Psychologie motiviert. Daraus leiteten sich Aufgabenstellungen zur Entwicklung von Messmethoden des Bereichs Psycho-Physik ab – messtechnische Aspekte und Modellierung begeisterten mich als Physiker stets mehr, als das Wagnis einer mir bis dato überwiegend fach-fremden Schwerpunktlegung in der Psychologie, zudem diese mir bei der Begutachtung am Fachbereich Physik hätte hinderlich sein können. Dennoch lernte ich mit viel Begeisterung psychologische Methoden schätzen und die umfangreichen Facetten der Wahrnehmungspsychologie kennen. Sehr beeindruckend waren in dieser Hinsicht auch die Erfahrungen bei der Arbeit mit den Patienten während der Kooperation mit dem Koma-Forschungszentrum in Lüttich/Belgien. Ebenso spannend war auch die Zusammenarbeit mit der Gruppe für architektonisches performance-integriertes Design vom EPFL in Lausanne/Schweiz, auch wegen meinem persönlichen Gefallen an Tageslicht während der Arbeit. Interdisziplinäre Kollaborationen sind eine super Möglichkeit, sich neues Wissen und fremde wissenschaftliche Ansätze anzueignen. Ich kann jedem ausdrücklich empfehlen, solche Gelegenheiten zu nutzen auch wenn damit eine gewisser Kontrollverlust bei der eigenen Projektplanung mit einhergeht.

Bei unseren Projekten war war neben Konzeptionierung der technischen Anforderungen an die jeweiligen experimentellen Aufbauten überwiegend Softwareentwicklung erforderlich. Ich implementierte unter anderem die Ansteuerung, Auslesung oder die Echtzeit-Positionierung von Video-Kameras und Bildverarbeitungsalgorithmen zur automatisierten Verarbeitung dieserr Videos. Desweiteren arbeitete ich auch intensiv mit computationalen Modellen für die Salienz-Bestimmung von visueller Aufmerksamkeit in natürlichen Szenen. Bei der Datenverarbeitung war ebenso Kreativität gefragt, um Daten von unabhängig arbeitenden Kameras/Messgeräten zu vereinen und innovative Lösungen zu konstruieren. Alle diese experimentellen Aspekte empfinde ich als wahre Spielwiese für den anwendungsorientierten Physiker. Sehr fruchtbare Berufserfahrungen bildeten die vielen extrem unterschiedlichen Herausforderungen, die ich während meiner Promotion meistern durfte.

Danksagung

Nach meinem Studium zum Diplom-Physiker mit Schwerpunktlegung Neurobiologie folgte ein Umzug ins beschauliche Mittelhessen, wo meine Frau Jennifer und ich mit der Vorort-Unterstützung durch die Schwiegerverwandtschaft möglichst flexibel unsere Karriereziele im Alltag mit unseren tollen Kindern, Moritz, Luna Josefine und Philomena, vereinbaren können. Ein ganz großes Dankeschön also an die ganze Familie - mein Quell der Energie!

Der bedeutendste Dank für die Verwirklichung dieser Dissertation gilt meinem Betreuer, Wolfgang Einhäuser, der mir etliche Freiheiten bei meiner zeitlichen Arbeitseinteilung und der Umsetzung meiner Projekte eingeräumt hat; vor allem aber dafür, dass er die Finanzierung meiner Doktorandenstelle und zahlreicher Dienstreisen für externe Kooperationen und zu Konferenzen gewährleistet hat. Besondere Wertschätzung gilt meinen Kollaboratoren, von denen ich Stefan - den großen Meister der EyeSeeCam-Software -, Camille - den Liebling aller Locked-in-Syndrom Patienten - und Mandana - der Sonnenengel der Büroarbeitsplätze, besonders hervorheben möchte, weil ohne sie keines meiner externen Projekte vollendet worden wäre.

Ein Extra-Danke möchte ich den Kollegen meiner AG aussprechen, besonders Alex - der gute AG-Geist und große Improvisator experimenteller Aufbauten; alle Leid-teilenden Doktoranden, Anna, Conny, Susanne, Svenja, Jan, Jonas, Matthias, Philipp, Stefan und Steffen. Hervorheben muss ich an dieser Stelle meine Zimmerkollegen Marnix und Marius - deren Konversationen auf holländisch ich stets schmunzelnd zuhörte aber nicht folgen konnte -, Peter, mit dem ich über wirklich alles reden konnte und finalement, merci Beaucoup et un Salut à Adrien pour ses moments de la culture française avec des repas du goût délicieux. Wie sehr habe ich durch Euch alle ein positives Arbeitsklima schätzen gelernt!

Wenn ich etwas für meinen körperlichen Ausgleich nötig hatte, ermöglichten mir das Bikepolo in Gießen und Frankfurt stets aufheiternde Stunden. Thus, a big hug to all people from german and european polo communities.

U Lai, also alle die nicht namentlich erwähnt sind: Thank you! Merci beaucoup! Danke schön! Merci vielmols!

Nummä net huudlä...

Dr. Feelgood

Inhaltsverzeichnis

Supplementary materials - demo movies:

Study II Visualizing the on-line fixation distance and resultant the correction of the parallaxes error being particularly prominent for fixation distances below $0.5m$:
`https://www.dropbox.com/s/9z036ysy0tp8z33/fixDemoSloMo_0.wmv?dl=0`

Study IV Exemplary difference in eye-movement behaviour between PSP- and IP-patients:
`http://journal.frontiersin.org/Article/DownloadFile/52120/octet-stream/Movie%201.AVI/7`

Study V Visual and auditory stimulus presentation to play *rock-paper-scissors*:
1. Raw footage of the eye:
`http://www.staff.uni-marburg.de/~wetgast/rps/Naber_Movie1.avi`
1. Abstract pupil representation:
`http://www.staff.uni-marburg.de/~wetgast/rps/Naber_Movie2.avi`

Study VI Showing the procedure of one yes/no-question on splitted screens:
`http://www.eurekalert.org/multimedia/pub/59814.php?from=245592`

Kapitel 1

Cumulus

1.1 Introduction

Characterizing and explaining gaze behaviour can be approached differently. On the one hand, from a computational view, gaze guidance seems to be strongly influenced by low-level features, such as luminance or color contrast within the present scene. This popular view, which is supported by numerous well tested models, is controversially debated, since on the other hand objects with their meaning are psychologically considered to be a fundamental unit of gaze guidance, while an observer is extracting the visual information of the present scene. We will provide evidence here that our gaze-behaviour is dominated by objects, which underlines their relevance in the context of natural-scene viewing.

Gaze-behaviour is measured by eye-movements and fixations and those yield an unobtrusive, sensitive real-time behavioural cue to the current visual and cognitive processing [25]. This processing allows the perception of the present scene, though its whole content is naturally perceived only after filtering. This means that barely more than subjectively relevant information attracts the focus of attention. The relationship between visual attention, task performance, and methods for measuring the gaze-behaviour, as well as decoding pupillary responses will be further investigated as main scientific topics in this dissertation.

1.1.1 Visual attention

Perception is the organization, identification, and interpretation of sensory information in order to represent and understand the environment, but is limited in regarding all available resources. Attention accounts for this bottleneck, takes over the selection of all present information, and appropriates contents to perception. The contents can for example lead to perception of the environment, of cognitive or emotional processes, or of one's own behaviour. Thus, attention allows the perceptual system to perceive stimuli or cognitive processes and focus of attention in turn ensures that other stimuli and cognitive processes are withheld from perception. Perception works serially and demands a rapid attention handling of all unconsciously occurring stimuli to achieve an efficient information processing. Perception retains its focus usually only for a short period, e.g. for a few seconds in case of two equivalently dominant stimuli, before switching to a different present stimulus.

This thesis focuses on visual attention, which separately considered, meaning independently from all other sensory modalities, operates on a huge amount of information from vision. Visual information again arrives serially at higher perceptual processing areas through the focus of attention, but only to a limited amount. The systems of visual recognition and attentional selection perform the essential task of perceptual categorisation the elements within the visual field. This categorisation is selective and is limited by the capacity of short-term memory buffer [12].

The control of visual attention features two different directions, called bottom-up and top-down attention. Bottom-up controlled attention means attraction by some conspicuous regions within the visual field, such that perception focusses upon there. Thus, bottom-up is defined by influence on attention through external stimuli. In contrast, top-down driven processes are individual, often currently conscious tasks or intentions. They can be driven by external goals and tasks or by the person's internal factors on visual attention. Nevertheless, bottom-up and top-down attentional control is not strictly separable since both mechanisms are obviously entangled [3]. For example, physical conspicuities in the visual field can receive amplification by certain internal expectations or inversely highly conspicuous items get ignored due to irrelevance regarding a current task [64].

As noticed, visual attention works selectively and only parts of the visual environment attain conscious perception. A strong relation between perception and the local focus of attention is provided by the line of sight, the gaze direction. At that point, the human visual organ's functional principle requires some consideration: The human eye's rotatable eye ball executes very frequent reorienting each after about 300 milliseconds through six different muscles. The eye's physical optics successively consist of a cornea, an aperture realised by the iris regulating the pupil size, a lens being able to accommodate (i.e., adapting its focal distance), a vitreous body containing intra-ocular fluid, and a photosensitive retina shaped in a spherical plane.

The retina detects an image of the visual field [32]. Photo receptors serving as detector elements are spread over the retina with varying area density numbers summing up to around $60 - 125$ millions of rod cells and $3.2 - 6.5$ millions of cone cells, thereby producing an image with distinctive spatial resolution. Additionally, subtypes of cone cells with different spectral sensitivity allow color perception with remarkable color contrast sensitivity.

The highest optical resolution regarding space ($8 - 10$arcmin, [69]) on the retina's image is created within the range of the fovea. The fovea spans a spatial angle of around one degree and is centred on the eye's optical axes, given the case of normal or corrected to normal vision (i.e., no strabismus) [68]. The optical resolution of the retina decreases with increased eccentricity related to that optical axes (see Figure 1.1.1, [47]). Gaze direction points to the spot, that is visually detected with highest image quality. This implies, which area of the visual field is perceived with maximal acuity for that instance. Thus, the focus of attention for visual perception is centred through gaze direction. However, there is a differentiation between overt and covert visual attention. Overt attention is the one strictly centred by gaze direction with a visual span over an angle of around $1 - 1.5^\circ$ [56], which is called the foveal visual field, whereas covert visual attention regards regions outside the foveal visual field. Covert attention mainly processes relevant information unconsciously for planning an eye movement towards this new region of interest [49] and thus reaching the conscious perception for further stimulus inspections. The feature-integration theory of visual attention [63] assumes, that visual attention needs to serially process all present stimuli for characterizing and differentiating the relation between their separable features. Thus, gaze directions are measured for localizing overt visual attention and for monitoring it time-resolved [17]. Gaze direction or fixations, as explained in the next paragraph, yield a proxy for overt visual attention and point to those contents of the visual field, that are relevant for attention.

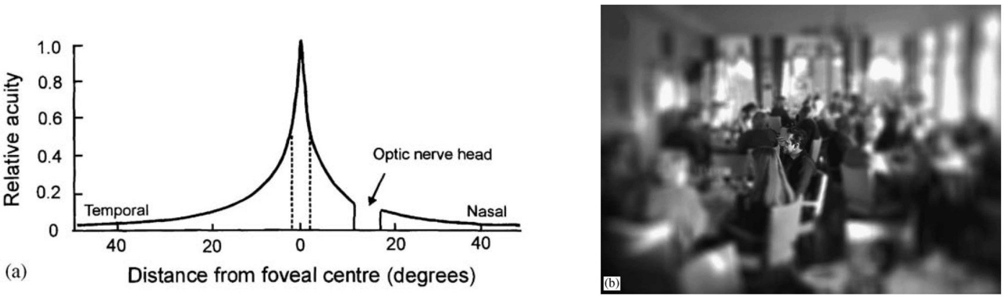


Abbildung 1.1: (a) Relative grating acuity across the horizontal visual field, FWHM by around the edge of the fovea (2° across: dotted lines). Based on data from Wertheim [68]. (b) Subsequent example image transmitted with retinal loss of resolution with eccentricity. Radially increasing blur, corresponding to the acuity decrease in (a), has been added to a photograph, whose field of view spans approximately over 20°. Picture by Benjamin T. Vincent from *Basic Vision, an Introduction to Visual Perception.* Oxford University Press (2006)

1.1.2 Eye movements in natural scenes and in real world

So far, we have used the term gaze behaviour in relation to gaze directions without any further specifying. This reduction ignored numerous differently characterized eye movements. The gaze direction relates to a common orientation of both eye balls upon a spot in one's visual field. Perceiving the content at that spot only occurs, if gaze holds there for at least about 200 milliseconds, meaning that a fixation happens – thus, the human eye moves more frequently than the heartbeats. Consequently, fixations are the relevant part within the path of eye movements and build the relationship in visual scene perception between overtly directed attention and the content of the current target. The transition between two consecutive fixations occurs by a very fast eye movement, whereby the eyeball accelerates with up to $9000°/s^2$ to velocities up to $700°/s$ [13]. Such eye movements called saccades purpose reorienting overt attention. Saccades happen roughly three to five times a second and on average induce differences of gaze-direction of about $6 - 7°$ [61]. The maximal human oculomotor span ranges as $\pm 55°$ horizontally and vertically [23]. Under laboratory condition, saccades and fixations are two main classes of eye movements, while in real-world slower eye-movements are also of relevance. Unlike for the laboratory definition, a fixation as condition of holding the gaze on a target in a natural three-dimensional environment does not require no movement to occur at all. In the real world, fixations do not imply that the eye is stationary with respect to the head, but rather that gaze is directed to a fixed point in the world despite concurrent head and body movements [35]. In the relationship to attention, when drawn overtly to a target in real world, possible relative movements of eye, head and target stimulus during a fixation can be regarded as real-world definition of fixations. Also included in this expanded definition are smooth-pursuit eye movements, which occur when a target stimulus is followed by the eye, while the stimulus moves relatively to the head. Smooth-pursuit eye movements belong to slow eye movements with velocities up to $100°/s$ and are merely controlled unconsciously. They happen in general, when a target stimulus moves steadily, like for example a trajectory of a flying object. Without pursuing a target stimulus, for instance in complete darkness, such slow eye movements do not appear. Without a movement of a target stimulus a slow eye movement still emerges after a large saccade, while the head usually carries out a retarded reorienting towards the target stimulus. In doing so, the eye-in-head coordination ensures a stable fixation on the target stimulus whilst an eye movement compensates for that head rotation.

The final class of eye movements that is relevant for the present thesis is associated with binocularity of human sense of vision, which means fusion of both eye’s visual inputs. This important function of the oculomotor system inhibits retinal disparity of both eyes – already slight disparities of above $0.25°$ create diplopia, restricted stereopsis, and distorted depth perception [41].

Consequently, both eyes focus specifically at different distances by accurate coordination, whereas the difference of the horizontal components of both eye’s gaze directions changes and increases for shorter fixation distances. This difference is called vergence angle and its adaptation vergence movement. Its delay after a saccade to a different fixation distance is below 100 milliseconds [65]. A specific measurement of vergence allows a useful estimate for fixation distances (see study II).

1.1.3 Early-salience models and maps

Since the work by Itti, Koch, and Niebur [31], models predicting attention guidance in visual scenes have continually gained increasing acceptance. The output of those models are so-called salience maps. These maps are arranged topologically regarding the visual scene and contain scalar values per pixel serving as measure of attracting attention. The peak within a salience map is used to predict the location of an unaffected observer's first fixation, after the visual scene belonging to that map was newly displayed on a monitor. Furthermore, the salience map constitutes a distribution, which could approximately represent the number of fixations after longer inspecting the belonging visual scene, if no bottom-up attentional factors would interfere.

The functional principle of those early-salience models are based on checking the presence of various characteristic patterns within an image of a scene and comparing them with each other to finally determine a gradual emphasizing of locations, where certain patterns stand out. Salience-models use computer-vision algorithms for filtering and processing the input image with the aim to possibly evaluate the whole diversity of image appearances in a physiologically plausible way and thereby predict fixation selection. However, this merely fulfils the demand of bottom-up attentional control mechanisms, which refers to influences from stimuli of the visual field.

Unfortunately, this scenario is only constructable under laboratory conditions and hardly gains relevance in natural environments. In everyday life, a person interacts with its real-world environment, perceives the context of a scene and the semantics of containing items. In such a context, tasks are often given, actions planned, or the context itself starts a thinking for further interactions, collectively considered as top-down attentional processes. These conditions are not covered by early-salience models, because their algorithms operate on pixel information of two-dimensional images of the visual environment while ignoring the influence on our gaze-behaviour through context, semantic or task. Consequently a salience-map's prediction power is limited and principally fails as soon as bottom-up attention is dominated by top-down attentional processes [19]. Of particular importance in this relationship is object recognition, which obligatorily occurs during scene inspection but which is not explicitly incorporated in early-salience models. In the following we come to the ground of the essential question, whether objects yield more prediction power for gaze-behaviour than early-salience models. These results cast doubt on the usefulness of salience-maps for predicting gaze-behaviour in real world.

1.1.4 Low-level Features

In general, low-level features serve as basis for computing salience maps. Extracting low-level features from pixel information reduces the dimensionality of the input data. The most basic image features are luminance, color hue, color saturation, as well as contrasts, created by edges, corners, and textures. They are essential for pattern generation and for emphasizing certain image regions. Early-salience models evaluate over the regarded features within the image a deterministic combination resulting in conspicuities and consequently in a distribution representing the salience map. Unfortunately, cognitive aspects induced by a natural visual scene are not or only implicitly regarded (but see: [42], [43]). Moreover, relevant processes like object detection or foreground-background segregation are not regarded. The bottom line is that the simplistic approaches of constructing salience maps from low-level features are limited in prediction power, since the gross information content of a scene is not fully apprehended.

1.1.5 Gaze behaviour and attention in natural scenes and real world

The explanation of gaze behaviour in natural scenes is implicitly affected by the question of attentional guidance and its causes. Which factors impact attention during inspection of a natural scene and which factors are added through interaction with the real world? Studies on gaze behaviour in natural scenes under laboratory conditions mean controlling the individual tasks set to the subjects. By contrast within the real world, many parameters are badly foreseeable or observable, such as the temporal course, the individual motor coordination and interaction, or light conditions, making it difficult to hold them constant over all subjects.

The visual bottom-up attention with the aim of purposeful information intake from a given scene regards mainly low-level features as well as detected objects. Anyhow, the control of gaze direction depends here from the relevance of the content in the region of a saccade target. This relevance involves also top-down attention and thus the choice of the targeted item varies strongly according to cognitively different situations. Hence, except for search tasks [22] and for those not in general [28], the superficial appearance and so the impact of low-level features matter less [59], than the actual task [34], the context of the environment [62] or the semantic of the present objects [29]. Early-salience models perform significantly above chance at predicting fixation targets in natural scenes presented on a monitor, if an observer explores it freely, but in case of a less artificial conditions the prediction power diminishes. Since such conditions inevitably appear at the transition into the real world, early-salience models hardly gain prediction power for gaze behaviour in real world [60].

Gaze behaviour in the real world also changes substantially in comparison with studies on natural scenes in head-fixed laboratory settings, because of the free mobility of the participants head and locomotor system. The head and eye-head-coordination play a crucial role for gaze allocation [23], [35], [20]: First, a head rotation necessarily precedes a saccade, if the planned alteration in gaze direction spans an angle bigger than the oculomotor range. Second, the head usually executes an unconscious reorienting towards the actual gaze direction also after smaller saccades within the oculomotor range. The locomotor system also requires visual attention for action planning, even if a person moves along as in everyday occurrence [39]. But the more difficult the coordination demands [1] in the environment (i.e., through rough irregular terrain), the more visual attention is absorbed for action planning [38], albeit such processes could still work unconsciously. This shows, that already unconsciously performed tasks distinctly affect gaze behaviour and that attention is essentially incorporated by daily routines [36]. However, the individual gaze behaviours appear to be pretty stereotypical during daily routines in real world, i.e., for actions like car driving, tea making, ball shots or other sports [34]. According to this, gaze behaviour does not perform free exploration of the surrounding, as long as actual tasks and interactions with the environment and thus top-down processes are the decisive factor for guiding visual attention. As underrated for a long time, the relevance of implicit and explicit tasks obtain a priority versus factors of purely visual scene appearance [19].

1.1.6 Evidence for object-based attention

Attentional processes are highly parallelized. Information is processed simultaneously across the visual field, first on the retina, in the lateral geniculate nucleus and then in the visual cortex. In doing so, recurrent influences benefit already the further processing of features, that are relevant within the actual context [14]. This is among others ([70], [30]) the most prominent neurophysiological connection, and because of this bottom-up controlled gaze allocation is not merely predictable through low-level features, so not sufficiently able to be modelled by early-salience maps. An alternative hypothesis uses for the prediction of fixation targets the localisation of objects without regarding low-level features at all. Fixation distributions within objects are modelled by a Gaussian with its mean centred on the object and are referred to as preferred viewing location (PVL, [45]). In detail, this object-based hypothesis models fixation distributions for each object given by the knowledge from PVL with a scaling factor reciprocally proportional to the area filled

[1] For imagination: Juggling and solving three Rubik's cubes - `https://www.youtube.com/watch?v=K_gHa2x2OQA`

by that object and sums over all present objects to result in a map for fixation predictions. While our investigation (study I) compared this object-based model against state-of-the-art early-salience models; prediction performance was at par and could not be optimized by combining both models' maps. We thus hypothesised that the prediction power from early-salience models is not directly explained through occurrence of low-level features, but rather through their prediction of object locations. For testing this object-based hypothesis, we manipulated low-level features by reducing their color saturation and luminance contrast explicitly in regions, where object-based maps would predict fixations with a probability threshold above the median over each image. With these manipulated stimuli we were able to test, whether low-level features would dominantly induce fixations in image region without increased objecthood or whether objects still stronger attract fixations even if their influence from low-level features is reduced. Now, if an early-salience model predicts fixations better than the object-based model, low-level would create more power in attentional guidance. But since observers' fixations on such manipulated images are significantly better predictable by the object-based model, we conclude, that objects independently from their low-level features attract the overt visual attention distinctly stronger than conspicuities through low-level features. Only for the case when several objects occur, that are equally ranked by an object-based map, low-level features would affect the fixation priority between these objects. This finding will be further explained in study I.

1.1.7 Pupillometry – pupillary responses and their measurement

While the visual cortex in the occipital lobe of the cerebrum processes the input information for visual perception, another older part of our central nervous system manages the continuous adaptation of the pupil size, more exactly the pretectal nucleus, the Edinger-Westphal nucleus (EW) emitting acetyl-choline, and the locus coeruleus (LC) being the major noradrenergic nucleus of the brain (see: Samuels & Szabadi, 2008, Fig. (2) [51]. EW belongs to the parasympathetic nervous system and LC belongs to the sympathetic nervous system. Both branches add together the autonomous nervous system being responsible among others for controlling the heart beat or breathing, as long as it is operating subconsciously. Also, the autonomous nervous system dictates the movement of the iris, like an aperture of a camera lens, for regulating the throughput of light through the pupil. The iris consists of two types of muscles: In a brightly illuminated surrounding, the sphincter muscle as a ring around the pupil constricts itself and thus the pupil up to a diameter of about 2 millimetres; in thick darkness, a set of radially arranged dilator muscles enlarge the pupil up to 8 millimetres.

Neurophysiological findings suggest, that changes in pupil diameter are strongly correlated to changes in activity in LC [2]. LC as major release site of the noradrenergic system emits the neurotransmitter norepinephrine (NE) and is considered as mediator of functional integration of the whole attentional system [52]. For this hypothesis, NE plays a decisive role in energising the cortical system and in exciting the appropriate activity levels for cognitive performance [48]. It is plausible within this relationship, that cognitive and emotional processes elicit also a constriction or dilation of the pupil. Albeit such effects on the pupil size occur with distinctly smaller changes than the pupil light reflex, the pupillary response is a convenient indicator for mental effort. While highly concentrated mental effort [27], [6] induces pupil dilations with diameter changes up to 2 millimetres, changes through cognitive and behavioural processes like emotions, arousal, attention, social interaction, et cetera range only up to 0.5 millimetres. Still, this dynamic is sufficient to be video-technically measurable and at least a mean over many trials shows a significant effect, if factors affecting light-adapted pupil size like brightness, color and distance [71] are well controlled.

Pupillometry can be realised technically quite simple with an infrared-light sensitive video camera. The camera operating at usual frame rates of $25fps$ is fast enough for measuring a sufficiently time-resolved pupillary response. A high image resolution is just as little required, but a good, low-noise image quality, that can be ensured by a proper infrared (IR) illumination. Under exposure of IR light, the iris, the eyeball, and the skin of the lids are highly reflective, but the pupil mostly transmits all IR light. Thus, the videos reveal a huge brightness contrast between pupil and the iris, without putting the retina at risk through overdosed artificial radiation intensity [15].

For detecting the pupil and measuring its diameter from video recordings, an image processing algorithm was implemented as follows in principal and applied in study V in experiment 1 and in study VI during pilot studies. The pupil was segregated by a luminance threshold from the raw image, the subsequent binary image was eroded for de-noising and finally an ellipse was fitted around the edge of the biggest dark blob. Twice the length of the resulting semi-major axis determines the pupil diameter in pixel, at what only the relative change is of relevance for decoding pupil dynamics. We ensured by the experimental procedure to exclude systematic errors [11], like that measuring the pupil diameter was not biased through major gaze-direction changes, since they would have induced foreshortening of the pupil on the camera image.

The time course is shaped by a signal, that is invoked by cognitive processes through tasks within a particular paradigm but also by noise from subconscious processes or from a certain state of distraction. The task-relevant processes reveal an increase of phasic LC activity and a “fast”, distinct pupil dilation with a fixed temporal delay. The tonic activity in LC should not be too low, as the pupil size begins to fluctuate considerably in that case [5], thereby also decreasing the signal-to-noise ratio. The tonic state correlates with the degree of vigilance, because these fluctuations appear prominently, if a person is sleepy or fatigued after sustained attention. Thus, vigilant persons yield more pronounced pupil signals.

1.1.8 Parkinson's disease

The Parkinson's disease, also called idiopathic parkinsonism (IP), is a neurodegenerative disorder of the central nervous system [50]. IP is with 75% the most frequent form of parkinsonian syndromes. The typical motor symptoms of IP are tremor (strong shaking of arms and legs), bradykinesia (slowness of movement), rigidity (increased muscle tone), and postural instability, while consciousness and intellectual performance stay fully intact. In a progressed stage, behavioural problems and dementia could appear, often accompanied by depressions. The symptoms result from still unresolved cell death of dopamine generating cells in the substantia nigra, a part of the basal ganglia in the midbrain. In most cases, IP emerges at an age of around 50 years. The current treatment options cannot heal the cause of IP, but mitigate the symptoms effectively up to a certain degree of the progressive disease. In particular Levodopa and dopamine agonists are the most common medication.

1.1.9 Progressive supranuclear palsy

Progressive supranuclear palsy (PSP) is a further neurodegenerative disease of the central nervous system, also belonging to the family of parkinsonian syndrome. It is the second most frequent occurrence of parkinsonism, that's why it is also called PSP parkinsonism. The symptoms are very similarly to IP, since the basal ganglia are also involved. Its diagnosis is often confused, though tremor is a seldom observed PSP symptom. Instead, an increasing difficulty in moving the eyes occurs with progressed disease, especially for vertical eye movements [55]. To date, there are no efficient drugs against symptoms, against the cause of disease, or to protect the affected nerve cells. Since drugs established for IP do not help or tend to induce adverse reactions, the differential diagnosis between IP and PSP should be improved. Therefor eye-tracking is a useful tool, because it enables to detect an impairment of eye movements in PSP objectively and with high sensitivity.

1.1.10 Locked-in Syndrome

The Locked-in-Syndrome (LiS) is a disease as consequence of a malfunctioning of the central nervous system, whereby consciousness is usually fully intact. According to clinical symptoms, we differentiate between total LiS with absolute complete immobility (quadriplegia – consciousness is registered via EEG signals), classical LiS with immobility except for vertical eye movements and eye blinks, and incomplete LiS with further remaining motor abilities [4]. Typical LiS arises from a lesion in the base of pons (brainstem), which in example occurs after strokes or a traumatic brain injury. Diseases like amyotrophic lateral sclerosis or multiple sclerosis can also lead to LiS, as well as snake bites.

For LiS patients, there are neither standardized treatment methods, nor a chance of recovery. The applied methods are solely symptomatic, in particular for chronic LiS patients. Nevertheless, there are cases of transient LiS with a spontaneous regeneration of at least parts of motor control and somatosensation. Electrical stimulation of muscle reflexes can also be helpful. Against disability of speech, eye movements, blinks or if any other body movements are employed in most cases of classical or incomplete LiS for encoding answers and thus facilitate communication. In study V, we investigate the possibility to communicate with LiS patients and show, that the pupillary response is useful for that purpose.

1.1.11 Daylight in workplaces and glare

The design of workplaces has an essential influence on health of its inhabitant spending a lot of time there. This is why workplaces fulfil norms, amongst which those for ergonomics [2] and workplace lighting [3] are the most remarkable, as they influence working comfort. A huge factor thereby is the visual comfort, which is characterized by several features [9]. So far, there is no scientific benchmark for visual comfort, but rather an individual subjective definition: *"That state of mind that expresses satisfaction with the visual environment."* [4]. Visual comfort is thus a subjective impression in regard of the amount and the spatial spreading and quality of light. The essential parameters are the illumination of the visual task, the luminance distribution within the room, the sight towards outside, the colour rendering and the colour of the light source, and absence of glare. Glare patches in the visual field are regarded as strongest interference on visual comfort while

[2] DIN EN ISO 9241, Info-Quelle: `http://www.ergo-online.de/site.aspx?url=html/service/gesetze_und_regelwerke/normen.htm`

[3] DIN 5034-1:2011 & DIN 5035

[4] Prof. Walter Grondzik, Architecture Faculty, Ball State University

working at an office workplace or in another indoor job, since they are mostly perceived there as discomfort glare. This type of glare creates a distracting uncomfortable situation, because of a disturbing effect on visual information uptake, though the ability of visually perceiving the working range is still given. Consequently, visual comfort gets subjectively degraded, though no direct physical or visual constraint is measurable within the task range. Possibly, the evaluation of subjective perception of visual comfort combined with eye tracking could constitute a method enabling an improved prediction of parameters creating visual discomfort.

1.1.12 Mobile eye-tracking

Mobile eye-trackers serve as device for measuring eye movements and allow recording visual details about the user's fixated items with a point-of-view camera (PoV camera) mounted beside the eye in central gaze direction. Essentially, one or two eyes are recorded by video-oculography (VOG) by employing IR-sensitive cameras and their positions in the camera image gets mapped into a gaze direction, while pupil detection yields a pixel position and an appropriate calibration determines the translation from image position to gaze direction. This measurement principle allows a high degree of free head and body motion and captures eye-in-head movements. Simultaneously, there arise higher requirements in gathering data, because the visual field in real world includes a third dimension. Admittedly, the PoV video allows recognizing and localizing relevant objects, events and coordinates [20], but analysing such videos requires either manual annotations with an enormous temporal effort or a sophisticated development process on algorithms for computer-vision controlled object recognition, which is easily limited by conditions of low image quality or enhanced complexity of the experimental environment. However, such computer-vision approaches advance evaluation routines and enable an efficient upscaling in data analysis. Study III shows an exemplary application with that strategy and copes with the higher degree of freedoms in a real-world setup by expanding the measurement of head positions to six instead of two dimensions (orientation and translation). Eye-head coordination is captured pretty precisely and additionally allow us to derive fixation targets in an automatic fashion, by analytically processing the eye-head coordinates combined with a 3D model of the experimental setup.

This proposes a novel concept for mobile eye-tracking studies in real-world environments.

1.2 Overview

In study I, evidence is provided that objects, rather than low-level features, control fixation guidance in natural scenes. As spatial properties play a crucial role in object recognition [40] and objects strongly influence gaze behaviour, a transition into real-world, spatial experimental setups is obvious for testing their influences on gaze behaviour.
Beside that argument, measurements in such real-world scenes are required for the feasibility of interactive tasks and thus also new eye-tracking methods. Related thereon, study II describes a novel three-dimensional eye-in-head calibration method. For study III, the eye-tracking measurement procedure was expanded by a head-positioning algorithm to include gaze-in-room coordinates and was thus completed for a universal approach in mobile eye tracking. Study IV was the first employment of study II's calibration method and demonstrated the practical benefits of mobile eye tracking for clinical usage.
Study V and VI required another experimental setup to develop, a mobile pupillometry device, with which the dynamics of the pupil size could be used to register cognitive processes and even for communication.

1.2.1 Gaze allocations on objects

Study I: Overt attention in natural scenes: Objects dominate features

The content in natural, visual scenes, like occurring in everyday life, is particularly perceived by for the current scene relevant objects. Concurrently, overt attention is also attracted by luminance and color contrasts. These are low-level features and were considered for over two decades to be responsible in bottom-up driven control of visual attention. Here, we aimed to reconcile the apparent conflict, whether low-level features of a scene or objects are predominantly causing the cognitive processing of time-dynamical gaze allocation and thus of planning saccades. Hence, our first study was motivated by combining arguments from Einhäuser, Spain, and Perona [21] and from Nuthmann and Henderson [45]; that is, we hypothesise object-based fixation selection and expand an object-locations model to account for preferred viewing locations (PVLs) within objects.
We measured gaze allocations of persons observing natural, object-containing scenes for three seconds each. The gathered fixations were statistically compared, whether they could be better predicted by our object-based fixation model, or by various state-of-the-art, early-salience models. Experiment 1 found, that our object-based model predicts fixations just as well as the best early-salience model. Since this result could not answer whether objects or low-level features dominate in causing fixations, another two experiments were performed with manipulated stimuli. We modified the natural scenes from ex-

periment 1 such that color saturation and luminance contrast, the two most essential low-level features for early-salience models, were considerably decreased in regions, where the object-based model predicts a strong attraction for fixations, to dissociate the processing of low-level features and objects during scene perception. The subsequent results show significantly, that objects predict fixation selection much better than early-salience models, though the objects' conspicuousness was decreased by the local image manipulations. Consequently, low-level features control our gaze behaviour only indirectly, since parallel processing object recognition apparently obtains higher priority in controlling eye-movements.

Thus we can conclude, that objects dominate features for fixation selection in natural scene viewing.

1.2.2 3D Mobile Eye-Tracking Methods for Real-World Applications

The original and to date most widely applied method to measure eye movements and gaze behaviour happens mostly in darkened, soundproofed rooms in front of a computer screen mainly head-fixed by a chin- and forehead-rest. This allows optimally controlled and reproducible laboratory conditions, also because the degree of freedoms in data gathering is substantially reduced. Thereby the data analysis is highly standardized and simplified. The investigation of gaze behaviour, respectively of psycho-physical signals of the eye in general is dependent on mobile eye-measurement methods, since not every situation or natural task is feasible in a laboratory, i.e., in front of a computer screen. 't Hart et al. [39] compared observers looking at identical visual stimuli under real-life and laboratory conditions. They found a different gaze behaviour for computer-screen presentation versus real-world environment. Furthermore, several other studies showed, that gaze behaviour is strongly influenced by actually given tasks or by motor or mental actions. Michael Land [37] tries to explain general rules for saccade planning with a *schema system*, which mainly controls gaze during the interaction with the environment. These findings require to adapt the experimental setup to the conditions of an aimed situation and that gaze or pupil are measured with mobile devices. Head and body movements should be freely possible, such that tasks are executable without restrictions. A natural gaze behaviour arises only by a three-dimensional environment with a full field of vision. One needs to consider, that depending on the respective experimental situation the whole range of the visual field as well as the distance to fixation targets can vary in their natural range, whereas certain gaze directions could dominate due to the experimental task. For example during haptic and visual inspection of a handled object, the field of view is focussed on the objects between the hands. These are usually raised toward the eyes within the near-field range.

Thus, gaze varies in a small range. Notably, that range differs from the straight-forward eye-in-head direction usually taken to centre the range of calibration and has to be adapted.

If additionally the fixation distance varies during the experiment within the near and the far field, the eye tracker is required to measure the fixation distance, which is realized in study II.

Study II: Mobile three dimensional gaze tracking

Fixation distance is computable by using the vergence between both eye directions. At the time study II was conducted, commercially available mobile eye-trackers lacked the accuracy to reliably measure a vergence signal. We achieve that goal by additionally giving the eye tracker information about the fixation distance during calibration and about the individual eye distance, i.e., distance between the centre of both eye balls. This novel calibration method allows instantaneous computing of the observer's fixation distance via horizontal eye positions' difference. The resulting data reveal three-dimensional eye-in-head coordinates.

Our mobile binocular eye-tracker, a prototypical EyeSeeCam [53] has another camera additionally to a common PoV camera, that is mounted manoeuvrable on a gimbal joint. This configuration is actuated within a latency below 10 milliseconds [16] to orient the camera's optical axis equal as the direction of the user's gaze. However a parallaxes problem results with an error dependent on variations of fixation distance, since the gaze-following camera (gaze-cam) is placed in front of the forehead. The error increases tremendously for shorter distances. By measuring three-dimensional eye-in-head coordinates, we get able to correct for the parallaxes error and to direct the gaze-cam's optical axis onto the 3D fixation point. For that purpose, an analytical relationship between the gaze-cam rotation and the positioning of the linear drives actuating the rotations was developed and implemented. This parametric gaze-cam positioning enables the system to drop an additional calibration for the gaze-cam positioning, which was alarmingly independent from the eye-in-head calibration and consequently could deviate from eye-position recordings. The accuracy of gaze-cam positioning is significantly improved and affects recordings particular beneficially within the near range, so the reach, grasp, or reading distance.

1.2.3 Measuring gaze-in-room

The three-dimensional calibration of study II is an essential methodological progress in mobile eye tracking, though it is restricted to head-referenced coordinates. By using a camera mounted on the forehead and thus recording PoV videos synchronized to VOG data, the user's gaze behaviour can be related to its surrounding. Fixations on objects, items, or other features are assignable in this way. This requires a cumbersome, mostly manual evaluation of PoV videos, though the effort is hardly manageable for measurement series with quite numerous participants. Hence a new method with improved efficiency is proposed here. An obvious approach is to include a 3D model of the experimental setup with all relevant details and further to add a measurement of head position as well as orientation to existing head-referenced coordinates captured by the mobile eye tracker. The actual head-positioning method is realized by first manually annotating some few PoV video frames each containing four key feature points known within the 3D model, with which the perspective-n-point (PnP) problem is solvable based on camera-lens parameters [72] allowing to re-project the feature points onto the 3D model in an iterative optimization process [46]. In a second step, temporally high-resolved sensor data from an inertial measurement unit (IMU) is employed and integrated between the absolute head positions extracted by the manual annotations. Our result is a six-dimensional time series with head-in-world position and orientation, which is synchronized to the eye tracking of the EyeSeeCam at 220 Hz. The superposition of these novel head-in-world and eye-in-head coordinates yields the requested gaze-in-room data, that can be projected now onto the experimental setup's 3D model for further data analysis.

Study III: Uncovering relationships between view-direction patterns and glare perception in a day-lit workspace

Up to now, factors from various daylight conditions were diagnosed merely subjectively based on psychological questionnaires. We contrast for the first time gaze behaviour as measure with the influence of low- and high-contrast lighting conditions, as findings from Vincent et al. [66] suggest a relationship between the allocation of overt attention and the level of light sources: medium luminance levels are preferred over low and high extremes (see also [44]). The innovative method for 3D gaze-data processing is applied in this study, where the influence of different lighting conditions on gaze behaviour is investigated at an office workplace during common tasks. In a cooperation with the *Interdisciplinary Laboratory of Performance-Integrated Design* (LIPID) from *École Polytechnique Fédérale de Lausanne* (EPFL), Switzerland and the *Fraunhofer Institute for Solar Energy Systems* (ISE), Freiburg i.Br., Germany, we are facing the following questions: How does accounting for gaze-direction and actual luminance distribution perceived by the eye enhance our understanding of visual comfort in indoor environments? Are there certain luminance levels that attract or distract our line of sight? For that purpose, a new approach in visual comfort assessments integrates both subjective and objective measures. Therefore, we recorded our participants' visual system responses during a real-task experiment in an office like test room upon the ISE's main building in Freiburg. By using eye-tracking methods while monitoring photometric quantities relevant to visual comfort measures using HDR imaging techniques, we achieve to account for actual luminance-fields perceived by the eye. The experimental setup for simulating office-like work essentially contains a desk with a computer screen, a keyboard, a telephone, paper writing material, and a window facade laterally to the desktop seat. Beside the mobile eye-tracking (EyeSeeCam) measurements, the photometric quantities were captured via HDR images with 2 CCD cameras with fisheye lenses spatially resolved over a 3π steradian (sr) unveiling the locations of glare sources. A specific variation of the sun angle (and/or view content) relative to the setup container is possible by rotating it into all cardinal direction with aid of a powered bogie. We protocol the whole procedure to pinpoint the gaze on the 3D modelled setup in order to derive the perceived luminance-field from simultaneous photometric records. Our results show, that participants tend to look out of the window during non-visual office tasks, but this behaviour is avoided under high-contrast lighting conditions potentially containing glare sources. Furthermore, we observe, like already hypothesised by Land [37], that gaze allocation is strongly focussed on the currently executed task regardless of lighting conditions.

Study IV: Validation of mobile eye tracking as novel and efficient means for differentiating progressive supranuclear palsy from Parkinson's disease

Our development on the EyeSeeCam, the mobile eye-tracker's three-dimensional VOG calibration is applied amongst others in this study. Here we investigate two groups of parkinsonian syndrome patients suffering either from PSP or from IP. Both diseases reveal pretty similar symptoms regarding pathological motor behaviour, rendering correct diagnosis of PSP and consequently adequate treatment difficult. As decisive symptom, PSP patients display an impairment of executing vertical saccades. This symptom is convenient to measure objectively with a technical device and to detect reliably subtile disorders in eye movements and thus for improving sensitivity and specificity of clinical diagnosis. Since earlier knowledge about this diagnostic procedure is based on restricting stationary measurement devices, we like to transfer that with aid of the EyeSeeCam into a substantially more practical measurement protocol. Thereby the eye-tracker's mobility and the rapid diagnostic procedure with a standardized fixation protocol within less than 20 seconds bring along serious diagnostic benefits. Furthermore we are interested to reveal abnormalities in oculomotor behaviour in both patient groups whilst they perform an un-

restricted every-day task that is walking along a corridor.
We find prominent impairment of both saccade velocities and amplitudes in the group of PSP patients, thus differentiating them clearly from Parkinson's disease patients and healthy controls. The differences are best pronounced in vertical saccades. Other eye movements hardly reveal distinct differences. The described effects arise stronger during the standardized fixation protocol than during an every-day task. To conclude, the combined analysis of saccades' velocity and amplitude during a rapid standardized protocol serves as a simple and reliable tool for differentiating clinically diagnosed PSP patients versus IP patients and healthy controls. Thus we offer a solid assistance for patients with uncertain diagnosis.

1.2.4 Pupil Communication

Measuring the dynamics of pupil size was established already in the early 1950s, whereby research in pupil dilation was out of fashion over several decades but recovered actuality around the turn of the millennium with the advent of comparable inexpensive video-based eye-tracking devices. Dilation and constriction of the pupil are controlled by the sympathetic and parasympathetic nervous system and underlie several unconscious as well as conscious causes. Cognitive arousal of any form elicits a pupil dilation with slight temporal delay. Our research interest overlaps mainly with pupil dynamics caused by cognitive decision-making processes, which also induce pupil dilation, since these are usually accompanied by a certain arousal. We aim to improve the capturing of decision-making signals through appropriate experimental procedures, that hold down unwanted factors. This leads to two goals, which are not practicable in common eye trackers. First, recording high-contrast videos of the pupil under IR illumination to serve as stimuli and as source for further abstracted stimuli conveying decision-signals. Second, realizing a mobile pupillometry setup consisting of a laptop and a CCD camera with a fast frame rate ($max.200fps$ at full ROI), which allows bedside measuring immobile patients. The resulting self-made pupillometry setup gathered capacity-saving videos ($8-bit$ greyscale packed in lossless JPG format) for study V and pupil-diameter data of all pilot studies for study VI. With that free-standing setup, we experienced a tendency to spasms with LiS patients impairing the data quality, so we switched later for study VI to a light-weight version of the EyeSeeCam with comfortable head-holder straps, though handling the EyeSeeCam with a lying patient was difficult do adjust. Thus, a pretty compact, easy employable pupillometry setup completely arranged upon a baseball cap (except for the monitor and cabling) is under development up to date for testing first of all to communicate with total LiS patients.

Study V: How to become a mentalist: Reading decisions from a competitor's pupil can be achieved without training but requires instruction

Einhäuser, Koch, and Carter [18] demonstrated, that a concrete cognitive decision process can be monitored by a pupil dilation as reaction and that the point in time of this phenomenon leads to infer the outcome of the decision. We like to further investigate the monitoring of intentionally induced pupil signals and test, whether one can learn to perceive an opponent's decision-making by observing its pupil and to take advantage thereof for a higher winning chance in the childhood game *rock-paper-scissors*. For this purpose we captured videos of the eye from three players, whilst playing a slightly modified Version of *rock-paper-scissors*. All further participants opted in a simulated *rock-paper-scissors* game for the auditory presented choices whereas only the visual stimulus of the opponent's eye was video replayed. For further reducing the information contained in the visual stimulus, a pupil-detection algorithm provided the time course of the pupil size. So we created videos with a black circle, whose diameter followed the temporal fluctuations of the real pupil videos, on a grey background.

Our results show, that both types of visual stimuli are appropriate to gather information during video replay for recognizing the opponent's choice between rock, paper, or scissor and for boosting its own chance to win. However, we could also demonstrate, that our test persons were only able to improve their winning performance, if they were explicitly instructed to regard the pupil dilation.

Study VI: Pupil responses allow communication in locked-in syndrome patients

This project raises the demand on pupillary response even higher than study V. It aims to realise a non-invasive user-friendly human-computer interface for communicating with LiS patients. On that account we started a cooperation with the coma-science group at the *Centre Hospitalier Universitaire de Liège*, Belgium. The pupil as information carrier brings the advantage for a human-computer interface, that no voluntary muscle control for a body reaction is needed. For the basic idea to communicate in the sense of information transfer, we are requested to decode the pupil size's temporal dynamics into a binary output, which allows us to classify answers on yes-no questions. The difficulty hereby is to ensure a potent generation of a pupil dilation and to achieve a robust algorithm for low-noise signal detection.

The breakthrough to reliably induce a pupil reaction with a high signal-to-noise ratio provides the integration of solving an arithmetic problem during serial auditory presentation of both yes-no answer options. In the meantime, the responding person should compute the arithmetic problem visible on a monitor during the period when the correct answer option is presented. The mental effort of problem solving induces a temporally slightly delayed robust monotonic increase in pupil size.

While recording the pupil size, we conducted guided interviews with LiS patients as well as with healthy controls, notably without allowing any active motion reaction. As result we bring the proof of principle that our method is applicable for decoding a pupillary response for communication to reliably reply to yes-no questions. Furthermore, we demonstrate that LiS patients also have control over the proposed communication paradigm. Remarkably, the same decoding algorithm settings are usable in all individuals, implying a principally unrestricted pupil reaction in LiS patients.

1.3 Discussion

1.3.1 Object dependent attention guidance

Study I: As fixations are a proxy for overt visual attention, their targets' investigation should reveal mechanisms driving attention. Within the last two decades, over 2000 studies on visual attention were published [14] and a prominent portion of these articles considers early salience as the crucial factor in generating bottom-up visual attention. Contrary to this view, our results postulate that attention guidance is dominated by objects, being detected as relevant items within a scene. Notably, fixations were predicted by using only all objects' locations, their size, and a general formulation for preferred viewing locations within an object, which was based on a fully independent study with a different stimulus set. The best early salience model to date could perform only similarly strong in predicting fixation locations in natural scenes due to an implicit relation between low-level features and object-hood in natural scenes. The decisive factor enabling this conception was to manipulate stimuli by depreciating low-level features specifically at locations where object maps predict fixations with augmented probability. In such stimuli, the observer's attention was drawn significantly stronger onto objects than by the conspicuousness through low-level features. In this context, early salience as purely low-level feature based concept loses its relevance for natural scenes. Admittedly, there are some findings showing benefits for object recognition from low-level features [58] and several computational approaches using low-level features for object detection [67], [26], [54]. However, it would be of interest to test such relations in real world environments and to compare whether depth cues are even stronger promoters for object detection.

1.3.2 Methodological developments

study II: Mobile eye-tracking devices allow studies in real world environments while users are freely behaving, but they all lacked the possibility of recording the fixation distance. So, we elaborated the existing calibration procedure of our binocular mobile eye-tracker, an EyeSeeCam prototype, for reliably measuring eye vergence and thus recording head-referenced eye positions in 3D. This novel 3D eye calibration is completed in about a third of the time needed in the former version and directs the EyeSeeCam's gaze-centred camera with increased accuracy not only for fixations at the distance, where the eyes were calibrated, but over the whole depth range of the visual field. The error of directing the gaze-centred camera is eminently reduced for very short fixation distances within the grasping range. Hence we became able to gather an object database reflecting natural behaviour in point of view perspective. It contains manipulating objects in the hands of the user, who carried the EyeSeeCam and thereby recorded gaze-centred videos. Such videos illustrate real-world stimuli appropriate for training object recognition models, which could imitate the human visual system under a natural condition of object handling. The EyeSeeCam's calibration was upgraded for registering fixation distance and so the depth perception in vision.

Gaze-in-room: While a mobile eye-tracker enables free behaviour, further degrees of freedom emerge in gaze coordination, which need to be measured for a full observation of eye-head-coordination, overt attention, fixation targeting, or particular task performances. Consequently, we aimed to additionally track head positions to gain full control over the visual scene during real-world tasks. We propose this rather complex approach, since further measurement equipment would have interfered with the behaviour during office-like work in study III. Head positions were computed on basis of the camera calibrated PoV-video from the EyeSeeCam and time-resolved by using EyeSeeCam's IMU data. Additionally, dimensions of the experimental setup with detailed information of a few widespread distinctive items are required to allow extracting absolute head positions from individual PoV images, though not every head orientation could be evaluated using this method. This processing step demanded very accurate image annotations, which is why they were done manually – feature detection algorithms turned out to be too unreliable. Our resulting head position traces synchronized to the VOG data and their superpositioning to gaze-in-room data yield promising visualizations, overlays on a 3D model – an office workplace 4.6.
Importantly, this method can be applied to every experimental setup, if individual views present to the EyeSeeCam user contain at least four appropriate items with known 3D coordinates for extracting absolute head positions; combined with the 3D eye calibration, mobile eye-tracking achieves recording of fully controlled eye, head, and gaze behaviour in real world experiments.

1.3.3 Applications

In **study III**, EyeSeeCam users performed a standardized sequence of typical office tasks relative to two different daylight conditions: low contrast condition with no direct sunlight as compared to high contrast condition with direct sunlight coming into the room. The $3\pi sr$ light field was captured. We aim to provide objective insights as to how luminance distribution in an office setting modulates our visual behaviour. 3D fixation distributions were computed and their variations were compared with discomfort glare source occurrence. Our results show that, while participants look more outside the window during a non-cognitive and non-visual office task, this effect is lower under the high contrast lighting conditions. As was expected by task-driven attentional guidance, fixations focus on the task area when the participants are performing a task involving visual and cognitive activities, but surprisingly this behaviour still remains to a smaller degree during contemplation and non-visual office tasks. We also found that eye and head were fundamentally differently affected by view and that this dependence was modulated by task and tool, unless participants' task was related to reading. Importantly, for some tasks head movements rather than eye movements dominated gaze allocation. Since head and body movements frequently remain unaddressed in eye-tracking studies, our data highlight the importance of unconstrained settings. Beyond assessing the interaction between top-down (task-related) and bottom-up (stimulus-related) factors for deploying gaze and attention under real-world conditions, such data are inevitable for realistic models of optimal workplace lighting and thus for the well-being of the office inhabitant.

Study IV: As already pointed out, users can complete the introduced 3D eye calibration in about a third of the time ($< 20s$) compared to the former version, what made its application particularly convenient for measuring patients with restrictions of ocular motion. This is the case with PSP patients, where we could employ the EyeSeeCam, though their ability to carry out vertical saccades is deteriorated. Thus, this eye-movement disturbance as key symptom of PSP in contrast to idiopathic Parkinson's syndrome could be reliably detected, allowing an improved sensitivity and specificity of the clinical diagnosis notably with an objective measurement device. Ocular movements were analysed during a standardized fixation protocol and in an unrestricted real-life scenario, while walking along a corridor. A prominent impairment of both saccade velocity and amplitude in PSP patients were detected enabling a reliable differentiation from PD and HCs. However, especially vertical saccades showed a stronger effect than other eye movements. Differences were more pronounced during a standardized protocol than in the real-life scenario. The standardized protocol prohibits head movements for targeting fixation points. Thus, such patients cope with their limited ocular movements in everyday life possibly by augmenting their eye-head coordination.

Our developments and findings demonstrate the practicability of wearable eye-tracking instead of restricted stationary VOG and prepare the ground for using rapid standardized fixation protocols in patients with uncertain diagnoses.

1.3.4 Pupillometry

In **study V**, we investigated perception and learning from the pupillary response of another individual during decision making, since pupil dilation is implicated as a marker of decision-making as well as of cognitive and emotional processes. During a modified version of *rock-paper-scissors*, only pre-recorded IR-videos containing a game-playing opponent's eye were presented to players as visual cue. We tested several conditions with following results:

(1) Players' win ratios (moving average, chance 33%) did not exhibit a learning curve, when they just observed the opponent's eye without specific instruction.

(2) When informed that the time of maximum pupil dilation was indicative of the opponents' choice, players performed significantly above chance on average.

(3) The learning probability apparently did not depend on other facial cues, since a reconstructed area of the pupil against a grey background let the players achieve similar performance. Thus, players indeed exploited the pupil. Over the whole stimulus footage, maximum pupil dilation was correct about the opponents' decision only in 60% of trials (chance 33%).

(4) A trial selection with only correct pupil indications was also presented to test, whether increasing this validity to 100% would allow spontaneous learning. In fact, half (5/10) of the players reached significance at an individual level in performing above chance.

These results suggest that people can perceive an opponent's cognitive decision via pupil dilation. Though for being able to spontaneously learn this relationship and conceive its utility, the visual cue seems to lack either salience or consistent appearance.
Still, explicit knowledge of pupillary response encourages to pay more attention to the dilation of another individual's pupil in everyday life instead of only performing subliminal judgements over one's personal sympathy for another's personality or concern. Even though the specific use in rock-paper-scissors has to be instructed, our results underline the possibility of pupil dilation as cue in real-life social interactions (cf. Harris et al., [24]).

Study VI explored, how pupil dilation as indicator for cognitive and emotional processes is usable to transmit a communication signal. Our aim is to construct a novel Brain-Computer-Interface for patients suffering from total LiS and thus having no other means of communication. The whole system contains a mobile pupillometry device, a

mini computer, speakers, and a monitor.
After trying different strategies to invoke a strong pupil dilation, occurring with sufficient reliability for detection on a trial by trial basis, we only succeeded with a very demanding task.
A variation of the decision timing paradigm from Einhäuser [18] and covert attention for oppositely bright half fields [7] were also tested for application, but could not sufficiently affect the pupillary response to result in a robust effect on a single trial basis. We interpret the occasionally insufficient signal-to-noise ratio as a result of the subjects' difficulty in sustaining a high level of attentional focus. When attention is not directed solely to the task, the pupil response becomes unreliable. This problem might also occur due to an intention to communicate, which leads to an interference with the decision encoding signal. Our proposed protocol for yes-no question tasks includes arithmetic problem solving for eliciting robust pupil dilations, as this method is advantageously usable without prior adaptation steps.

- Neither training nor adjustment of our system's decoding parameters were required.
- Decoding performance of all HCs' pupillary response was above 80% and reached ceiling level for the half $(3/6)$ demonstrating the reliable measurability of the targeted pupil dilation.
- Classical LiS patients' $(4/7)$ pupil decoding achieved also significant performance.

This shows that pupillary response to cognitive processes is still intact in LiS and supports the concept of pupil communication. Furthermore, we measured a patient in a minimally conscious state with the same protocol and demonstrated his ability of command-following, suggesting that our system has potential as a diagnostic tool for patients whose state of consciousness is in question.
The aim to find a total LiS patient, who is highly vigilant and induces a prominent pupil dilation while accomplishing the yes-no arithmetic task is still pending to date of this thesis' publication.

1.3.5 Collaborations beyond the Philipps-University Marburg:

- Antje Nuthmann - School of Philosophy, Psychology and Language Sciences, Psychology Department, University of Edinburgh, UK.
- Camille Chatelle & Steven Laureys - Coma Science Group, Cyclotron Research Centre, University and University Hospital of Liège, Belgium.
- Christof Koch - Allen Institute for Brain Science, Seattle, USA.
- Erich Schneider & Stefan Kohlbecher - Institute for Clinical Neurosciences, University Hospital, Munich, Germany.
- Jan Wienold & Sandra Mende - Fraunhofer Institute for Solar Energy Systems (ISE), Freiburg, Germany.
- Mandana Sarey Khanie & Marilyn Andersen - Interdisciplinary Laboratory of Performance-Integrated Design (LIPID), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
- Olivia Carter - Psychological Sciences, University of Melbourne, Parkville, Australia.

1.3.6 Author contributions

Study I **Overt attention guidance: Objects dominate features**
Conceived and designed the experiments: JS, AN, and WE;
Performed the experiments: JS and MT;
Analyzed the data: JS, WE, and AN;
Wrote the manuscript: JS, AN, and WE, all authors proof-read the article.

Study II **Mobile Three Dimensional Gaze Tracking**
Conceived the study: JS;
Software development: JS and SK;
Performed the experiment: JS and SM;
Analyzed the data: JS;
Wrote the manuscript: JS, WE, all authors proof-read the article.

Study III **Gaze behaviour: Day lighting at office work**
Conceived and designed the experiments: JS, MS, WE, and MA;
Provided technical support: JS, JW;
Performed the experiments: JS, MS, and SM;
Analyzed the data: JS, MS, and JW;
Wrote the manuscript: JS, MS, WE, MA, all authors proof-read the article.

Study IV **Ocular motor analysis in PSP patients**
Conceived the study: SM, MS, FB, WHO, GUH, and WE;

Gave access to patients: GR, MS, WHO, and GUH;
Performed the experiments: SM, GR, MS, and SD;
Provided technical support: JS;
Analyzed the data: SM, SD, and JS;
Wrote the manuscript: GR, SM, and WE, all authors proof-read the article.

Study V **How to become a mentalist**
Conceived and designed the experiments: MN, OC, and WE;
Performed the experiments: MN and JS. Analyzed the data: MN;
Contributed reagents/materials/analysis tools: MN, JS, WE, and OC;
Wrote the manuscript: MN, WE, and OC, all authors proof-read the article.

Study VI **Pupil Communication in LiS patients**
Conceived and designed the experiments: JS, OC, CK and WE;
Performed the experiments: JS and CC; Gave access to patients: SL;
Analyzed the data: JS;
Performed the clinical diagnostics: CC;
Wrote the manuscript: JS, CC, OC, and WE, all authors proof-read the article.

Fulfilled experimental methodological tasks - overview:

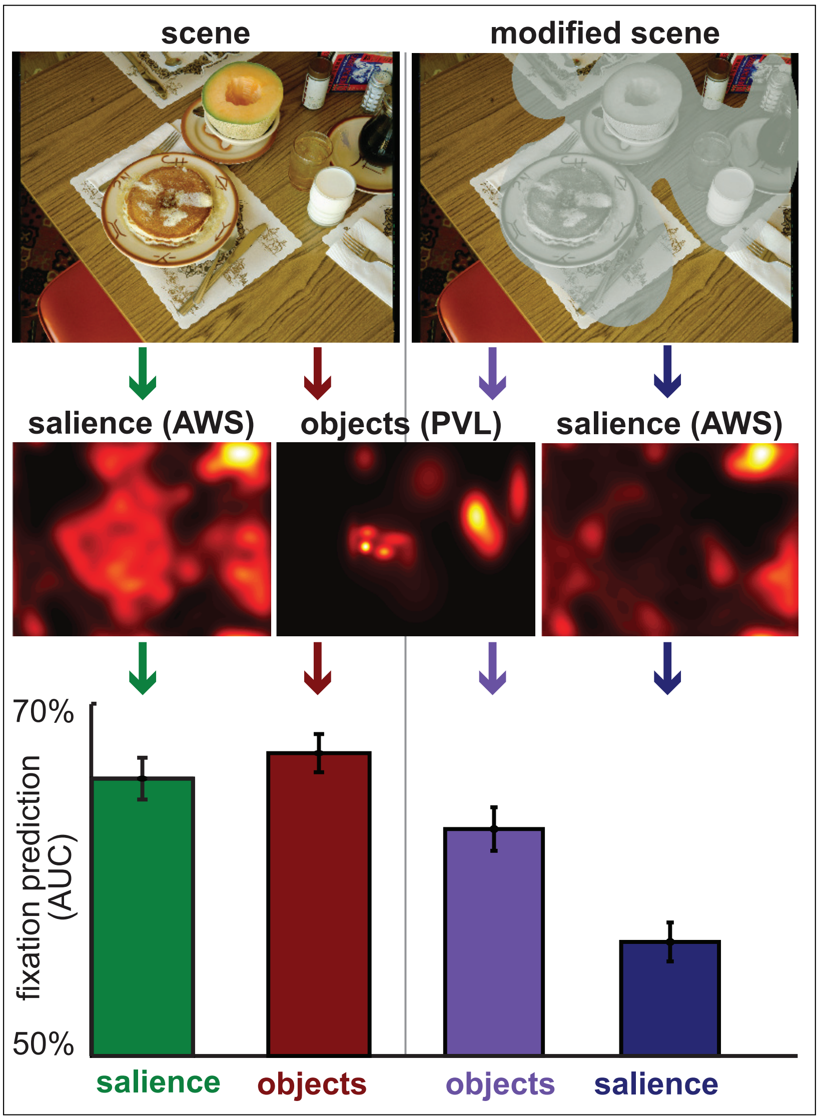
- Classical eye-tracking experiments implemented with Matlab/Psychtoolbox [10] and an Eyelink 1000 setup (SR Research [5]) as well as with EyeSeeCam software [6].
- Measuring vergence, extract fixation distance, thus 3D eye-in-head and apply result on gaze-camera positioning controller to accurately record gaze-centred videos.
- Head positioning, thus measuring gaze-in-room coordinates. Subsequently, we can directly allocate gaze to objects/targets, if those are mapped in a 3D model.
- Set up a custom camera with frame grabbing software from scratch with an integrated on-line pupil detection algorithm and synchronized stimulus presentation. Purpose:
  1. Recording VOG-videos in a customized AVI-format.
  2. Measuring patients at bedside with custimized camera-lens and IR-lighting.

[5] http://www.sr-research.com/EL_1000.html

[6] www.eyeseecam.com

Kapitel 2

Study I: Overt attention guidance: Objects dominate features



Vision Research 107 (2015) 36–48

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Overt attention in natural scenes: Objects dominate features

Josef Stoll [a], Michael Thrun [a], Antje Nuthmann [b], Wolfgang Einhäuser [a,*]

[a] Neurophysics, Philipps-University Marburg, Germany
[b] School of Philosophy, Psychology and Language Sciences, Psychology Department, University of Edinburgh, UK

ARTICLE INFO

Article history:
Received 19 June 2014
Received in revised form 4 November 2014
Available online 3 December 2014

Keywords:
Attention
Fixation
Eye movements
Salience
Proto-objects
Natural scenes

ABSTRACT

Whether overt attention in natural scenes is guided by object content or by low-level stimulus features has become a matter of intense debate. Experimental evidence seemed to indicate that once object locations in a scene are known, salience models provide little extra explanatory power. This approach has recently been criticized for using inadequate models of early salience; and indeed, state-of-the-art salience models outperform trivial object-based models that assume a uniform distribution of fixations on objects. Here we propose to use object-based models that take a preferred viewing location (PVL) close to the centre of objects into account. In experiment 1, we demonstrate that, when including this comparably subtle modification, object-based models again are at par with state-of-the-art salience models in predicting fixations in natural scenes. One possible interpretation of these results is that objects rather than early salience dominate attentional guidance. In this view, early-salience models predict fixations through the correlation of their features with object locations. To test this hypothesis directly, in two additional experiments we reduced low-level salience in image areas of high object content. For these modified stimuli, the object-based model predicted fixations significantly better than early salience. This finding held in an object-naming task (experiment 2) and a free-viewing task (experiment 3). These results provide further evidence for object-based fixation selection – and by inference object-based attentional guidance – in natural scenes.

1. Introduction

Is attention guided by objects or by the features constituting them? For simple stimuli and covert shifts of attention, evidence for object-based attention arises mainly from the attentional costs associated with switching between objects as compared to shifting attention within an object (Egly, Driver, & Rafal, 1994; Moore, Yantis, & Vaughan, 1998). Such benefits extend to search in visual scenes with 3D objects (Enns & Rensink, 1991). For more natural situations, however, the question as to when a cluster of features constitutes an "object" does not necessarily have a unique answer (Scholl, 2001) and it may depend on the context and task. In the context of visual working memory, Rensink (2000) suggested that "proto-objects" form pre-attentively and gain their objecthood ("coherence") through attention. Extending the notion of objects to include such proto-objects, attention can be guided by "objects", even if more attentional demanding object processing has not yet been completed.

While for covert attention an object-based component to attention seems rather undisputed, for the case of overt attention, defined as fixation selection, in natural scenes two seemingly conflicting views have emerged, referred to as the "salience-view" and the "object-view". The "salience-view" states that fixated locations are selected directly based on a salience map (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985) that is computed from low-level feature contrasts. The term "salience" or "early salience" in this context is used in a restrictive sense to denote feature-based effects, and is thus not equivalent, but contained in "bottom-up", "stimulus-driven" or "physical" salience (Awh, Belopolsky, & Theeuwes, 2012). Put to the extreme, the salience-view assumes that these features drive attention irrespective of objecthood (Borji, Sihite, & Itti, 2013). The salience-view appears to be supported by the good prediction performance of salience-map models (Peters et al., 2005) and the fact that features included in the model (e.g., luminance contrasts) indeed correlate with fixation probability in natural scenes (Krieger et al., 2000; Reinagel & Zador, 1999). The "object-view", in turn, states that objects are the primary driver of fixations in natural scenes (Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010). As a corollary of this view, the manipulation of an object's features should leave

* Corresponding author at: Philipps-Universität Marburg, AG Neurophysik, Karl-von-Frisch-Str. 8a, 35032 Marburg, Germany. Fax: +49 6421 2824168.
E-mail address: wet@physik.uni-marburg.de (W. Einhäuser).

http://dx.doi.org/10.1016/j.visres.2014.11.006
0042-6989/© 2014 The Authors. Published by Elsevier Ltd.

the pattern of preferably fixated locations unaffected, as long as the impression of objecthood is preserved.

The object-view is supported by two independent lines of evidence. One of them is based on the prediction of fixated locations within a scene, whereas the second one derives from distributional analyses of eye fixations within objects in a scene. With regard to the former, it is important to note that the robust correlation between fixations and low-level features, which seem to argue in favour of the salience-view, does not imply causality. Indeed, when lowering local contrast to an extent that the local change obtains an object-like quality, the reduced contrast attracts fixations rather than repelling them (Einhäuser & König, 2003), arguing against a causal role of contrast. Even though this specific result can be explained in terms of second-order features (texture contrasts, Parkhurst & Niebur, 2004), objects attract fixations and once object locations are known, early (low-level) salience provides little additional information about fixated locations (Einhäuser, Spain, & Perona, 2008). Together with the finding that object locations correlate with high values in salience maps (Elazary & Itti, 2008; Spain & Perona, 2011), it seems that salience does not drive fixations directly, but rather that salience models predict the locations of objects, which in turn attract fixations. This support for the object-view has, however, recently been challenged. In a careful analysis of earlier data, Borji, Sihite, and Itti (2013) showed that more recent models of early salience outperform the naïve object-based model of Einhäuser, Spain, and Perona (2008). This raises the question whether a slightly more realistic object-based model is again at par with early-salience models.

The second line of evidence for the "object-view" arises from the analysis of fixations relative to objects. Models of early salience typically predict that fixations target regions of high contrasts (luminance-contrasts, colour-contrasts, etc.), which occur on the edges of objects with high probability. Although the density of edges in a local surround indeed is a good low-level predictor of fixations (Mannan, Ruddock, & Wooding, 1996) and even explains away effects of contrast as such (Baddeley & Tatler, 2006; Nuthmann & Einhäuser, submitted for publication), fixations do *not* preferentially target object edges. Rather, fixations are biased towards the centre of objects (Foulsham & Kingstone, 2013; Nuthmann & Henderson, 2010; Pajak & Nuthmann, 2013). As a consequence of this bias, for edge-based early-salience models fixation prediction improves when maps are smoothed (Borji, Sihite, & Itti, 2013) and thus relatively more weight is put from the edges to the objects' centre (Einhäuser, 2013). Quantitatively, the distribution of fixations within an object is well-described by a 2-dimensional Gaussian distribution (Nuthmann & Henderson, 2010). The distribution has a mean close to the object centre, quantifying the so-called preferred viewing location (PVL), and a standard deviation of about a third of the respective object dimension (i.e., width or height). Since a PVL close to object centre in natural-scene viewing parallels a PVL close to word centre in reading (McConkie et al., 1998; Rayner, 1979), it seems likely that the PVL is a general consequence of eye-guidance optimizing fixation locations with respect to visual processing – at least when no action on the object is required: fixating the centre of an object (or word) maximizes the fraction of the object perceived with high visual acuity. A possible source for the variability in target position, as quantified by the variance or standard deviation of the PVL's Gaussian distribution, is noise in saccade programming (McConkie et al., 1998; Nuthmann & Henderson, 2010). Taken together, the existence of a pronounced PVL for objects in scenes suggests that fixation selection, and by inference attentional guidance, is object based.

Both lines of evidence for the object-view assume that object *locations* are known prior to deploying attention and selecting fixation locations. This does not require objects to be *recognized* prior to attentional deployment. Rather, a coarse parcellation of the scene into "proto-objects" could be computed pre-attentively (Rensink, 2000). If models of early salience in fact predict the location of objects or proto-objects, they could reach indistinguishable performance from object-based models, even if attention is entirely object based. The explanatory power of low-level feature models, like Itti, Koch, and Niebur (1998) salience, would then be explained by them incidentally modelling the location of objects or proto-objects. In turn, the existence of a PVL would be a critical test as to whether proto-objects as predicted by a model indeed constitute proto-objects that can guide attention in an object-based way. An early model that computed proto-objects in natural scenes explicitly in terms of salience (Walther & Koch, 2006) failed this test and showed no PVL for proto-objects, except for the trivial case in which proto-objects overlapped with real objects and the observed weak tendency for a central PVL for these proto-objects was driven by the real objects (Nuthmann & Henderson, 2010). In a more recent approach along these lines, Russell et al. (2014) developed a proto-object model that directly implements Gestalt principles and excels most existing models with respect to fixation prediction. Although a direct comparison of this model with real objects is still open, Russell et al.'s approach shows how object-based salience can act through proto-objects and can thus be computed bottom-up (and possibly pre-attentively) from scene properties.

In the present study, we test the object-view against the salience-view for overt attention in natural scenes. Two predictions follow from the object-view hypothesis.

(I) A model of fixation locations that has full knowledge of object locations in a scene and adequately models the distribution of fixations within objects ("PVL-model") does not leave any additional explanatory power for early salience. That is, salience-based models cannot outperform object-based models.

(II) Early-salience models that reach the level of object-based models do so, because they predict object (or proto-object) locations rather than guiding attention per se. Under the object-view hypothesis, any manipulation of low-level features that neither affects the perceived objecthood nor the location of the objects in the scene, will decrement the performance of the early-salience model more dramatically than that of the object-based model.

Here we test these predictions directly: using the object maps from Einhäuser, Spain, and Perona (2008) and a canonical PVL distribution from Nuthmann and Henderson (2010) we predict fixated locations for the images of the Einhäuser, Spain, and Perona (2008) stimulus set (S. Shore, uncommon places, Shore, Tillman, & Schmidt-Wulffen, 2004). In a first experiment, prediction (I) is tested on an independent dataset of fixations from 24 new observers who viewed the same Shore, Tillman, and Schmidt-Wulffen (2004) images. We compare an object-based model that incorporates the within-object PVL (PVL map) to the prediction of the Adaptive Whitening Salience Model (AWS, Garcia-Diaz et al., 2012a, 2012b), which is the best-performing model identified in the study by Borji, Sihite, and Itti (2013). In a second experiment, prediction (II) is tested by reducing saturation and contrast of the objects and testing how PVL map and AWS predict fixations of 8 new observers viewing these modified stimuli. In experiment 3, we repeat experiment 2 with a free-viewing task to rule out that object-based instructions biased the results in experiment 2.

2. Materials and methods

Stimuli for all experiments were based on 72 images from the Steven Shore "Uncommon places" collection (Shore, Tillman, & Schmidt-Wulffen, 2004; Fig. 1A), which constitute a subset of the 93 images used in Einhäuser, Spain, and Perona (2008) and correspond to the subset used in an earlier study ('t Hart et al., 2013). Our object-based modelling used the annotation data from the original study by Einhäuser, Spain, and Perona (2008), while all fixation data was obtained from an independent set of 40 new observers (24 in experiment 1, 8 in experiments 2 and 3, see Sections 2.2–2.4).

2.1. Models

All object-based models were computed based on the keywords provided by the 8 observers of Einhäuser, Spain, and Perona (2008) and the object outlines created in the context of this study. A list of all objects is available at http://www.staff.uni-marburg.de/~einhaeus/download/ObjectRecall.csv. From the outlines, bounding boxes were computed as the minimal rectangle that fully encompassed an object. In case an object had more than one disjoint part or more than one instantiation within the scene, separate bounding boxes were defined for each part and/or instantiation (Fig. 1B). Hereafter, both cases will be referred to as object "parts" for simplicity of notation. In total, the 72 images used for the present study contained 785 annotated objects consisting of a total of 2450 parts.

2.1.1. Original object map (OOM)

To test whether we could replicate the finding by Borji, Sihite, and Itti (2013) that AWS outperformed the trivial object map representation proposed in Einhäuser, Spain, and Perona (2008) on our new fixation dataset, we used the maps as defined there: the interior of each object named by any observer received a value of 1, the

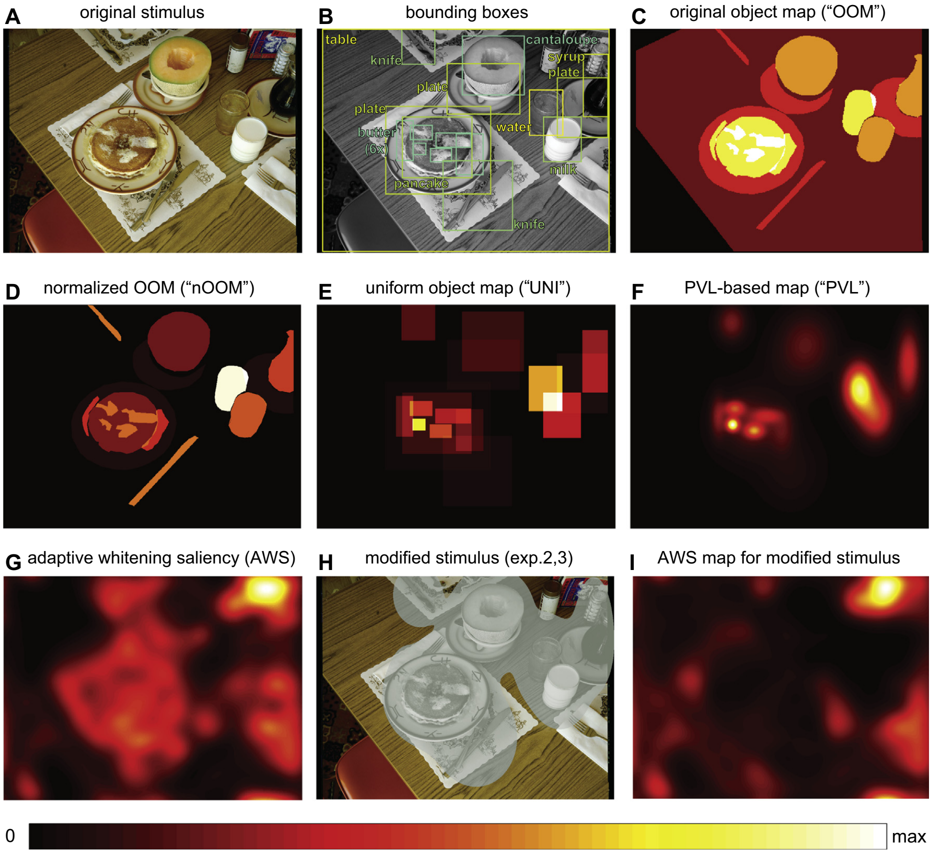


Fig. 1. Stimulus example and maps. (A) Example stimulus from the S. Shore set. (B) Bounding boxes of annotated objects, based on annotation data from Einhäuser, Spain, and Perona (2008). Same colour indicates same object. Note that some objects have multiple instances (knife, plate) or multiple disjoint parts (butter). (C) Object map as used in main analysis of Einhäuser, Spain, and Perona (2008) and Borji, Sihite, and Itti (2013). (D) Normalized version of the map in (C) ("nOOM"), in which each object is normalized to unit integral. (E) Object map based on bounding boxes from panel (B) with uniform sampling inside each object or object part. (F) Object map based on bounding boxes from panel (B) with sampling within each object or object part according to the Gaussian distribution of preferred viewing locations using the parameters from Nuthmann and Henderson (2010). (G) Salience map according to the AWS algorithm by Garcia-Diaz et al. (2012a, 2012b) for example image of panel (A). (H) Modified version of the stimulus of panel (A), as used in experiments 2 and 3. (I) AWS map for the modified stimulus. In panels (C) through (G) and (I), "hotter" colours indicate more weight, all maps are scaled to the same dynamic range for illustration. Note that maps in (D), (E), and (F) normalize each *object* to unit integral, hence large objects carry less weight per pixel, rendering the object "table" indistinguishable from background black at the colour-depth used for this illustration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

exterior a value of 0 and the resulting object maps were added per image. That is, each pixel provides a count of how many objects cover this pixel (Fig. 1C). Here and for the following models, we ignored how many observers had named an object in the original study. This "unweighted" use of the maps had the rationale that the original data serves to provide a more or less exhaustive representation of objects in the scene, rather than providing their "representativeness" for the scene. Using "weighted" maps, however, yielded qualitatively similar results.

2.1.2. Normalized original object map (nOOM)

For comparison with the other models described in Sections 2.1.3 and 2.1.4, we normalized each object of the original object map to unit integral by dividing its contribution to the OOM by the number of pixels covered by it. This resulted in a normalized original object map (nOOM, Fig. 1D).

2.1.3. Uniform object map (UNI)

To provide a baseline for the PVL-based maps as described below (Section 2.1.4), we modelled the distribution of fixations within each bounding box to be uniform. To compute a uniform object map for each image, for each object we assigned each pixel within its bounding box the value of $1/A$, where A denotes the area in pixels (i.e., bounding box width w times its height h). If an object o consists of P_o parts, the contribution of each part was in addition multiplied by $1/P_o$. The maps obtained for each object were then added to obtain the map for the scene (Fig. 1E). By definition, the sum over all pixels of an object is 1, irrespective of the number of its parts; and each object makes the same contribution to the map, irrespective of its size or number of parts.

2.1.4. PVL-based object maps (PVL)

To model fixations within an object adequately, we started with the observation that fixations can be described by a 2-dimensional Gaussian distribution (Nuthmann & Henderson, 2010). We modelled the fixation distribution for each object as a Gaussian centred at the bounding box centre, with a vertical standard deviation of 34% of the bounding box height and a horizontal standard deviation of 29% of the bounding box width, using the numbers provided in Table 2 of (Nuthmann & Henderson, 2010). Following the procedure for the uniform maps, the Gaussians for each object were normalized to unit integral or – if there were P_o parts per object – the Gaussian for each part was normalized to integral $1/P_o$. Maps of the objects within each image were then added (Fig. 1F). As with the uniform maps, each object makes the same contribution to the map, irrespective of its size or number of parts.

2.1.5. Formal description of the models

Formally, we can write the description of Sections 2.1.1–2.1.4 as follows. For the OOM, we have

$$\mathrm{OOM}(x,y) = \sum_{o=1}^{N} S_o(x,y)$$

where S_o denotes the surface of the object o and N the number of objects in the image.

$$S_o(x,y) = \begin{cases} 1 & \text{if } (x,y) \text{ falls in the boundary of object } o \\ 0 & \text{otherwise} \end{cases}$$

For the nOOM, we have

$$\mathrm{nOOM}(x,y) = \sum_{o=1}^{N} \frac{S_o(x,y)}{A_o}$$

with A_o denoting the number of pixels in object o.

For the UNI maps, we first defined for each part p of object o

$$U_{o,p}(x,y) = \frac{1}{wh} B_{o,p}(x,y)$$

where $B_{o,p}$ denotes the bounding box of part p of object o, w is the bounding box width and h the bounding box height (indices of w and h have been omitted for simplicity, but w and h are to be understood to depend on o and p).

$$B_{o,p}(x,y) = \begin{cases} 1 & \text{if } (x,y) \text{ falls in the bounding box of part } p \text{ of object } o \\ 0 & \text{otherwise} \end{cases}$$

Then we summed over all objects and parts

$$\mathrm{UNI}(x,y) = \sum_{o=1}^{N} \frac{1}{P_o} \sum_{p=1}^{P_o} B_{o,p}(x,y)$$

Similarly, for the PVL maps, we first computed for each part p of object o

$$G_{o,p}(x,y) = \frac{\exp\left[-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right]}{2\pi\sigma_x\sigma_y}$$

where (x_0, y_0) is the bounding box centre. Except for the analysis of Section 3.1.5, the standard deviations followed the Nuthmann and Henderson (2010) data: $\sigma_x = 0.29w$ and $\sigma_y = 0.34h$ with h and w denoting bounding box width and height for the respective object part.

Then we summed as above to obtain the PVL-based object map.

$$\mathrm{PVL}(x,y) = \sum_{o=1}^{N} \frac{1}{P_o} \sum_{p=1}^{P_o} G_{o,p}(x,y)$$

2.1.6. Adaptive whitening salience (AWS)

As an early-salience model we used the model that Borji, Sihite, and Itti (2013) identified to achieve best performance on our earlier data: adaptive whitening salience (AWS; Garcia-Diaz et al., 2012a, 2012b). We applied the matlab implementation as provided by the authors at http://persoal.citius.usc.es/xose.vidal/research/aws/AWSmodel.html using a scaling factor of 1.0 to the unmodified version of each image in its pixel intensity representation (Fig. 1G). Except for the scaling factor, which has a default value of 0.5 to reduce computation time for large images, default parameters as set in the authors' implementation were used. The effect of decreasing the scaling factor is explored in Appendix B.

2.1.7. Combined maps

To test whether adding an early-salience model to the object-based model improved fixation prediction, we combined normalized versions of the PVL and the AWS map. For each image, we computed a set of combined maps as

$$\mathrm{COM}_\alpha(x,y) = \alpha \frac{\mathrm{AWS}(x,y)}{\sum_{x,y} \mathrm{AWS}(x,y)} + (1-\alpha) \frac{\mathrm{PVL}(x,y)}{\sum_{x,y} PVL(x,y)}$$

In this equation α parameterizes the weight given to the early-salience model, with $\alpha = 0$ corresponding to the pure PVL map and $\alpha = 1$ to the pure AWS map.

For comparison, we also tested a multiplicative interaction between the maps

$$\left(\frac{\mathrm{AWS}(x,y)}{\sum_{x,y} \mathrm{AWS}(x,y)}\right)\left(\frac{\mathrm{PVL}(x,y)}{\sum_{x,y} \mathrm{PVL}(x,y)}\right)$$

2.2. Experiment 1

To test the models described herein, we recorded a new eye-tracking dataset using 24 observers (mean age: 24.6 years; 13

female, 11 male). Images were used in 3 different conditions, in their original colour (Fig. 1A) and in two colour-modified versions. Each observer viewed each of the 72 images once, 24 stimuli in each condition (24 unmodified, 24 with clockwise colour rotation and 24 with counter-clockwise colour rotation). Each condition of each image was in turn viewed by 8 observers. For the present study, only the unmodified images were analyzed; for completeness, the details of the colour modification and the main analysis for the colour modified stimuli are given in Appendix A. Stimuli were presented centrally at a resolution of 1024 × 768 pixels on a grey background (18 cd/m^2) using a 19′ EIZO FlexScan F77S CRT monitor running at 1152 × 864 pixel resolution and 100 Hz refresh rate, which was located in 73 cm distance from the observer. Eye position was recorded at 1000 Hz with an Eyelink-1000 (SR Research, Ottawa, Canada) infrared eye-tracking device, and for fixation detection the Eyelink's built-in software with default settings (saccade thresholds of 35 deg/s for velocity and 9500 deg/s^2 for acceleration) was used. Observers started a trial by fixating centrally, before the image was presented for 3 s. The initial central (0th) fixation was excluded from analysis. After each presentation, observers were asked to rate the aesthetics of the preceding image and to provide five keywords describing the scene afterwards. Neither keywords nor ratings were analyzed for the present purposes. All participants gave written informed consent and all procedures were in accordance with the Declaration of Helsinki and approved by the local ethics committee (Ethikkommission FB04, Philipps-University Marburg).

2.3. Experiment 2 – modified stimuli, original task

In natural scenes, low-level salience and object presence tend to be correlated, which presents one possible explanation for good performance of salience models. To dissociate low-level salience from object presence, in experiment 2 we used a modified version of each stimulus. Specifically, we calculated the median value of the PVL map, and all pixels in the stimulus that exceeded this median (i.e., half of the image with largest PVL map values) were desaturated (transformed to greyscale) and halved in luminance contrast, while keeping the mean luminance unchanged (Fig. 1H). For these stimuli, any salience model that is based on low-level features such as colour-contrasts or luminance-contrast will therefore predict fixations in the unmodified (normal saturation, high luminance contrast) area or at the boundaries between saturated and unmodified regions (Fig. 1I). The PVL map remains unchanged by the experimental manipulation (all objects remain visible). Therefore, the salience model and the PVL-based object model now differ in their predictions with regard to fixation selection in scenes. Eight new observers participated in experiment 2 (mean age: 26.5, 4 male, 4 female). Other than using the modified stimuli, the experimental methods were identical to experiment 1.

2.4. Experiment 3 – modified stimuli, free viewing

Experiment 3 was identical to experiment 2 with the exception that observers were not asked to provide keywords after each stimulus, but the next trial started with a central fixation on a blank screen after each stimulus presentation. Observers received no specific instructions except that they were free to look wherever they liked as soon as the stimulus appeared. Eight new observers participated in experiment 3 (mean age 25.3; 6 female, 2 male).

2.5. Data analysis

To quantify how well a given map (AWS, OOM, nOOM, UNI, PVL) predicted fixation locations irrespective of spatial biases, we used a measure from signal-detection theory (SDT). For a given image *i*, we pooled the fixations of all observers and measured the values of the map at these locations. This defined the "positive set" for this image. We then pooled the fixations from all other images and measured the values at these locations for the map of image *i*. This defined the "negative" set for image *i*. This negative set includes all biases that are not specific for image *i*, and thus presents a conservative baseline. In some analyses, we restricted the dataset (e.g., to one colour condition, or to the nth fixation). In these cases, restrictions were applied to positive and negative set alike. To quantify how well the negative set could be discriminated from the positive set, we computed the receiver operating characteristic (ROC) and used the area under the ROC curve (AUC) as measure of prediction quality. Importantly, this measure is invariant under any strictly monotonic scaling of the maps, such that – except for combining maps – no map-wise normalization scheme needed to be employed for making the different maps comparable.

AUCs were obtained independently for each of the 72 images. Since AUCs were obtained by pooling over observers, all statistical analysis in the main text (ANOVAs, *t*-tests) was done "by-item". This by-item analysis allows for a robust computation of AUCs pooled across observers; however, for completeness we also report "by-subject" analyses (Appendix C).

In addition to parametric tests (ANOVAs, *t*-tests), for the main comparisons we also performed a sign test. The sign test is a non-parametric test that makes no assumptions on the distributions of AUCs across images. It tests whether the sign of an effect (i.e., is model "A" or model "B" better for a given image) is consistent across images, but ignores the size of the effect (i.e., by how much is model "A" better than model "B").

To analyze the effect of salience on object selection (Section 3.1.6), we used a generalized linear mixed model (GLMM). For the GLMMs we report *z*-values, that is, the ratio of regression coefficients to their standard errors ($z = b/\mathrm{SE}$). Predictors were centred and scaled.

For the GLMM analysis we used the R system for statistical computing (version 3.1; R Core Team, 2014) with the glmer programme of the lme4 package (version 1.1–7; Bates et al., 2014), with the bobyqa optimizer. Data processing and all other analyses were performed using Matlab (Mathworks, Natick, MA, USA).

3. Results

3.1. Experiment 1

3.1.1. Object maps that consider the PVL are at par with the best early-salience model

Using all data from the fixation dataset, we test whether the prediction of fixated locations depended on the map used (AWS, OOM, nOOM, UNI, PVL). We find a significant effect of map type ($F(4,284) = 37.0$, $p < 0.001$, rm-ANOVA) on prediction. Post-hoc tests show that there are significant differences between all pairs of maps (all $ts(71) > 3.7$, all $ps < 0.001$) except between AWS and PVL ($t(71) = 1.29$, $p = 0.20$) and between AWS and UNI ($t(71) = 1.54$, $p = 0.13$). This confirms that AWS significantly outperforms naïve object-based maps. However, once the within-object PVL is taken into account, object-based maps are at par with state-of-the-art early salience maps; numerically, they even show a slightly better performance, though this is not statistically significant for the by-item analysis (Fig. 2A).

3.1.2. PVL is a necessary factor for the prediction of fixations

Already the UNI maps achieve better performance than the OOMs and reach indistinguishable performance from AWS. This

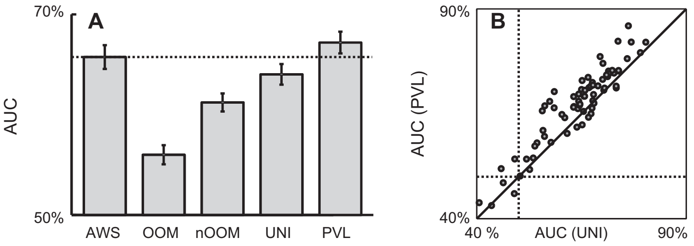


Fig. 2. PVL-based object map outperforms other maps. (A) Area under the ROC curve (AUC) as measure of fixation prediction by the five maps depicted in Fig. 1. Bars indicate mean AUC over images, errorbars s.e.m.s over the 72 images. (B) Image-wise comparison between AUCs for "PVL" and "UNI". The PVL-based map predicts fixations better than uniform sampling on the object in 59 images (points above diagonal), and worse only in 13 (points below diagonal).

raises the question whether the bulk of the benefit compared to the naïve object model arises from using bounding boxes rather than object outlines or from the normalization by object area. In other words, is there any true benefit of modelling the PVL within an object in detail? To address this question, we compare UNI maps to PVL maps image by image. We find that in 59/72 images, the maps that take the PVL into account outperform the UNI maps, with the reverse being true in 13/72 cases only (Fig. 2B). This fraction of images is significant ($p < 0.001$, sign-test). Similarly, the PVL outperforms the nOOM map (57/72, $p = 0.003$) and the OOM map (63/72, $p < 0.001$) for the vast majority of images. This shows that the benefit of considering the within-object PVL is robust across the vast majority of images.

3.1.3. Early salience provides little extra explanatory power, once object locations are known

Provided the location of objects are known, how much extra information does early salience add with regard to fixated locations? To address this question, we combine the PVL and AWS maps additively. We screen all possible relative weights ("α") of AWS relative to PVL in steps of 1%. When enforcing the same α for all images, as would be required for a model generalizing to unknown images, we find that even at the best combination ($\alpha = 52\%$, AWS adds only 2.2 percentage points to the PVL performance alone (69.4% as compared to 67.2%, Fig. 3). Even when allowing for adapting the weight for each image separately, the maximum AUC reaches 71.2%, such that the maximum possible gain by adding AWS to PVL is less than 4%. A multiplicative interaction (i.e., PVL "gating" AWS) is in the same range as the additive models (69.1% AUC).

3.1.4. Object salience and early salience are similar from the first free fixation on

In order to test whether the relative contributions of objects and early salience vary during prolonged viewing, we measured the fixation prediction of AWS and PVL separated by fixation number (Fig. 4). Even using a liberal criterion (uncorrected paired *t*-tests at each fixation number), we do not find any difference between PVL and AWS for any fixation number (all $ps > 0.17$; for fixation 0–9: all $ts(71) < 1.38$; for fixation 10 only 69 images contributed data: $t(68) = 0.82$). Neither do we find any clear main effect of fixation number (excluding fixation 0 and fixation 10) on the AUC for either PVL ($F(8,568) = 1.73$, $p = 0.09$) or AWS ($F(8,568) = 1.69$, $p = 0.10$). Thus, there is little evidence that fixation prediction by either early salience or objects changes over the course of a trial, and there is no evidence of any difference between PVL and AWS at any point in the trial.

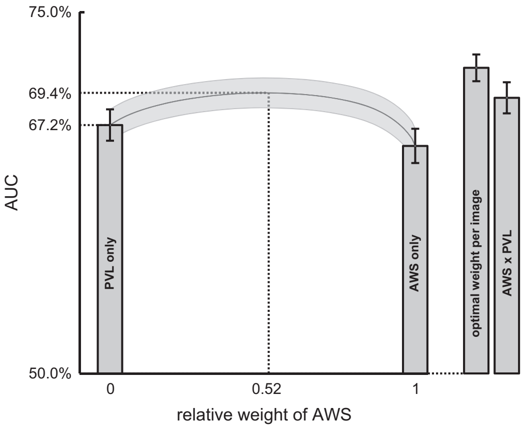


Fig. 3. Combination of PVL and AWS. AUC values for various combinations of PVL and AWS maps; *solid line*: AUCs for additively combined maps αAWS + (1 − α)PVL against weight α, mean over images (same weight for all images); *shaded area*: s.e.m. over images; *left bar*: PVL result ($\alpha = 0$; cf. Fig. 2A), *second bar from left*: AWS result ($\alpha = 1$, cf. Fig. 2A); *third bar from left*: AUC for choosing optimal weight per image, mean and s.e.m. over images; *right bar*: multiplicative interaction of AWS and PVL maps, mean and s.e.m. over images. Horizontal dashed lines indicate maximal gain in AUC for combining AWS and PVL compared to PVL alone: 2.2 percentage points.

3.1.5. Optimal PVL parameters generalize across datasets

For the analysis so far, we used the parameters of Nuthmann and Henderson (2010) to model the phenomenon of a PVL within objects. These were obtained on an entirely different stimulus set with observers performing distinct tasks (memory, search and preference judgements) and with a different setup. This raises the question, whether the average PVL generalizes across datasets. To test this, we modelled the PVL by 2-dimensional Gaussians with horizontal standard deviations ranging from 0.10 to 0.60 of bounding box width and vertical standard deviations with the same fraction of bounding box height, varied in 0.01 steps (i.e., we set $\sigma_x = \beta w$ and $\sigma_y = \beta h$, with h and w denoting bounding box height and width, and varied β). We find that prediction indeed reaches a maximum around $\beta = 0.31$ (Fig. 5A), in line with the values found in Nuthmann and Henderson (2010) and used throughout this paper. Interestingly, even the optimum value (67.15% AUC at an sd of 33% of bounding box dimensions) is very close to but slightly below the 67.19% found for the anisotropic Nuthmann & Henderson values. To test whether the result improves further for anisotropic (relative to the bounding box) PVL distributions, we vary σ_x and σ_y independently ($\sigma_x = \beta_x w$ and $\sigma_y = \beta_y h$) in 0.01 steps

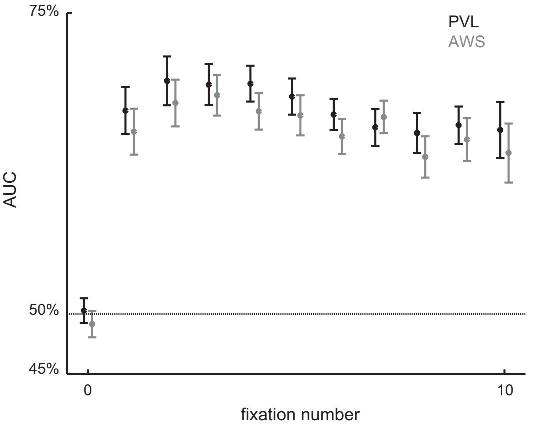


Fig. 4. Development of AUC over course of a trial. AUCs for AWS and PVL separated by fixation number, 0th fixation denotes the initial central fixation, which is not used for any other analysis.

around the Nuthmann and Henderson (2010) default value of $\beta_x = 0.29$ and $\beta_y = 0.34$. We find the AUC has its maximum off the diagonal (at $\beta_x = 0.29$, $\beta_y = 0.35$, Fig. 5B, circle) indicating an optimal PVL that is slightly anisotropic relative to the object's bounding box. This optimal AUC value is only 0.01% (percentage points) larger than the result for the original parameters (0.29, 0.34). Since these values were obtained on a different data set and experimental setup, it is tempting to speculate that the fraction of about 1/3 of bounding box dimensions, possible with a slight anisotropy, for the PVL might reflect a universal constant for object-viewing.

3.1.6. Prioritisation of objects by low-level salience

Is the prioritisation as to which object is selected, once the objects are given, biased by early salience? First, we test whether the probability that an object is fixated at all is related to its salience. To avoid obvious confounds with object area, we quantify an object's low-level salience by the maximum value of the normalized AWS map within the object surface ("peak AWS"). To keep all measures well defined, we restrict analysis to those 366 objects that consist of only one part. In addition to peak AWS, we consider two object properties, which could potentially confound peak AWS effects: object size (the number of pixels constituting the object), and object eccentricity (the distance of the object's centre of mass from the image's centre). For each observer, we allocate a "1" to a fixated object and a "0" to a non-fixated object, yielding a binary matrix with 366 × 8 (number of objects × number of observers, who viewed the respective image in unmodified colour) entries. We use a GLMM to determine the impact of the object properties on the thus defined fixation probability (cf., Nuthmann & Einhäuser, submitted for publication). The model includes the three object properties as fixed effects. With regard to the random effects structure, the model includes random intercepts for subjects and items as well as and by-item random slopes for all three fixed effects. We find a significant effect of peak AWS ($z = 4.55$, $p < 0.001$) above and beyond the effects of object size ($z = 5.09$, $p < 0.001$) and eccentricity ($z = -4.70$, $p < 0.001$). This indicates that among all objects, the objects with higher low-level salience are preferentially selected.

The analysis so far asks whether an object is fixated at all in the course of a 3 s presentation. For an infinite presentation duration, it seems likely that all objects would be eventually fixated; in turn, the salience of an object may be especially predictive for fixations early in the trial. To quantify this, we modify the analysis such that, rather than assigning a 1 to an object that is fixated at any time in the trial, we assign a 1 only to objects that are fixated at or before a given fixation number n. Computing the same GLMM for this definition and for each n, we find a significant prediction of fixation duration for each n (all $z > 3.3$, all $p < 0.001$). Z-values tend to increase with fixation number (i.e., with increasing n; Table 1). The effects of object size and eccentricity are also significant for all n, with the effect of eccentricity declining over fixation number (Table 1). These results offer a role for early salience that complements object-based fixation selection: attention is guided to objects, but among all the objects in a scene those with higher early salience may be preferentially selected.

3.2. Experiments 2 and 3 – modified natural stimuli

The PVL map performs as well as AWS, but it does not outperform the salience-based model. While this result already invali-

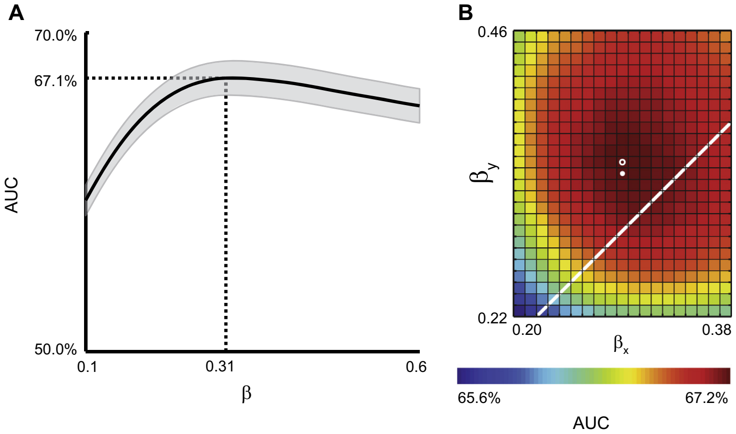


Fig. 5. Varying PVL-model. (A) AUC for varying size of the Gaussian that models the PVL within an object simultaneously for horizontal and vertical dimension, using the same ratio (β) relative to the respective bounding box dimension. Mean and s.e.m. over images; dashed lines indicate maximum. (B) Independent variation of horizontal and vertical standard deviation of the Gaussian that models the PVL, white line indicates diagonal (matching the values of panel A), white circle marks peak (0.29/0.38), white star the values by Nuthmann and Henderson (2010) used throughout the present paper (0.29/0.34).

Table 1
GLMM results: *z*-values for the fixed effects peak AWS, object size and object eccentricity on the probability of fixating labelled objects in scenes. Each column, labelled as fixation number *n*, reports data for a model that considers objects that are fixated at or before a given fixation number *n*.

Fixation number	1	2	3	4	5	6	7	8	9	10	Any
Peak AWS	3.551	3.348	4.206	4.345	4.596	4.558	4.206	4.348	4.552	4.682	4.554
Object size	3.915	4.284	4.482	4.462	4.362	4.741	5.215	5.272	5.214	5.079	5.088
Object eccentricity	−6.296	−7.279	−7.468	−6.223	−5.571	−5.389	−5.085	−5.267	−5.010	−4.850	−4.702

dates a key argument put forward against object-based fixation selection, namely that AWS was better than the naïve object-based approach (Borji, Sihite, & Itti, 2013), experiment 1 alone does not show a superiority of object-based models. In experiments 2 and 3, we dissociate the effect of objects from the effect of the features that constitute objects by manipulating low-level features at object locations.

3.2.1. Modification de-correlates AWS and PVL maps

The object-view states that early-salience models predict fixation selection through correlations of their features with object locations. To test this, we measure the correlation between AWS and PVL map values. To obtain sufficiently independent samples, we sample values from both maps on a central 11 × 8 grid of pixels 100 pixels apart (i.e., at (13,35), (13,135), … (13,735), (113,35), … (1013,735)) for each image. For the original images as used in experiment 1, AWS and PVL map values are indeed positively correlated ($r(6334) = 0.16$, $p < 0.001$).

The aim of the experimental manipulation in experiments 2 and 3 is to disentangle AWS from PVL predictions by reducing this correlation. We therefore generated new stimuli by halving contrast and removing saturation from the half of the image in which PVL was highest (Fig. 1H). This manipulation was effective in that the correlation between PVL and AWS for the modified stimuli is now negative ($r(6334) = -0.06$, $p < 0.001$) and – when individual images are considered – smaller than for the original image in 71/72 cases.

3.2.2. On modified images, PVL outperforms AWS

Using the fixation data of experiment 2 and computing AWS on the modified stimuli used, the PVL map now significantly outperforms AWS with respect to fixation prediction (AUC: 63.0 ± 1.3% vs. 56.6 ± 1.1% (mean ± s.e.m.); Fig. 6A, $t(71) = 3.41$, $p = 0.001$). On the level of individual images, the prediction of PVL is better for 51/72 images, a significant fraction ($p < 0.001$, sign-test). In the free-viewing task of experiment 3, the prediction by the PVL map remains virtually unchanged (AUC: 63.1 ± 1.3%) and is significantly better than the AWS performance (AUC: 58.1 ± 1.2%) both on average (Fig. 6B, $t(71) = 2.43$, $p = 0.02$) and for individual images (46/72, $p = 0.02$, sign-test). Both experiments show that when the prediction of an object-based model and a salience-based model are dissociated by experimentally manipulating the correlation between early salience and objecthood, object-based models outperform early salience. The result of experiment 3, in which observers had no specific instruction, furthermore rules out that the precedence of object-based fixation selection over low-level salience is a mere consequence of an object-related task.

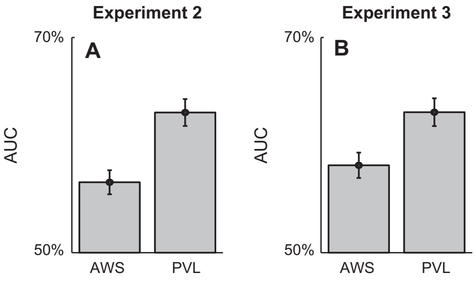


Fig. 6. Modified stimuli – PVL based map outperforms AWS. (A) Experiment 2 (object naming) and (B) experiment 3 (free viewing). Notation as in Fig. 2.

3.2.3. Dependence on fixation number

As for experiment 1, we analyzed the time course of PVL and AWS predictions. In experiment 2 (Fig. 7A), with the exception of the initial (0th) fixation, prediction is above chance for all fixations and both maps (all *ps* < 0.007, all *ts* > 2.8). Excluding the initial (0th) fixation and including all fixation numbers for which data from all images is available (1st through 9th), we find no effect of fixation number on AUC, neither for AWS ($F(8,568) = 0.53$, $p = 0.83$) nor for PVL ($F(8,568) = 1.37$, $p = 0.31$). In experiment 3, fixation durations were longer than in the other two experiments (270.6 ± 2.0 ms vs. 244.6 ± 1.8 ms and 243.3 ± 1.5 ms, excluding the initial fixation), such that from the 9th fixation on, data for some images are missing, and we only analyze fixations 1 through 8 further. For those fixations, AUCs are significantly different from chance for both maps (all *ps* < 0.001, all *ts* > 4.0). Again, we find no main effect of fixation number for AWS ($F(7,497) = 0.45$, $p = 0.87$). However, we find a main effect of fixation number for the PVL map ($F(7,497) = 4.79$, $p < 0.001$). Surprisingly, however, the prediction is best for the early fixations (Fig. 7B). The PVL model performs significantly better than AWS only for the 2nd and 3rd fixation ($t(71) = 3.30$, $p = 0.002$ in both cases), while for the other fixations performance is indistinguishable from AWS (*ts* < 1.9; *ps* > 0.06). Hence, especially early fixations, though not the first one, are guided rather by objects than by low-level salience if no object-related task is to be performed. At no time point during viewing a fixation is guided primarily by low-level salience.

3.2.4. AWS as object model

The object-view explains the performance of early-salience models by the correlation of their features to objects in natural scenes. Hence, if an experimental manipulation dissociates objects from their natural low-level features – like in our experiments 2 and 3 – the prediction performance of early-salience models should drop. Notably, we can derive an additional prediction from the object-view hypothesis: if the early-salience model is computed on the original (i.e., unmodified) stimulus, it predicts object locations. These object locations remain unaffected by the experimental manipulation. Consequently, the early-salience model computed on the unmodified image should still predict fixations on the modified image. We tested this hypothesis and found that AWS applied to the original image indeed predicts fixations on the modified image in experiment 2 better than AWS applied to the modified image itself (AUC: 66.0 ± 1.2% $t(71) = 7.41$, $p < 0.001$). The same holds for experiment 3 (AUC: 66.8 ± 1.2%; $t(71) = 6.15$; $p < 0.001$). This shows that the AWS model incidentally captures attention-guiding properties of natural scenes that still predict fixations when their correlation to the low-level features that are captured by low-level salience are removed.

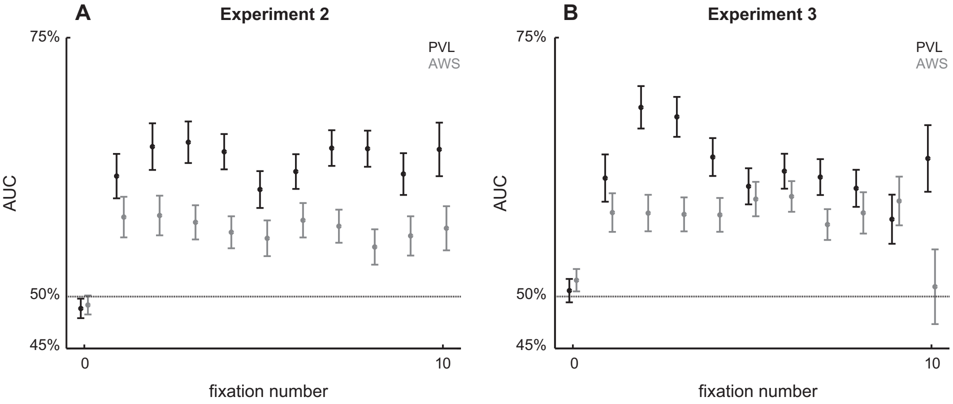


Fig. 7. Modified stimuli – development of AUC over course of a trial. (A) Experiment 2 and (B) experiment 3. Notation as in Fig. 4. Note that for the 10th fixation of panel (A), and for the 9th and 10th fixation in panel (B) not all images contributed data. Hence, statistical analysis was restricted to fixations 1 through 9 (experiment 2) and 1 through 8 (experiment 3).

4. Discussion

In this study we show that an object-based model that adequately models fixation distributions within objects (i.e., the preferred viewing location, PVL) performs at par with the best available model of early salience (AWS). The prediction by the object-based model is robust to small variations of the PVL's standard deviation and not substantially improved by any combination with the AWS model. Notably, when low-level features are manipulated while keeping objecthood intact, the object-based model outperforms the early-salience model. Together, these findings provide further support for the object-view of fixation selection: objects guide fixations and the prediction by early salience is mediated through its correlation with object locations.

If attention is indeed object-based, the question arises up to which level of detail object processing has to be performed prior to fixation selection and how such information can be extracted from the visual stimulus. The degree of object-knowledge required prior to attentional deployment has frequently been associated with "proto-objects" (Rensink, 2000). In the context of salience maps, Walther and Koch (2006) define proto-objects by extending the peaks of the Itti, Koch, and Niebur (1998) saliency map into locally similar regions. In this conception, proto-objects are a function of saliency (Russell et al., 2014). In principle, proto-objects can guide attention in two ways. First, proto-objects can be a proxy for real objects that is computed from stimulus properties. In this case, proto-objects, just like low-level salience, predict fixations through their correlation with object locations. Alternatively, proto-objects could constitute a "higher-level" feature that is causal in driving attention. Yu, Samaras, and Zelinsky (2014) provide indirect evidence for the latter view by showing that proto-objects are a proxy for clutter, and clutter is a possible higher-level feature for attention guidance (Nuthmann & Einhäuser, submitted for publication). In the present study, we show, however, that PVL-based maps outperform other object maps. For proto-objects that do not exhibit a PVL it seems therefore unlikely that they predict fixations better than real objects. An analysis testing proto-objects as defined by Walther and Koch (2006) showed that there was little evidence for a PVL for human fixations within these proto-objects (Nuthmann & Henderson, 2010). Importantly, there was no evidence for a PVL when only saliency proto-objects that did not spatially overlap with annotated real objects were analyzed. Therefore, proto-objects of that sort are not a suitable candidate for the unit of fixation selection in real-world scenes. In addition, AWS generates some notion of objecthood and can be used to extract proto-objects from a scene (Garcia-Diaz, 2011), presumably since the whitening aids figure–ground segmentation (see Russell et al., 2014, for a detailed discussion of this issue). Again, as shown by experiments 2 and 3, the features of AWS are dominated by object-based selection (PVL-based object maps), indicating that the implicit "proto-objects" of AWS do not match real objects with respect to fixation prediction. It is conceivable that the phenomenon of a PVL indeed constitutes an important property that distinguishes proto-objects from real objects, at least with respect to fixation selection. Consequently, the question whether proto-objects, whose computation is stimulus-driven, but not based exclusively on low-level features (Russell et al., 2014; Yu, Samaras, & Zelinsky, 2014), exhibit a PVL is an interesting question for future research.

Attention is likely to act in parallel with object processing rather than being a mere "pre-processing" step. There is a high structural similarity of salience-map models and hierarchical models of object recognition. Already the archetypes of such models, Koch and Ullman's (1985) salience map and Fukushima's (1980) Neocognitron, shared the notion of cascading linear filters and non-linear processing stages in analogy to simple and complex cells of primary visual cortex (Hubel & Wiesel, 1962). The computational implementation of the salience map (Itti, Koch, & Niebur, 1998) and the extension of the Neocognitron idea into a multi-stage hierarchical model (HMAX, Riesenhuber & Poggio, 1999) allowed both models to extend their realm to complex, natural scenes. Given the similarity between the salience map and HMAX, it is not surprising that more recent descendants of salience-map models, such as Itti and Baldi's (2005) "surprise", model human object recognition (Einhäuser et al., 2007) to a similar extent as HMAX itself (Serre, Oliva, & Poggio, 2007), and that in turn HMAX is a decisive ingredient in a state-of-the-art model of attentional guidance in categorical search tasks (Zelinsky et al., 2013). This modelling perspective – together with its roots in cortical physiology – argues that attentional selection and object recognition are not separated, sequential processes, but rather object processing and attention are tightly interwoven.

A challenge for both model testing and experimental research is that an object is not necessarily a static entity, but rather a perceptual and hierarchical construct that can change depending on the task and mindset of the observer. In the present study, we took a

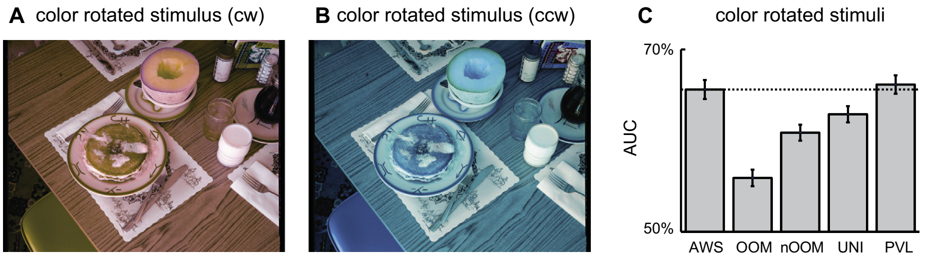


Fig. A.1. Effects of colour modification in experiment 1. (A) Colour-modified (“rotation clockwise”) version of example stimulus in Fig. 1A. (B) Colour-modified (“rotation counter-clockwise”) version of stimulus in Fig. 1A. (C) AUCs for the models of Fig. 2 for colour-modified images (48 per observer). Notation as in Fig. 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

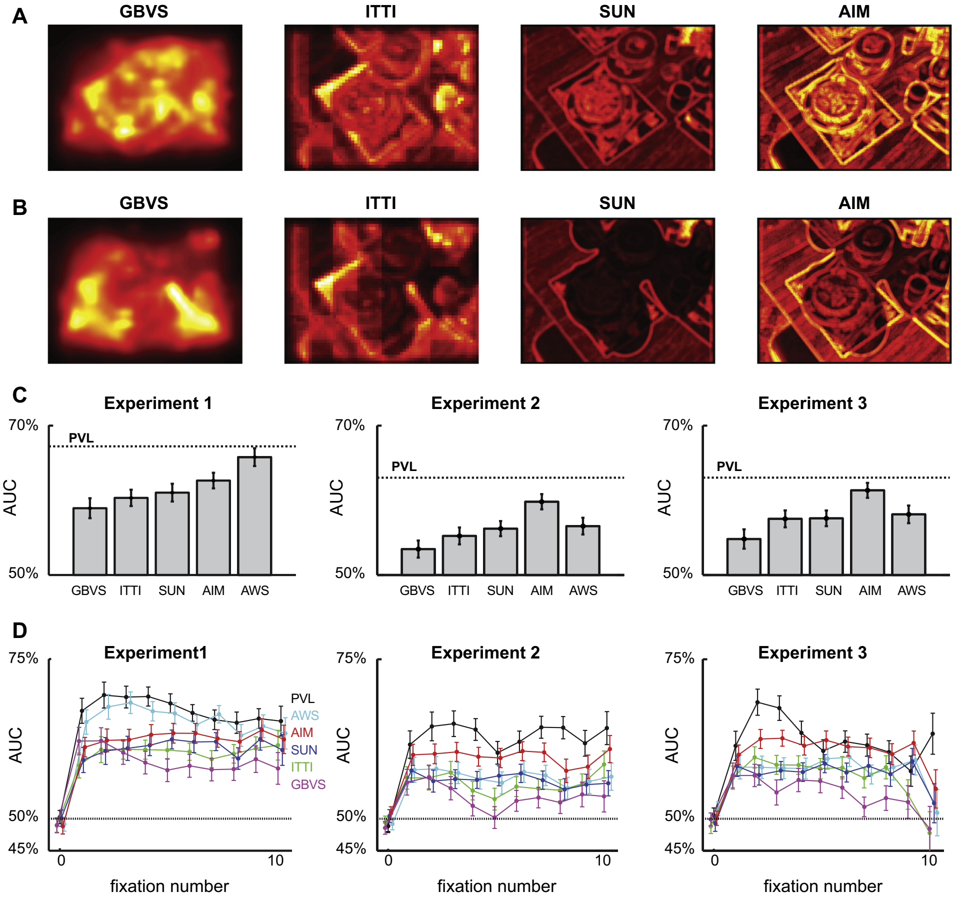


Fig. B.1. Other early-salience models. (A) Output of 4 different salience models (see text for details) on the example image of Fig. 1A. (B) Output of the models for the modified version of the stimulus used in experiments 2 and 3. Colourbar as in Fig. 1. (C) AUC for the 4 models, in comparison to PVL (dashed line) and AWS (right bar) for the 3 experiments. Notation as in Figs. 2 and 6. (D) AUC for the 4 models, AWS and PVL by fixation number. Notation as in Figs. 4 and 7. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pragmatic approach, using all the keywords provided by at least one of the 8 observers in our original study (Einhäuser, Spain, & Perona, 2008). These ranged from large background objects (“sky”, “grass”, “road”, “table”) over mid-level objects (“car”, “house”, “woman”, “cantaloupe”) to objects that are part of other objects (“roof”, “window”, “purse”, “door”). Treating all of the objects equally, as done here, makes several simplifying assumptions. First, it assumes that the parameters of the PVL are indepen-

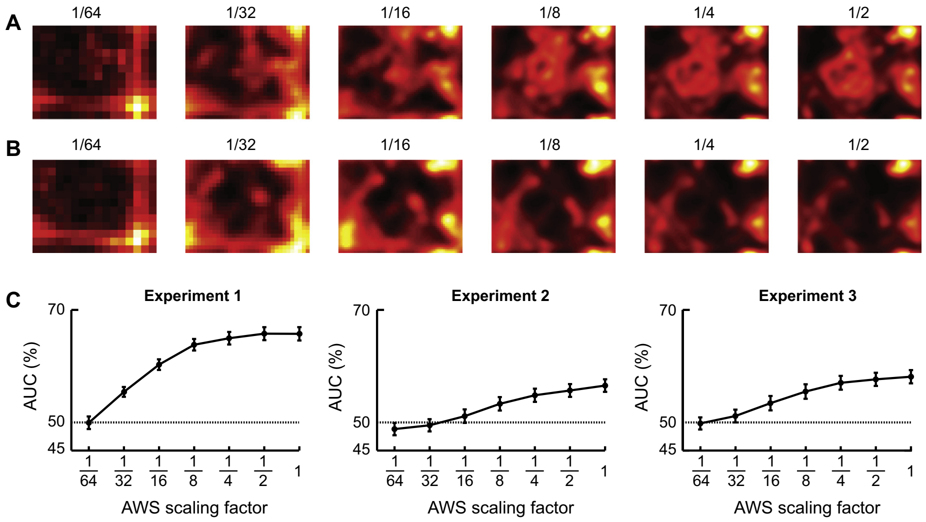


Fig. B.2. Effect of scaling factor in AWS model. (A) AWS map of example stimulus (Fig. 1A) at different scaling factors of the AWS model (given above each panel); scaling factor 1 (no scaling) is depicted in Fig. 1G. (B) AWS map of example modified stimulus (Fig. 1H). Scaling factors as in panel (A); scaling factor 1 is depicted in Fig. 1I. (C) AUC for scaling factors 1/64, 1/32, . . . , 1 for the three experiments. Rightmost datapoint in each panel (factor 1) corresponds to AWS data of the main text (Fig. 2A, Fig. 6A and Fig. 6B for exp. 1, 2 and 3, respectively).

dent of object size. Second, it puts more weight to objects that consist of multiple parts, provided that parts and object are named. Third, objects that are disjoint by occlusion are treated as separate objects. Forth, it does not respect any hierarchy of parts, objects or scene.

With regard to object size, Pajak and Nuthmann (2013) reported wider distributions of within-object fixation locations (i.e., larger variance) for smaller objects. Since – especially for very small and very large objects – the details of the presentation and measurement conditions may also have an effect on the exact distribution, we refrained from modelling this size dependence explicitly here. Since the PVL results are rather robust against the exact choice of the width of the Gaussian distribution, it is unlikely that the effects would be substantial, and – if anything – they should improve fixation prediction by the PVL maps further.

Putting more weight on objects with multiple named parts seems reasonable, at least as long as no clear hierarchy between parts and objects is established and both are likely to follow similar geometric rules to gain objecthood. By normalizing each object to unit integral, very large background objects in any case have a comparably small contribution, except in regions where no other (foreground) objects are present. In an extreme case, where the background object spans virtually the whole scene, the PVL for the background resolves to a model of the central fixation bias (Tatler, 2007), which in this view corresponds to a PVL at scene level. Indeed, the central bias is fit well by an anisotropic Gaussian for a variety of datasets (Clarke & Tatler, 2014). Note, however, that the present analysis is unaffected by generic biases through its choice of baseline.

Since disjoining objects by occlusions is rare in the present data set, this is more a technical issue than a conceptual one. Whether, from the perspective of fixation selection, occlusions are processed prior to attentional deployment (e.g., by means of estimating coarse scene layout prior to any object processing, cf. Hoiem, Efros, & Hebert, 2007; Schyns & Oliva, 2004) remains, however, an interesting question for further research in databases with substantial occurrences of such occlusions.

Finally, the issue concerning the relation between parts and objects has frequently been addressed in parallel in computational and human vision. Dating back to the works of Biederman (1987), human object recognition is thought to respect a hierarchy of parts. On the computational side, mid-level features seem ideal for object recognition (Ullman, Vidal-Naquet, & Sali, 2002), and many algorithms model objects as constellation (Weber, Welling, & Perona, 2000) or compositions (Ommer & Buhmann, 2010) of generic parts. The interplay between objects and parts is paralleled on the superordinate levels of scene and object: Humans can estimate scene layout extremely rapidly and prior to object content (Schyns & Oliva, 2004) and scene layout estimation aids subsequent computational object recognition (Hoiem, Efros, & Hebert, 2007). For human vision, this provides support for a "reverse hierarchy" (Hochstein & Ahissar, 2002) of coarse to fine processing after an initial quick feed-forward sweep (Bar, 2009). Transferring these results to the question of attentional guidance and fixation selection in natural scenes might provide grounds for some reconciliation between a pure "salience-view" and a pure "object-view". It is well conceivable that several scales and several categorical levels (scene, object, proto-objects, parts, features) contribute to attentional guidance. Indeed, recent evidence shows that the intended level of processing (superordinate, subordinate) biases fixation strategies (Malcolm, Nuthmann, & Schyns, 2014). The appropriate hierarchical level might then be dynamically adapted, and – for sufficiently realistic scenarios – be controlled by task demands and behavioural goals. The present data show, however, that for a default condition of comparatively natural viewing conditions, object-based attention supersedes early salience.

Acknowledgments

The research was supported by the German Research Foundation (DFG; grants: EI 852/1 and SFB/TRR 135). We thank the authors of the tested salience models for making their code publically available, D. Walper for support in running the experiments, and the Centre for Interdisciplinary Research Bielefeld (ZiF; FG

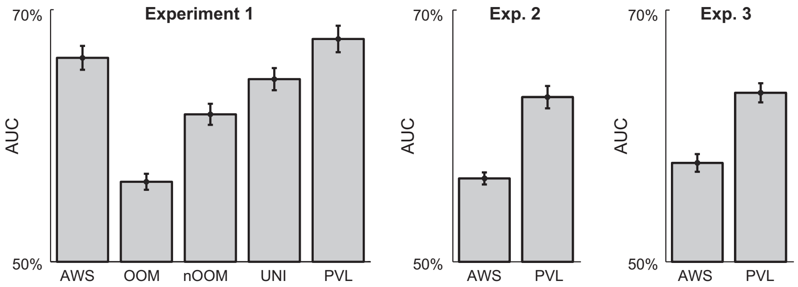


Fig. C.1. By-observer analysis. Comparison akin to Figs. 2 and 7 between object-based models and AWS, but first averaging across images and then analysing data by observer. Consequently, errorbars denote s.e.m. across observers ($N = 24$ for experiment 1, $N = 8$ for experiments 2 and 3).

"Priority") for providing the inspiring environment in which part of this study was conceived.

Appendix A. Colour conditions in experiment 1

Experiment 1 used stimuli in 3 different colour conditions: in their original colour (Fig. 1A) and in two colour-modified versions (Fig. A.1A and B): for the colour modification, images were transformed to DKL colour space (Derrington, Krauskopf, & Lennie, 1984) and each pixel was "rotated" by 90° (either clockwise or counter clockwise) around the luminance ($L + M$) axis. This manipulation (Frey, König, & Einhäuser, 2007) changes the hue of each pixel, but keeps saturation (or rather chroma) and luminance unchanged. Effectively, the manipulation swaps the $S - (L + M)$ axis (roughly: "blue–yellow") with the $L - M$ axis (roughly: "red–green") and therefore keeps the sum over these two "colour-contrasts", as used in most models of early salience, intact. That is, while the global appearance of the stimuli changes dramatically, for most commonly used salience models (including AWS) the effect of the modification is by definition negligible or absent. As the modification neither affected the AWS maps nor the PVL maps, we only used data from the unmodified stimuli for the present study. For completeness, we repeated the main analysis for the modified stimuli. As expected, the results are qualitatively very similar (Fig. A.1C) to the original colour data (Fig. 2A). This indicates that modifications to hue, at least if saturation and luminance are preserved, has little effect on the selection of fixated locations.

Appendix B. Other models and AWS parameter

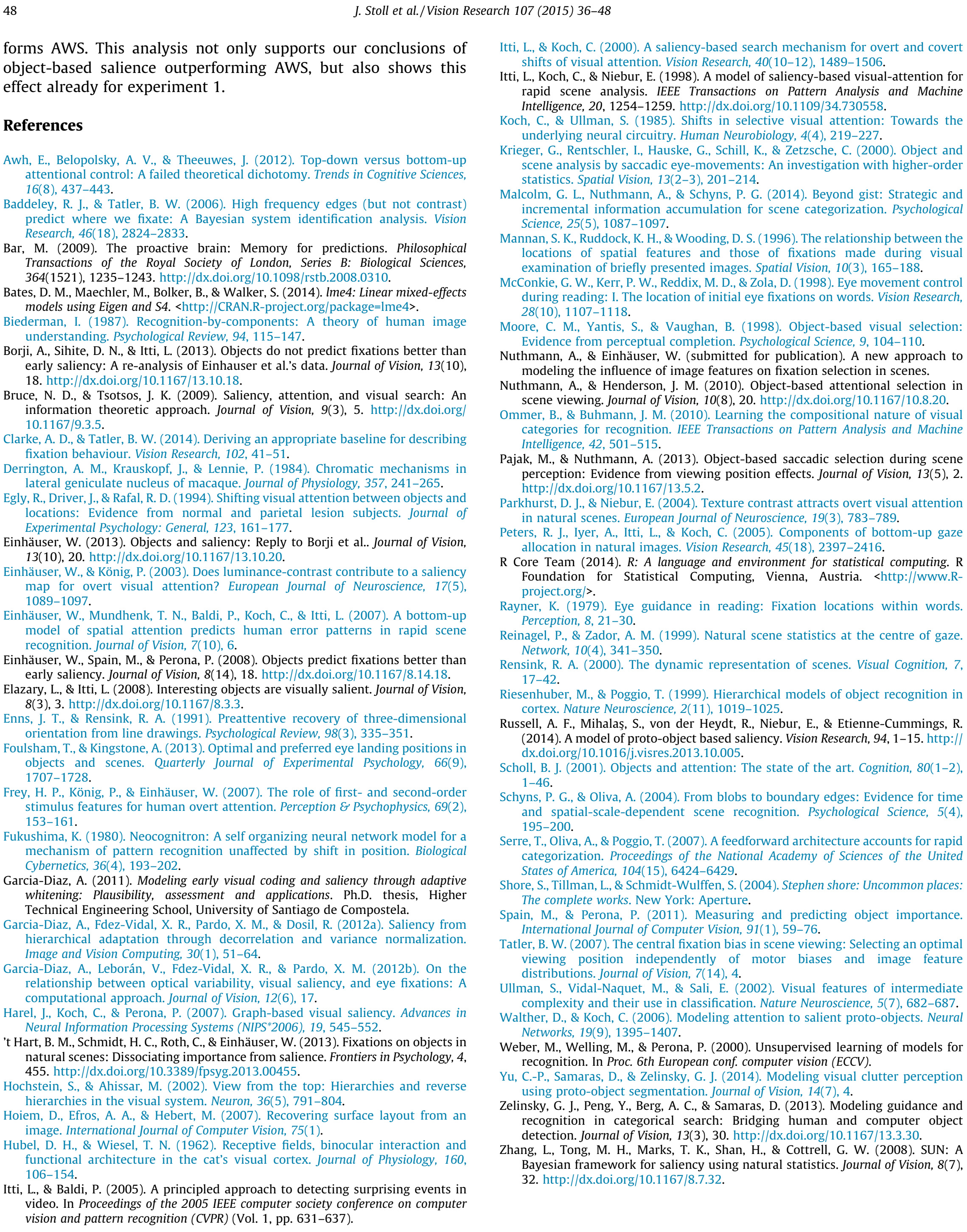
For the main analysis, AWS has been chosen as reference, since Borji, Sihite, and Itti (2013) had identified it as best performing low-level salience model on the Einhäuser, Spain, and Perona (2008) data. Our data, especially the result that the AWS model applied to the original image predicts fixations better than the model applied to the actual modified stimulus (Section 3.2.4), however, casts doubt on the characterization of AWS as an "early" salience model. We therefore tested a series of other models (Fig. B.1A and B): Graph based visual saliency (GBVS; Harel, Koch, & Perona, 2007), saliency maps following the Itti, Koch, and Niebur (1998) model in the latest (as of September 2014) implementation available at http://ilab.usc.edu ("ITTI"), the "SUN" model (Zhang et al., 2008) and the "AIM" model (Bruce & Tsotsos, 2009). Since optimizing the model parameters for our dataset is not within the scope of the present paper, we used the default parameter settings as suggested by the respective authors throughout.

With the exception of the AWS model for experiment 1 ($t(71) = 1.29$, $p = 0.20$) and the AIM model for experiment 3 ($t(71) = 1.03$, $p = 0.31$), all models perform significantly worse than the PVL map (Fig. B.1C; all other $ts > 4.0$, $ps < 0.001$). However, even in experiment 3 and similar to AWS (Fig. 7), the AIM model still performs significantly worse than PVL for the 2nd and 3rd fixation (Fig. B.1D, right; $t(71) = 2.30$, $p = 0.02$ and $t(71) = 2.06$, $p = 0.04$, respectively). This indicates that our results are not specific to the AWS model, and further supports the view that early in the trial fixation selection is object-based even in the free-viewing task of experiment 3.

The implementation of the AWS model has one parameter, the factor by which the input image is scaled (Fig. B.2). For the unmodified images of experiment 1 (Fig. B.2A), a reduction by a factor of 0.5 does not change the prediction ($t(71) = 0.14$, $p = 0.89$), if anything, it yields a tiny improvement. With further reduction of the scaling factor, prediction performance monotonically decreases, but remains above chance ($ts(71) > 6.2$, $p < 0.001$) for all tested scales down to 1/32 (Fig. B.2C). At a factor of 1/64, prediction is indistinguishable from chance ($t(71) = 0.05$, $p = 0.96$). In experiments 2 and 3, a similar picture emerges: prediction performance decreases monotonically with decreasing scaling factor and becomes indistinguishable from chance at factors of 1/16 (exp. 2) or 1/32 (exp. 3, Fig. B.2C).

Appendix C. Alternative analyses by subject

For the main analysis, we first pool fixations within an image across all observers and then perform a "by-item" analysis, with the images being the items. Pooling over observers allows us to obtain a robust estimate of AUCs for each image. This is especially critical for those analyses that separate data by fixation number, as without pooling over observers the "positive set" for the AUC would contain only a single data point. For the main analysis, which aggregates over fixations, we alternatively could compute the AUC individually for each observer, then average over images and finally perform the statistical analysis over observers for these means. For completeness, we tested this "by-subject" analysis for the comparison between AWS and PVL for all three experiments as well as for all the object models for experiment 1 (Fig. C.1). The pattern of data looks similar to the main analysis (Figs. 2 and 7). For experiment 1, there is a significant effect of object model ($F(4,92) = 19.4$, $p < 0.001$, rmANOVA). Unlike in the main analysis, all pairwise comparisons, including the one between AWS and PVL, show significant differences (all $t(23) > 3.3$, all $ps < 0.003$): PVL performs better than any other model, followed by AWS, UNI, nOOM and OOM (Fig. C.1). Similarly, the difference between PVL and AWS for experiment 2 and 3 is significant (exp. 2: $t(7) = 8.96$, $p < 0.001$; exp. 3: $t(7) = 4.00$, $p = 0.005$): PVL outper-



forms AWS. This analysis not only supports our conclusions of object-based salience outperforming AWS, but also shows this effect already for experiment 1.

References

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences, 16*(8), 437–443.

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46*(18), 2824–2833.

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 364*(1521), 1235–1243. http://dx.doi.org/10.1098/rstb.2008.0310.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. <http://CRAN.R-project.org/package=lme4>.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Borji, A., Sihite, D. N., & Itti, L. (2013). Objects do not predict fixations better than early saliency: A re-analysis of Einhauser et al.'s data. *Journal of Vision, 13*(10), 18. http://dx.doi.org/10.1167/13.10.18.

Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision, 9*(3), 5. http://dx.doi.org/10.1167/9.3.5.

Clarke, A. D., & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research, 102*, 41–51.

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology, 357*, 241–265.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*, 161–177.

Einhäuser, W. (2013). Objects and saliency: Reply to Borji et al.. *Journal of Vision, 13*(10), 20. http://dx.doi.org/10.1167/13.10.20.

Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience, 17*(5), 1089–1097.

Einhäuser, W., Mundhenk, T. N., Baldi, P., Koch, C., & Itti, L. (2007). A bottom-up model of spatial attention predicts human error patterns in rapid scene recognition. *Journal of Vision, 7*(10), 6.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14), 18. http://dx.doi.org/10.1167/8.14.18.

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision, 8*(3), 3. http://dx.doi.org/10.1167/8.3.3.

Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 98*(3), 335–351.

Foulsham, T., & Kingstone, A. (2013). Optimal and preferred eye landing positions in objects and scenes. *Quarterly Journal of Experimental Psychology, 66*(9), 1707–1728.

Frey, H. P., König, P., & Einhäuser, W. (2007). The role of first- and second-order stimulus features for human overt attention. *Perception & Psychophysics, 69*(2), 153–161.

Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36*(4), 193–202.

Garcia-Diaz, A. (2011). *Modeling early visual coding and saliency through adaptive whitening: Plausibility, assessment and applications*. Ph.D. thesis, Higher Technical Engineering School, University of Santiago de Compostela.

Garcia-Diaz, A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012a). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing, 30*(1), 51–64.

Garcia-Diaz, A., Leborán, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012b). On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. *Journal of Vision, 12*(6), 17.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems (NIPS*2006), 19*, 545–552.

't Hart, B. M., Schmidt, H. C., Roth, C., & Einhäuser, W. (2013). Fixations on objects in natural scenes: Dissociating importance from salience. *Frontiers in Psychology, 4*, 455. http://dx.doi.org/10.3389/fpsyg.2013.00455.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804.

Hoiem, D., Efros, A. A., & Hebert, M. (2007). Recovering surface layout from an image. *International Journal of Computer Vision, 75*(1).

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Itti, L., & Baldi, P. (2005). A principled approach to detecting surprising events in video. In *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR)* (Vol. 1, pp. 631–637).

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual-attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254–1259. http://dx.doi.org/10.1109/34.730558.

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

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision, 13*(2–3), 201–214.

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categorization. *Psychological Science, 25*(5), 1087–1097.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision, 10*(3), 165–188.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1998). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research, 28*(10), 1107–1118.

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science, 9*, 104–110.

Nuthmann, A., & Einhäuser, W. (submitted for publication). A new approach to modeling the influence of image features on fixation selection in scenes.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8), 20. http://dx.doi.org/10.1167/10.8.20.

Ommer, B., & Buhmann, J. M. (2010). Learning the compositional nature of visual categories for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*, 501–515.

Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception: Evidence from viewing position effects. *Journal of Vision, 13*(5), 2. http://dx.doi.org/10.1167/13.5.2.

Parkhurst, D. J., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience, 19*(3), 783–789.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45*(18), 2397–2416.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8*, 21–30.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10*(4), 341–350.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7*, 17–42.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.

Russell, A. F., Mihalaş, S., von der Heydt, R., Niebur, E., & Etienne-Cummings, R. (2014). A model of proto-object based saliency. *Vision Research, 94*, 1–15. http://dx.doi.org/10.1016/j.visres.2013.10.005.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*(1–2), 1–46.

Schyns, P. G., & Oliva, A. (2004). From blobs to boundary edges: Evidence for time and spatial-scale-dependent scene recognition. *Psychological Science, 5*(4), 195–200.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6424–6429.

Shore, S., Tillman, L., & Schmidt-Wulffen, S. (2004). *Stephen shore: Uncommon places: The complete works*. New York: Aperture.

Spain, M., & Perona, P. (2011). Measuring and predicting object importance. *International Journal of Computer Vision, 91*(1), 59–76.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 4.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*(7), 682–687.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395–1407.

Weber, M., Welling, M., & Perona, P. (2000). Unsupervised learning of models for recognition. In *Proc. 6th European conf. computer vision (ECCV)*.

Yu, C.-P., Samaras, D., & Zelinsky, G. J. (2014). Modeling visual clutter perception using proto-object segmentation. *Journal of Vision, 14*(7), 4.

Zelinsky, G. J., Peng, Y., Berg, A. C., & Samaras, D. (2013). Modeling guidance and recognition in categorical search: Bridging human and computer object detection. *Journal of Vision, 13*(3), 30. http://dx.doi.org/10.1167/13.3.30.

Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision, 8*(7), 32. http://dx.doi.org/10.1167/8.7.32.

Kapitel 3

Study II: Mobile Three Dimensional Gaze Tracking

Mobile Three Dimensional Gaze Tracking

Josef STOLL [a,1], Stefan KOHLBECHER [b], Svenja MARX [a], Erich SCHNEIDER [b]
and Wolfgang EINHÄUSER [a]
[a] *Neurophysics, Philipps-University Marburg, Germany*
[b] *Institute for Clinical Neurosciences, University Hospital, Munich, Germany*

Abstract. Mobile eyetracking is a recent method enabling research on attention during real-life behavior. With the *EyeSeeCam*, we have recently presented a mobile eye-tracking device, whose camera-motion device (*gazecam*) records movies orientated in user's direction of gaze. Here we show that the EyeSeeCam can extract a reliable vergence signal, to measure the fixation distance. We extend the system to determine not only the direction of gaze for short distances more precisely, but also the fixation point in 3 dimensions (3D). Such information is vital, if gaze-tracking shall be combined with tasks requiring 3D information in the peri-personal space, such as grasping. Hence our method substantially extends the application range for mobile gaze-tracking devices and makes a decisive step towards their routine application in standardized clinical settings.

Keywords. Mobile eyetracking, 3D gaze calibration, vergence eye movements

Introduction

Gaze allocation in natural scenes has been a subject of research for nearly a century [1,2]. Possible applications reach from advertisement [1,3,4], over basic research to clinical applications [5,6,7]. Most experimental studies, however, measure eye movements in constrained laboratory settings. While such data have some predictive quality for gaze allocation in real-world environments, plenty of qualitative features remain unexplorable for principled reasons [8]. Recently, we have introduced a wearable eye-tracking device (*EyeSeeCam*) that allows recording gaze-centered movies while an observer pursues natural tasks in a natural environment [9]. Unlike stimuli presented on a screen, however, the real world is inherently 3D. Despite of research in virtual reality (VR), where eye trackers have been coupled with VR goggles [10,11] and in remote eye-tracking applications [12], most of today's commercial eye tracking systems ignore this fact and restrict their calibration to one plane or use a recording setup that avoids parallax errors[2]. Here we propose a solution that in addition yields distance information.
To achieve robust 3D gaze-tracking, each eye needs to be represented in its own coordinate system under the constraint that the gaze directions of both eyes converge onto the fixation point. Fulfilling this condition allows the measurement of disjunctive eye movements, yielding a vergence signal for depth measurement. Here we present an extension

[1] Corresponding Author: Josef Stoll, AG Neurophysik, Philipps-Universität Marburg, Karl-von-Frisch-Str. 8a, 35032 Marburg, Germany; E-mail: stollj@physik.uni-marburg.de.

[2] e.g., ISCAN, Woburn MA, USA, http://www.iscaninc.com

of the EyeSeeCam software that allows calibration in depth. Besides the identification of fixated targets in space, this allows the system to compensate for inevitable errors of parallax arising from the distance between the gaze-controlled camera-motion device (*gazecam*) and eyes. We quantify advantages in calibration accuracy and provide a proof of principle that eye-tracking can be used to tag objects in 3D space.

1. Methods

EyeSeeCam Hardware The basic design of the EyeSeeCam has been described previously [13]. In brief: The EyeSeeCam consists of a binocular video-oculography (VOG) device and a head-mounted camera-motion device (*gazecam*), that is continuously oriented to the user's point of regard by the eye movement signals . The gazecam captures images nearly identical to the user's retina-centered visual stimulus, thus operating as an artificial eye. The whole apparatus is lightweight, mobile, battery-driven and controlled and powered by one laptop computer (Apple, MacBook). Altogether four digital cameras (Point Grey, Firefly MV) are integrated in the EyeSeeCam (Figure 1,B). The gazecam reaches the dynamic properties of the human ocular motor system - velocities above 500 deg/s and accelerations of up to 5000 deg/s^2 - with a latency of 10 ms. The workspace lies in the range of ± 30 deg for horizontal and vertical movements [9]. For minimal user restriction and high orientation accuracy, a compact, light-weight, noiseless, and precise system is realized by a parallel kinematics setup with small piezo linear motors (Physik Instrumente, Germany). The camera platform, which is connected by a gimbal joint to the supporting head mount, gets rotated via two universal joints and push rods through two parallel displaced sleds, driven by the piezo actuators (modelled in figure 1C).

Model-Free Gazecam Calibration In routine usage, the direction of the gazecam is aligned to the observer's direction of view by the following procedure. The gazecam is moved towards 25 pre-defined locations on a 5x5 grid, whose central point is approximately aligned with the user's gaze straight ahead. The user is asked to fixate the location a laser aligned with the optical axis of the gazecam pointer indicates. The mapping between known camera commands and VOG signal is interpolated and extrapolated to

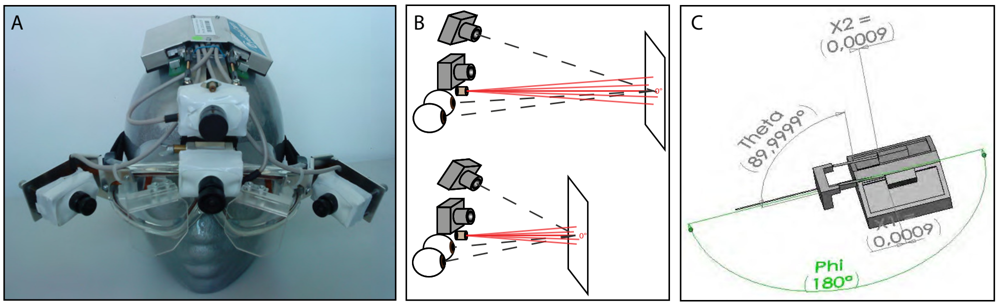


Figure 1. Setup A) EyeSeeCam device; VOG cameras (600Hz, low-res) are visible to behind the hot mirrors; gazecam(752x480 pixels, 60Hz) on top; wide-angle head fixed camera (below gaze cam) is not used in the present experiment. B) Optical axes under two distance conditions; note the difference in eye vergence and parallax between gazecam and eyes. C) Simplified mechanical model simulated by CAD software; dependence between gazecam orientation and piezo sled position. By symmetry, the perpendicular position of the platform joint plane generates its zero position; actual shift by 0.0009 is the consequence from the push rod's inclination.

arrive at a mapping valid for most of the visual field. Since the whole setup (including the laser pointer) moves with the head, this calibration method is insensitive to head movements and thus allows an accurate alignment of gaze and gazecam.

Limitations of Model-Free Calibration In the model-free calibration, the mapping between VOG and camera-motion commands is computed directly, but no information is obtained on the angular orientation of the eye in its orbit. The validity of the calibration is restricted to the distance the laser pointer is projected to. This is tolerable if the distance is always large (i.e., virtually infinite) or all operations happen near this plane. Real-world tasks, however, often require switching between peri-personal and far space, such that a depth-corrected calibration is required. Similarly, parallax errors due to the inevitable misalignment of eye and gazecam need to be corrected in near space. Since such depth corrections require vergence information, thus orientation of the eye, the model-free calibration approach is insufficient for parallax correction and 3D calibration.

Eye-in-Head Calibration In routine usage, the gazecam calibration is complemented by an independent calibration for eye-in-head orientation [14]. This has to be done for each individual and session, as the device's position relative to the observer's head may vary when the device is taken off and back on. For this calibration, an additional head-fixed laser is mounted between the eyes. The laser is equipped with a diffraction grating that generates a face-centered grid of dots on, e.g., a nearby wall (Figure 1B). The 4 first-order diffraction maxima have an angular distance of 8.5° to the central zeroth-order maximum. This central laser projection defines the primary position, i.e., the origin of the gaze angle coordinates. By use of an eyeball model, these 5 gaze directions are integrated to map the VOG signal to the coordinates over the full calibration plane. So far, eye-in-head calibration and model-free calibration were performed independently. Although this allowed mutual verification of the two methods, a depth-correct allocation of the gazecam was impossible. Thus we here present a novel method that combines both strategies to arrive at a calibration and distance estimation in 3D space.

Offset Correction To align the gazecam with the calibrated eye position, a second laser pointer is mounted in parallel to the optical axis of the gazecam. The offset of the gazecam is then adjusted by the experimenter through the graphical user interface such that this pointer matches the central (0th order) maximum of the projected calibration grid. This represents the uniquely adjusted origin of the gazecam coordinates plus a parallax correction matching the calibration distance. This procedure is independent from eye-in-head calibration. Only after offset correction is performed, the gazecam is set to follow gaze (i.e., is in *tracking* mode). During this usage of the EyeSeeCam the pointer can be used to verify the calibration against drift and to *tag* items of interest.

3D Calibration The 3D-information about the user's fixation point is observable, if both eyes are calibrated in the same reference (coordinate) system. At fixation in infinite distance, the zero direction of each eye is parallel to the calibration laser grid's central ray. For fixation at finite distances, a vergence eye movement adds to the zero direction. Since calibration must be performed at finite distances, we correct for the vergence angle occurring at each calibration point. By adding this angle, we adjust the coordinate systems such that both eyes point in parallel directions if their measured gaze coordinates are equal. For this correction, the relative positions between the eyes and the calibration points are needed and thus the projection distance and the relative position between the eyes and the source of the laser grid are required. While the distance between pupils has to be adjusted individually, inter-individual differences in the distance to the source of the diffraction pattern, i.e., the laser, are negligible. Given the distance of the eyes to the laser and the application of the offset correction, the eye-in-head calibration yields a vergence angle as the difference of the spherical angles of both eyes according to the following calculation. Equal polar angles imply fixation at infinite distance. Any vergence angle greater than zero implies that rays originating from both eyes cross. The vergence angle is the inclination in the fixation point that is equal to the difference in both eyes polar angles. Due to the symmetry of the problem, the azimuth angle does not influence this computation. Using the polar angles and the interpupillary distance, we construct two linear equations whose solution is the fixation point in 3D space.

Gazecam Vergence Correction Now we change the axis of symmetry and ignore the polar angle for the parallax correction. The gaze-direction in azimuth, the fixation distance b and the relative position of the gaze-camera's gimbal joint allow spanning a triangle, whose unknown angle γ is the difference in azimuth between eyes and gazecam rays and equals the parallax correction - the gimbal joint lies on the optical axis of the gazecam. The distance between the center of the eyeballs and the gimbal joint is used in a. γ is the difference of the averaged gaze azimuth and the angle included by gaze zero direction and the direction from eyeballs center to the gimbal joint (Figure 1B). This problem is solved in plain trigonometry by formulas valid for oblique triangles, where two sides a, b and their included angle γ are available. Thus, $\tan(\frac{\alpha-\beta}{2}) = \frac{a-b}{a+b}\cot\left(\frac{\gamma}{2}\right)$ and $\alpha + \beta + \gamma = \pi$ yield the parallax correction:
$\alpha = \frac{\pi-\gamma}{2}\arctan(\frac{a-b}{a+b}\cot\frac{\gamma}{2})$.

Parameterized Gazecam Positioning Distance variations of a user's fixation imply an adaptation of the angle, which corrects for the parallax error of gazecam orientation. This requires the positioning of the gazecam to be implemented as a function of spherical angles that stays constant given the present mechanical conditions. It needs only the geometry of the systems to be known and thus is a one time measurement. The gimbal joint as the center of rotation coincides with the optical axis of the gazecam. This facilitates the mapping from a gazecam direction to the linear positions of the two piezo actuator sleds. The function is built on the holonomic constraints of the three-point coupling behind the camera platform, whose two universal joints are displaced by the piezo actuators via push rods.

The transformation from angle to linear sled position replaces the previously employed point-by-point matching, but not the origin-offset adjustment, which depends on the individual fit of the EyeSeeCam on the user's head. This means, the novel gazecam control still would need the angle relative to its zero position, in which the gazecam is oriented parallel to the primary eye position. The issue of offset-independent positioning is solved by a separation of the orientation into a vertical rotation followed by a horizontal rotation. First, the tilt given by the azimuth is virtually executed by a symmetric displacement of both piezo actuator bars, which is related by a simple sine mapping. Then the new positions of the pivots mounted on the camera platform are computed by projecting them on the plane of polar rotation, the pan. The resulting triangle gets rotated by the polar angle and the final positions of the universal joints are reached. Their projection on a linear parallel to the linear motor direction provides the asymmetric sled displacement. Additionally taking into account their perpendicular projection enables correction pushrod inclination. This approach simplifies the previous solution [15] and is easily adaptable to future systems with different geometries due to its parametric nature. The resulting mapping from direction angles to piezo actuator bar positions ensures an accurate rotation of the gazecam. We verified the analysis with a 3D mechanical CAD program (SolidWorks, Dassault Systèmes, France).

2. Results

To compare the novel calibration method to the previous one, we performed an experiment with predefined fixation targets. The projection distance during calibration was held constant at 2 m, alike in normal usage.The measurement process included fixations at 1, 2 and 4 m with a pattern of fixation targets, whose dot size and extension was scaled proportional to the distance. The dots were distributed on the corners and edge midpoints of squares, whose edge midpoints are oriented in the cardinal directions and have an angular distance of 2.4° - like the corners of the board game mill.
To compare the accuracy of both calibration methods, the images from gazecam recordings were analyzed for deviations of the fixated target from image center. To quantify the parallax error, the statistics from the absolute vertical pixel deviation were exploited and plotted in angular degrees. We opposed the parallax error resulting from the former method to the vergence corrected error remaining when 3D calibrated. The old method shows a clear decrease of the vertical error with increasing distances (Figure 2A bottom, red diamonds). This parallax error could be diminished substantially, as the level of the vertical error is around 1° for all measured fixation distances (black circles). These results provide the proof of concept for a systematical vergence correction and mark a considerable increase for the accuracy with respect to the parallax error. To evaluate the overall performance of both calibration methods, the absolute value (direction independent) of the target-image-center deviation was analyzed. The mean over each eccentricity is plotted separately in Figure 2A, top. 3D calibration increases accuracy for the peri-personal range and also at large distances.

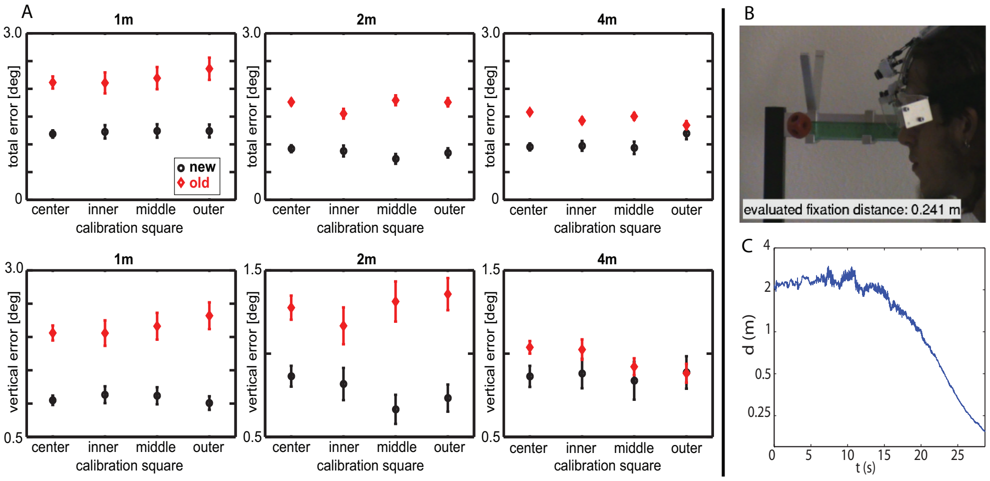


Figure 2. A) **Accuracy** Absolute error measured at validation points for different fixation distances (top of each panel) for four eccentricities (x–axes). Top: Absolute error; bottom: vertical component of error. Red diamonds: old calibration method; black circles: novel approach. Mean and sem over users (n=2), trials (n=3) and fixated points (n=8 for inner, middle and outer calibration square, 23 for center). Note: different scales for total and vertical error. B) User approaches a cube in space while fixating it, online display of distance, note the vergence correction; http://www.staff.uni-marburg.de/~einhaeus/ESC_3D.html for video C) estimated log fixation distance vs. walking time.

To investigate the 3D calibration in a truly 3D setting, we hang a small cube in space and ask one user to walk towards it. We observe an accurate measurement of fixated locations in 3D space (Figure 2B). The distances represent a usable estimator with an uncertainty of around 10% in the peri-personal range. During the approach the time course of estimated distance gets rather smooth for distances below 1m (Figure 2C). Above 1.5m, when eyes are almost parallel, the distance estimate gets increasingly instable and eventually looses in accuracy. Nonetheless, the robust and accurate estimation of 3D fixation distance in near space further validates our setup for use in tasks requiring operation in 3D space.

3. Discussion

We present a novel calibration that uses angle calibrated binocular eye movements for 3D gaze tracking and extracts a reliable vergence signal that defines a 3D point for the gazecam's target direction. In addition to fixation determination in 3D, the new procedure increases accuracy for the near field by analytic correction of parallax induced errors. The procedure simplifies calibration and makes it more efficient despite improvements in reliability and accuracy, which is of particular importance in clinical settings where large cohorts of patients have to be measured under strict time limitations.

The uncertainty of fixation estimation is inversely proportional to the distance: distance is reliably estimated in the peri-personal range and gets imprecise only above 2m. Although far distances are worse to measure based on the vergence signal, this error does not affect gaze accuracy, as the vergence correction is then converging to zero. The nearer the fixation point, in turn, the lower is the uncertainty for the distance. Over all distances, performance is sufficient to capture human behavior as device accuracy (in terms of standard error) is below usual human variations over fixations, which are about 1° [16]. In addition, the human vergence system itself seems to be rather variable - the vergence angle varies between fixations, although the target position is the same. This variation seems to be increased after large shifts in depth of fixation, where vergence movements during slow disjunctive saccades arise. Humans are expected to have their fixation point sharply tuned only after about 800ms [17]. Thus, eye-movement dynamics influence the evaluable fixation distance considerably. By measuring vergence movements, we can apply our system to address vergence dynamics in a variety of realistic scenarios.

Interacting with others in real-world space requires prediction of their intentions, which is often achieved by an estimate of the other's gaze [18]; impairment of this function is fundamental to several clinical conditions in which social interactions or the perception of others' intentions are impaired [19,20]. Using gaze direction as clinical tool [7] will thus be greatly fostered if combined with standardized clinical paradigms. As those often require action in peri-personal space or switching from peri-personal to far space, the 3D calibration presented here is inevitable for an eventual combination of gaze tracking with other tasks in standardized clinical settings. First tests in clinical populations with neuro pathological disorders (e.g., schizophrenia, Parkinson's disease) have demonstrated the applicability of the EyeSeeCam in clinical settings. In addition, the EyeSeeCam has successfully been used as a training tool for surgeons by recording gaze-centered videos from a first person perspective [21]. Such videos intuitively visualize the experienced surgeon's action planning. By applying the new calibration method the camera may be actively focused at the correct distance thereby further improve the first-person feeling. Eventually, such application as teaching tool can extend well beyond surgery - for example towards dentists and anaesthesists [22] and even to other professions, like mechanics and engineers.

Acknowledgements

We gratefully acknowledge support by the DFG under excellence cluster CoTeSys, BMBF (No. 01EO0901) (SK), grant EI 852/1 (WE), and research training unit 885 (JS).

References

[1] G.T. Buswell, *How People Look at Pictures: A Study of The Psychology of Perception in Art*, The University of Chicago Press, 1935.

[2] A.L. Yarbus, *Eye movements and vision*, Plenum Press, New York, 1967.

[3] R. Carmi, L. Itti, The Role of Memory in Guiding Attention during Natural Vision, *J Vis* ***9*** (2006), 898-914.

[4] N. Höning et al., GoodGaze: Eine Technologie aus der Hirnforschung analysiert Webseiten auf ihre Aufmerksamkeitswirkung, *up08* Lübeck, 2008.

[5] J. Vockeroth, K. Bartl, S. Pfanzelt, E. Schneider, Medical documentation using a gaze-driven camera, *Stud Health Tech Informat* ***142*** (2009), 413-416. Published by *IOS Press*.

[6] R.J. Leigh, C. Kennard, Using saccades as a research tool in the clinical neurosciences, *Brain* ***127*** (2004), 460-477.

[7] E.H. Pinkhardt et al., Differential diagnostic value of eye movement recording in PSP-parkinsonism, Richardson's syndrome, and idiopathic Parkinson's disease, *J Neurol* ***255*** (2008), 1916-1925.

[8] B.M. 't Hart et al., Gaze allocation in natural stimuli: comparing free exploration to head-fixed viewing conditions, *Vis Cog* ***17(6)*** (2009), 1132-1158.

[9] E. Schneider et al., EyeSeeCam: An eye movement-driven head camera for the examination of natural visual exploration, *Ann N Y Acad Sci* ***1164*** (2009), 461-467.

[10] A.T. Duchowski et al., Binocular eye tracking in VR for visual inspection training, *VRST* ***8*** (2001).

[11] G.P. Mylonas et al., Gaze-contingent soft tissue deformation tracking for minimally invasive robotic surgery, *MICCAI* (2005), 843-850.

[12] C. Hennessey, P. Lawrence, 3d point-of-gaze estimation on a volumetric display, *Proceedings of the 2008 symposium on Eye tracking research & applications* ***59***.

[13] J. Vockeroth et al., The combination of a mobile gaze-driven and a head-mounted camera in a hybrid perspective setup, *IEEE SMC* (2007), 2576-2581.

[14] T. Dera, G. Boning, S. Bardins, E. Schneider, Low-latency video tracking of horizontal, vertical, and torsional eye movements as a basis for 3dof realtime motion control of a head-mounted camera. *IEEE SMC* (2006).

[15] T. Villgrattner, H. Ulbrich, Piezo-driven two-degree-of-freedom camera orientation system. *IEEE ICIT* (2008), 1-6.

[16] T. Eggert, Eye movement recordings: methods, *Dev Ophthalmol* ***40*** (2007), 15-34.

[17] C. Rashbass, G. Westheimer, Disjunctive eye movements, *J Physiol* ***159*** (1961), 339-360.

[18] R. Stiefelhagen, J. Yang, A. Waibel, Estimating Focus of Attention Based on Gaze and Sound, *ACM Int Conf Proc Series* ***15*** (2001).

[19] A. Frischen, A.P. Bayliss, S.P. Tipper, Gaze Cueing of Attention, *Psychol Bull* ***133(4)*** (2007), 694-724.

[20] M.J. Green, J.H. Waldron, M. Coltheart, Eye Movements Reflect Aberrant Processing of Social Context in Schizophrenia, *Schizophr Bull* ***31(2)*** (2005) 470.

[21] E. Schneider et al., Documentation and teaching of surgery with an eye movement driven head-mounted camera: see what the surgeon sees and does, *Stud Health Tech Informat* ***119*** (2006), 486-90.

[22] C. Schulz et al., Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment, *Br. J. Anaesth.* first published online October 30, (2010) doi:10.1093/bja/aeq307

Kapitel 4

Study III: Gaze behaviour: Day lighting at office work

Sarey Khanie et al. (2013). Uncovering relationships between view-direction patterns and glare perception in a day-lit workspace,
In Proc. Of " *LUXEUROPA* ", Sep 17-19, 2013, Krakow, Poland

UNCOVERING RELATIONSHIPS BETWEEN VIEW-DIRECTION PATTERNS AND GLARE PERCEPTION IN A DAY-LIT WORKSPACE

Sarey Khanie, M.*[1], Stoll, J.[2], Mende, S.[3], Wienold, J.[3], Einhäuser, W.[2,4], Andersen, M.[1]
[1]Interdisciplinary Laboratory of Performance-Integrated Design (LIPID), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), SWITZERLAND,
[2]Neurophysics Department, Philipps-Universität Marburg, GERMANY, [3]Fraunhofer Institute for Solar Energy Systems (ISE), Freiburg i.Br., GERMANY, [4]Center for Interdisciplinary Research (ZiF), Bielefeld, GERMANY
*mandana.sareykhanie@epfl.ch

ABSTRACT

This paper presents the results of an experimental study that aims to provide objective insights as to how luminance distribution in an office setting modulates our view direction (VD) in a day-lit workspace while performing office tasks. Using the office-like test facility at Fraunhofer ISE (Freiburg i.Br., Germany) to create a range of controlled daylighting conditions, and a wearable mobile eye-tracker to measure eye and head orientation, we assessed VD distributions for subjects performing a standardized sequence of typical office tasks relative to two different daylight conditions: low contrast condition with no direct sunlight as compared to high contrast condition with direct sunlight coming into the room. Our results show that while the participants look more outside the window during a non-cognitive and non-visual office task, this effect is lower under the high contrast lighting conditions. Moreover, the focus of the VDs is on the task area when the participants are performing a task involving visual and cognitive activities.

Keywords: Discomfort glare, Eye-tracking methods, Office space lighting

4.1 INTRODUCTION

Daylight plays a determining role of how a workspace will ultimately correspond to its occupants' needs, expectations, visual comfort and appraisal of the space (VEITCH 2001, NEWSHAM, et al. 2005, OSTERHAUS 2005, WEBB 2006). Considering the substantial proportion of daily office tasks involving mainly visual activities, there is a strong need for optimized daylighting strategies that support visually comfortable workspace design (CORREIA DA SILVA, et al. 2012). One of the main challenges in this regard is maximizing daylight access while maintaining a glare-free indoor environment.

The risk of discomfort glare can be quantified using one of at least seven recognized indices conceived for that very purpose. One index of note is the Daylight Glare Probability (WIENOLD, CRISTOFFERSEN 2006), which is derived from day-lit conditions. While there are many situations where these indices disagree with each other, most are drawn upon the same four physical quantities: the glare source luminance, size and position in the field of view (FOV), and the adapted luminance.

A major limitation, shared by all known glare indices, is that they ignore the glare perception dependencies on VD, which ultimately causes ambiguities in glare indices applications (CLEAR 2012). VD is where we direct our gaze by combined shift of eye, head and body movements. The perception of discomfort glare differs greatly depending on the locations of the glare source in the FOV with respect to the VD line (LUKIESH, GUTH 1949, IWATA, et al. 1991, KIM, et al. 2009). Existing glare evaluation models – whether derived from field study methods or from High Dynamic Range image analysis – assume that the occupants' VD is fixed and is directed towards the office task area. The extension of VD to a pre-defined, static range has been proposed lately (JAKUBIEC 2012) to account for probable head and eye movements (*adaptive zone* concept) though without considering the actual VDs of the occupants. Natural visual behaviour associated with glare indicates that by blinking and changing our VD we are able to avoid glare (BOYCE 2004). As we have none or slight conscious knowledge of where our eyes are fixating at any instant, observing this natural visual behaviour in relation with glare in realistic scenarios can provide us an objective insight in understanding this phenomenon. This type of observation is now more applicable due to recent advances in eye-tracking methods (HUBALEK, SCHIERZ 2005). However, so far the relation between the VD distributions and the perception of discomfort glare has not been investigated. Based on observation on actual VD distributions, the correct angular displacement of the glare source and the actual adaptation luminance present in the FOV can be integrated into the discomfort glare assessments.

In this article, we present the results of a recent set of experiments with focus on VD distributions in relation to low and high contrast lighting conditions. Our earlier investigations showed that the change of position of VDs is towards the "view outside the window" while the participants are not performing a visually focused office task under low contrast day-lit condition (SAREY KHANIE, et al. 2013). Here, the VD distributions under two day-lit conditions were measured and compared.

4.2 METHODOLOGY

Based on our pilot study, where we saw a clear effect of four daylight conditions on VD distributions (SAREY KHANIE, et al. 2011), we set up a new series of experiments to further investigate this effect. Using the office-like rotatable test facilities, the objective in that initial study was to test a range of conditions for view outside the window and daylight conditions. Towards this end, we first investigated the VD distributions in relation to two views outside the window. The two views fell into the category that is most appreciated by the office workers (HELLINGA, 2010, TAUYCHAROEN, 2007). The results showed that the participants' VDs did not change in function to which of the two views was displayed (SAREY KHANIE, et al. 2013).

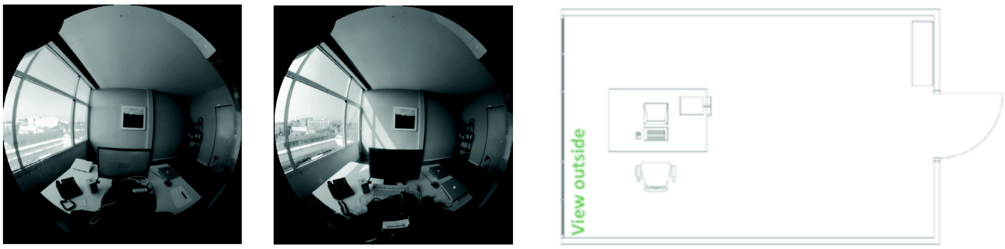


Abbildung 4.1: a) Low contrast with no direct sunlight b) High contrast with direct sunlight inside, c) The room layout

Second, we compared the view-direction distributions under two lighting conditions: low contrast with no direct sun in the room and high contrast with direct sun in the room (Fig. 4.2a,b) that are discussed in further detail in the next sections.

4.2.1 Experimental setting

The experiments were done in an office-like side-lit module, located on top of a four-story building at Fraunhofer ISE (Freiburg i.Br., Germany). The module is 360° rotatable so as to allow repeatable experiments for varying sun positions. The office layout is a single workstation pertinent to standard space requirements (NEUFERT, 2012) (Fig. 4.2c). The glazing type is a double glass with a light transmission of 54 %, a U-value of $1.1W/m^2K$, and a total solar energy transmission of 29 %. Indoor luminance variations were recorded every 30 second using luminance mapping with HDR imaging techniques with two calibrated luminance cameras equipped with fish-eye lenses. The cameras were situated above the participants' head and were adjusted according to each participant's height when seated. The eye movements were measured by means of a mobile eye-tracker, EyeSeeCam (SCHNEIDER, et al. 2009), that records both eye and head movements for accurate VD positions in the 3D space.

4.2.2 Test Procedure

Each test was divided into three task blocks where three different task supports (monitor screen, paper, phone) were used to aid the office task (Fig. 4.2.2a, b and c). In each task block the participant performed a standardized (ISO/FDIS 9241-303 2008, LEGGE 2006, SIVAK 1989, ÖSTBERG, et al. 1975) office task consisting of four main phases: "Input", "Thinking", "Response" and "Interaction". The office task sequence was designed to allow for a combination of visually highly demanding (Input) and of non-visual office task activity (Thinking), while maintaining a realistic flow. In office worker's age group, 33 participants, among them 5 females and 28 males, were recruited under consent from the Fraunhofer-ISE staff to participate in the experiment. At the beginning of each trial, to allow for a similar visual adaptation interval to the indoor light, the participant entered from the outside first through the adjacent room, and then to the test scene. Thereafter, the eye-tracker was calibrated for each participant and demographic data was gathered. Each participant's head position in the room was measured in order to obtain an accurate VD measure in the 3D space.

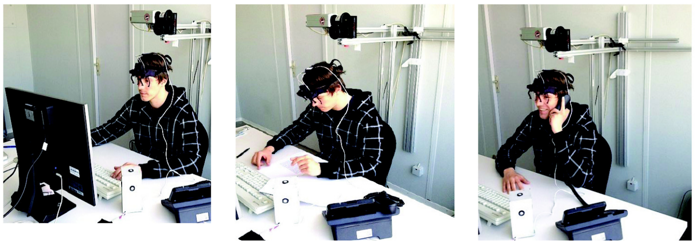


Abbildung 4.2: Task supports: a) monitor screen, b) paper and, c) phone

4.2.3 Daylight conditions

The experiments were performed under two different daylight conditions (Fig. 4.2a, b). The low contrast condition is defined as a situation where there was no direct sunlight coming into the room and the sky condition was either overcast or clear. The high contrast condition is defined as a situation where there was direct sunlight coming into the room and the sky condition was clear. Each measurement set was categorized based on observations made at the time of the experiment. Thereafter, based on the photometric measurement evaluations using *Evalglare* (WIENOLD, CHRISTOFFERSEN 2006) we refined the groupings of the participants (Fig. 4.2.3a,b, and c).

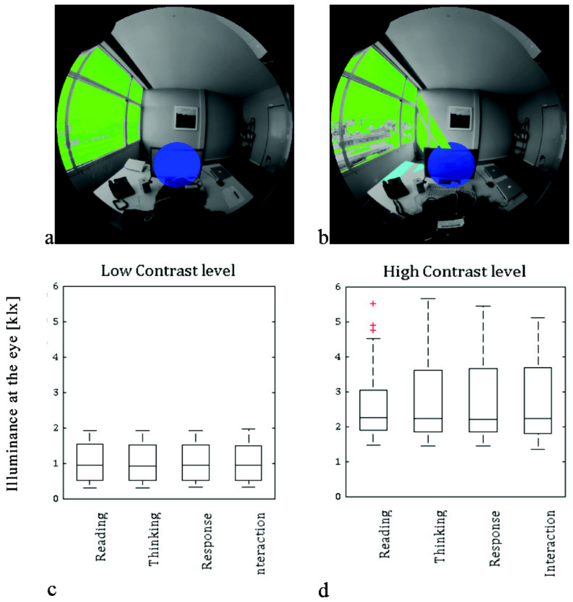


Abbildung 4.3: Daylight conditions: Glare evaluations were used for refining the groupings: a) Low contrast, b) High contrast; c, d) The box-plot shows illuminance variations reaching the eye at a vertical level.

4.3 RESULTS

4.3.1 Eye-tracking results

We studied the effect of the two daylight conditions in the analysis of the VD data. The radial standard deviation of the VD distributions, which is an appropriate measure to demonstrate the general spatial tendencies, was used in the analysis.

To quantify the effects of the independent factors we made an Analysis of variance (ANOVA) on the radial standard deviation of the VDs under each task phase. The factors were the lighting conditions (low contrast, high contrast) and the task supports (monitor screen, paper, phone). During the "Input" phase the effect of lighting condition is small (F=0.76, $p<0.05$), though the task-support's effect is apparent (F=17.16, p=0). There was an effect of lighting condition under the "Thinking" phase (F= 4.58, $p<0.05$). There was neither any effect of lighting conditions nor of the task support during the response phase. The last phase was the interaction phase where we found an effect of task-support (F=8.66, p<0.001).

4.3.2 Dominant view direction (VD) distributions

The dominant VDs represent the direction that the participant has looked at the most during the task phase. This was determined by organizing the VD data in matrix of bins of 5° spread and selecting the maximum values. Here we compare the two distinct phases of "Input" and "Thinking" being respectively the most and the least visually demanding phases. As shown in the ANOVA results, the VDs were mainly determined by the task-support during the "Input" phase under both light conditions (Fig. 4.3.1). The VDs during the thinking phase are more dispersed (Fig. 4.3.1). Our results also show that the focus on the task area is less apparent for the telephone call. Even though the phone is situated on the left hand side of the participant, there is a clear tendency towards the inside of the room (right) under the high contrast lighting conditions (Fig. 4.3.1a). During the monitor and phone task blocks, the tendency of the VDs under the low contrast levels is mainly towards the view from the window but on the other hand under the high contrast level lighting conditions, this tendency is lower (Fig. 4.3.1b). This result does not apply to the paper task block due to the presence and availability of the task support at all times.

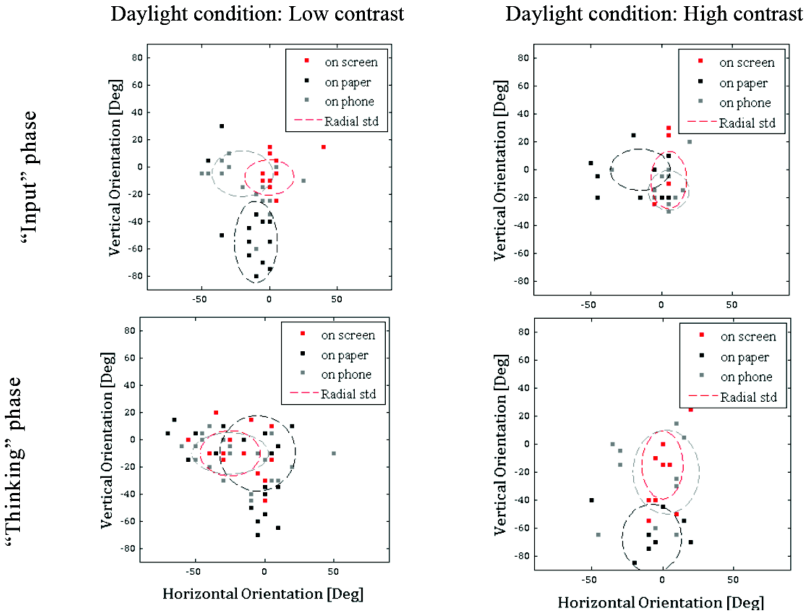


Abbildung 4.4: During the "Input" phase the VD is more focused on the task area under both light conditions, where as during the "Thinking" phase the VD is more dispersed over the space.

4.4 DISCUSSION AND CONCLUSION

The proposed work seeks to start guiding the design of workspaces in a new direction with regard to visual comfort by integrating eye-tracking methods to – ultimately – construct a dynamic model

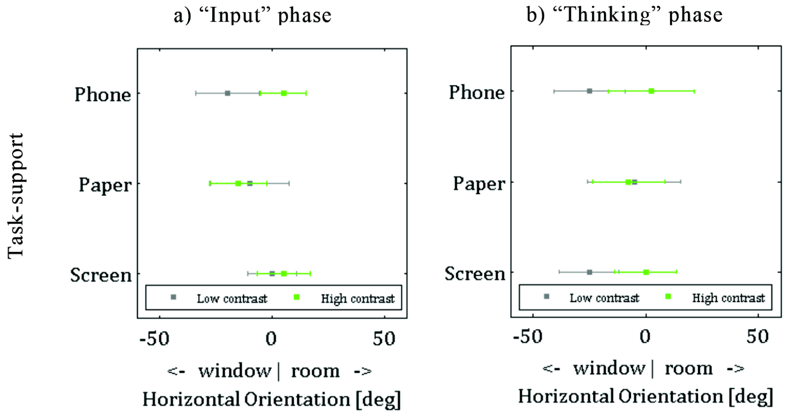


Abbildung 4.5: The mean and radial standard deviations are compared here during: a) the "Input", b) the "Thinking" phase.

of glare through an uncovering of inter-dependencies between perceived comfort, VD and lighting conditions. The hypothesis is that a certain range of luminance affects the VD, thus creating predictable patterns over a luminous space.

In this study, we investigated VD distributions under two different lighting conditions while the participants were performing standardized office tasks. The comparisons show that when the participants are not engaged in any visually focused task and the presence of the task support is minimal, the VDs are inclined towards the view outside the window under the low contrast lighting conditions, but this tendency is less apparent and sways more towards the inside of the room under high contrast lighting conditions. Based on our current findings, we will run a second series of experiments including a larger number of participants for statistical validity in which we will include participants' subjective assessments of the lighting conditions. Finally, analysis needs to be done to integrate dynamic VDs with participants' subjective assessments in the discomfort glare assessments.

ACKNOWLEDGEMENT

The authors were supported by the École Polytechnique Fédérale de Lausanne (EPFL), and the German Research Foundation (DFG) through grant EI 852/1 (WE). The authors would like to thank Fraunhofer ISE for granting access to the test room module and equipment.

REFERENCES

BOYCE, PR. 2003. Human Factors in Lighting, 2nd ed. New York: Taylor and Francis;

CORREIA DA SILVA, P. LEAL, V. ANDERSEN, M. Influence of shading control patterns on the energy assessment of office spaces, Energy and Buildings 2012, 50, 35-48;

LUCKIESH, M. GUTH, SK. Brightness in Visual Field at Borderline Between Comfort and Discomfort (BCD). Illuminating Engineering Society 1949, 44, 650-670;

ISO/FDIS 9241-303. Ergonomics of human-system interaction)—Part 303: Requirements for electronic visual displays 2008;

HUBALEK, S., SCHIERZ, C. LichtBlick – photometrical situation and eye movements VDU work places. Lux Europa. Berlin, 2005;

IWATA, T., SOMEKAWA, N., TOKURA, M. SHUKUYA, M. KIMURA, K. Subjective responses on discomfort glare caused by windows. Proceeding of 22nd Session of the CIE Division 3. Melbourne, Australia 1991,108-109;

JAKUBIEC,A. REINHART,C. The' adaptive zone' -A concept for assessing discomfort glare throughout day-lit spaces. Lighting Research and Technology 2012, 44,149-170;

LEGGE, GE. Psychophysics of reading in normal and low vision. Mahwah, New Jersey 2006: Lawrence Erlbaum Associates. Inc. Publishers;

SIVAK, M. FLANNAGAN, M. ENSING, M. SIMMONS, C. Discomfort Glare is Task Dependent. Transportation research institute, Michigan, Report No. UMTRI-89-27, 1989;

KIM, W. HAN, H. KIM, JT. The position index of a glare source at the borderline between comfort and discomfort (BCD) in the whole visual field. Building and Environment 2009, 44, 1017-1023;

NEWSHAM, G. ARSENAULT, C. VEITCH, J. TOSCO, A. DUVAL, C. Task lighting effects on office worker satisfaction and performance, and energy efficiency.LEUKOS 2005,1,4,7-26;

NEUFERT, E. NEUFERT P. 2012. Architects' Data 235-236. Blackwell publishing Ltd. OSTERHAUS, W. Discomfort glare assessment and prevention for daylight applications in office environments. Solar Energy 2005, 79,148-158;

CLEAR, R. Discomfort glare: What do we actually know? Lighitng research and technology 2012, 0, 1-18;

SCHNEIDER, E. VILLGRATTNER, T. VOCKEROTH, J. BARTL, K. KOHLBECHER, S. BARDINS, S. EyeSeeCam: an eye movement-driven head camera for the examination of natural visual exploration. Annals Of The New York Academy Of Sciences 2009, 1164, 461-467;

STONE, P. A model for the explanation of discomfort and pain in the eye caused by light. Lighting research and technology 2009, 41, 109-121;

SAREYKHANIE, M. 'T HART2, B.M. STOLL, J. ANDERSEN, M. EINHÄUSER, W. Integration of eye-tracking methods in visual comfort assessments, Proceedings of CISBAT 11: CleanTech for Sustainable Buildings - From Nano to Urban Scale. Lausanne, Switzerland. 2011, 14-15;

SAREY KHANIE, M. STOLL, J. MENDE, S. WIENOLD, J. EINHÄUSER, W. ANDERSEN, M. Investigation of gaze patterns in day-lit workplaces: using eye-tracking methods to objectify VD as a function of lighting conditions. Proceedings of CIE Centenary Conference "Towards a New Century of Light", Paris, France, 2013, 250-259;

TUAYCHAROEN, N. TREGANZA, P. View and discomfort glare from windows. Lighting Research Technology 2007, 39, 2, 185-200;

VEITCH, J. 2001. Psychological processes influencing lighting. Illuminating Engineering Society 2001, 30, 124-140;

VOS, J. Reflections on glare. Lighting research technology 2003, 163-176;

WEBB, A. Consideration for lighting in the built environment: Non-visual effects of light. Energy and building 2006, 38, 721-727;

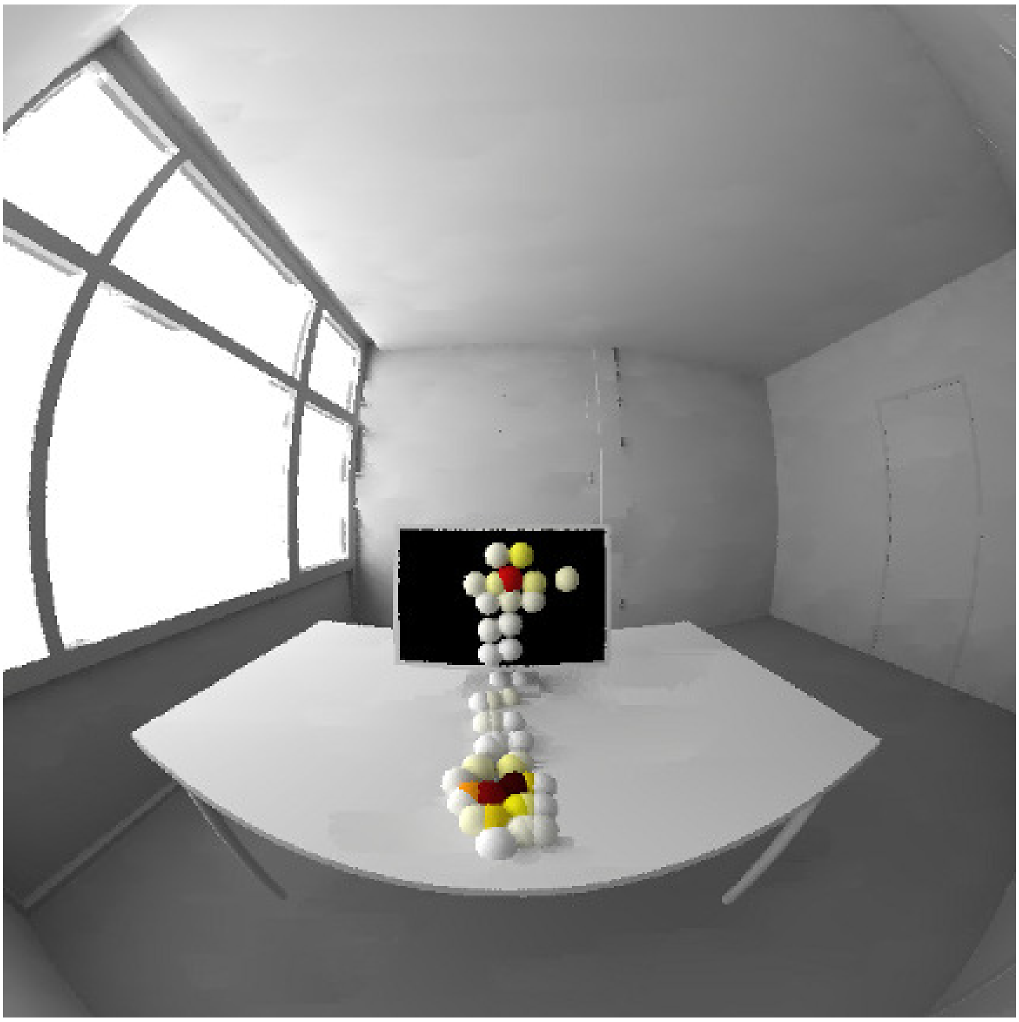
WIENOLD, J. CHRISTOFFERSEN, J. 2006. Evaluations methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. Energy and Buildings 2006, 38, 743-757;

ÖSTBERG, W. STONE, P. , BENSON, R. Free magnitude estimation of discomfort glare and working task difficulty. University of Göteborg: Göteborg psychological reports SWEDEN,Göteborg, 5. No.15. 1975;

4.5 Gaze-in-room approach

The following describes the approach to extract room-referenced gaze directions developed further compared to the proceedings paper in 4.2. Further data processing required a joint data format for gaze and luminance data, for what a 3D model appeared to be most suitable. Accordingly, gaze-in-room data are depicted as overlay on a 3D model (see figure 4.6).

Abbildung 4.6: Visualization of gaze-in-room data as overlay on a 3D model. Affixed balls illustrate the occurence of fixations at respective locations with their frequency color-coded for each location - the darkest ball marks the most dominant fixation location. Performed Task: Writing a text on the computer.



4.5.1 EyeSeeCam: technical features

The EyeSeeCam (ESC) is a state-of-the-art lightweight head mounted eye tracker [53] equipped with three cameras. This eye-tracker measures both eyes pupil positions with two IR-sensitive cameras, i.e. video-oculography (VOG). The ESC software transforms these pupil detection data in real-time into head-referenced eye positions, which is referred to as eye-in-head coordinates in this paper. Additionally, an integrated inertia measurement unit (IMU) records infinitesimal head-in-room movements, i.e. rotation velocities and translative accelerations. A camera located on the forehead records point of view (PoV) videos to an AVI format.All data streams are synchronised, so that they are with the PoV video, such that their informations are mutually combinable. The mobile ESC minimally limits the participant's movements and allows for natural exploration of the scene. In essence, we processed the following raw data:

- Eye-in-head positions from VOG at a frame rate of $221Hz$ which rely on a prior calibration [57] with a 5 point laser diffraction pattern that was performed successfully for each participant.

- A scene video recorded with a wide angle PoV camera, that is configured to a resolution of 366×216 pixels at a sampling rate of $25Hz$. A camera calibration [8] and the image pixel where the centre point of the 5 point calibration pattern was projected to connect the eye-in-head data to the scene PoV.

- Angle velocities and translative accelerations from the IMU deliver relative head rotations and up to a certain limit relative head translations, both at synchronised $221Hz$. These are applied to interpolate between absolute head orientations and positions, that we extracted approximately every 2nd second by 3D-pose estimation from the PoV camera.

Based on these data, a conversion to room referenced coordinates for head and gaze is realised.

4.5.2 ESC coordinate system

The ESC coordinate system has to cope with coordinates for eye orientation in head together with the position and orientation of the head relative to the set-up. The eye-in-head referenced coordinates describe gaze directions with respect to the participants head. This is what the ESC records as eye positions and will be related to the room coordinates of the measurement set-up in the following. Eye-in-head coordinates are represented in spherical

coordinates with a radius representing the fixation distance and two polar angles, whereas a radius of 0 would point in between both eye balls and the origin direction $(0°, 0°)$ is set by the projection direction of the central dot in the calibration pattern.Thus the gaze origin directs straight ahead from the participant and deviates only little but well controlled from the PoV camera's optical axis as we localized the central dot in the video from each participants calibration and evaluated the offset angle between both with the camera calibration toolbox function *normalize()*. The offset angle is added to the eye-in-head angles for further gaze-in-room processing.
The head-in-room referenced coordinates are aligned with the set-up such that a head orientation straight ahead when sitting at the desk with the PoV camera's optical axis parallel to the y room reference axis represents the origin direction.Head orientations are notated analogously to Euler angles with the polar angle ϕ between $-180°$ and $180°$ for head rotations around the vertical room axis, the azimuth angle θ between $-90°$ and $90°$ for down and upward directing and the angle ψ for sidewards head tilts.Thus, for example while setting each other angle to zero, ϕ maps rotations around the z room reference axis, θ around the x axis and ψ around the y axis respectively.

4.5.3 Absolute head orientation and position determined by video analysis

The process of obtaining room referenced head- and gaze orientations constitute of several steps. The first step was to derive the three-dimensional position and orientations of the head-in-room coordinates from the PoV videos. We manually labelled 4 key points within each processed frame. The key points were taken from a selection of 12 prominent corners, whose 3D room coordinates are known from the CAD model of the setup. The key points are for example the corners of the computer screen, the desk or the window. Furthermore these key points covered well the $3\pi sr$ field of the LMK cameras and they were individually chosen from our selection to be not too far eccentric in the scene to accounting for the error through a slightly deficient camera calibration, that increases with eccentricity. The image pixel positions of the 4 key points were acquired with the Matlab function *ginput()*. The resulting 2D points and their corresponding 3D room coordinates were feed into the camera calibration function *computeExtrinsic()*, that iteratively finds the camera pose from 3D-2D point correspondences by regarding the given intrinsic camera calibration parameters (i.e. solve PnP-problem, [1]). The output of *computeExtrinsic()* contains a translation vector $\vec{t}_{Head}$ and a rotation matrix $\vec{R}_{Head}$. The 3 angles of the absolute head orientation in the set-up, as defined in 4.5.2, are extracted from $\vec{R}_{Head}$. The absolute head position is computed through $-\vec{R}^{\intercal}_{Head} \cdot \vec{t}_{Head}$. This procedure is not applicable for head orientations where the image frames of the PoV video either does not contain enough key reference points, which also implies that the participants oriented the head without focusing its attention on one of the relevant objects in the set-up during that moment. Consequently, the scene videos had to be screened for suitable images. In the final run for the present data conversion, the video screening for suitable images were done with a one image per two seconds resolution. After measuring the pixel properties of the first image including the reference point, the video screening continued with an image that was recorded 30 second after the last measured image.

4.5.4 Relative head orientation through integrated angle velocities

The head orientations resulting from scene video analysis are essential for determining absolute heading and gaze directions, but they are only produced for a few time points, relative to the number of data points recorded by the ESC. For a continuous time series from head orientations, the angle velocities measured given by the IMU were integrated over time. Their sampling rate is adjusted to match the pretty fast sampling of the VOG cameras, which allows a direct bin-wise synchronisation of head and eye data. The angle velocities represent a incremental rotation around the respective cardinal axes. Such a rotation applied to a head orientation of the former time bin is one integration step over time.

For a convenient integration of rotations, the given angle velocity vector $\vec{w}$ is transformed into a quaternion rotation operator **Q** [33]:

$$\mathbf{Q} = (cos(\mid \vec{w} \mid 2), \frac{\vec{w}}{\mid \vec{w} \mid 2} \cdot sin \mid \vec{w} \mid 2) \tag{4.1}$$

The quaternions are multiplied incrementally by using the Matlab function *quatmultiply()* with absolute head orientations (4.5.3) as marginal conditions. Absolute head orientations are transformed between the ESC coordinates and the quaternion formalism by using the Matlab functions *angle2quat()* or *quat2angle()* respectively. IMU sensors are generally not perfectly calibrated and tend to produce grave drift errors for integrated signals over longer time periods. Thus, the drift correction is very important for accuracy and applied like following. Head orientations were measured from scene video at multiple time points. The result of the quaternion integration at the time of the next available scene video measure was compared with the given absolute head orientation. Their difference was divided through the number of integration steps and added as additional rotation term between each step in a second run of integration over time. Hence, the drift error could be reduced to a negligible amount.

4.5.5 Relative head position through integrated acceleration data

The IMU acceleration data are used for estimating the head translations in between the PoV video analysed head position measures. The IMU accelerations are represented in head referenced Cartesian 3D vectors, which are first rotated into set-up referenced coordinates by using the previously derived head orientations. Now, we subtract earth's gravitational pull from the set-up referenced data since the IMU sensor registers all inertial forces. Unfortunately, this is barely accurate because the IMU sensor is not mounted in the PoV camera case, but one of the VOG camera cases and thus can't be exactly trued up based on the PoV video. Consequently, drift errors after time integration over some seconds get tremendous and we chose to bandpass-filter the acceleration data before integrating them over time. However deficient translation trajectories resulting from that process are corrected linearly under the constraint that the absolute head positions from 4.5.3 are matched. Given the experimental procedure, at what the participants are constantly sitting on a swivel chair, the flaw through a weakly resolved translations dynamic seem to be negligible.

Computing gaze-in-room directions

The final step in order to compute the gaze-in-room orientations is to transform the eye-in-head data to room-referenced data in the ESC coordinate system. Before superimposing them with the head-in-room data, the head tilt needs to be considered for the eye-in-head directions, since the head fixed cardinal axes are tilted with respect to room-referenced horizontal and vertical axes. For compensation we applied a rotation around the zero eye-in-head direction by the negative tilt angle $-\psi$ on the eye-in-head data whereas eye-in-head angles were mapped before into a Cartesian norm vector and transformed back afterwards into angles, ϕ_{Eye} and θ_{Eye}. These resulting eye-in-head angles have the same rotation axes as the head-in-room angles ϕ and θ and are superposed together to obtain gaze-in-room directions. An example epoch, while a participant is writing a text on the computer and checks its writing on the monitor, is visualized in figure 4.6.

Kapitel 5

Study IV: Ocular motor analysis in PSP patients

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 13 December 2012
doi: 10.3389/fnbeh.2012.00088

Validation of mobile eye-tracking as novel and efficient means for differentiating progressive supranuclear palsy from Parkinson's disease

***Svenja Marx*[1*†], *Gesine Respondek*[2,3†], *Maria Stamelou*[2,4], *Stefan Dowiasch*[1], *Josef Stoll*[1], *Frank Bremmer*[1], *Wolfgang H. Oertel*[2], *Günter U. Höglinger*[2,3,5‡] and *Wolfgang Einhäuser*[1,6‡]**

[1] *Department of Neurophysics, Philipps-University, Marburg, Germany*
[2] *Department of Neurology, Philipps-University, Marburg, Germany*
[3] *Department of Neurology, Technische Universität München, Munich, Germany*
[4] *Sobell Department for Motor Neurosciences and Movement Disorders, Institute of Neurology, University College London, London, UK*
[5] *German Center for Neurodegenerative Diseases (DZNE), Munich, Germany*
[6] *Center for Interdisciplinary Research, Bielefeld University, Bielefeld, Germany*

Edited by:
Markus Lappe, Universität Münster, Germany

Reviewed by:
Florian Ostendorf, Charité - Universitätsmedizin Berlin, Germany
Uwe Ilg, Hertie-Institute for Clinical Brain Research, Germany

****Correspondence:***
Svenja Marx, Department of Neurophysics, Philipps-University, AG Neurophysik, Karl-von-Frisch Straße 8a, 35032 Marburg, Germany.
e-mail: svenja.marx@physik.uni-marburg.de

[†,‡] These authors equally contributed to this work.

Background: The decreased ability to carry out vertical saccades is a key symptom of Progressive Supranuclear Palsy (PSP). Objective measurement devices can help to reliably detect subtle eye movement disturbances to improve sensitivity and specificity of the clinical diagnosis. The present study aims at transferring findings from restricted stationary video-oculography (VOG) to a wearable head-mounted device, which can be readily applied in clinical practice. **Methods:** We investigated the eye movements in 10 possible or probable PSP patients, 11 Parkinson's disease (PD) patients, and 10 age-matched healthy controls (HCs) using a mobile, gaze-driven video camera setup (EyeSeeCam). Ocular movements were analyzed during a standardized fixation protocol and in an unrestricted real-life scenario while walking along a corridor. **Results:** The EyeSeeCam detected prominent impairment of both saccade velocity and amplitude in PSP patients, differentiating them from PD and HCs. Differences were particularly evident for saccades in the vertical plane, and stronger for saccades than for other eye movements. Differences were more pronounced during the standardized protocol than in the real-life scenario. **Conclusions:** Combined analysis of saccade velocity and saccade amplitude during the fixation protocol with the EyeSeeCam provides a simple, rapid (<20 s), and reliable tool to differentiate clinically established PSP patients from PD and HCs. As such, our findings prepare the ground for using wearable eye-tracking in patients with uncertain diagnoses.

Keywords: progressive supranuclear palsy, mobile eye-tracking, eye movements, Parkinson's disease, video-oculography

INTRODUCTION

Eye movement abnormalities are an essential clinical feature of Progressive Supranuclear Palsy (PSP). Vertical supranuclear gaze palsy or decreased velocities of vertical saccades are a key to the clinical diagnosis of PSP (Litvan et al., 1996). Besides their role as diagnostic signs, eye movement abnormalities disable PSP patients in their daily routine.

Stationary video-oculography (VOG) during head-fixed viewing shows that virtually all forms of eye movements are affected in PSP, with saccadic eye movements being most prominently impaired. Particularly vertical saccades show reduced amplitude and peak velocity when compared to Parkinson's disease (PD) patients and healthy controls (HCs) (Pinkhardt et al., 2008; Chen et al., 2010; Pinkhardt and Kassubek, 2011). Vergence movements and the associated modulation of the linear vestibuloocular reflex are also considerably affected (Chen et al., 2010). The presence of horizontal square wave jerks during attempted fixation of stationary targets is characteristic of PSP (Chen et al., 2010; Otero-Millan et al., 2011). Among these deficits, saccadic peak velocity in the vertical plane shows the sharpest contrast between PSP and PD (Pinkhardt and Kassubek, 2011).

These PSP-specific eye movement abnormalities make clinical investigation of eye movements in patients with Parkinsonian syndromes of great value for differential diagnosis. Correct diagnosis of PSP remains challenging, especially in its early stages (Burn and Lees, 2002). Eye movement abnormalities are not always easy to detect clinically. Particularly, slowing of saccades is a characteristic symptom that can be missed by less experienced neurologists.

Objective measurement devices aid detection of subtle eye movement disturbances. Stationary VOG setups typically require careful calibration, need patient collaboration, and are thus largely impractical for clinical routine. Head-fixed viewing lacks vestibular and other cross-modal information, leaving the relevance of observed eye movement impairment for real-life behavior open. As a first step toward the development of an objective, easy-to-use method for eye movement-based diagnosis, we here tested if recording eye movements with the versatile, head-mounted EyeSeeCam (Brandt et al., 2006; Schneider et al., 2006, 2009) in a brief and simple fixation protocol can differentiate between patients with clinically established PSP as compared to established PD and HCs, and measured gaze in these groups during free behavior. We aimed at establishing the EyeSeeCam's usage in PD and PSP cases and validating its discriminative power between these groups. The parameters established in the present study in clinically established patients shall pave the way for prospective studies with uncertain diagnoses.

MATERIALS AND METHODS

PARTICIPANTS

Patients examined in the Department of Neurology of the University of Marburg qualified for participation in the study, if they had clinically possible or probable PSP (Litvan et al., 1996) and were not more advanced than Hoehn and Yahr stage IV (Golbe and Ohman-Strickland, 2007). As defined by the NINDS-SPSP criteria (Litvan et al., 1996), all patients had supranuclear gaze palsy or slowing of vertical saccades at the time of examination, as evidenced by an examiner specialized in the clinical evaluation of ocular movements.

As controls, we included patients with clinically probable PD (Gibb and Lees, 1988) and HCs. HCs were free of neurologic, systemic, or psychiatric diseases, including alcohol or substance abuse, as verified by detailed evaluation of their medical histories and a comprehensive physical examination.

Further exclusion criteria were other neurological disorders, dementia (mini mental status examination <24), presently active psychiatric disorder (e.g., depression or psychosis), structural brain lesion (e.g., brain surgery, stroke with persistent neurological deficit), cataract, or other neuro-ophthalmological disorders leading to functionally relevant impairment. Since glasses cannot be worn with the EyeSeeCam, people requiring visual correction by glasses stronger than ±2 dpt were also excluded.

Before inclusion into the study, participants gave their informed written consent. All procedures conformed to the Declaration of Helsinki and were approved by the local ethics committee (Ethikkommission FB20, Philipps-Universität Marburg).

EYE AND HEAD MOVEMENT RECORDINGS

We used a mobile VOG setup (EyeSeeCam) to record the participants' eye and head movements. Participants accustomed themselves to wearing the device, while the experimental procedure was explained.

The head-mounted device consists of a head-fixed camera to record the perspective of the head, two high-speed cameras tracking eye-in-head movements, and a camera, which is automatically aligned with the observer's direction of gaze. Gaze- and head-centered videos are recorded at 25 Hz (**Figure 1A**; Movie 1 in supplementary material); eye movements at 300 Hz.

According to manufacturer's specifications, the spatial resolution of the eye-tracking device is given to 0.02° and the precision (relative error) on the order of 0.1° ("maximal resolution error," Schneider et al., 2009). The accuracy (absolute error) of the device under ideal conditions is about 0.5° according to specifications, and can substantially worsen if the goggles move relative to the head during prolonged measurements without recalibration. Hence, all analysis reported here only use relative measures, which are unaffected by these drifts, such as velocities and saccade amplitudes.

Being not concerned with absolute gaze orientation (i.e., with high accuracy) comes at the advantage that the device may be operated using an internal ("default") model of ocular geometry for all participants. In this mode of operation, the mapping from eye measurements on gaze direction does not require a subject-specific calibration, which is in particular beneficial in patients with limited ocular motor control or limited compliance with instructions. Although this sacrifices some precision (depending on the actual head shape compared to the default

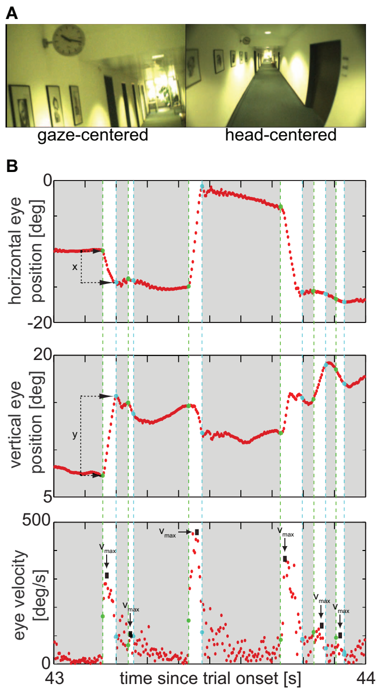


FIGURE 1 | (A) Example frame at 43.81 s in the real-life measurement, while a PD patient was looking at the clock; left: gaze camera, right: head camera. The movie of this scene including velocity histograms is shown as supplemental online Movie 1. **(B)** Eye-traces of the scene. Upper panel: Horizontal eye position, indicating the horizontal amplitude of saccades; starting and end points of saccades are marked by green and cyan dashes lines, respectively; durations of saccades are highlighted by a white background. Middle panel: Vertical eye position, indicating the vertical amplitude of the same saccades. Lower panel: Absolute eye velocity, arrows mark saccade peak velocities, used for analysis.

model), no systematic effect on the measures analyzed here can be expected. For the "fixation protocol" (see below), the default model was used in all participants; for the "real-life measurements" (see below), in those participants, in whom it was possible, the subject-specific model obtained from the fixation protocol was used; for the remainder the default model was also used in real-life measurements. Since the subject-specific adaptation of the model represents a calibration procedure for absolute position, for the real-life measurement, these participants will be referred to as successfully and unsuccessfully calibrated, respectively.

When extracting head movements from the head fixed camera, for the analysis conducted here, the spatial resolution is limited by the pixel width of about 0.3°, even though sub-pixel analysis would be possible in principle. When analysis is based on subsequent frames, this limits the resolution for head movements to about 7.5°/s. While integration over multiple frames would be possible to lower this number, this would come at the cost of lower temporal resolution and thus possibly lumping distinct head movements into one.

Fixation protocol

To test the utility of the EyeSeeCam as diagnostic tool, we employed a "fixation protocol." In addition to being the first experimental part, this protocol also served to refine the EyeSeeCam's calibration for the subsequent real-life experiments by adapting the system's internal eye model to the individual. During the fixation protocol, the participants' heads were unrestrained, but they were asked to avoid head movements as far as possible. They were instructed to move their eyes to look successively at 5 laser dots projected onto a wall straight ahead, a central dot and four at 8.5° in the cardinal directions. An experimenter pointed with a finger at the dot the participant should look at. To give the participant the possibility to self-pace their fixations, presentation of the dots in time was to some degree flexible and not exactly clocked. However, the participant had to look at each dot for 2 s at least once in a time span of approximately 20 s. While this procedure is far less constrained and standardized than usual laboratory measurements, it is still more controlled than the real-life conditions of the present study. This flexible and efficient procedure makes the participation of very severely affected patients possible, presenting a clear advantage over more constrained settings.

Real-life behavior

For measuring a large range of gaze behaviors as occurring in real-life situations, we asked participants to perform a series of tasks, while spontaneous eye and head movements were recorded. First, free-exploration behavior was assessed by asking participants to walk along a 50 m corridor. Right before the participant turned around at the end of the corridor, an experimenter laid two paper spots on the floor to assess tracking behavior. Participants were asked to track the dots with their eyes, while walking back toward them. Finally, participants took the elevator and descended one-level to test a situation without active movement in a confined visual environment with subtle vestibular input. Those two PSP and PD patients who were wheelchair-dependent were wheeled throughout the whole procedure by an experimenter instead of actively walking.

The objective of the real-life measurement was to provide a naturalistic set of behaviors, while differences between real-life conditions were not at the focus of the current study. Consequently, all data of real-life measurement were pooled per participant. The real-life measurement lasted less than 10 min per participant.

DATA ANALYSIS AND STATISTICAL EVALUATION

Eye movements

Raw eye-position data were processed offline using MATLAB (Matlab 7.10, The MathWorks, Natick, MA), which was also used for statistical analysis. We calculated eye velocity by differentiation of the horizontal and vertical eye position (**Figure 1B**). Absolute speed was then calculated as the square root of the sum of the squared horizontal and squared vertical velocity componentss.

All phases faster than 60°/s and lasting longer than 10 ms are referred to as "saccades," irrespective of whether they were actual saccades or fast phases of reflexive movements (**Figure 1B**). This threshold is higher than those typically used in laboratory settings, as signals obtained during real-life measurements contain rich eye movement dynamics and are typically noisier than under constrained settings. The conservative choice is, however, consistent with previous research on eye movements in PSP patients: for example, judging from the figures in Pinkhardt et al. (2008), their patients had their 5% percentile of peak saccade velocities around or above 60°/s, meaning that we can still expect to include about 95% of actual saccades with our comparably conservative criterion. Since this criterion could also be employed in practice, it will not affect any conclusion on the discriminability of patient groups. Nonetheless, for the general questions pertaining to eye movement disturbances in PSP and PD, the fact that any threshold must remain arbitrary motivates to add an analysis that does not classify eye movements in saccade/non-saccade, but uses the unclassified (i.e., raw) eye movement data (see below and section "Unclassified Eye Movements").

Parameters to describe saccades were their direction, peak velocity, amplitude, and duration (**Figure 1B**). Since peak velocity, saccade amplitude, and duration are typically not independent, the functional relationship of amplitude and peak velocity and of amplitude and saccade duration, the so-called main sequence (Bahill et al., 1975), was also considered for real-life data: we fitted the relation with a power function of the form velocity $= a \times \text{amplitude}^b$ or duration $= a \times \text{amplitude}^b$, respectively (cf. Garbutt et al., 2003), and considered only the fit parameters a and b further. Since reliable fits of main sequences require substantial amounts of data, this analysis was only performed for the real-life measurements.

To test whether there is an abundance of one saccade direction in a group, we coarsely classified saccades into equally spaced 45° wedges: horizontal (±22.5° from the horizontal), vertical (±22.5° from the vertical), and oblique (the remaining $4 \times 45° = 180°$).

For analysis of raw ("unclassified") eye data (i.e., all data irrespective of whether defined as saccade or not), two-dimensional histograms were used. Each bin of the histograms used for analysis corresponds to a velocity interval of 15°/s in each direction (horizontal and vertical); the central bin ranges from −7.5°/s to +7.5°/s in each direction. The number of samples in each bin is color-coded.

Head movements

Head movements were computed from the video of the head-fixed camera at 25 Hz. To obtain head position, the same stationary point of the environment was marked in each video-frame. From this point's position in the camera's field of view relative head orientation in the world was computed. Head velocity was obtained by differentiation of this signal, and was thus independent of this choice of origin. All quantitative analysis was therefore based on velocities. Unlike for eye movements and due to the low spatial and temporal resolution (section "Eye and Head Movement Recordings" top), we could not classify head movements in distinct classes (e.g., fast/slow) with the data at hand. Therefore, all analysis was based on overall velocity distributions for each individual.

Statistical analysis

Data are presented as mean ± standard deviation. Statistical evaluation used non-parametric tests for raw eye data, such as amplitude and peak velocity of each saccade (Kruskal–Wallis when three groups were compared and Mann–Whitney-*U*-Test for two groups). To compare these parameters in an exploratory manner across participants, the individual distributions are described by their medians as robust measure (since the distributions are either leptokurtic or prone to outliers). Since these medians can be assumed to follow a normal distribution across participants, the group effects were analyzed by parametric tests; that is, ANOVAs for three group comparisons and two-tailed *t*-tests for two-group comparisons and *post-hoc* tests.

Signal-detection-theory measures

For assessing the performance of the classifiers between PSP and PD patients, we performed signal-detection analysis by computing the Receiver-Operating-Characteristic (ROC). The ROC is quantified by its area under the curve (AUC), the cut-off point for maximal specificity and sensitivity, and the corresponding values of specificity and sensitivity. Values are reported such that all values of patients classified as PSP patients are strictly smaller than this cut-off value.

RESULTS

PARTICIPANT CHARACTERISTICS

We investigated 10 PSP patients (6 probable, 4 possible), 11 PD patients and 10 HCs (**Table 1**). All patients were under treatment in the University Hospital in Marburg. There were no significant differences regarding age, disease duration, and gender between the groups. For all patients Hoehn and Yahr stage was assessed in off-state and, as expected, the stages differed significantly between PSP and PD patients (**Table 1**).

Eye velocities and relative eye positions (e.g., saccade amplitudes) require only minimal subject–specific adjustment

Table 1 | Clinical characteristics of the participants in this study: overview.

	PSP	PD	HC
N	10	11	10
Age (years)	65.9 ± 4.6	65.5 ± 12.7	68.3 ± 9.1
Gender (F/M)	3/7	3/8	6/4
DD (years)	3.9 ± 2.7	6.2 ± 4.7	–
H&Y	3.9 ± 0.4	2.5 ± 0.4	
Wheelchair	2/10	2/11	0/10
Real-life measurement time	304.3 ± 114.4 s	242.2 ± 78.5 s	202.8 ± 35.3 s

Details

Patient ID/gender/age [years]	Onset	Exam. date	H&Y	Medication
PSP01/F/67	2004	08/2010	4	Levodopa
PSP02/M/70	2008	08/2010	3	Amantadine
PSP03/F/63	2007	08/2010	4	Levodopa, Amantadine
PSP04/M/70	2007	08/2010	4	Levodopa, Amantadine, Piribedil
PSP05/F/65	2007	08/2010	3	Amantadine, Rotigotine
PSP06/M/67	2000	08/2010	4	Levodopa
PSP07/M/62	2008	02/2011	4	Levodopa
PSP08/M/74	2005	05/2011	4	Levodopa, Amantadine
PSP09/M/59	2010	10/2011	3	Levodopa
PSP10/M/62	2009	11/2011	3	Levodopa
PD01/M/61	2007	09/2010	2	Rotigotine
PD02/M/75	1995	09/2010	3	Levodopa
PD03/M/75	2007	02/2011	1	Ropinirole
PD04/M/64	2000	07/2011	3	Levodopa, Amantadine, Pramipexole, Rasagiline
PD05/M/67	2007	07/2011	1	Levodopa, Ropinirole, Rasagiline
PD06/M/51	2010	09/2011	2	Levodopa, Rasagiline
PD07/F/62	2007	10/2011	3	Levodopa, Rasagiline, Piribedil
PD08/M/38	2010	10/2011	2	Pramipexole
PD09/M/78	2007	12/2011	3	Levodopa
PD10/F/82	2001	12/2011	3	Levodopa, Amantadine, Ropinirole
PD11/F/68	2000	12/2011	3	Levodopa, Amantadine, Pramipexole
HC01/F/58		08/2010		
HC02/M/71		08/2010		

(Continued)

Table 1 | Continued

Patient ID/gender/age [years]	Onset	Exam. date	H&Y	Medication
HC03/F/53		02/2011		
HC04/F/63		02/2011		
HC05/M/73		03/2011		
HC06/F/64		03/2011		
HC07/F/69		03/2011		
HC08/F/74		09/2011		
HC09/M/85		12/2011		
HC10/M/73		12/2011		

PSP, progressive supranuclear palsy; PD, Parkinson's disease; HC, healthy controls; DD, disease duration; H&Y, Hoehn and Yahr Stage. H&Y stage is significantly different between PD and PSP [$t_{(19)} = 4.12$, $p < 0.001$]; real-life measurement duration differs significantly between PSP and HC ($p = 0.02$ post-hoc test); all other comparisons do not show a significant difference ($p > 0.05$).

and could thus be measured accurately in all participants. However, individual-specific calibration of absolute eye-position failed in eight PSP and two PD patients as a consequence of their inability to steadily fixate instructed targets over a 2-s integration window. Interestingly, this inability did not primarily result from square-wave jerks, which were robustly observed only in 1 out of the 10 PSP patients under our experimental conditions. As a consequence of the calibration failures for absolute position, all quantitative analysis hereafter is based on relative eye-position and velocities only.

SACCADES

Fixation protocol

All participants performed a standard fixation protocol, as described in the "Materials and Methods" section, which was also used for individual calibration refinement. Irrespective of whether this absolute-position calibration was successful or not, these measurements provided a sufficient number of visually-guided saccades to analyze differences between PSP patients and PD patients or HCs (**Figure 2**).

Averaged median saccadic peak velocity was 135.1 ± 43.8°/s for PSP, 220.1 ± 31.5°/s for PD patients and 233.0 ± 44.4°/s for HCs. A One-Way ANOVA revealed a significant main effect [$F_{(2,\ 28)} = 17.81$, $p < 0.001$, **Figure 2B**] and *post-hoc t*-tests showed that PSP patients generated saccades with significantly slower median peak velocity than PD patients [$t_{(19)} = 5.14$, $p < 0.001$] and HCs [$t_{(18)} = 4.96$, $p < 0.001$]. There were also significant differences in the vertical components of saccade peak velocity. Averaged vertical saccade peak velocity was 54.9 ± 28.0°/s for PSP patients, 158.5 ± 47.9°/s for PD patients and 151.1 ± 60.3°/s for HCs [$F_{(2,\ 28)} = 14.53$, $p < 0.001$; PSP-PD: $t_{(19)} = 5.83$, $p < 0.001$; PSP-HC: $t_{(18)} = 4.51$, $p < 0.001$, **Figure 2C**].

Saccade amplitudes also differed significantly between groups [$F_{(2,\ 28)} = 18.26$, $p < 0.001$, PSP-PD: $t_{(19)} = 4.26$, $p < 0.001$, PSP-HC: $t_{(18)} = 6.60$, $p < 0.001$, **Figure 2B**]. Averaged median amplitudes were 1.88 ± 0.72° for PSP patients, 4.16 ± 1.53° for PD patients and 5.42 ± 1.53° for HCs. Vertical saccade amplitude was 0.52 ± 0.37° for PSP patients, 2.89 ± 1.62° for PD patients and 3.03 ± 2.16° for HCs and thus also differed significantly [$F_{(2,\ 28)} = 7.76$, $p = 0.002$; PSP-PD: $t_{(19)} = 4.37$, $p < 0.001$; PSP-HC: $t_{(18)} = 3.57$, $p = 0.002$, **Figure 2C**].

We did not find significant main effects for the horizontal components of peak velocity [$F_{(2,\ 28)} = 2.12$, $p = 0.14$, ANOVA; **Figure 2D**] and amplitude [$F_{(2,\ 28)} = 1.69$, $p = 0.20$, **Figure 2D**].

The ROC comparing saccade peak velocity of PSP and PD patients showed an AUC of 0.95. Specificity was 11/11 and sensitivity was 9/10 for a cut-off value of 189.8°/s (i.e., all patients having slower peak velocities than this value were classified as PSP) patients. For the comparison of vertical saccade peak velocities, the AUC was 1 and for the cut-off value 111.7°/s, specificity was 11/11 and sensitivity was 10/10. The AUC for the comparison of saccade amplitude was 0.97 with a specificity of 11/11 and a sensitivity of 9/10 for a cut-off value of 2.79°. For the vertical component, AUC was 0.99 and the ROC analysis showed a specificity of 10/11 and a sensitivity of 10/10 for the cut-off value 1.68°.

For completeness, we also analyzed saccade duration in all groups. We found a significant main effect between groups [PSP: 19.6 ± 7.2 ms, PD: 26.2 ± 6.3 ms, HC: 32.7 ± 6.5 ms, $F_{(2,\ 28)} = 9.6$, $p < 0.001$, see **Figures 2E,F**]. *Post-hoc t*-test revealed significant differences between all groups [PSP-PD: $t_{(19)} = 2.25$, $p = 0.037$; PSP-HC: $t_{(18)} = 4.27$, $p < 0.001$; PD-HC: $t_{(19)} = 2.30$, $p = 0.033$]. Sensitivity was 7/10 and specificity was 9/11 for the cut-off value 21.6 ms, the AUC was 0.77. These values are much lower than for amplitude and peak velocity and thus less informative where differential diagnosis is concerned. Hence, we hereafter focus most analysis on peak velocity and amplitude.

Real-life

Since the eye movement impairment in PSP was evident during the fixation protocol, we next analyzed their relevance for real-life situations. Hence, we measured the spontaneous ocular motor behavior in a real-life, minimally restrained scenario, comprising self-paced walking in a corridor, tracking of a stationary target, and taking an elevator. Self-paced walking implies speed differences between participants. ANOVA revealed a significant main effect for differences in real-life measurement duration [$F_{(2,\ 28)} = 3.85$, $p = 0.03$, **Table 1**]; the difference was not significant between PSP and PD patients, but for HCs the measurement lasted significantly shorter than for PSP patients [$t_{(18)} = 2.68$, $p = 0.02$]. Aggregating over the whole real-life measurement, we assessed the same parameters as during the fixation protocol (**Figure 3**).

All groups had the same fraction of vertical [PSP: 24.1% ± 15.4%, PD: 28.9% ± 10.4%, HC: 31.7% ± 7.1%, $F_{(2,\ 28)} = 1.14$, $p = 0.33$], horizontal [PSP: 21.7% ± 9.0%, PD: 18.5% ± 8.1%, HC: 18.3% ± 5.9%, $F_{(2,\ 28)} = 0.64$, $p = 0.53$] and oblique [PSP: 54.3% ± 10.1%, PD: 52.6% ± 6.1%, HC: 50.0% ± 3.7%, $F_{(2,\ 28)} = 0.92$, $p = 0.41$] saccades.

The medians of saccade peak velocity differed significantly between the groups [$F_{(2,\ 28)} = 5.47$, $p = 0.01$, **Figure 3B**]. PSP

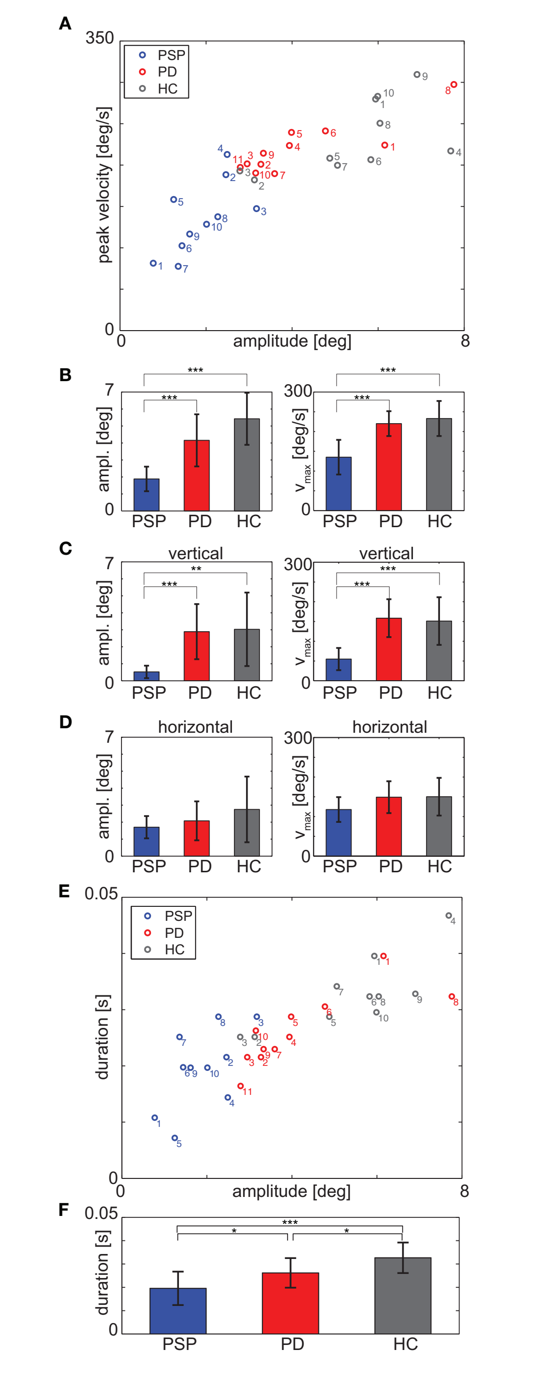


FIGURE 2 | Continued

FIGURE 2 | (A) Medians of saccade peak velocity and amplitude for each participant during the fixation protocol. **(B)** Mean over participants of median amplitude (left panel) and median peak velocity (right panel) for each group. **(C)** Vertical component and **(D)** horizontal component of the data of panel **(B)**; **(E)** Medians of saccade duration and amplitude for each participant during fixation protocol; note that the duration is discretized due to sampling frequency **(F)**. Mean over participants of median duration. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

patients' averaged median saccade peak velocity was 131.1 ± 29.0°/s and thus slower than those of PD patients [163.1 ± 25.8°/s; $t_{(19)} = 2.68$, $p = 0.002$] and HCs [160.2 ± 15.4°/s; $t_{(18)} = 2.80$, $p = 0.01$]. The vertical component of saccade peak velocity (PSP: 71.9 ± 15.5°/s, PD: 89.6 ± 11.5°/s, HC: 89.5 ± 9.6°/s) also differed significantly [$F_{(2,\ 28)} = 6.88$, $p = 0.004$, PSP-PD: $t_{(19)} = 3.00$, $p = 0.007$; PSP-HC: $t_{(18)} = 3.05$, $p = 0.007$, **Figure 3C**], whereas there was no significant difference between means of the horizontal component of peak velocity [$F_{(2,\ 28)} = 1.66$, $p = 0.21$, **Figure 3D**] between groups.

ANOVA did not reveal a significant main effect for saccade amplitude [$F_{(2,\ 28)} = 2.55$, $p = 0.10$, **Figure 3B**], but the vertical component of saccade amplitude differed significantly [$F_{(2,\ 28)} = 3.46$, $p = 0.045$, **Figure 3C**]; *post-hoc t*-tests revealed that PSP patients' vertical component of saccade amplitude was significantly shorter (0.79 ± 0.36°) than PD patients' [1.12 ± 0.33°; $t_{(19)} = 2.12$, $p = 0.047$] and HCs' [1.06 ± 0.13°; $t_{(18)} = 2.16$, $p = 0.04$]. There was no significant difference between medians of the horizontal components of amplitudes [$F_{(2,\ 28)} = 0.25$, $p = 0.78$, **Figure 3D**].

The AUC was 0.84 for peak velocity with a sensitivity of 8/10 and a specificity of 9/11 for the cut-off value 139.9°/s. For vertical peak velocity, the AUC was 0.82 and for a cut-off value of 83.2°/s sensitivity was 7/10 and specificity was 8/11. For analysis of saccade amplitudes, the AUC was 0.80 with a sensitivity of 8/10 and a specificity of 8/11 for a cut-off value of 1.85°. The AUC for comparison of vertical components was 0.75 with a sensitivity of 6/10 and a specificity of 11/11 for the cut-off value 0.69°.

Differences in medians of saccade duration were not significantly different between groups [PSP: 25.5 ± 3.7 ms, PD: 27.6 ± 4.0 ms, HC: 25.6 ± 2.3 ms, $F_{(2,\ 28)} = 1.31$, $p = 0.29$; see **Figures 3E,F**].

Correlation between fixation protocol and real-life

Median of peak velocity and its vertical component in the fixation protocol and during real-life measurement correlated significantly ($N = 31$, $r = 0.39$, $p = 0.03$; vertical: $r = 0.50$, $p = 0.004$). Thus, the data collected during the fixation protocol not only differentiated between PSP and PD patients, but also in part predicted real-life performance.

Main-sequence analysis

Peak velocity and duration were plotted as a function of amplitude for each saccade of every participant. We fitted this main sequence with a power function (**Figure 4A**) and compared the fit parameters between groups. There were no significant differences between groups with respect to the value of fit parameters *a*

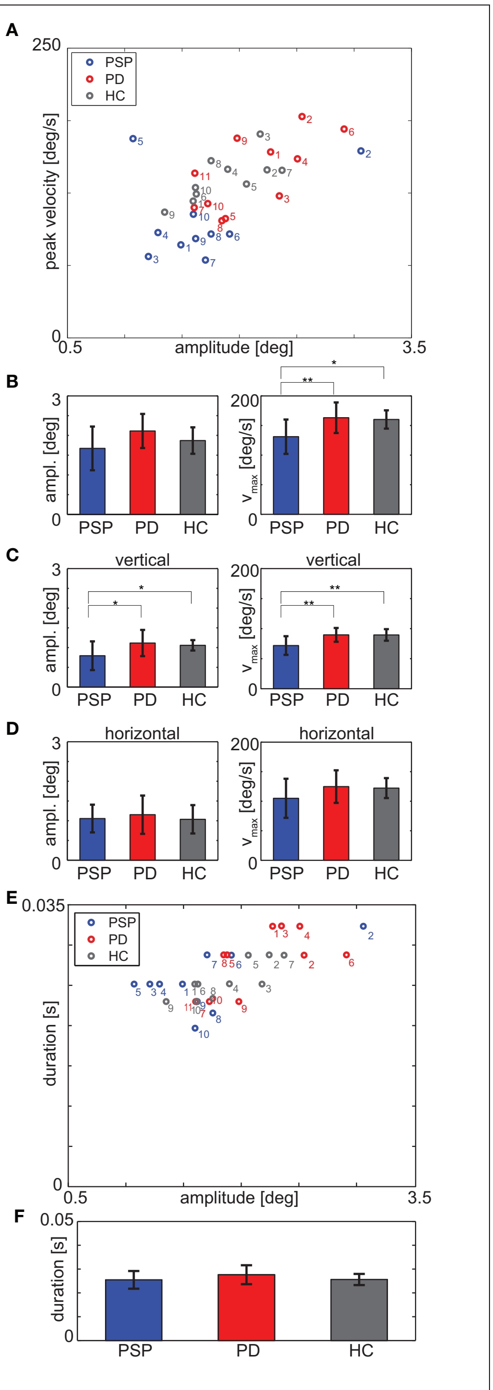


FIGURE 3 | Continued

FIGURE 3 | (A) Medians of saccade peak velocity and amplitude for each participant during real-life measurement. **(B)** Mean over participants of median amplitude (left panel) and median peak velocity (right panel) for each group. **(C)** Vertical component and **(D)** horizontal component of the data of panel **(B)**. **(E)** Medians of saccade duration and amplitude for each participant during real-life measurement; note that the duration is discretized due to sampling frequency **(F)**. Mean over participants of median duration. $^{*}p < 0.05$; $^{**}p < 0.01$.

[$F_{(2,\ 28)} = 1.69$, $p = 0.20$, **Figure 4B**] and b [$F_{(2,\ 28)} = 1.38$, $p = 0.27$, **Figure 4C**]. There were also no differences between groups in the vertical component of saccades [value of a: $F_{(2,\ 28)} = 2.54$, $p = 0.097$, **Figure 4D**; value of b: $F_{(2,\ 28)} = 1.08$, $p = 0.35$, **Figure 4E**] and in the value of the fit parameter a of the functional relationship between duration and amplitude [$F_{(2,\ 28)} = 0.02$, $p = 0.98$, **Figure 4F**]. There was a significant main effect for the values of b in that case [$F_{(2,\ 28)} = 4.11$, $p = 0.027$, **Figure 4G**] but *post-hoc* t-tests did not reveal significant differences between PSP and PD patients [$t_{(19)} = 1.77$, $p = 0.09$] or PD patients and HCs [$t_{(19)} = 1.24$, $p = 0.23$]. The only significant difference was found between PSP patients and HCs [$t_{(18)} = 2.43$, $p = 0.026$].

UNCLASSIFIED EYE MOVEMENTS

Under real-life conditions, fast eye movement phases (saccades), as analyzed above, accounted for only a small amount of the entire measurement time (PSP: 7.6 ± 3.8%, PD: 11.7% ± 7.9%, HC: 10.4% ± 2.8%). To compare saccade-based analysis to all eye movements, we generated 2-dimensional velocity histograms for saccades only (**Figure 5A**) and for all eye movements ("unclassified movements," **Figure 5B**) during the entire real-life measuring time. The histograms show pooled data from all participants of each group, normalized such that each participant contributes with equal weight to the respective histograms. In the distribution of saccade peak velocities (**Figure 5A**), a preference for horizontal movements is evident in all groups, which is particularly pronounced in PSP patients, reflecting their prominent reduction in vertical peak velocity. Interestingly, this difference between groups was less evident when analyzing all eye movements (**Figure 5B**). We quantified the spread in each direction by standard deviation. When considering all unclassified eye movements, there were no significant differences among the groups [vertical: $F_{(2,\ 28)} = 1.74$, $p = 0.19$; horizontal: $F_{(2,\ 28)} = 1.86$, $p = 0.18$]. When instead considering saccades only (**Figure 5A**), a picture consistent with the analysis above (section "Real-Life") emerged: the standard deviation of saccade peak velocities yielded highly significant differences between the groups [vertical: $F_{(2,\ 28)} = 8.53$, $p = 0.001$; horizontal: $F_{(2,\ 28)} = 12.42$, $p < 0.001$]. Significant differences appeared between PSP and PD patients [vertical: $t_{(19)} = 3.38$, $p = 0.003$; horizontal: $t_{(19)} = 4.34$, $p < 0.001$] as well as between PSP patients and HCs [vertical: $t_{(18)} = 3.41$, $p = 0.003$; horizontal: $t_{(18)} = 3.75$, $p = 0.002$]. Moreover, when testing analogous measures to those that yielded significant differences and high diagnostic power between patient groups for saccades (**Figures 2** and **3**), no significant effects were found for the full, unclassified eye movement data. For example, the medians of all velocities were not significantly different between the groups

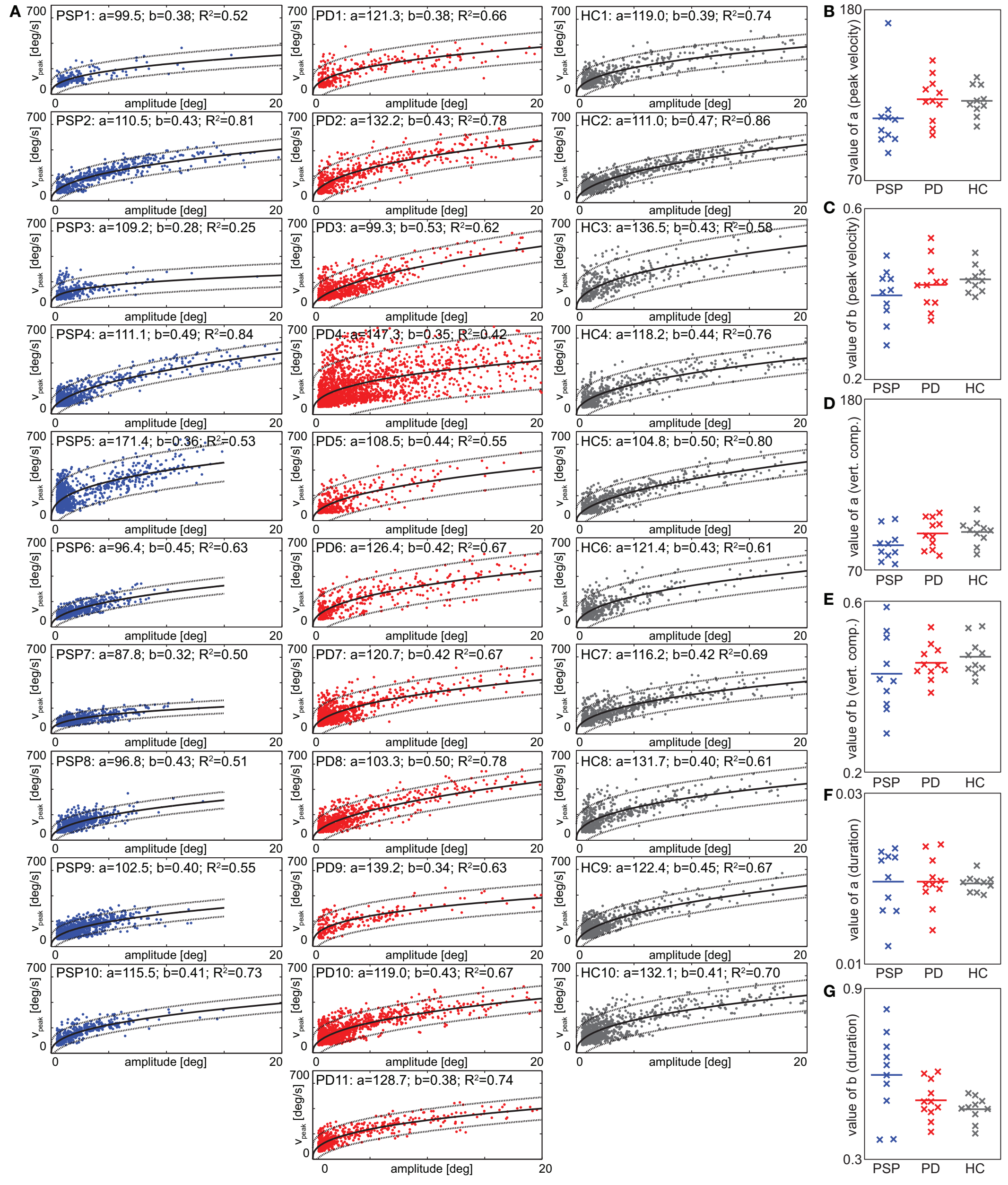


FIGURE 4 | (A) Saccade peak velocity plotted against saccade amplitude ("main sequence") for all individuals. Each data-point corresponds to one saccade (note the cutoff to the bottom and left, given by the thresholds on velocity and duration). Solid black line denotes best fitting power functions (see section "Eye Movements") in a least-squares sense, dotted lines 5 and 95% confidence intervals. Fit parameters and R^2 are given in the panel headers. **(B,C)** Main-sequence fit parameters of the functional relationship between peak velocity and amplitude. **(D,E)** Main-sequence fit parameters for the functional relationship between vertical component of peak velocity and amplitude. **(F,G)** Main-sequence fit parameters for the functional relationship between saccade duration and amplitude.

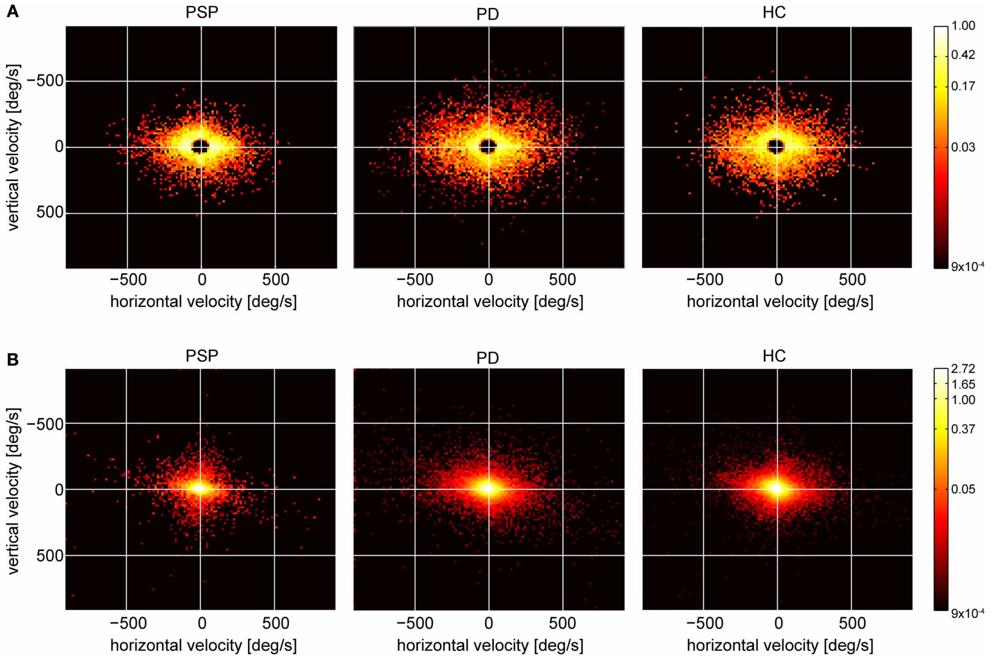


FIGURE 5 | (A) Peak velocity histograms of cumulated saccades of all participants in each group (range: −900 to 900°/s in the cardinal directions, bin size: 15°/s × 15°/s). **(B)** Velocity histograms of raw eye velocities of all cumulated data points of all participants in each group, same range, and bin size as in **(A)**.

[$F_{(2,\,28)} = 1.01, p = 0.38$]. Notwithstanding some degree of arbitrariness in the definition of saccade thresholds, this indicates that—at least under our recording conditions—the described effects are best observed in fast movements.

HEAD MOVEMENTS

For 26 participants (9 PSP, 7 PD, and 10 HC) we successfully obtained head data during the fixation protocol, for 27 (9 PSP, 9 PD, and 9 HC) during walking along the corridor without target tracking, and for 29 (9 PSP, 10 PD, and 10 HC) while they tracked the stationary target. In the remaining participants, head orientation was not recorded or recording was unsuccessful for technical reasons. We chose to split walking the corridor into periods with tracking and without tracking for head-in-world data considered here, as we expected higher consistency with respect to the overall head movements.

During the fixation protocol, all but one participant deviated less than 2° from their average gaze orientation, 22/26 even less than 1°. Thus, head movements were small and rare, and the median head velocity was below 2°/s in all but one participant. While this implies that participants complied with the instruction to avoid head movements, it also means insufficient movements to obtain robust velocity data.

During tracking, spread (quantified as standard deviations) of head velocities was not significantly different between groups [vertical: $F_{(2,\,26)} = 0.49, p = 0.62$, **Figure 6A**; horizontal: $F_{(2,\,26)} = 0.63$, $p = 0.54$, **Figure 6B**]. During walking without tracking, the vertical spread in velocity showed no dependence on group [$F_{(2,\,24)} = 0.51$, $p = 0.61$, **Figure 6C**], either. In contrast, horizontal spread showed a significant group dependence [$F_{(2,\,24)} = 3.67, p = 0.04$, **Figure 6D**], indicating that the absence of an effect during tracking, where less participants contributed, was not due to a lack of power. Importantly, this group dependence resulted from a difference between PSP patients and HCs [PSP-HC: $t_{(16)} = 3.41$, $p = 0.004$], but not from a difference between patient groups [PSP-PD: $t_{(16)} = 0.01, p = 0.99$] or between PD patients and HCs [PD-HC: $t_{(16)} = 2.07, p = 0.055$]. In sum, neither head orientation nor head velocity—to the extent they could be analyzed with the present device—could offer any parameters that might serve to discriminate PSP from PD.

DISCUSSION

In the present study we used a novel, wearable eye-tracking device to assess gaze behavior in PD, PSP, and HCs. First, we demonstrate that wearable eye-tracking distinguishes PSP from PD with high sensitivity and specificity. Second, we show that these differences in gaze behavior are most prominent for saccades in a brief fixation protocol and less pronounced in activities of daily living.

The observed differences between saccadic peak velocities in the fixation protocol are highly consistent with earlier findings

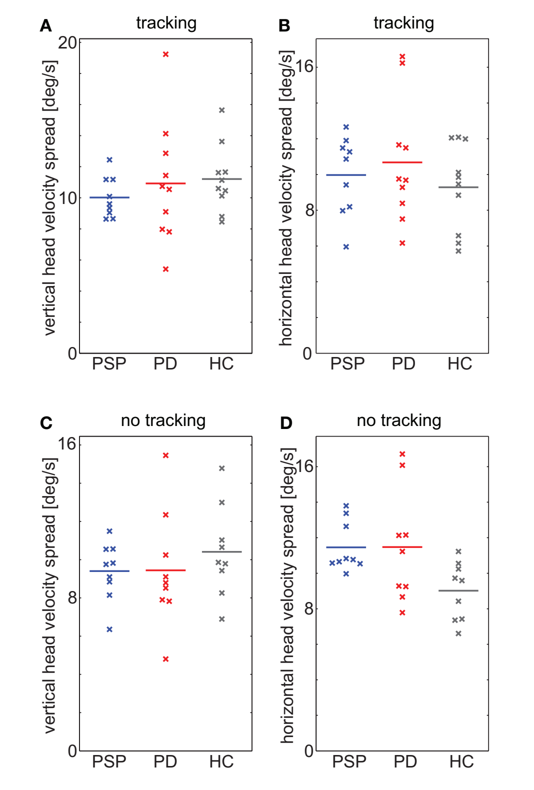


FIGURE 6 | (A,B) Spread of head velocity in **(A)** vertical and **(B)** horizontal direction during walking while tracking a stationary target. **(C,D)** Spread of head velocity in **(C)** vertical and **(D)** horizontal direction during walking without specific instruction. Horizontal displacement of datapoints within groups is for improved visibility only.

(Pinkhardt and Kassubek, 2011; Boxer et al., 2012). Similarly, the lack of evidence for a difference in peak velocities between the PD group and HCs are in line with previous data (Tanyeri et al., 1989; Pinkhardt and Kassubek, 2011). As such, our data extend earlier findings obtained using visually-guided saccades in standard laboratory setups to wearable eye-tracking, which allows efficient assessment of these parameters in less restrained conditions. Even though many sorts of eye movements are affected by PSP, we focused on saccadic peak velocity and amplitude for reasons of efficiency. Duration of saccades as conceivable alternative turned out to have less diagnostic power, despite some difference in the average. Although amplitude, peak velocity, and duration are not independent, but coupled through the "main sequence," the functional fit does not provide any additional diagnostic power in real-life data, and requires more data than available from the 20-s fixation protocol, such that amplitude and peak velocity remain as the main diagnostic markers for this rapid assessment. Still, if these two parameters should turn out to be insufficient for differential diagnosis in a patients with clinically uncertain diagnosis, other eye movements like vergence and the linear vestibuloocular reflex can also be measured with the EyeSeeCam.

The comparison between raw data and data filtered for saccades allows three main conclusions. First, it stresses the specifically prominent impairment of the saccade system for PSP patients as compared to other eye movement systems (Chen et al., 2010). Second, it underlines the importance of objective measurement devices to reliably detect potentially subtle eye movement-related disease markers (Bartl et al., 2009). Finally, the comparably mild differences in overall gaze orienting behavior might point to a strategy how the specific deficits may be compensated for and thus offers a promising path for carefully quantifiable therapeutic intervention (Zampieri and Di Fabio, 2008).

The reduced differences in gaze behavior during activities of daily living indicate that patients at least in part compensate for their ocular motor deficits. Analysis of head movements, however, suggests substantial inter-individual differences, indicating that compensation strategies are largely idiosyncratic. Predicting such compensation behaviors and relating them to other parameters, such as disease progression, will be an interesting issue for further research in larger, heterogeneous PSP cohorts. In a longitudinal study, the precise quantification of compensatory behavior might then also aid the efficient monitoring of treatment success. For differential diagnosis, the free exploration paradigm is clearly less valuable, demonstrating the importance of a flexible, but at the same time standardized fixation protocol for clinical use. Nonetheless, the free exploration data may yield important information on compensation mechanisms and the consequences of the disease on everyday life.

In contrast to eye movements, the parameters considered for head movements did not allow a significant dissociation between patient groups under any of the tested tasks. This could be due to the low spatial and temporal resolution of the head movement measurements as compared to eye movement measurements. It is conceivable that with an improved measurement device for head movements, with different instructions or tasks, or when effects on eye-head coordination are measured with sufficient spatial and temporal accuracy and precision, head movements might eventually become useful and could augment a PSP/PD discrimination system. However, with the present technology and based on the tasks used in the present study, eye velocity and amplitude during the fixation protocol present a most promising candidate for dissociating PSP from PD also in subclinical populations.

This study is to be regarded as a first step toward establishing a new method as a diagnostic tool. Prospective studies measuring eye movements of still unclassified patients are needed to prove that subclinical oculomotor disturbances can be detected prior to the establishment of the clinical diagnosis. Also, square wave jerks which are characteristic of PSP patients could only be detected in one PSP patient, even by careful visual inspection of all eye movement traces. While beyond the scope of the present study, the question as to whether their absence from the measured data is a technical limitation or a true effect of the population and condition at hand remains an important issue for future research.

Importantly for a possible application in diagnosis and treatment monitoring, the usage of the wearable eye-tracking device is efficient, requiring less than 20-s for the fixation protocol and virtually no device-specific training. While wearable eye-tracking has recently been suggested as tool in a variety of ocular motor and vestibular conditions (Hayhoe and Ballard, 2005; Schumann et al., 2008), the present study demonstrates that wearable eye-tracking also lends itself for efficient clinical use in the context of more complex syndromes, such as typical and atypical Parkinsonism. Whether or not wearable eye-tracking will allow diagnosis beyond the current gold standard obviously can only be established in a long-term longitudinal prospective study, which will apply the criteria found herein already early during disease, when current clinical criteria are not yet clear cut.

ACKNOWLEDGMENTS

Financial support by the German Research Foundation (DFG) through grant EI 852/1 (Wolfgang Einhäuser), the Swiss National Science Foundation (SNF) and DFG through grant 135687/EI 852/3 (Wolfgang Einhäuser), and the DFG through grant HO2402/6-1 (Günter U. Höglinger) is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http:// www.frontiersin.org/Behavioral_Neuroscience/10.3389/fnbeh.2012.00088/abstract

Movie 1 | Example movies of two participants, PD07 and PSP09, showing a part of the real-life measurement. Histograms picture eye velocity (left panel, range: −500 to 500°/s in the cardinal directions, bin size for this movie: 5°/s × 5°/s) and head velocity (right panel, range: −60 to 60°/s in the cardinal directions, bin size: 3°/s × 3°/s).

REFERENCES

Bahill, A. T., Clark, M. R., and Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Math. Biosci.* 24, 191–204.

Bartl, K., Lehnen, N., Kohlbecher, S., and Schneider, E. (2009). Head impulse testing using video-oculography. *Ann. N.Y. Acad. Sci.* 1164, 331–333.

Boxer, A. L., Garbutt, S., Seeley, W. W., Jafari, A., Heuer, H. W., Mirsky, J., et al. (2012). Saccade abnormalities in autopsy-confirmed frontotemporal lobar degeneration and Alzheimer disease. *Arch. Neurol.* 69, 509–517.

Brandt, T., Glasauer, S., and Schneider, E. (2006). A third eye for the surgeon. *J. Neurol. Neurosurg. Psychiatry* 77, 278.

Burn, D. J., and Lees, A. J. (2002). Progressive supranuclear palsy: where are we now? *Lancet Neurol.* 1, 359–369.

Chen, A. L., Riley, D. E., King, S. A., Joshi, A. C., Serra, A., Liao, K., et al. (2010). The disturbance of gaze in progressive supranuclear palsy: implications for pathogenesis. *Front. Neur.* 1:147. doi: 10.3389/fneur.2010.00147

Garbutt, S., Harwood, M. R., Kumar, A. N., Han, Y. H., and Leigh, R. J. (2003). Evaluating small eye movements in patients with saccadic palsies. *Ann. N.Y. Acad. Sci.* 1004, 337–346.

Gibb, W. R., and Lees, A. J. (1988). The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 51, 745–752.

Golbe, L. I., and Ohman-Strickland, P. A. (2007). A clinical rating scale for progressive supranuclear palsy. *Brain* 130, 1552–1565.

Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci.* 9, 188–194.

Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R. C., et al. (1996). Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome). *Neurology* 47, 1–9.

Otero-Millan, J., Serra, A., Leigh, R. J., Troncoso, X. G., Macknik, S. L., and Martinez-Conde, S. (2011). Distinctive features of saccadic intrusions and microsaccades in progressive supranuclear palsy. *J. Neurosci.* 31, 4379–4387.

Pinkhardt, E. H., Jurgens, R., Becker, W., Valdarno, F., Ludolph, A. C., and Kassubek, J. (2008). Differential diagnostic value of eye movement recording in PSP-parkinsonism, Richardson's syndrome, and idiopathic Parkinson's disease. *J. Neurol.* 255, 1916–1925.

Pinkhardt, E. H., and Kassubek, J. (2011). Ocular motor abnormalities in Parkinsonian syndromes. *Parkinsonism Relat. Disord.* 17, 223–230.

Schneider, E., Bartl, K., Bardins, S., Dera, T., Boning, G., and Brandt, T. (2006). Documentation and teaching of surgery with an eye movement driven head-mounted camera: see what the surgeon sees and does. *Stud. Health Technol. Inform.* 119, 486–490.

Schneider, E., Villgrattner, T., Vockeroth, J., Bartl, K., Kohlbecher, S., Bardins, S., et al. (2009). Eyeseecam: an eye movement-driven head camera for the examination of natural visual exploration. *Ann. N.Y. Acad. Sci.* 1164, 461–467.

Schumann, F., Einhäuser, W., Vockeroth, J., Bartl, K., Schneider, E., and König, P. (2008). Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *J. Vis.* 8, 12.1–12.17.

Tanyeri, S., Lueck, C. J., Crawford, T. J., and Kennard, C. (1989). Vertical and horizontal saccadic eye movements in parkinson's disease. *Neuroophthalmology* 9, 165–177.

Zampieri, C., and Di Fabio, R. P. (2008). Balance and eye movement training to improve gait in people with progressive supranuclear palsy: quasi-randomized clinical trial. *Phys. Ther.* 88, 1460–1473.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 October 2012; accepted: 27 November 2012; published online: 13 December 2012.

Citation: Marx S, Respondek G, Stamelou M, Dowiasch S, Stoll J, Bremmer F, Oertel WH, Höglinger GU and Einhäuser W (2012) Validation of mobile eye-tracking as novel and efficient means for differentiating progressive supranuclear palsy from Parkinson's disease. Front. Behav. Neurosci. ***6****:88. doi: 10.3389/fnbeh.2012.00088*

Kapitel 6

Study V: How to become a mentalist

OPEN ACCESS Freely available online PLOS ONE

How to Become a Mentalist: Reading Decisions from a Competitor's Pupil Can Be Achieved without Training but Requires Instruction

Marnix Naber[1,2,3,4*], **Josef Stoll**[1], **Wolfgang Einhäuser**[1,5], **Olivia Carter**[2]

1 Neurophysics, Philipps-University, Marburg, Germany, **2** School of Psychological Sciences, University of Melbourne, Parkville, Victoria, Australia, **3** Vision Sciences Laboratory, Harvard University, Cambridge, Massachusetts, United States of America, **4** Cognitive Psychology Unit, Leiden University, Leiden, The Netherlands, **5** Center for Interdisciplinary Research (ZiF), Bielefeld, Germany

Abstract

Pupil dilation is implicated as a marker of decision-making as well as of cognitive and emotional processes. Here we tested whether individuals can exploit another's pupil to their advantage. We first recorded the eyes of 3 "opponents", while they were playing a modified version of the "rock-paper-scissors" childhood game. The recorded videos served as stimuli to a second set of participants. These "players" played rock-paper-scissors against the pre-recorded opponents in a variety of conditions. When players just observed the opponents' eyes without specific instruction their probability of winning was at chance. When informed that the time of maximum pupil dilation was indicative of the opponents' choice, however, players raised their winning probability significantly above chance. When just watching the reconstructed area of the pupil against a gray background, players achieved similar performance, showing that players indeed exploited the pupil, rather than other facial cues. Since maximum pupil dilation was correct about the opponents' decision only in 60% of trials (chance 33%), we finally tested whether increasing this validity to 100% would allow spontaneous learning. Indeed, when players were given no information, but the pupil was informative about the opponent's response in all trials, players performed significantly above chance on average and half (5/10) reached significance at an individual level. Together these results suggest that people can in principle use the pupil to detect cognitive decisions in another individual, but that most people have neither explicit knowledge of the pupil's utility nor have they learnt to use it despite a lifetime of exposure.

Citation: Naber M, Stoll J, Einhäuser W, Carter O (2013) How to Become a Mentalist: Reading Decisions from a Competitor's Pupil Can Be Achieved without Training but Requires Instruction. PLoS ONE 8(8): e73302. doi:10.1371/journal.pone.0073302

Editor: Cristina Becchio, University of Turin and the Italian Institute of Technology, Italy

Received April 17, 2013; **Accepted** July 19, 2013; **Published** August 26, 2013

Funding: This work was supported by the German Research Foundation (DFG) through Research Training Group 885 "NeuroAct" (MN), grant EI852/1 (JS & WE), Netherlands Organisation for Scientific Research (MN), and by an Australian National Health and Medical Research Council fellowship # 628590 (OC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: marnixnaber@gmail.com

Introduction

The notion of "mind-reading" has long been a theme within popular culture but with the development of new brain imaging methods for "decoding" what a person is seeing or thinking (e.g., [1,2]), "mind-reading" has begun to move into the scientific mainstream. The success of these methods has also renewed interest in the question to what extent subtle facial signals may provide clues into one's private thoughts. It is well-established that facial expressions and gaze direction can reveal some information about an individual's emotional state or intention [3,4]. One question that remains open, however, is whether pupil dilation can also be used to gain strategic insights into another's mind.

It has been shown that observers tend to mirror their pupil size to those of others [5] suggesting that humans have an autonomic system that is responsible for implicit (subconscious) monitoring of others' pupil sizes [6,7]. These systems may further play a specific role during the unconscious processing of socially relevant cues such as emotions [5,8] and attractiveness [9–15]. These studies thus imply that the pupil is a potential source of information during interpersonal interactions, at least in an emotional context. However, it remains unknown whether observers can exploit pupil dynamics in other, non-emotional circumstances.

Pupil dilation is known to accompany a wide range of behaviors and mental processes, including load [16], arousal [17], alertness [18], working memory load [19–21], attention [22,23], familiarity [24], emotions [25–28], high-level visual processing [29], and making a conscious decision [30–34]. It has also been demonstrated recently that eye tracking cameras can capture small changes in pupil size that predict

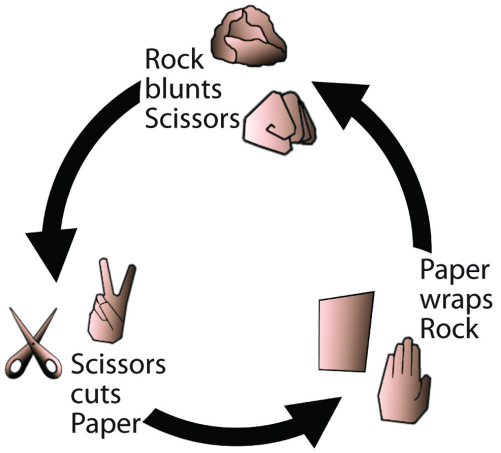


Figure 1. Schematic illustration of the rock-paper-scissors game. In a game of rock-paper-scissors, the relative choices of the two competitors determine the outcome.
doi: 10.1371/journal.pone.0073302.g001

cognitive events such as an act of deception [35] and the timing of decisions [33]. Here we examine whether human observers can - and do - extract similar information about another individual's cognitive decisions from these subtle changes in pupil diameter.

To test this we modified the popular childhood game Rock-Paper-Scissors (RPS) (see Figure 1). In the original version of the game participants are required to make their decisions simultaneously. One feature of the traditional version is that it is possible to gain an advantage by taking past decisions into account because people have difficulty generating random response sequences in general [36] and in the context of RPS [37]. To ensure that the prior history of events was not informative in the context of this experiment, we pre-recorded a sequence of 75 games played by 3 "opponents" (25 games each) and played them back to a new set of players in a random order. Critically, the response of the pre-recorded opponent remained concealed during the game and was revealed immediately after the player had indicated their own choice of Rock, Paper or Scissors. Therefore, while the temporal sequence of the game was radically altered, this modification allowed us to maintain the critical elements of the game (concealed mutual decision in a competitive game environment), while randomizing the order of game presentation to ensure that the prior sequence of decisions was uninformative.

Using this modified version of RPS, here we demonstrate for the first time that people can use the pupil to detect cognitive decisions in another individual, but that most people have to first be made explicitly aware of the strategic information provided by pupil dilation.

Materials and Methods

2.1: Ethics Statement

The experiments were approved by the ethics committees of University of Melbourne and Philipps-University, Marburg's department of psychology, and conformed to the ethical principles of the Declaration of Helsinki.

2.2: Participants

In total 33 volunteers (age: 19-30) participated in the study. Three of them served as "opponents" in that their responses were used to generate the stimuli used in the main experiments. These opponents were recruited from the Philipps-University Marburg, where the filming was conducted and the stimuli were generated. The remaining 30, filling the role of "players", were recruited and tested in Melbourne, Australia. The 30 players each participated in one or more of the experimental conditions as indicated in Figure 2. All participants were naïve to the purpose of the experiments, gave written-informed consent before participation, and received payment for participation in addition to performance-dependent reward ($10-$20).

2.3: Procedure

2.3.1: Rock-paper-scissors rules. The two-player game had a straightforward rule structure, which all participants understood with no training. The outcome (win/loss/draw) of each game was determined by the relative choices. Rock wins against scissors, scissors against paper, paper against rock (Figure 1). If both competitors chose the same option, the game was drawn.

2.3.2: Stimulus construction (Opponents' games). In a first phase, we constructed stimuli by recording videos of the pupil dilations of 3 participants ("opponents") who each played 25 games of rock-paper-scissors. The opponents played against a computer in a room with low ambient light levels. The words "rock", "paper", "scissors" were read out by a text-to-speech-voice converter and presented through computer speakers at comfortable listening volume. Opponents were presented the audio track of "rock", "paper", "scissors" in random order with 4-s intervals between each word onset, and were asked to indicate their choice by pressing a button immediately after the respective audio word was presented. At the completion of each game (i.e., 4s after the last option was read out), the computer's (random) choice was shown on the screen and feedback about the resulting outcome (win/loss/draw), and the associated monetary reward was provided. Throughout the rest of the game the screen remained a blank uniform grey. During each game, video of their left eye was recorded by a Grasshopper GRAS-03K2M camera (Point Grey Research, Richmond, BC, Canada) at 120 Hz and 640x480 resolution and stored together with the presented audio track (Movie S1). It was these recorded movies of the opponents left eye that served as stimulus material for the main experiments.

To verify whether the time of maximal pupil dilation was indeed informative about an opponent's choice, we first analyzed the opponent's responses and their corresponding

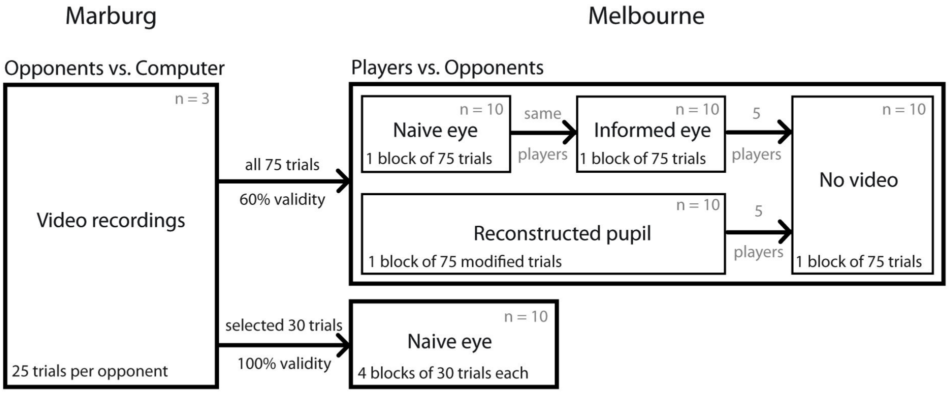


Figure 2. Participant breakdown. Three opponents each played 25 modified RPS games against a computer. During these games video of the opponents' eyes was recorded along with the audio produced by the computer. For all conditions, except "100% validity", all 75 of these games were used. For the 100% validity condition, 30 (out of 45) videos in which maximum pupil dilation indeed followed the choice of the opponent were selected with 10 games for each of the 3 intervals respectively. These valid games were then randomly presented as in the original "naïve eye" condition. Ten players participated first in the "naïve eye" and subsequently in the "informed eye" condition to allow a within-subject comparison. A distinct set of 10 players participated in the reconstructed pupil condition. Half (5) of each group participated in addition in the "no video" control. Finally, a distinct set of 10 players participated in the 100% validity condition.
doi: 10.1371/journal.pone.0073302.g002

pupil dilation. Opponent's choices were spread approximately evenly over the 3 intervals with 28, 25 and 22 selections for first, second and third option, respectively. Analysis showed that the opponents' pupil size varied substantially throughout their games despite minimal variability in external light sources (opponent's viewed a blank screen in a dimly lit room during all games). On average the pupil measured 4.6mm during the games, and the difference between minimum and maximum in each game amounted to 3.8mm on average (SD across games: 1.0mm). In 45 out of 75 games, maximum pupil dilation followed the selected word (Figure 3). Hence, the time of maximum dilation was a significant marker for the opponent's choice with a validity of 60% (45/75 games, compared to chance level of 25/75 – indicated by the dashed lines in Figure 4), which is compatible with earlier data obtained in a free choice scenario [33].

2.3.3: Conditions (Players' games). A separate group of 30 observers ("players") - distinct from the "opponents" - played rock-paper-scissors against the video recordings of the opponent's games under a number of conditions. Players sat in a room with low ambient light levels at a distance of about 120cm from the screen with their head stabilized by a chin rest. Videos (8-bit grayscale, 640x480pixels@60Hz on a 1920x1200 pixel, 52x32.5cm wide TFT screen) were presented centrally subtending a visual angle of about 4° x3°, with the pupil covering about 0.25° to mimic real-life conversation distance of about 50cm [38]. In all conditions, each game consisted of the same audio track ("rock", "paper", "scissors" in random order with 4-s intervals) that had been presented to the respective opponent during the corresponding game.

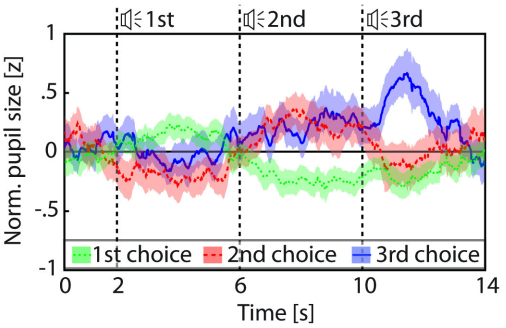


Figure 3. Pupil responses to cognitive decisions. Mean pupil size (diameter, normalized to z-scores within each opponent) and its standard error (shading) for games in which opponents selected the first, second or third option; average pupil dilation peaks shortly after the presentation of the selected word (dashed vertical lines).
doi: 10.1371/journal.pone.0073302.g003

Players were instructed to report their choice ("rock", "paper" or "scissors") via button response at the end of each game (i.e., approximately 4s after the last option had been read out). Immediately after the players' choice had been indicated, their opponent's selection was shown on the screen and feedback about the resulting outcome (win/loss/draw), and the

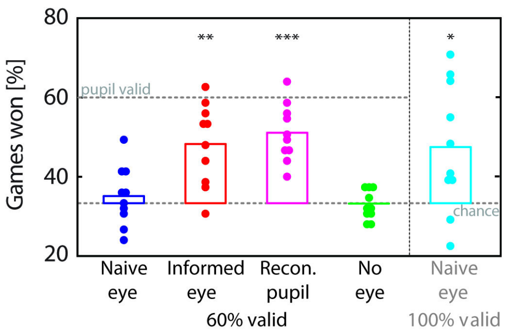


Figure 4. Group performance across conditions. Percentage of games won on average (bar) and by individual players (circles). For the 60%-validity conditions upper dashed line indicates maximum performance if a player always selected the interval signaled by the pupil (correct in 45/75 games).

doi: 10.1371/journal.pone.0073302.g004

associated monetary reward was provided. Importantly, as the opponent's choices were recorded during the video recording phase, the feedback provided to players was accurate with respect to the original opponent's games.

In all conditions, the games were presented in random order such that no information was provided by the previous pattern of choices, and players were correctly informed about this fact prior to the experiment. The visual information available to the player depended on condition:

- (1) *Naïve-eye* condition: To assess whether naïve participants had any explicit or implicit knowledge about the utility of the pupil, players were shown the full videos of their opponents' left eye and were instructed simply to "look for any behavioral signs that could reveal the opponent's decision."
- (2) *Informed-eye* condition: To determine whether pupil dilation could, in principle, be used to gain a strategic advantage, players were shown the full videos of their opponents' left eye and were informed that "the largest pupil dilation should follow the presentation of the word selected by the recorded opponent".
- (3) *Reconstructed pupil* condition: The video was replaced by a black disk on a gray background that matched the pupil size at each point in time (Movie S2). As this condition was otherwise identical to the "informed-eye" condition, it served as an important control to rule out any contribution of non-pupil factors (such as blinks or facial movements) in the players' performance.
- (4) *No-video* condition: Players were also presented the same audio track of the RPS games while viewing a blank grey screen (in all other respects the basic procedure was the same). Consistent with the previous conditions the players were asked to indicate their selection immediately after the presentation of the audio track and they were provided correct feedback after each game. This condition served as control to test that indeed no information was available from the audio sequence or any biases/patterns in the opponent's choices.
- (5) *Naïve-eye* (*100% validity*) condition: This condition aimed to determine whether observers could ever learn to use the signal, by using the most optimal conditions. This served as a control to show that there was no technical or perceptual limitation that prevented players from using the pupil in the naïve-eye 60% validity condition. Unlike in the four aforementioned conditions, which used all the games recorded from the opponents (i.e., irrespective of whether the pupil correctly signaled the option chosen by the opponent), here 30 games were selected. These were selected by first identifying the 45 "valid" games - for which maximum pupil dilation followed the selected option. From these 45 games we selected the first 10 games in chronological order that corresponded to each of the first, second or third interval. Instructions were identical to the original naïve-eye condition. This condition allowed us to provide optimal feedback to test whether the players could ever learn to exploit the pupil signal in our RPS game scenario. Since only 30 trials were available, these were repeated four times in separate blocks and randomized independently within each block.

2.4: Data analysis

Average performance in each condition was compared against chance (33%) by a two-sided t-test. Average performance between original *naïve-eye* condition and *informed-eye* condition (same set of players) was compared by a paired two-sided t-test, other comparisons between conditions (partially or fully distinct set of players – see Figure 2 for participant breakdown) by unpaired two-sided t-tests. The unpaired tests, which for the comparison to the "no eye" condition are less powerful, and thus more conservative with respect to the hypotheses considered.

To test whether individuals performed significantly above chance in a condition, irrespective of whether the group on average succeeded in doing so, we compared individual performance over the course of each experiment against chance by means of a Binomial test. A Binomial test is an exact statistical test that provides the probability (p) that for a certain total number of games(n), a certain amount of wins (k) or more are to be expected by chance. For the present case of 1/3 chance, the (one-sided, as learning can be expected to improve performance) p-value is thereby given to $p = \sum_{k \leq i \leq n} \binom{n}{i} \left(\frac{1}{3}\right)^{i} \left(\frac{2}{3}\right)^{(n-i)}$. Since 10 tests have to be performed in each condition, we refer to a result as "significant", only if the p-value is below 0.005, corresponding to a Bonferroni-corrected alpha-level of 5%.

Results

3.1: Players can readily exploit opponent's pupil signal, but most require instruction

3.1.1: Average performance. Ten players competed against the original unedited movies of the opponents' games, including the original "rock"-"paper"-"scissors" audio track and the corresponding video of their left eye (Movie S1). When no instruction was given regarding the pupil (naïve-eye condition), their average performance was indistinguishable from chance (Figure 4, dark blue; $t(9) = 0.73$, $p = 0.48$). When the same 10 players repeated the experiment with instruction to look for the maximum pupil dilation ("informed-pupil"), their performance improved to significantly better than both chance (Figure 4, red; $t(9) = 4.56$, $p = 0.001$) and performance in the naïve-eye condition (t(9) = 3.03; $p = 0.01$).

A group of 10 new players also performed clearly above chance when they were presented movies of the "reconstructed pupil" represented by an expanding/constricting black disk on gray a background (Movie S2), (Figure 4, magenta; $t(9) = 7.71$, $p = 3.02\text{x}10^{-5}$). This condition merely served to test whether participants were not exploiting any potentially useful signals other than the pupil. As observers performed as well as in the "instructed-pupil" condition in the absence of all other cues, this suggests that the pupil is at least as informative as in combination with other facial information.

To ensure that none of the players could exploit any possible sequential order of the opponent's choices, in all conditions the order of games had been randomized. To further ensure that there was no possible strategy involved, we tested 5 players of each player group in an additional control condition ("no eye"), in which only the audio track was presented for the identical collection of games tested in the previous conditions. In the absence of any visual information, performance was indistinguishable from chance ($t(9) = 0.46$, $p = 0.66$, Figure 4, green) and significantly worse than in the "informed-eye" ($t(18) = 4.45$, $p = 3.10\text{x}10^{-4}$) and the "reconstructed-pupil" ($t(18) = 7.09$, $p = 1.32\text{x}10^{-6}$) conditions, but indistinguishable from the naïve-eye condition ($t(18) = 0.86$, $p = 0.40$). This control verified that there was indeed no information beyond the video signal that was being exploited by players.

3.1.2: Individual performance. While there was no indication that the average *naïve-eye* player could exploit pupil dilation, the question remained whether a "lucky few" individuals could learn to use the pupil without explicit instruction. Testing individual performance over the course of the naïve-eye experiment by means of a Binomial test, we found one individual in the naïve-eye condition that indeed increased their chance of winning significantly above chance (37/75 wins, p=0.003, Figure 5a). In comparison, by the end of the 75 games, 7/10 players reached significance at an individual level for "informed-eye" (all 7p<0.0005, Figure 5b) and 9/10 for reconstructed-pupil (all 9p<0.0005, Figure 5c). In the control "no-eye" condition, no individual showed any sign of learning (Figure 5d, all p>0.05).

3.2: More players learn to exploit the pupil without instruction, if its validity is increased

While the timing of maximum dilation, which observers were instructed to look for, was informative about an opponent's choice in the majority of games, such 60%-validity is far from perfect. Most players reported in debriefing after the naïve-eye condition that they had tried using pupil-related cues, but varied strategies between looking for constrictions and dilations, and intermixed information from other facial signals, events like blinks, eye-brow twitches, and head movements. This raised the question as to whether the lack of learning in most individuals was a consequence of the relative proportion of valid and invalid games. To test whether players could theoretically ever spontaneously learn to use the information conveyed by the pupil, we selected 30 games (10 from each response interval) for which maximum pupil accompanied correct choice (i.e., 100% validity). A fresh set of 10 naïve players played 4 randomized blocks of these 30 games. In all other respects this "100%-validity" naïve-eye condition was identical to the original naïve-eye condition. Unlike in the original 60% validity naïve-eye condition, players significantly performed above chance on average (Figure 4, cyan; t(9) = 2.77, $p = 0.022$) and better than in the *no-video* condition ($t(18) = 5.22$, $p = 5.77\text{x}10^{-5}$). On an individual level, 5 out of 10 individuals significantly performed above chance, showing clear signs of learning (Figure 5e; $p < 0.0005$). This shows that, in the absence of any instruction, people can – in principle – spontaneously learn to use information signaled by another individual's pupil. Nonetheless, as such high validity is unlikely to occur in real-world situations this condition serves as a control to verify that there is no principled inability to extract useful information for learning, when the pupil is embedded in its natural context. Indeed, the fact that the players show no clear ability to use this signal in the naïve condition, despite a life-time of exposure to other people's pupils, suggests that it is extremely unlikely that they are currently utilizing this signal in daily life.

Discussion

In the present paper we used an adapted version of the childhood game of rock-paper-scissors to demonstrate four key results. i) We filmed the left eye of a group of opponents as they played against a computer and confirmed earlier findings that pupil dilation increases at the time of a conscious choice. ii) We show, for the first time, that competitors can exploit the signal conveyed by another individual's pupil. iii) As performance was equivalent for games in which the players viewed the unedited footage of their opponent's eye or a reconstructed movie showing only the pupil diameter, it can be concluded that the information in the pupil must be at least as informative as that conveyed by the pupil plus other facial features. iv) It was found that – in principle – the information conveyed by the pupil can be learned without any explicit instruction. However, such learning requires conditions that are so constrained that a general implicit knowledge and use of this signal in daily life appears unlikely.

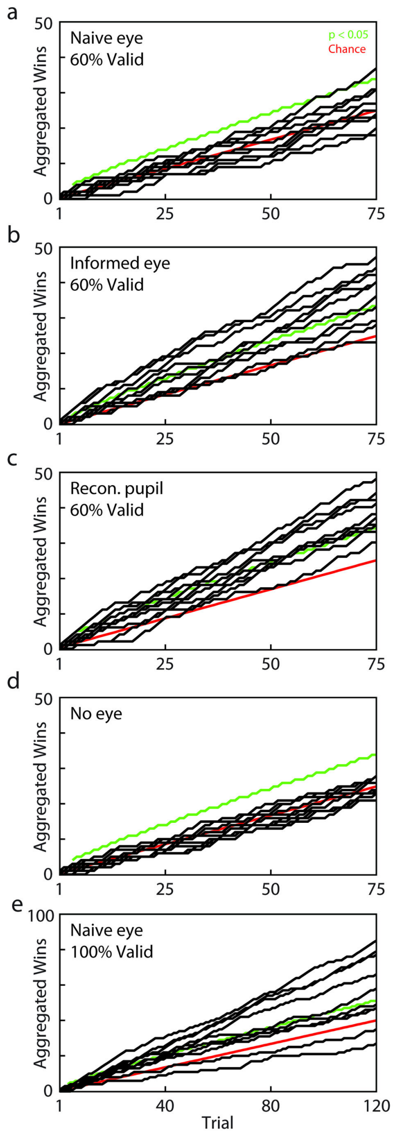


Figure 5. Individual performance across conditions. Learning curves for the 5 different experimental conditions (*Black lines*: aggregate number of wins for each individual; *red line*: chance level; *green line*: 5% significance level for Binomial test at the given number of games). Note that all preceding games are considered in this computation, such that learned information has to persist longer if learning starts later. **a**) naïve-eye **b**) informed-eye **c**) reconstructed-pupil d) no-eye control condition **e**) naïve-eye (100% validity).

doi: 10.1371/journal.pone.0073302.g005

Pupil dilation in the context of a decision has been known for over half a century [39]. As we confirm here, the link between dilation and decision is timed with sufficient precision to convey a covert decision [33], suggesting that careful monitoring of the pupil may assist the detection of deceptive acts [35]. Evidence also shows that the observation of another individual's pupil can influence the perception of emotions, such as sadness [5,8]. Such observed pupil dilation is known to modulate neural activity related to emotion processing, presumably without explicit awareness [5–8]. Despite this demonstrated link between pupil response and decision-making and the use of pupil dilation in an emotional context, no study has yet addressed whether competitors can *explicitly* exploit an opponent's pupil dilation in a competitive scenario. Our results on the explicit use of pupil dilation are two-fold. On the one hand, we find that explicit usage of the pupil signal is clearly possible. Furthermore – as shown by the "reconstructed pupil" condition – the pupil alone is at least as informative as its combination with other facial features. On the other hand, however, nearly all naïve observers require explicit instruction to exploit the information the pupil conveys.

It is possible that the constrained nature of the experimental design has limited the relevance of this study to more realistic scenarios. The use of prerecorded rather than live opponents and the presentation being limited to the eye region rather than full faces might have created an unusual situation that failed to engage an implicit ability to link pupil dilations to decision-making. Although all of these factors need to be considered in future studies on the role of pupil dynamics in interactive contexts, we consider it unlikely that the procedural constraints undermined people's awareness of the pupil's cognitive signals in this instance. Firstly, players were able to instantly use the pupil once they were made aware of the association between dilations and decision-making. This shows that their inability to pick up the pupil cue prior to instruction – despite exposure and opportunity to learn from others' pupils across a lifetime – was not a consequence of technical or perceptual limitations associated with our stimuli.

While different limitations could in principal apply to *using* the signal as compared *to learning to use* it, the finding that observers could learn to use the same signal under conditions of increased "100% validity" again argues against the artificial nature of our paradigm obscuring any underling implicit capacity to *use* or *learn* the value of the pupil signal in more realistic, lower validity, conditions. This makes an interesting distinction between the apparent inability to exploit the pupil in a competitive situation to its apparent implicit use in emotion processing [5–8].

Because the "naïve-eye" condition always preceded the "informed-eye" condition, and we did see some evidence of learning in the 100% validity condition, one limitation of this design is that we cannot rule out a contribution of learning in the dramatic improvement of players in the "informed eye" condition. We consider this unlikely, however, as we would expect to see some traces of learning, such as a gradual increase in correct trials at the end of the "naïve eye" block. With the possible exception of one player, such an effect was, however, not observed in general (Figure 5a). Furthermore, most participants in the "reconstructed pupil", who performed no other condition before, were also able to utilize the pupil dilations immediately (Figure 5c).

While we chose rock-paper-scissors for the present study because its intuitive rule structure allowed naïve participants to play without training, one strength of the game is that it shares some elements with more elaborate social exchanges and competitive situations. Pupil dilation is clearly sufficiently salient to be detected and could potentially be used in a range of social contexts. The question remains open, however, whether there exist any situations in which people are currently using pupil dilation or could be instructed on how to use the pupil to their advantage in more realistic scenarios.

In conclusion, our results show on the one hand that people could use pupil size as a cue in competitive interactions, but on the other hand render it unlikely that pupil dilation is being used in this way in everyday life. Although we cannot rule out that such situations exist, this seems in sharp contrast to emotional processing, where perceived pupil size modulates emotion perception and its neural substrate [5,8]. Given recent claims of "mind-reading" or "brain-reading" in the context of brain imaging (e.g., [1,2]), it remains remarkable a comparably simple physiological signal allows similar degrees of "mind-reading" in real-time. Even more remarkable, such "mind-reading" seems possible to nearly anyone using standard video equipment and the naked eye. This makes pupil dilation a signal utilizable for communication, which is of particular interest to patients with severe motor impairments [40].

Supporting Information

Movie S1. Example movie of an opponent's pupil.
Video depicting three games as the players viewed it in the informed-eye and naïve-eye conditions condition. If you want to try the experiment yourself, watch the movies and pick the best option to beat your opponent. The audio track consists of the words "rock", "paper", and "scissors", presented in 4-s intervals in the identical (randomly ordered) sequence the opponent heard when the video was recorded. In case the video format does not work on your computer, various formats are available at http://www.staff.uni-marburg.de/~wetgast/rps/ (Correct answers for both movies: Opponent 1 selected "Rock" (3rd option). Opponent 2 selected "Paper" (1st option). Opponent 3 selected "Scissors" (2nd option). Winning options thus were "paper" in the first game, "Scissors" in the second and "rock" in the last).
(MOV)

Movie S2. Example movie of an opponent's reconstructed pupil.
Video depicting the same three games shown in Movie S1 for the reconstructed-pupil condition.
(MOV)

Author Contributions

Conceived and designed the experiments: MN OC WE. Performed the experiments: MN JS. Analyzed the data: MN. Contributed reagents/materials/analysis tools: MN JS WE OC. Wrote the manuscript: MN WE OC.

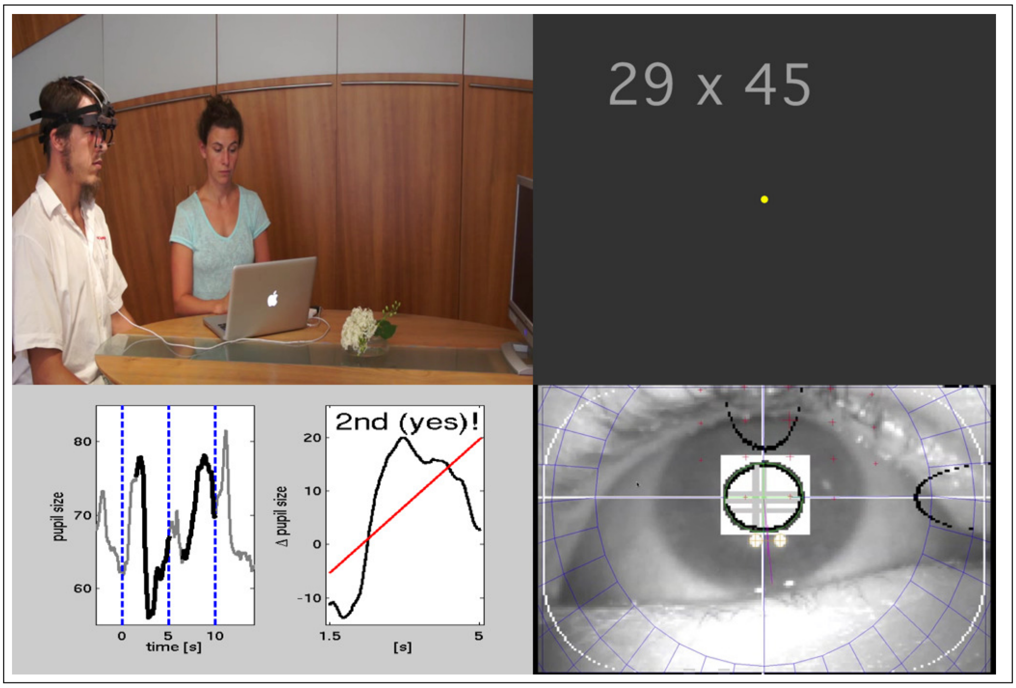
References

1. Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7: 523-534. doi:10.1038/nrn1931. PubMed: 16791142.
2. Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B et al. (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21: 1641-1646. doi:10.1016/j.cub.2011.08.031. PubMed: 21945275.
3. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I (2001) The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. J Child Psychol Psychiatry 42: 241-251. doi: 10.1111/1469-7610.00715. PubMed: 11280420.
4. Frith CD, Frith U (2008) Implicit and explicit processes in social cognition. Neuron 60: 503-510. doi:10.1016/j.neuron.2008.10.032. PubMed: 18995826.
5. Harrison NA, Singer T, Rotshtein P, Dolan RJ, Critchley HD (2006) Pupillary contagion: central mechanisms engaged in sadness processing. Soc Cogn Affect Neurosci 1: 5-17. doi:10.1093/scan/nsl006. PubMed: 17186063.
6. Harrison NA, Gray MA, Critchley HD (2009) Dynamic pupillary exchange engages brain regions encoding social salience. Soc Neurosci 4: 233-243. doi:10.1080/17470910802553508. PubMed: 19048432.
7. Amemiya S, Ohtomo K (2012) Effect of the observed pupil size on the amygdala of the beholders. Soc Cogn Affect Neurosci 7: 332-341. doi: 10.1093/scan/nsr013. PubMed: 21421732.
8. Harrison NA, Wilson CE, Critchley HD (2007) Processing of observed pupil size modulates perception of sadness and predicts empathy. Emotion 7: 724-729. doi:10.1037/1528-3542.7.4.724. PubMed: 18039039.
9. Simms TM (1967) Pupillary response of male and female subjects to pupillary difference in male and female pictures. Percept Psychophys 2: 553-555. doi:10.3758/BF03210265.
10. Stass J, Willis F (1967) Eye contact, pupil dilation, and personal preference. Psychon Sci 7: 375-376.
11. Hamel RF (1974) Female subjective and pupillary reaction to nude male and female figures. J Psychol 87: 171-175. doi: 10.1080/00223980.1974.9915687. PubMed: 4443952.
12. Tomlinson N, Hicks R, Pellegrini R (1978) Attributions of female collegestudents to variations in pupil size. Bull Psychon Soc 12: 447-478.
13. Bull R, Shead G (1979) Pupil dilation, sex of stimulus, and age and sex of observer. Percept Mot Skills 49: 27-30. doi:10.2466/pms.1979.49.1.27. PubMed: 503746.
14. Tombs S, Silverman I (2004) Pupillometry: Asexual selection approach. Hum Behavior 25: 221-228. doi:10.1016/j.evolhumbehav.2004.05.001.
15. Demos KE, Kelley WM, Ryan SL, Davis FC, Whalen PJ (2008) Human amygdala sensitivity to the pupil size of others. Cereb Cortex 18: 2729-2734. doi:10.1093/cercor/bhn034. PubMed: 18372291.
16. Hess EH, Polt JM (1964) Pupil Size in Relation to Mental Activity during Simple Problem-Solving. Science 143: 1190-1192. doi:10.1126/science.143.3611.1190. PubMed: 17833905.
17. Bradshaw J (1967) Pupil size as a measure of arousal during information processing. Nature 216: 515-516. doi:10.1038/216515a0. PubMed: 6057275.
18. Yoss RE, Moyer NJ, Hollenhorst RW (1970) Pupil size and spontaneous pupillary waves associated with alertness, drowsiness, and sleep. Neurology 20: 545-554. doi:10.1212/WNL.20.6.545. PubMed: 5463609.

19. Granholm E, Asarnow RF, Sarkin AJ, Dykes KL (1996) Pupillary responses index cognitive resource limitations. Psychophysiology 33: 457-461. doi:10.1111/j.1469-8986.1996.tb01071.x. PubMed: 8753946.
20. Poock GK (1973) Information processing vs pupil diameter. Percept Mot Skills 37: 1000-1002. doi:10.2466/pms.1973.37.3.1000. PubMed: 4764491.
21. Kahneman D, Beatty J (1966) Pupil diameter and load on memory. Science 154: 1583-1585. doi:10.1126/science.154.3756.1583. PubMed: 5924930.
22. Kahneman D (1973) Attention and effort. New Jersey, USA: Prentice Hall.
23. Binda P, Pereverzeva M, Murray SO (2013) Attention to bright surfaces enhances the pupillary light reflex. J Neurosci 33: 2199-2204. doi:10.1523/JNEUROSCI.3440-12.2013. PubMed: 23365255.
24. Naber M, Frässle S, Rutishauser U, Einhäuser W (2013) Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. J Vis 13: 11-. PubMed: 23397036.
25. Simpson HM, Molloy FM (1971) Effects of audience anxiety on pupil size. Psychophysiology 8: 491-496. doi:10.1111/j.1469-8986.1971.tb00481.x. PubMed: 5094932.
26. Steinhauer SR, Boller F, Zubin J, Pearlman S (1983) Pupillary dilation to emotional visual stimuli revisited. Psychophysiology 20: 472.
27. Harrison NA, Singer T, Rotshtein P, Dolan RJ, Critchley HD (2006) Pupillary contagion: central mechanisms engaged in sadness processing. Soc Cogn Affect Neurosci 1: 5-17. doi:10.1093/scan/nsl006. PubMed: 17186063.
28. Naber M, Hilger M, Einhäuser W (2012) Animal detection and identification in natural scenes: image statistics and emotional valence. J Vis 12: ([MedlinePgn:]) PubMed: 22281692.
29. Naber M, Nakayama K (2013) Pupil responses to high-level image content. J Vis 13: ([MedlinePgn:]) PubMed: 23685390.
30. Bradshaw J (1967) Pupil size as a measure of arousal during information processing. Nature 216: 515-516. doi:10.1038/216515a0. PubMed: 6057275.
31. Hess EH, Polt JM (1964) Pupil Size in Relation to Mental Activity during Simple Problem-Solving. Science 143: 1190-1192. doi:10.1126/science.143.3611.1190. PubMed: 17833905.
32. Beatty J, Wagoner BL (1978) Pupillometric signs of brain activation vary with level of cognitive processing. Science 199: 1216-1218. doi:10.1126/science.628837. PubMed: 628837.
33. Einhäuser W, Koch C, Carter O (2010) Pupil dilation betrays the timing of decisions. Front Hum Neurosci 4: 18. PubMed: 20204145.
34. Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD (2010) Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci 10: 252-269. doi:10.3758/CABN.10.2.252. PubMed: 20498349.
35. Wang JT, Spezio M, Camerer CF (2006) Pinocchio's Pupil: Using Eyetracking and Pupil Dilation To Understand Truth-telling and Deception in Games. Am Econ Rev 100: 984-1007.
36. Wagenaar WA (1972) Generation of random sequences by human subjects: A critical survey of literature. Psychol Bull 77: 65-72. doi:10.1037/h0032060.
37. Vickery TJ, Chun MM, Lee D (2011) Ubiquity and specificity of reinforcement signals throughout the human brain. Neuron 72: 166-177. doi:10.1016/j.neuron.2011.08.011. PubMed: 21982377.
38. Hall ET (1959) The silent language. New York: Doubleday Publishing Group.
39. Simpson HM, Hale SM (1969) Pupillary changes during a decision-making task. Percept Mot Skills 29: 495-498. doi:10.2466/pms.1969.29.2.495. PubMed: 5361713.
40. Stoll J, Chatelle C, Carter O, Koch C, Laureys S et al. (. (2013)) Pupil responses allow communication in locked-in syndrome patients. Curr Biol (. (2013)) PubMed: 23928079.

Kapitel 7

Study VI: Pupil responses allow communication in locked-in syndrome patients



Graphical abstract demonstrating the communication procedure.

Correspondences

Pupil responses allow communication in locked-in syndrome patients

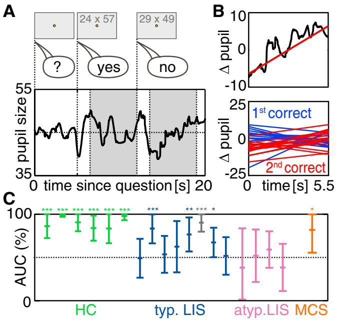
Josef Stoll[1,7], Camille Chatelle[2,7], Olivia Carter[3], Christof Koch[4,5], Steven Laureys[2], and Wolfgang Einhäuser[1,6,*]

For patients with severe motor disabilities, a robust means of communication is a crucial factor for their well-being [1]. We report here that pupil size measured by a bedside camera can be used to communicate with patients with locked-in syndrome. With the same protocol we demonstrate command-following for a patient in a minimally conscious state, suggesting its potential as a diagnostic tool for patients whose state of consciousness is in question. Importantly, neither training nor individual adjustment of our system's decoding parameters were required for successful decoding of patients' responses.

Pupil size is controlled by the complementary activity of muscles innervated by parasympathetic and sympathetic projections. In addition to the pupil dilation known to accompany emotionally arousing events, more subtle pupil dilation events have been linked to a variety of mental functions [2], including decision-making [3]. Our paradigm used mental arithmetic as a tool for patients to control and maximize their pupil dilation to signal their responses [4]. Each trial had the following structure (Figure 1A): an experimenter read out a factual question with a clear yes/no answer, such as "Is your age 20?". The correct answer was known for all questions, and 'correct decoding' refers to this ground truth. Five seconds after the question ended, a computer voice read out the first answer option: "yes" in half of the trials, "no" in the other half. Simultaneously with the onset of this read-out, a calculation task was presented in large font (150 pt) on a computer screen placed approximately one to two metres from the participants, and remained visible for a fixed duration. After this 'first calculation interval', the computer voice read out the alternative option ("no"/"yes"), and simultaneously a second calculation task was presented on the screen. This calculation remained visible for the same duration as the first ('second calculation interval'). Participants were asked to perform the calculation presented in the interval that accompanied the correct answer, and to ignore the calculation accompanying the incorrect answer. Throughout a session, each question was asked twice with the order of answers reversed. All trials were treated independently for all analyses except for the assessment of consistency.

The duration of the calculation intervals and the difficulty of the arithmetic problems were set for each individual prior to each experiment and remained fixed thereafter (see Supplemental Experimental Procedures). All other experimental parameters and analysis protocols were first established in healthy participants and then remained unchanged for all patients. In particular, the first 1.5 seconds of each calculation interval were discarded in all patients based on healthy participant data to reduce any possible impact of pupil responses to changes in the visual stimulus depicting the calculation task.

To reduce the pupil dynamics across a trial to a single value, we first subtracted pupil size in the first calculation interval from pupil size in the second. To the resulting difference trace (Figure 1B, black trace), a linear regression was then fit (Figure 1B, red line). All further analysis was based on the slope of this regression line ('pupil slope'). Pupil slope is by definition larger if the pupil dilates predominantly in the second interval, and smaller if pupil dilates predominantly in the first interval. Hence, if pupil control through mental arithmetic is successful [4], large pupil slopes correspond to the answer option presented second (Figure 1B, red), small pupil slopes to the option presented first (Figure 1B, blue). Success of decoding based on pupil slope is quantified in each individual by the area under receiver operator characteristics curve (AUC). Six healthy participants performed 30 trials each. For each individual, decoding of responses based on pupil slope was near ceiling (AUC range: 84–99%; Figure 1C, green) and each individual showed an AUC significantly different from chance (50%).



Current Biology

Figure 1. Pupil-size communication with locked-in syndrome patients.
(A) Example trial. *Top*: trial layout, *bottom*: corresponding pupil trace of typical locked-in syndrome patient #2. Calculations were presented above a fixation mark, which remained visible on the screen throughout the trial to minimize refocusing and thus re-accommodation; grey shading indicates part of the calculation intervals used for analysis; the unit of pupil size is pixel^2 (see Supplemental Experimental Procedures). (B) *Top*: the slope of the difference signal (black trace) is determined by linear regression (red line), and referred to as pupil slope of the trial; *bottom*: all pupil slopes for the example patient split by questions for which the correct answer corresponded to the first interval (blue) or the second (red). (C) Areas under the receiver operator characteristics curve (AUCs) with 95% confidence intervals for decoding of response from pupil size (***$p<0.001$; **$p<0.01$; *$p<0.05$). *Green*: healthy participants (HC), *blue*: patients with typical locked-in syndrome, second session of patient #5 in gray, *magenta*: patients with atypical locked-in syndrome, *orange*: patient in a minimal conscious state performing command following.

The system was then tested in seven 'typical' locked-in syndrome patients with brainstem stroke etiology (normal cognitive function, no supratentorial lesions, Supplemental Table S1). Five patients performed 30 trials, one patient (#4) 18 trials and another (#6) 40 trials. Three patients were significantly different from chance at an individual level (AUCs: 67%, 77%, 84%; Figure 1C, blue). For a further three patients decoding performance was above chance but failed to reach significance. Post-hoc analysis indicated that two out of these three would also reach significance individually (AUCs: 71%, 73%) by adjusting a single parameter — the onset of the interval used to compute the pupil slope (see ‘Alternative measures’ in the Supplemental Results). A single patient (#5) was retested on a different day (Figure 1C, grey), and showed improved performance in the second session (90%) as compared to the first (77%).

Four ‘atypical’ locked-in syndrome patients with supratentorial brain injuries were tested. Of those only one (#3) completed all 30 trials, while the others performed 10 (#1) or 18 (#2, #4) trials before showing obvious signs of fatigue. Decoding performance for all four supratentorial locked-in syndrome patients fluctuated around chance (AUC range: 38–59%; Figure 1C, magenta) and none of them reached significance individually.

When comparing the decoded answer for the two occurrences a patient is asked the same question, inconsistent responses are by definition uninformative, even if (as in practical application) ground truth is unknown. For consistently decoded questions — the system indicated either “yes” or “no” both times for a fixed cut-off point; for details see ‘Inter-block consistency’ in the Supplemental Results —decoding was *always* correct with respect to ground truth in the three significantly decoded patients and, with one exception, in all healthy participants. Of the other four locked-in syndrome patients that were asked all questions twice, but did not show significant decoding, two still had all (4/4) or all but one (6/7) of the consistently decoded questions correct, though these results in the absence of statistical significance in decoding performance need to be considered with caution. Nonetheless, for our sample of questions in significantly decoded patients, by simply asking the same question twice and considering both jointly, the system’s decoding was either correct or known to be uninformative, but never incorrect.

In the case of a non-communicative minimally conscious state patient, it became evident during the first couple of trials that he did not follow task instructions. Hence, instead of relying on the patient’s free choice of interval (answer), he was instructed by one experimenter when he had to perform the calculation. Of the 13 trials he performed in this command-following mode, the response could be decoded from the pupil slope at an AUC of 82%. Despite the low number of trials, this result was significantly above chance level (Figure 1C, orange).

Our data provide proof-of-principle for pupil dilation as means of communication in severely motor-impaired patients. With no training and no parameter adjustment to the individual, up to 90% decoding performance was reached. Rather than utilizing the response to a decision as such [3], which could, for example, be confounded with the difficulty of the decision, our paradigm allows patients to actively control their pupil dilation by modulating their mental effort. Whether or not patients actually solve the problem posed is of little relevance. Rather, mental arithmetic provides one robust way (amongst many) to manipulate one’s pupil dilation, even if — as in the minimally conscious state patients — no active engagement in the task is otherwise apparent.

Typical brain-computer-interfaces either use invasive methods [5] or EEG in combination with machine-learning techniques [6] to measure neural activity. Besides the risk of the surgical procedure or the maintenance demands (for example, electrode cleaning), with few exceptions [7] training with the individual is required. Pupil size controlled through mental effort offers an alternative path, reflecting (neuro-)physiological activity that is easy and inexpensive to measure in daily life, requiring nothing but a bedside camera. Furthermore, in cases of complete locked-in syndrome, approaches that require some residual volitional movements, such as sniffing [8] or blinking [9], are by definition unsuitable [10]. In contrast, our system may in principle be tested in patients in complete locked-in syndrome without training prior to an acute insult. Finally, the minimally conscious state data demonstrates our system’s potential usefulness as an additional diagnostic tool to assess a patient’s state of consciousness.

Supplemental Information

Supplemental Information includes results, experimental procedures, two figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.06.011.

Acknowledgements

We thank the patients, their families and the Association Locked In Syndrome (ALIS; France) and particularly F. Pellas, V. Sol and V. Blandin who helped with patient recruitment. This study was supported by the FNRS, European Commission (FP7-247919 DECODER to S.L.), James S. McDonnell Foundation, Mind Science Foundation, French Speaking Community Concerted Research Action, DFG (EI 852/1, EI 852/3 to W.E.), the Australian NHMRC (628590 to O.C.), the G. Harold & Leila Y. Mathers Foundations (to C.K.). We thank E. Schneider and his team for support with the EyeSeeCam; their contribution was supported by the BMBF (01 EO 0901).

References

1. Bruno, M.A., Bernheim, J.L., Ledoux, D., Pellas, F., Demertzi, A., and Laureys, S. (2011). A survey on self-assessed well-being in a cohort of chronic locked-in syndrome patients: happy majority, miserable minority. BMJ Open *1*, e000039.
2. Laeng, B., Sirois, S., and Gredbäck, G. (2012). Pupillometry - a window to the preconscious? Perspect. Psychol. Sci. *7*, 18–27.
3. Simpson, H.M., and Hale, S.M. (1969). Pupillary changes during a decision-making task. Percept. Mot. Skills *29*, 495–498.
4. Hess, E.H., and Polt, J.M. (1964). Pupil size in relation to mental activity during simple problem-solving. Science *143*, 1190–1192.
5. Brumberg, J.S., Nieto-Castanon, A., Kennedy, P.R., and Guenther, F.H. (2010). Brain-computer interfaces for speech communication. Speech Commun. *52*, 367–379.
6. Birbaumer, N., Murguialday, A.R., and Cohen, L. (2008). Brain-computer interface in paralysis. Curr. Opin. Neurol. *21*, 634–638.
7. Hill, N.J., Lal, T.N., Schröder, M., Hinterberger, T., Wilhelm, B., Nijboer F, Mochty, U., Widman, G., Elger, C., Schölkopf, B., *et al*. (2006). Classifying EEG and ECoG signals without subject training for fast BCI implementation: comparison of nonparalyzed and completely paralyzed subjects. IEEE Trans. Neural. Sys. Rehabil. Eng. *14*, 183–186.
8. Plotkin, A., Sela, L., Weissbrod, A., Kahana, R., Haviv, L., Yeshurun, Y., Soroker, N., and Sobel, N. (2010). Sniffing enables communication and environmental control for the severely disabled. Proc. Natl. Acad. Sci. USA. *107*, 14413–14418.
9. Bruno, M.A., Schnakers, C., Damas, F., Pellas, F., Lutte, I., Bernheim, J., Majerus, S., Moonen, G., Goldman, S., and Laureys, S. (2009). Locked-in syndrome in children: report of five cases and review of the literature. Pediatr. Neurol. *41*, 237–246.
10. Birbaumer, N., Piccione, F., Silvoni, S., and Wildgruber, M. (2012). Ideomotor silence: the case of complete paralysis and brain-computer interfaces (BCI). Psychol. Res. *76*, 183–191.

[1]Neurophysics, Philipps-University Marburg, Germany. [2]Coma Science Group, Cyclotron Research Centre, University and University Hospital of Liège, Belgium. [3]Psychological Sciences, University of Melbourne, Parkville, Australia. [4]Allen Institute for Brain Science, Seattle, USA. [5]Division of Biology, California Institute of Technology, Pasadena, USA. [6]Center for Interdisciplinary Research (ZiF), Bielefeld University, Germany.
[7]These authors contributed equally to the work.
*E-mail: wet@physik.uni-marburg.de

7.2 Supplemental Information

Supplemental Information: Pupil responses allow communication in locked-in syndrome

Josef Stoll, Camille Chatelle, Olivia Carter, Christof Koch, Steven Laureys, Wolfgang Einhäuser

Supplemental Results

Inter-block consistency

When possible the same question was used twice in each session, once in the first "block", once in the second (see Supplemental Experimental Procedures below). While for the main analyses all trials were treated as independent, this design allowed us to exclude some potential confounds. First, since the order of responses was reversed between blocks (if "yes" corresponded to the first interval when the question was asked first, "no" corresponded to the first interval when the questions was asked the second time and vice versa), we can fully exclude the possibility that the pupil responses reflected a preference for the first or second interval (rather than selection of the "yes" or "no" option). If this would have been the case, the respective question would have been decoded inconsistently (i.e., once correctly, once incorrectly) and thus overall performance would have been at chance. Second, the design allowed us to test whether the questions that were decoded consistently (i.e., "yes" in both blocks or "no" in both blocks for a given cut-off point), and thus would receive high confidence in practical use, were decoded correctly. Since the ROC includes all possible criteria (cut-off points) to define an answer as belonging to the first or second interval, respectively, we for this consistency analysis used the cut-off point for which the sum of both types of errors was minimal (1st interval decoded as 2nd plus 2nd interval decoded as 1st). Any slope larger than the cut-off point will be decoded as response to the second interval, any slope smaller as the first interval. This treats both errors symmetrically and is analogous to defining a cut-off point in clinical situations at maximum specificity and sensitivity. At these cut-off points, the significantly decoded LIS patients had 87% (26/30, #2), 80% (24/30, #5, 1st session), 83% (25/30, #5, 2nd session) and 68% (27/40) correctly decoded responses, which is close to the expected values as given by the AUCs (84%, 77%, 90%, 67%). For typical LIS patient #2, 11 of the 15 questions were decoded with the same response both times they were asked (either "yes" in both blocks, or "no" in both blocks). Of these 11, all decoding results (11/11) were correct with respect to ground truth. The same applied to the other LIS patients with significant decoding: the questions that were decoded consistently both times they were asked were always decoded correctly - 9/9 in #5's first session, 10/10 in #5's second session, and 7/7 (of 20 questions total) in #6. This result also held for all HCs, with the exception of a single question in HC #5. In the patients who did not complete at least 30 trials (typical LIS #4, atypical LIS #1, #2, #4, MCS patient), none or very few questions were repeated, precluding inter-block consistency analysis. Of the 4 patients, who were asked all 15 questions twice, but did not show significant decoding, two still had all (4/4, atypical LIS #3) or all but one (6/7, typical LIS #1) of their consistently decoded questions correct, though it should be noted that the definition of the cut-off point is more brittle in these cases, when the ROC fluctuates around chance. In any case, our results imply that in all LIS patients with AUCs significantly different from chance and in nearly all HCs, those questions that were decoded consistently, and thus would receive high confidence in practical use, were always decoded correctly.

Alternative measures

Pupil slope as defined here is only one of many possible scalar measures that could be used to quantify the pupil response of each trial. We chose this measure, as we had used it before in other paradigms [S1] and it was successful in healthy controls under the present paradigm. Keeping parameters unchanged is essential when transferring to a distinct population, as it avoids the problem of over-fitting for the relevant data (*here:* the patient data). However, we can ask post-hoc, whether simple alternative measures exist. One straightforward possibility is mean pupil size in a given part of the calculation intervals. Although it is conceivable that such mean-related measures exist for each individual and could be extracted with a sufficient number of trials using standard machine-learning techniques, across individuals they turned out to be less robust. Interestingly, when considering the time course of the mean over trials, some participants showed effects that were nearly reversed relative to each other. This is best exemplified for the two patients that showed best decoding performance for the pupil slope (typical LIS #2 and typical LIS #5, 2nd session, Figure 1C, blue): While one patient had a clearly larger mean in the first interval when the second interval contained the correct response, and this difference reduced in the second interval (after a short reversal immediately after the second option was presented, typical LIS #2, Supplemental Figure S2A), the other showed the reversed pattern in the first interval and a larger difference in the second interval (typical LIS #5, second session; Supplemental Figure S2B). This reversal in the difference between both answer options is indeed consistent on a trial-by-trial basis. When decoding is based on the mean pupil size in the part of the calculation intervals that are also used for computing pupil slopes in the main analysis, both patients can also be decoded significantly different from chance (Supplemental Figure S2C, D). However, in one case the AUC is significantly below, in the other significantly above chance. This still means that both patients could be decoded well (since there is no a prior assumption whether the mean should be smaller or larger in the interval following the correct response), if the system is trained and adjusted to the individual. Hence the mean in the interval can be an alternative (or even additional) measure if training with the patient is possible. However, when the system needs to be used without training or adjustment, the slope remains the preferable measure. The idiosyncrasies in the pupil absolute response make the inter- and intra-individual robustness of the pupil slope (compare Supplemental Figure S2E,F to Figure 1B) even more remarkable and make it a likely candidate not only for communication with LIS patients, but also – as suggested by the MCS data – possibly for improving the diagnosis of patients with disorders of consciousness.

In turn, idiosyncratic measures, such as the average size in a given interval, open the possibility that repeated training with the same patient may further improve results, in particular if online feedback is given and parameters are iteratively adjusted. Importantly, for the single patient who was tested twice (typical LIS #5), we saw no evidence of degradation in decoding performance, which would prohibit long-term use. While clearly beyond the proof-of-principle sought here, individual adjustment of parameters and individual training thus have the potential to further increase the system's efficiency, not only in terms of reliability but also in terms of bit-rate. While any bit-rate larger than 0 is of use for patients without established communication or whose state of consciousness is in question, such adaptations could make the proposed system a clinically useful application even for those LIS patients who already have an established mode of communication and environmental control.

One parameter that can easily be modified is the part of the calculation interval actually used for computing the pupil slope. Based on the HC data we chose to skip the first 1.5s for all data analysis. Testing the patients' data post-hoc shows that with adjustment for the skipped interval in 0.5s steps 5 out of 7 individuals in typical LIS would show significant decoding individually. The AUCs of all 7 patients then ranged from 64 to 85%, and patients #5‘s second session reached 98%. In contrast, despite some improvement, no atypical LIS patient reached significant decoding performance on an individual level (AUC range: 58-71%), even after such adjustment. Nonetheless, the fact that a straightforward adjustment to the individual improves decoding together with the improvement seen for typical LIS patient #5 in his second session supports the notion that the present system can be readily developed into a system that reaches stable performance on a level appropriate for daily use.

Supplemental Experimental Procedures

Participants

Healthy controls (HCs; age: 20-23, 3 males) were recruited from the Philipps-University Marburg (Germany) student body. Inclusion criteria were: age older than 18 and normal or corrected-to-normal visual acuity; exclusion criteria were: (1) history of psychiatric or neurologic illness, (2) requirement of visual correction that inferred with the measurement equipment.

Patients were recruited via the University Hospital of Liège (Belgium) and the Association for Locked-In Syndrome (ALIS, France). Experiments were conducted in the patients' homes throughout Belgium and France. Inclusion criteria were: (1) patients older than 18 and (2) presence of operational criteria for classical (i.e., total immobility except for vertical eye movements or blinking) or incomplete (i.e., permitting remnants of voluntary motion) LIS following brainstem stroke and without supratentorial brain lesions (i.e., typical etiology) or following severe brain injury with supratentorial brain lesions (i.e., atypical etiology). Exclusion criteria were: (1) documented history of prior brain injury; (2) premorbid history of developmental, psychiatric or neurologic illness resulting in documented functional disability up to time of the injury; (3) visual problems in both eyes. The Coma Recovery Scale-Revised was administered on the day of the assessment [S2, S3]. The CRS-R is a standardized and validated behavioral assessment scale to determine patients' level of consciousness. It assesses auditory, visual, verbal and motor functions as well as communication and arousal level. The total score ranges between 0 (coma) and 23 (emergence from the MCS; see CRS-R subscores for each patient in Supplemental Table S1).

The study was approved by the Ethics Committee of the Faculty of Medicine of the University of Liège (patients), and the Ethics Committee of the Department of Psychology of the Philipps-University Marburg (healthy controls); written informed consent was obtained from all patients and all healthy participants prior to the experiment. The study conformed to the Declaration of Helsinki.

Based on the aforementioned criteria, seven patients who had a brainstem stroke (five males; aged between 40 and 74 yr; between 3 and 19 years post insult, see Supplemental Table S1) and five patients with severe brain injury (five males; aged between 21 and 56 yr; between 2 and 13 years post insult, see Supplemental Table S1) were included in the study (two traumatic, two ischemic/hypoxic encephalopathy and one cardiovascular accident). They all had the clinical consensus diagnosis of LIS. When tested on the day of assessment, all LIS patients could communicate either via yes-no coded eye or head movements, use an eye-controlled spelling-board or use a computer-controlled letter speller (see Supplemental Table S1). The MCS patient (male, stroke, 50 yr, 2.1 years post onset) did not show any functional communication or functional object use.

All participants (patients and healthy controls) took part in one experimental session, with the exception of typical LIS patient #5, who took part in two sessions.

Patients with severe brain injury or LIS often take a variety of centrally acting drugs with potential anticholinergic or sympaticomimetic effects, as was also the case in the studied convenience sample (Supplemental Table S1). For obvious medical and ethical reasons these drugs could not be withdrawn for the current study. Our results show that despite this potentially confounding factor, several patients were able to show measurable pupil-related responses to the employed paradigm, illustrating its possible clinical use.

Setup

Pupil size of both eyes was measured non-invasively by a mobile head-free video-oculographic device (EyeSeeCam, [S4]) at a sampling rate of 221Hz. In brief, the device illuminates the eyes with infrared LEDs embedded in swimming goggles and records videos of both eyes with attached cameras. The device's software uses an adaptive thresholding procedure to determine pixels belonging to the pupil. Based on these pixels, the device fits the pupil and computes a measure that is proportional to the actual pupil area. By design, the measure is insensitive to partial obstruction of the eye by its lid, and tolerates deviations of gaze from the straight-ahead in the range of relevance to the present study (i.e., the size of the presentation screen). Since all analysis is insensitive to scaling, the raw values in pixel^2 are used throughout. The transformation of pixels to actual size varies between observers, but 40 pixel^2 typically corresponds to a pupil diameter of about 3.2 mm, and the fluctuations reported here would be clearly visible by the naked eye.

Visual and auditory stimulation were run on a laptop computer (MacBookPro, Apple Inc, Cupertino, CA, USA) that also recorded the EyeSeeCam data. Stimuli were generated using Matlab (Mathworks, Nattick, MA, USA) and its Psychophysics toolbox extension [S5, S6], which was controlled by and thus synchronized to the EyeSeeCam software. At the end of each trial, the software provided the pupil trace for visual inspection to one experimenter, who was unaware of the correct response, to verify the correct functioning of the setup and participant's cooperation. Visual stimuli were displayed on a 19' TFT monitor connected to the laptop; auditory stimuli were generated using the default tts (text-to-speech) voice of Mac OS X for French (patients) and German (healthy controls), respectively.

Procedure – Preparation and details

To familiarize participants with the task and setup, they were presented calculation tasks of varying levels of difficulty prior to each session. During this preparation phase, pupil size was monitored to set the difficulty of calculations for the main experiment and the duration of the calculation interval (5s in HCs, typical LIS patient #6 and 1st session of typical LIS patient #5, 10s in atypical LIS patient #3, 7s otherwise). Each session consisted of two blocks. Each block included 15 trials. The questions of the first block were repeated in the second block, with the order of response alternatives reversed as compared to the same question in the first block. In some patients (see main text), the number of trials had to be reduced, as patients showed obvious signs of fatigue, one patient (typical LIS #6) performed an additional two blocks of 5 questions each directly after the first two blocks (40 trials / 20 distinct questions in total) . In the first trials conducted with the MCS patient, it became evident that he did not follow task instructions (e.g., ignoring the screen, not showing any substantial pupil response to either interval), which prompted the use of command-following as described in the main text, instead.

Preprocessing of pupil data

Pupil data, as recorded by the EyeSeeCam device was interpolated through times of blinks using cubic spline interpolation. To suppress extreme outliers, a 50 ms median filter was applied to the resulting signal. No other pre-processing or normalization was performed, such that the information used in the present study is available at each trial's end. In each trial the data of the eye with the better fit (smaller mean-squared error) for the pupil slope (Figure 1B) was retained.

Signal-detection analysis, statistics for individuals

To compute confidence intervals for the AUC of each individual using non-parametric statistics [S7], we used the implementation of R's pROC package [S8]. Reported p-values (at alpha levels of <0.05, <0.01 and <0.001) correspond to two-sided statistics (AUC different from chance level).

Supplemental References

S1. Paulus, F.M., Blanke, M., Krach, S., Belke, M., Roth, C., Rosenow, F., Menzler, K., Sonntag, J., Sommer, J., Kircher, T., Jansen, A., Bremmer, F., Einhäuser, W., and Knake, S. (2012) Dysfunctions in frontal circuitries during empathy for pain in juvenile myoclonic epilepsy (JME). Hum. Brain. Map. 2012 [Abstract].

S2. Giacino, J., Kalmar, K., and Whyte, J. (2004). The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. Arch. Phys. Med. Rehabil. 85, 2020-2029.

S3. Schnakers, C., Giacino, J., Kalmar, K., Piret, S., Lopez, E., Boly, M., Malone, R, and Laureys, S. (2006). Does the FOUR score correctly diagnose the vegetative and minimally conscious states? Ann. Neurol. 60, 744-745.

S4. Schneider, E., Villgrattner, T., Vockeroth, J., Bartl, K., Kohlbecher, S., Bardins, S., Ulbrich, H., and Brandt, T. (2009). EyeSeeCam: an eye movement-driven head camera for the examination of natural visual exploration. Ann. N. Y. Acad. Sci. 1164, 461-467.

S5. Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. 10, 433-436.

S6. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat. Vis. 10, 437-442.

S7. DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 44, 837-845.

S8. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77.

Supplemental Figures

Figure S1

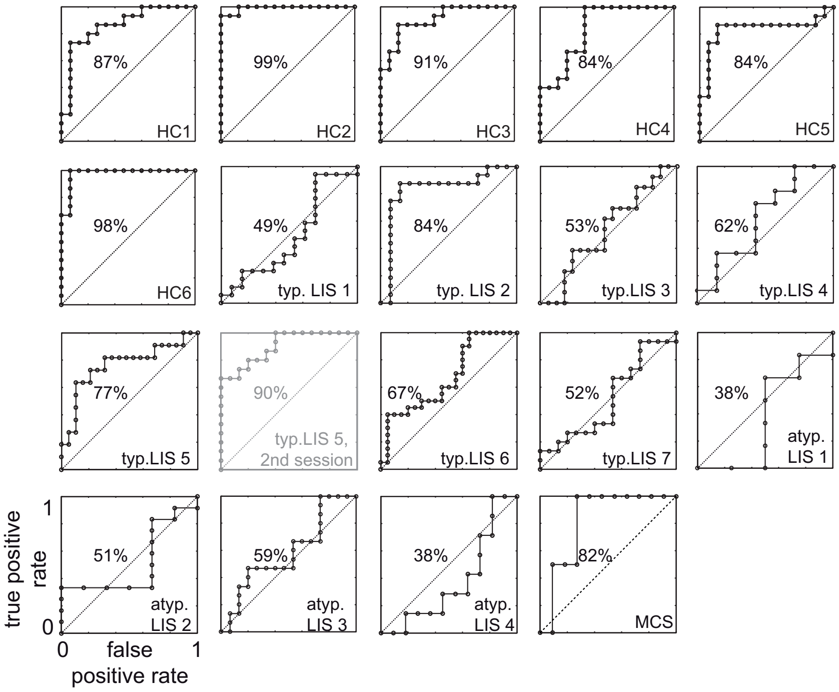


Figure S1. Raw ROC curves corresponding to data in Figure 1C. In all panels false positive rate plotted on the x-axis, true positive rate on the y-axis. Ideal decoding performance (100% AUC) would correspond to a curve in the upper left corner, chance performance (50% AUC) to a ROC curve around the diagonal. AUCs are given in each plot. Number of data points varies as some participants had a reduced number of trials (see main text). HC=healthy control, typ. LIS = typical LIS patient, atyp. LIS = atypical LIS patient, MCS = MCS patient.

Figure S2

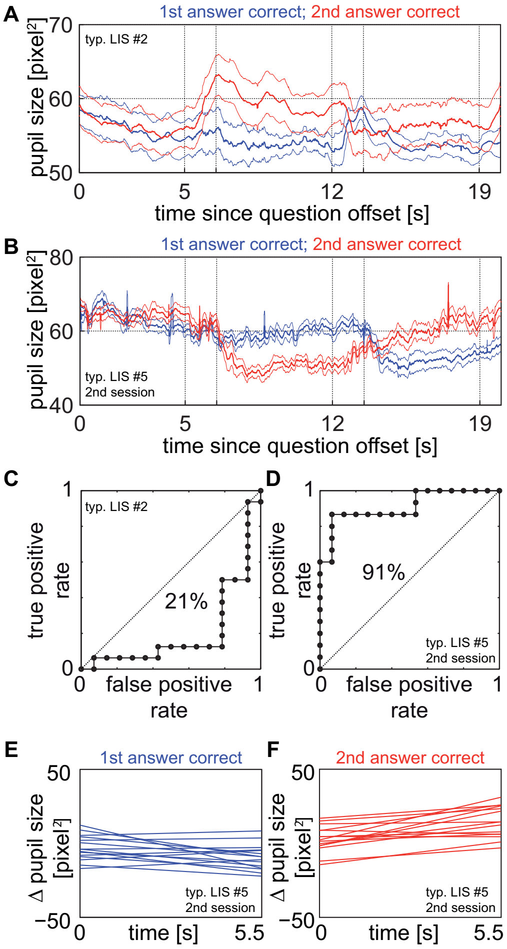


Figure S2 Mean traces for the two best decoded patients. Time course of average pupil size for trials in which the first answer was correct (*blue*) or the second answer was correct (*red*), mean and standard error over trials at each time point. **A)** typical LIS patient #2, **B)** typical LIS patient #5, 2nd session. Calculation intervals start at 5s and 12s after trial onset. For the main analysis, pupil slope was analyzed between 6.5s and 12s as well as between 13.5s and 19s. The responses are distinct for the two calculation intervals in either individual, but the patterns are reversed in the first interval, which yields a reversed effect when differences between means of each interval would be considered rather than slopes. **C, D)** ROC curves based on means in the calculation intervals. In both patients there would be decoding significantly different from chance (AUC: 21%, 91%); however, the direction (does a larger or a small mean correspond to an answer in the respective interval) is reversed. **E,F)** Pupil slopes for all trials of typical LIS patient #5's 2nd session, for which the first answer was correct (panel E) and those for which the second answer was correct (panel F). Compared to the data of typical LIS patient #2 (panel A, and Figure 1B), the baselines are reversed, but the direction of the slopes (positive for 2nd interval correct, negative for 1st interval correct) is qualitatively similar.

Supplemental Table

Patient	Gender	Age	Etiology	Time since onset (years)	Structural brain lesions	CRS-R subscores	Centrally acting drugs	Communication code
typical LIS #1	M	74	Brainstem stroke	2.7	Brainstem	A4V5M0O2C2Ar2	escitalopram 10mg, zopiclone 5mg	Small Yes-No head movement
typical LIS #2	M	40	Brainstem stroke	18.6	Brainstem	A4V5M0O1C2Ar2	baclofen 40mg	Eye blinks (virtual keyboard for letter spelling)
typical LIS #3	M	46	Brainstem stroke	12.1	Brainstem	A4V5M1O2C2Ar2	nihil	Yes-no head movements and vocalizations
typical LIS #4	M	47	Brainstem stroke	3.3	Cerebellum and brainstem	A4V5M1O2C2Ar2	lamotrigine 100mg, levocetirizine 5mg	Yes-no head movements and vocalizations via tracheostomy
typical LIS #5	F	45	Brainstem stroke	3.7	Brainstem and middle cerebellar peduncles	A4V4M2O2C2Ar2	amantadine 100mg, duloxetine 60mg ; pregabalin 75, moxonidine 0,4mg, ipratropium bromide	Small Yes-No head movement
typical LIS #6	M	49	Brainstem stroke	11.1	Brainstem	A4V5M1O2C2Ar2	baclofen 60mg, zopiclone 5mg, lorazepam 2,5mg, acedicone 5 mg, ipratropium bromide	Small Yes-No head movement
typical LIS #7	F	46	Brainstem stroke	17.3	Brainstem	A4V5M1O2C2Ar2	bromazepam 6mg, fluoxetine chlorhydrate 20mg, trimebutine 100mg	Yes-no head movements and vocalizations and finger-controlled letter speller
atypical LIS #1	M	36	Ischemic/hypoxic encephalopathy	7.7	Bilateral basal ganglia lesions, diffuse cortical atrophy	A4V4M2O1C2Ar2	baclofen 40mg, tianeptine 25mg, alfuzosin 10mg (α1 adrenergic receptor antagonist)	Yes-no head movements and finger-controlled letter speller (very limited attention span)
atypical LIS #2	M	25	Ischemic/hypoxic encephalopathy	12.8	Lenticular nuclei and thalamus, rolandic fissure and head of the left caudate nucleus	A4V5M3O2C2Ar2	dantrolene 25mg, baclofen 10mg, pregabalin 50mg	Yes-no head movements and vocalizations (very limited attention span)
atypical LIS #3	M	21	TBI	5.8	Right cerebellar, right prefrontal cortex and left lenticular. Diffuse cerebral atrophy.	A4V5M1O2C2Ar2	nihil	Eye-controlled letter speller
atypical LIS #4	M	56	TBI	4.9	Frontal, left temporal and occipital.	A4V2M1O1C2Ar2	domperidone 10mg, citalopram 40 mg, cetirizine 10 mg, pregabalin 25mg, intrathecal baclofen	Yes-no eye-code
MCS	M	50	CVA	2.1	Massive right temporo-occipital, left temporal and brainstem lesions	A3V3M2O0C0Ar1	acebutolol 200mg, baclofen 30mg, clonazepam 2mg, dantrolene 100mg, domperidone 10mg, milnacipran 100mg, levodopa benzerazide 100/25mg, tramadol hydrochloride 250mg	None

Table S1. Clinical information for locked-in syndrome (LIS) patients with brainstem stroke (typical), with severe brain injury (atypical) and the minimally conscious (MCS) patient. CRS-R = Coma Recovery Scale-Revised, A = auditory subscale, V = visual subscale, M = motor subscale, O = oromotor verbal subscale, C = communication subscale, Ar = Arousal subscale.

Literaturverzeichnis

[1] Helder Araújo, Rodrigo L Carceroni, and Christopher M Brown. A fully projective formulation to improve the accuracy of lowe's pose-estimation algorithm. *Computer Vision and Image Understanding*, 70(2):227–238, 1998.

[2] Gary Aston-Jones and Jonathan D Cohen. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, 28:403–450, 2005.

[3] Edward Awh, Artem V. Belopolsky, and Jan Theeuwes. Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in cognitive sciences*, 16(8):437–443, August 2012.

[4] G. Bauer, F. Gerstenbrand, and E. Rumpl. Varieties of the locked-in syndrome. *Journal of Neurology*, 221(2):77–91, 1979.

[5] Jackson Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2):276, 1982.

[6] Jackson Beatty and Daniel Kahneman. Pupillary changes in two memory tasks. *Psychonomic Science*, 5(10):371–372, 1966.

[7] Paola Binda, Maria Pereverzeva, and Scott O. Murray. Attention to bright surfaces enhances the pupillary light reflex. *The Journal of Neuroscience*, 33(5):2199–2204, 2013.

[8] Jean-Yves Bouguet. Complete camera calibration toolbox for matlab, 2004.

[9] P.R. Boyce. *Human Factors in Lighting, Third Edition*. Taylor & Francis, 2014.

[10] David H Brainard. The psychophysics toolbox. *Spatial vision*, 10:433–436, 1997.

[11] Julie Brisson, Marc Mainville, Dominique Mailloux, Christelle Beaulieu, Josette Serres, and Sylvain Sirois. Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior research methods*, 45(4):1322–1331, 2013.

[12] Claus Bundesen. A theory of visual attention. *Psychological Review*, 97(4):523–547, October 1990.

[13] Roger HS Carpenter. *Movements of the eyes (2nd rev*. Pion Limited, 1988.

[14] Marisa Carrasco. Visual attention: The past 25 years. *Vision Research*, 51(13):1484 – 1525, 2011. Vision Research 50th Anniversary Issue: Part 2.

[15] Berufsgenossenschaft der Feinmechanik. Elektrotechnik: Expositionsgrenzwerte für künstliche optische strahlung. *Berufsgenossenschaftliche Informationen für Sicherheit und Gesundheit bei der Arbeit*, 2004.

[16] Thomas Dera, Guido Boning, Stanislavs Bardins, and Erich Schneider. Low-latency video tracking of horizontal, vertical, and torsional eye movements as a basis for 3dof realtime motion control of a head-mounted camera. In *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*, volume 6, pages 5191–5196. IEEE, 2006.

[17] Heiner Deubel and Werner X Schneider. Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision research*, 36(12):1827–1837, 1996.

[18] Wolfgang Einhäuser, Christof Koch, and Olivia Carter. Pupil dilation betrays the timing of decisions. *Frontiers in human neuroscience*, 4:18, 2010.

[19] Wolfgang Einhäuser, Ueli Rutishauser, and Christof Koch. Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, 8(2):2, 2008.

[20] Wolfgang Einhäuser, Frank Schumann, Stanislavs Bardins, Klaus Bartl, Guido Böning, Erich Schneider, and Peter König. Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, 18(3):267–297, 2007.

[21] Wolfgang Einhäuser, Merrielle Spain, and Pietro Perona. Objects predict fixations better than early saliency. *Journal of Vision*, 8(14):18, 2008.

[22] Tom Foulsham and Alan Kingstone. Goal-driven and bottom-up gaze in an active real-world search task. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 189–192, New York, NY, USA, 2012. ACM.

[23] D Guitton and M Volle. Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of neurophysiology*, 58(3):427–459, 1987.

[24] Neil A Harrison, C Ellie Wilson, and Hugo D Critchley. Processing of observed pupil size modulates perception of sadness and predicts empathy. *Emotion*, 7(4):724, 2007.

[25] John M. Henderson. Human gaze control during real-world scene perception. *Trends in cognitive sciences*, 7(11):498–504, November 2003.

[26] Adam Herout, Jiri Havel, Lukas Polok, Michal Hradis, Pavel Zemcik, Radovan Josth, and Roman Juranek. *Low-Level image features for Real-Time object detection*. INTECH Open Access Publisher, 2010.

[27] Eckhard H Hess and James M Polt. Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611):1190–1192, 1964.

[28] Alex D Hwang, Emily C Higgins, and Marc Pomplun. A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, 9(5):25, 2009.

[29] Alex D Hwang, Hsueh-Cheng Wang, and Marc Pomplun. Semantic guidance of eye movements in real-world scenes. *Vision research*, 51(10):1192–1205, 2011.

[30] Alla Ignashchenkova, Peter W Dicke, Thomas Haarmeier, and Peter Thier. Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nature neuroscience*, 7(1):56–64, 2004.

[31] Laurent Itti, Christof Koch, and Ernst Niebuhr. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, November 1998.

[32] Eric R. Kandel, J. H. Schwartz, and Thomas M. Jessell. *Principles of Neural Science*. McGraw-Hill Medical, 4th edition, July 2000.

[33] Jack B Kuipers. *Quaternions and rotation sequences*, volume 66. Princeton university press Princeton, 1999.

[34] Michael Land. *Looking and acting: vision and eye movements in natural behaviour*. Oxford University Press, 2009.

[35] Michael F Land. Predictable eye-head coordination during driving. *Nature*, 1992.

[36] Michael F Land. Eye movements and the control of actions in everyday life. *Progress in retinal and eye research*, 25(3):296–324, 2006.

[37] Michael F Land. Vision, eye movements, and natural behavior. *Visual neuroscience*, 26(01):51–62, 2009.

[38] Bernard Marius ’t Hart and Wolfgang Einhäuser. Mind the step: complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental brain research*, 223(2):233–249, 2012.

[39] Bernard Marius 't Hart, Johannes Vockeroth, Frank Schumann, Klaus Bartl, Erich Schneider, Peter Koenig, and Wolfgang Einhäuser. Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, 17(6-7):1132–1158, 2009.

[40] David Marr and Herbert Keith Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences*, 200(1140):269–294, 1978.

[41] James S Maxwell and Clifton M Schor. The coordination of binocular eye movements: vertical and torsional alignment. *Vision research*, 46(21):3537–3548, 2006.

[42] Vidhya Navalpakkam and Laurent Itti. Modeling the influence of task on attention. *Vision research*, 45(2):205–231, 2005.

[43] Vidhya Navalpakkam and Laurent Itti. Search goal tunes visual features optimally. *Neuron*, 53(4):605–617, 2007.

[44] Antje Nuthmann and Wolfgang Einhäuser. A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences*, In press, 2015.

[45] Antje Nuthmann and John M. Henderson. Object-based attentional selection in scene viewing. *Journal of vision*, 10(8), 2010.

[46] Denis Oberkampf, Daniel F DeMenthon, and Larry S Davis. Iterative pose estimation using coplanar points. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 626–627. IEEE, 1993.

[47] Gustav Osterberg. *Topography of the layer of rods and cones in the human retina*. Nyt Nordisk Forlag, 1935.

[48] Michael I Posner and Jin Fan. Attention as an organ system. *Topics in integrative neuroscience*, pages 31–61, 2008.

[49] Giacomo Rizzolatti, Lucia Riggio, Isabella Dascola, and Carlo Umiltá. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25(1):31–40, 1987.

[50] Ali Samii, John G Nutt, and Bruce R Ransom. Parkinson's disease. *The Lancet*, 363(9423):1783–1793, 2015/01/30 2004.

[51] ER Samuels and E Szabadi. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part ii: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Current neuropharmacology*, 6(3):254, 2008.

[52] Susan J Sara. The locus coeruleus and noradrenergic modulation of cognition. *Nature reviews neuroscience*, 10(3):211–223, 2009.

[53] Erich Schneider, Thomas Villgrattner, Johannes Vockeroth, Klaus Bartl, Stefan Kohlbecher, Stanislavs Bardins, Heinz Ulbrich, and Thomas Brandt. Eyeseecam: An eye movement–driven head camera for the examination of natural visual exploration. *Annals of the New York Academy of Sciences*, 1164(1):461–467, 2009.

[54] Xiaohui Shen and Ying Wu. A unified approach to salient object detection via low rank matrix recovery. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 853–860. IEEE, 2012.

[55] Maria Stamelou, Rohan de Silva, Oscar Arias-Carrión, Evangelia Boura, Matthias Höllerhage, Wolfgang H Oertel, Ulrich Müller, and Günter U Höglinger. Rational therapeutic approaches to progressive supranuclear palsy. *Brain*, 133(6):1578–1590, 2010.

[56] Robert M Steinman, Zygmunt Pizlo, Tatiana I Forofonova, and Julie Epelboim. One fixates accurately in order to see clearly not because one sees clearly. *Spatial vision*, 16(3):225–241, 2003.

[57] Josef Stoll, Stefan Kohlbecher, Svenja Marx, Erich Schneider, and Wolfgang Einhäuser. Mobile three dimensional gaze tracking. *Studies in health technology and informatics*, 163:616–622, 2010.

[58] Bernard Marius ’t Hart, Hannah Schmidt, Ingo Klein-Harmeyer, Christine Roth, and Wolfgang Einhäuser. The role of low-level features for rapid object detection and guidance of gaze in natural scenes. *Journal of Vision*, 12(9):807–807, 2012.

[59] Benjamin W. Tatler. Current understanding of eye guidance. *Visual Cognition*, 17(6-7):777–789, 2009.

[60] Benjamin W Tatler, Mary M Hayhoe, Michael F Land, and Dana H Ballard. Eye guidance in natural vision: Reinterpreting salience. *Journal of vision*, 11(5):5, 2011.

[61] Benjamin W Tatler and Benjamin T Vincent. Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2):1–18, 2008.

[62] Antonio Torralba, Aude Oliva, Monica S Castelhano, and John M Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review*, 113(4):766, 2006.

[63] Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, January 1980.

[64] Jochen Triesch, Dana H. Ballard, Mary M. Hayhoe, and Brian T. Sullivan. What you see is what you need. *Journal of Vision*, 3(1):86–94, 2003.

[65] Christopher W. Tyler, Anas M. Elsaid, Lora T. Likova, Navdeep Gill, and Spero C. Nicholas. Analysis of human vergence dynamics. *Journal of Vision*, 12(11), 2012.

[66] Benjamin T. Vincent, Roland Baddeley, Alessia Correani, Tom Troscianko, and Ute Leonards. Do we look at lights? using mixture modelling to distinguish between low- and high-level factors in natural image viewing. *Visual Cognition*, 17(6-7):856–879, 2009.

[67] Paul Viola and Michael Jones. Robust Real-time Object Detection. *International Journal of Computer Vision - to appear*, # 2002.

[68] T Wertheim. Über die indirekte sehschärfe. *Zeitschrift für Psychologie & Physiologie der Sinnesorgane*, 7:172–187, 1894.

[69] G. Westheimer. Relative localization in the human fovea: radial–tangential anisotropy. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1471):995–999, 2001.

[70] Brian J White and Douglas P Munoz. Separate visual signals for saccade initiation during target selection in the primate superior colliculus. *The Journal of Neuroscience*, 31(5):1570–1578, 2011.

[71] B Winn, D Whitaker, D B Elliott, and N J Phillips. Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, 35(3):1132–7, 1994.

[72] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.

Kapitel 8

Summary - Zusammenfassung

8.1 Summary

This dissertation contains studies on visual attention, as measured by gaze orientation, and the use of mobile eye-tracking and pupillometry in applications. It combines the development of methods for mobile eye-tracking (studies II and III) with experimental studies on gaze guidance and pupillary responses in patients (studies IV and VI) and healthy observers (studies I and V).

Object based attention

Study I

What is the main factor of fixation guidance in natural scenes? Low-level features or objects? We developed a fixation-predicting model, which regards preferred viewing locations (PVL) per object and combines these distributions over the entirety of existing objects in the scene. Object-based fixation predictions for natural scene viewing perform at par with the best early salience model, that are based on low-level features. However, when stimuli are manipulated so that low-level features and objects are dissociated, the greater prediction power of saliency models diminishes. Thus, we dare to claim, that highly developed saliency models implicitly obtain object-hood and that fixation selection is mainly influenced by objects and much less by low-level features. Consequently, attention guidance in natural scenes is object-based.

3D tracking

Study II

The second study focussed on improving calibration procedures for eye-in-head positions with a mobile eye-tracker. We used a mobile eye-tracker prototype, the EyeSeeCam with a high video-oculography (VOG) sampling rate and the technical gadget to follow the users gaze direction instantaneously with a rotatable camera.
For a better accuracy in eye-positioning, we explored a refinement in the implementation of the eye-in-head calibration that yields a measure for fixation distance, which led to a mobile eye-tracker 3D calibration. Additionally, by developing the analytical mechanics for parametrically reorienting the gaze-centred camera, the 3D calibration could be applied to reliably record gaze-centred videos. Such videos are suitable as stimuli for investigating gaze-behaviour during object manipulation or object recognition in real worlds

point-of-view (PoV) perspective. In fact, the 3D calibration produces a higher accuracy in positioning the gaze-centred camera over the whole 3D visual range.

Study III, eye-tracking methods

With a further development on the EyeSeeCam we achieved to record gaze-in-world data, by superposing eye-in-head and head-in-world coordinates. This novel approach uses a combination of few absolute head-positions extracted manually from the PoV video and of relative head-shifts integrated over angular velocities and translational accelerations, both given by an inertia measurement unit (IMU) synchronized to the VOG data. Gaze-in-world data consist of room-referenced gaze directions and their origins within the environment. They easily allow to assign fixation targets by using a 3D model of the measuring environment – a strong rationalisation regarding fixation analysis.

Applications

Study III

Daylight is an important perceptual factor for visual comfort, but can also create discomfort glare situations during office work, so we developed to measure its behavioural influences. We achieve to compare luminance distributions and fixations in a real-world setting, by also recording indoor luminance variations time-resolved using luminance maps of a scenery spanning over a $3\pi sr$. Luminance evaluations in the workplace environment yield a well controlled categorisation of different lighting conditions and a localisation as well as a brightness measure of glare sources. We used common tasks like reading, typing on a computer, a phone call and thinking about a subject. The 3D model gives the possibility to test for gaze distribution shifts in the presence of glare patches and for variations between lighting conditions.

Here, a low contrast lighting condition with no sun inside and a high contrast lighting condition with direct sunlight inside were compared. When the participants are not engaged in any visually focused task and the presence of the task support is minimal, the dominant view directions are inclined towards the view outside the window under the low contrast lighting conditions, but this tendency is less apparent and sways more towards the inside of the room under the high contrast lighting condition. This result implicates an avoidance of glare sources in gaze behaviour. In a second more extensive series of experiments, the participants' subjective assessments of the lighting conditions will be included. Thus, the influence of glare can be analysed in more detail and tested whether visual discomfort judgements are correlated in differences in gaze-behaviour.

Study IV

The advanced eye-tracker calibration found application in several following projects and included in this dissertation is an investigation with patients suffering either from idiopathic Parkinson's disease or from progressive supranuclear palsy (PSP) syndrome. PSP's key symptom is the decreased ability to carry out vertical saccades and thus the main diagnostic feature for differentiating between the two forms of Parkinson's syndrome. By measuring ocular movements during a rapid (< 20s) procedure with a standardized fixation protocol, we could successfully differentiate pre-diagnosed patients between idiopathic Parkinson's disease and PSP, thus between PSP patients and HCs too. In PSP patients, the EyeSeeCam detected prominent impairment of both saccade velocity and amplitude. Furthermore, we show the benefits of a mobile eye-tracking device for application in clinical practice.

Study V

Decision-making is one of the basic cognitive processes of human behaviours and thus, also evokes a pupil dilation. Since this dilation reflects a marker for the temporal occurrence of the decision, we wondered whether individuals can read decisions from another's pupil and thus become a mentalist. For this purpose, a modified version of the *rock-paper-scissors* childhood game was played with 3 prototypical opponents, while their eyes were video taped. These videos served as stimuli for further persons, who competed in *rock-paper-scissors*. Our results show, that reading decisions from a competitor's pupil can be achieved and players can raise their winning probability significantly above chance. This ability does not require training but the instruction, that the time of maximum pupil dilation was indicative of the opponent's choice. Therefore we conclude, that people could use the pupil to detect cognitive decisions in another individual, if they get explicit knowledge of the pupil's utility.

Study VI

For patients with severe motor disabilities, a robust mean of communication is a crucial factor for well-being. Locked-in-Syndrome (LiS) patients suffer from quadriplegia and lack the ability of articulating their voice, though their consciousness is fully intact. While classic and incomplete LiS allows at least voluntary vertical eye movements or blinks to be used for communication, total LiS patients are not able to perform such movements. What remains, are involuntarily evoked muscle reactions, like it is the case with the pupillary response. The pupil dilation reflects enhanced cognitive or emotional processing, which we successfully observed in LiS patients. Furthermore, we created a communication system based on yes-no questions combined with the task of solving arithmetic problems during matching answer intervals, that yet invokes the most solid pupil dilation usable on a trial-by-trial basis for decoding yes or no as answers. Applied to HCs and patients with various severe motor disabilities, we provide the proof of principle that pupil responses allow communication for all tested HCs and 4/7 typical LiS patients.

Résumé

Together, the methods established within this thesis are promising advances in measuring visual attention allocation with 3D eye-tracking in real world and in the use of pupillometry as on-line measurement of cognitive processes. The two most outstanding findings are the possibility to communicate with complete LiS patients and further a conclusive evidence that objects are the primary unit of fixation selection in natural scenes.

8.2 Zusammenfassung

Diese Dissertation enthält Studien über visuelle Aufmerksamkeit, welche mittels Blickrichtungen gemessen wird, wobei mobile Augen- und Pupillenmessungen angewendet werden. Sie kombiniert Methodenentwicklungen für mobile Augenbewegungsmessungen (Studien II und III) mit experimentellen Studien zu Blicksteuerung und Pupillenreaktionen in Patienten (Studien IV und V) sowie gesunden Probanden (Studien I und VI).

Objekt-basierte Aufmerksamkeit

Studie I

Welches ist der ausschlaggebende Faktor für die Blickpositionssteuerung in natürlichen Szenen? Einfache Merkmale oder Objekte? Wir erarbeiteten dafür ein Modell zur Fixationsvorhersage, welches bevorzugte Blickpunkte pro Objekt berücksichtigt und mit der Gesamtheit aller Objekte in der Szene kombiniert. Dessen Performanz steht dem besten Salienzmodell, welches ausschließlich niedere Merkmale verwertet, bei natürlichen Szenen in nichts nach. Wenn aber durch Bildmanipulation niedere Merkmale von Objekten dissoziiert werden, verliert das Salienzmodell deutlich mehr an Vorhersagekraft. Somit haben wir gezeigt, dass bessere Salienzmodelle implizit die Objektrelevanzen einer Szene miteinbeziehen und die Fixationsauswahl stärker von Objekten abhängt, als von niederen Merkmalen. Folglich arbeitet auch die Aufmerksamkeitslenkung in natürlichen Szenen objekt-basiert.

3D Blickvermessung

Studie II

An einem mobilen Augenbewegungsmessgerät sollte hier das Kalibrationsverfahren für Aug-in-Kopf-Positionen optimiert werden. Wir verwendeten hierzu den Prototypen eines mobilen Augenbewegungsmessgeräts, der EyeSeeCam, die sich durch eine hohe Samplingrate bei der Video-Okulographie (VOG) und den technischen Zusatz einer schwenkbaren Kamera, welche der Blickrichtung des Nutzers praktisch instantan folgt, auszeichnet. Für eine höhere Genauigkeit der Augenpositionsbestimmung arbeiteten wir eine Verfeinerung der Implementierung der Auge-in-Kopf-Kalibration aus, welche nun die Messung der Fixationsdistanz liefert und somit zu einer 3D-Kalibration des mobilen Augenbewegungsmessgeräts führt. Zusätzlich entwickelten wir eine analytische Mechanik zur parametrischen Ansteuerung des Schwenkapparates für die blickzentrierte Kamera wofür auch 3D-Kalibration zwingend war. Tatsächlich resultiert aus der 3D-Kalibration eine höhere Genauigkeit bei der Positionierung der blickgesteuerten Kamera über das komplette Sehfeld in 3D. Die blickzentrierten Videos sind besonders geeignet als Stimuli zur Untersuchung von Blickverhalten bei Objektmanipulation oder zur Objekterkennungsmodellierung in echter Point-of-View-Perspektive (PoV).

Studie III, Augenmessmethodik

Mit einer weiteren Entwicklung an der EyeSeeCam erreichten wir durch die Superposition von Aug-in-Kopf- und Kopf-im-Raum-Koordinaten die Aufnahme von Blick-im-Raum-Daten. Dieser neuartige Ansatz verwendet eine Kombination aus wenigen absoluten Kopfpositionen, welche manuell aus PoV-Videos extrahiert wurden, und aus relativen Kopf-Verschiebungen, integriert über Winkelgeschwindigkeit und Translationsbeschleunigung, welche beide von einer zur VOG synchronisierten Inertialmesseinheit gegeben sind. Blick-im-Raum-Daten bestehen aus raumbezogenen Blickrichtungen und deren jeweiligem Ursprung in der Messumgebung. Sie erlauben durch die Verwendung eines 3D-Modells der Messumgebung eine einfache Auswertung von Fixationszielen – eine starke Rationalisierung hinsichtlich von Fixationsanalysen.

Anwendungen

Studie III

Tageslicht hat einen wichtigen Einfluss auf die Wahrnehmung von visuellem Komfort, aber es kann auch unbehagliche Blendungen während der Büroarbeit hervorrufen. Deren Verhaltenseinflüsse zu messen, streben wir hier an. Wir erreichen dies, indem wir Fixationen mit Helligkeitsverteilungen in einer realen Messumgebung vergleichen, für welche wir Variationen der Helligkeit über einen 3π-Raumwinkel zeit-aufgelöst in Luminanz-Karten festhalten. Die Helligkeitsauswertung der Schreibtischumgebung ergibt eine wohl kontrollierte Kategorisierung von unterschiedlichen Lichtbedingungen und die Lokalisierung sowie das Helligkeitsmaß von Blendung-Quellen. Wir ließen die Probanden gewöhnliche Aufgaben wie Lesen, Tippen am Computer, Telefonieren und Nachdenken über ein Thema ausüben. Mithilfe des 3D-Modells haben wir die Möglichkeit, das Blickverhalten auf Verschiebungen seiner Verteilungen in Gegenwart von Blendung-Flecken und durch Variation der Lichtbedingungen zu testen. Hier vergleichen wir ein nieder-kontrastierte

Lichtbedingung ohne direkte Sonneneinstrahlung gegenüber einer hoch-kontrastierten Lichtbedingung mit direkter Sonneneinstrahlung in den Raum. Wenn die Teilnehmer nicht auf irgendeine visuelle Aufgabe fokussiert sind und die Präsenz des Aufgabenmediums minimal ist, sind die Hauptblickrichtungen unter der nieder-kontrastierten Bedingung tendenziell zur Aussicht aus dem Fenster geneigt, aber bei der hoch-kontrastierten Bedingung schwindet diese Tendenz und die Hauptblickrichtung schwenkt eher in den Innenraum. Dieses Ergebnis impliziert ein Ausweichen der Blendung-Quellen beim Blick-Verhalten. In einer zweiten umfangreicheren Messreihe werden auch subjektive Bewertungen der Lichtbedingungen von den Teilnehmern mitberücksichtigt. Somit kann der Einfluss durch Blendungen detaillierter analysiert und darauf getestet werden, inwieweit Beurteilungen über visuelles Unbehagen mit Unterschieden im Blickverhalten korrelieren.

Studie IV

Die weiterentwickelte Kalibration des Augenbewegungsmessgeräts fand Anwendung in mehreren Folgeprojekten und hier ist eine Untersuchung mit Patienten miteinbezogen, welche entweder an Morbus Parkinson oder an progressiver supranukleärer Blickparese (PSP) leiden. PSP's Schlüsselsymptom ist die stetig schwindende Fähigkeit, senkrechte Sakkaden auszuführen und daher das Hauptdiagnosemerkmal zur Differenzierung zwischen den beiden Formen von Parkinson-Syndromen. Durch eine zügige Prozedur von Augenbewegungsmessung mit einem Standardprotokoll konnten wir vor-diagnostizierte Patienten erfolgreich zwischen Morbus Parkinson und PSP, sowie auch zwischen PSP und Kontrollprobanden, differenzieren. Bei PSP Patienten detektierte die EyeSeeCam eine prominente Verminderung von beiden Sakkaden-Geschwindigkeit und – Amplitude. Außerdem zeigen wir damit, dass die Verwendung mobiler Augenbewegungsmessgeräte für die klinische Praxis sehr nützlich ist.

Studie V

Entscheidungsbildung ist ein äußerst grundlegender Prozess menschlichen Verhaltens und erzeugt eine Pupillenerweiterung. Da diese Erweiterung ein Kennzeichen des zeitlichen Auftretens der Entscheidung wiedergibt, untersuchten wir, ob Personen aus der Pupille eines Gegenübers Entscheidungen ausmachen und damit zum "Gedankenleser"werden können. Zu diesem Zweck spielten 3 Gegner eine modifizierte Version des Kinderspiels *Schere-Stein-Papier*, während ihr Auge gefilmt wurde. Diese Videos dienten als Stimuli für weitere Personen, welche gegen die ersten in *Schere-Stein-Papier* wetteiferten. Unsere Ergebnisse zeigen, dass das Entscheidungs-Lesen aus eines Gegner's Pupille machbar ist und dass man damit die eigene Gewinnwahrscheinlichkeit deutlich steigern kann. Diese Fähigkeit bedarf keines Trainings aber der Instruktion, dass der Zeitpunkt der größten Pupillenerweiterung als Hinweis für des Gegners Wahl dient. Unsere Schlussfolgerung lautet, dass Personen die Pupille eines Anderen zur Erkennung seiner Entscheidungsbildung nutzen könnten, wenn sie die ausdrückliche Kenntnis über die Nützlichkeit der Pupille haben.

Studie VI

Für Patienten mit schweren Bewegungseinschränkungen ist ein stabiles Kommunikationsmittel ein entscheidender Faktor für das Wohlbefinden. Locked-in-Syndrom- (LiS) Patienten sind Tetraplegiker und unfähig zu sprechen, obschon ihr Bewusstsein vollkommen intakt ist. Während klassische und unvollständige LiS-Patienten zumindest durch gewollte senkrechte Augenbewegungen oder Wimpernschläge kommunizieren, können totale LiS-Patienten nicht einmal solche Bewegungen ausführen. Übrigbleibend sind nur willenlos ausgelöste Muskelreaktionen, wie dies auch bei Pupillenantworten der Fall ist. Die Pupillenerweiterung spiegelt erhöhte kognitive oder emotionale Tätigkeit wider und wir konnten diese bei LiS-Patienten erfolgreich beobachten. Des Weiteren entwarfen wir ein Kommunikationssystem, welches Ja-Nein-Fragen mit dem Lösen von Rechenaufgaben während der passenden Antwortoption kombiniert. Damit konnte die bislang solideste Pupillenerweiterung zur robusten Dekodierung von einzelnen Durchgängen realisiert werden. Angewandt auf Kontrollpersonen und auf Patienten mit verschiedenen akuten Bewegungseinschränkungen erweisen die Resultate den Machbarkeitsbeweis, dass Pupillenantworten als Kommunikationsmittel bei allen Kontrollpersonen und bei 4 von 7 typischen LiS Patienten dienen können.

Fazit

Die in dieser Dissertation eingeführten Methoden stellen zukunftsträchtige Fortschritte zur Messung von visueller Aufmerksamkeitszuordnung mit 3D-Augenmessungen in der echten Welt bzw. bei der Verwendung von Pupillometrie zur Online-Messung von kognitiven Prozessen dar. Die zwei herausragendsten Erkenntnisse sind die Möglichkeit zur Kommunikation mit komplette LiS-Patienten und der schlüssige Beweis, dass Objekte die ursprüngliche Einheit zur Fixationswahl in natürlichen Szenen sind.

8.3 Scientific Career

Studies

Abitur	**1998, Klettgau-Gymnasium Tiengen**	
Basic Studies	**2002-2003, Allgemeine Physik, prediploma**	
University	**Technische Universität München**	Munich, DE
Trinational	**2002-2003, Biotechnology, 1st year**	
Elite-Course	**Degree: Equivalent to prediploma in biology**	
University	**Ecole Superieur de Biotechnologie a Strasbourg**	FR
Internship	**6 weeks, Solare Wasserdesinfektion - SODIS**	
Institute	**EAWAG, ETH Zurich**	CH
Subject	Mikrobiology: *Spektrale Abhängigkeit der Sonneneinstrahlung auf die Keimreduktion in wässrigen Lösungen.*	
Supervisor	Prof. Thomas Egli	
Main Studies	**2003-2008, Allgemeine Physik, Diplom**	
University	**Albert-Ludwig-Universität Freiburg**	DE
Diploma Thesis	*Coherent activity in an between motoric M1- and premotoric F5-cortical signals during reach to grasp.*	
Supervisor	Prof. A. Aertsen, Dr. Carsten Mehring, Prof. R. Lemon	
Research	**2008, BCCN**	Freiburg i.Br., DE
Assistant	**Group: Brain Machining Interfacing initiative (BMI)**	
Subject	Data analysis of neurophysiological signals (SUA, LFP)	

Scientific Career

Dissertation

PHD STUDIES	**2008-2015, AG Neurophysik, FB Physik**
TITLE	**Measuring gaze and pupil in the real world: object-based attention, 3D tracking and applications**
SUBJECTS	Visual attention, models, eye-tracking, pupillary response

Conference talk

2011 J. Stoll, S. Kohlbecher, et al.: *Mobile three dimensional gaze tracking.*
Medicine Meets Virtual Reality, Long Beach, CA, USA

Conference abstracts

2013 J. Stoll, M. Sarey Khanie, et al.: *Real-world tasks with full control over the visual scene: combining mobile gaze tracking and 4pi light-field measurements.*
Talk, 17th European Conference on Eye Movements, Lund, SE.

2013 J. Stoll, M. Sarey Khanie, et al.: *Combining wearable eye-tracking with* 4π *light-field measurements: towards controlling all bottom-up and top-down factors driving overt attention during real-world tasks.* Poster,
10th Göttingen Meeting of the German Neuroscience Society, Göttingen, DE.

2010 J.Stoll, W. Einhäuser: *Object recognition in natural stimuli combining saliency and temporal coherence.*
Poster, autumn school: *Space, Time, and Numbers*, Kloster Seeon, BY, DE.

Reviewing

2014 Articles for European Conference on Eye Movements, ECEM 2013, Lund, SE.

2009 Articles for International Conference on Computer Vision, ICCV 2009, Kyoto, JP.

Teaching

LECTURE SUBSTITUTION:	Introduction to experimental physics (6 x 2hours)
STUDENT-SUPERVISING:	4 diploma-, 1 master-, and 1 bachelor-theses (5 theory loaded projects); *computational physics traineeship* (2 x 1 week)
TUTORING:	Experimental physics: *E-Lehre u. Thermodynamik* (1 term), physics beginners internship (6 terms)