Guided Viewing: An Eye Tracking Approach to Increase Memory and Reduce Anxiety

A cumulative dissertation

Submitted to the Faculty of Psychology, University of Basel,
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

M.Sc. Bernhard Fehlmann

from Schöftland, Switzerland

Basel, 2020

First supervisor: Prof. Dr. med. Dominique J.-F. de Quervain

Second supervisor: Prof. Dr. med. Andreas Papassotiropoulos

Approved by the Faculty of Psychology at the request of

Prof. Dr. med. Dominique J.-F. de Quervain

Prof. Dr. med. Andreas Papassotiropoulos

Basel,

Prof. Dr. phil. Jens Gaab

Abstract

Visual exploration, the way we strategically guide our gaze through the environment, is greatly affected by cognitive states like the current focus of attention, goals and knowledge. The causal link between cognition and viewing has first been described and demonstrated in humans in the 1960s, and researchers have tried to further characterize it ever since. The advent of functional magnetic resonance imaging (fMRI) has – in combination with eye tracking – considerably benefitted the understanding of this link, as it allowed to study its neuronal underpinnings. A key role has thereby been attributed to regions centered around the hippocampus as part of the medial temporal lobe (MTL), orchestrating visual exploration based on previous experience.

However, while it is well-established that visual exploration depends on cognitive states, it is unclear if cognitive states equally depend on visual exploration. To close this gap, the aim of this thesis was to investigate if viewing can be guided to affect aspects of cognition, and more specifically, if this can be used to (1) increase memory and (2) decrease anxiety.

In a first study, Fehlmann, Coynel et al. (2020), we analyzed data of a picture encoding task performed by 967 healthy subjects during fMRI and simultaneous eye tracking. We replicated and generalized the finding of a triadic correlation between individual visual exploration patterns (i.e. eye fixation frequency and location), brain activation in the MTL and subsequent memory performance. In a second experiment, we experimentally altered visual exploration patterns in an independent population of 64 subjects. We thereby showed that both the fixation frequency and location can be causally manipulated by guided viewing conditions to affect memory performance.

In a second study, Fehlmann, Müller, et al. (2020), we investigated the intervention potential of guided viewing to reduce fear in 89 participants suffering from public speaking anxiety (PSA). We thereby targeted gaze avoidance, a potential key factor in the etiology and maintenance of the condition. The repeated use of a stand-alone, smartphone- and virtual reality (VR)-based mutual gaze training was effective in reducing gaze avoidance as well as the fear of public speaking in real-life speech situations.

In conclusion, the thesis showcases two studies that used guided viewing as a tool to affect cognitive states. The gained insights add to the knowledge about the interplay between viewing and cognition in general, and the causal effect of viewing on cognition in particular. The described phenomenon has great relevance for neuroscientific research and great potential for the clinical practice.

Acknowledgements

I wish to express my deep gratitude to Professor Dominique de Quervain. As my supervisor, he has fundamentally shaped my academic and personal development. He has continually inspired me to seek for clarity and simplicity even in the most complex matters, which appears to be the essence of science. To work under his mentorship has been a genuine honor and pleasure. My sincere appreciation further goes to my co-supervisor, Professor Andreas Papassotiropoulos. His inspiring work and thoughts helped me to integrate my everyday efforts into a bigger picture. He has convinced me that not only in biology but also in psychology 'nothing […] makes sense except in the light of evolution' (Dobzhansky, 1973). I am also indebted to Professor Hennric Jokeit, who triggered my fascination for neuroscience during my master's studies and encouraged me to embark on this journey. I thank him for his expert advice on my research projects and for his support beyond the PhD.

Thanks to all my fantastic current and former colleagues and friends at the Transfaculty Research Platform Molecular and Cognitive Neurosciences (MCN) of the University of Basel for amazing years at the Birmannsgasse 8, but also in the nearby cafés, the Rhine and the Swiss Alps. In particular, I would like to thank David Coynel, who taught me almost everything I know about eye tracking and fMRI and, with his insight and knowledge, steered me through this project. Special thanks to Eva Loos and Ehssan Amini – I could not have wished for more clever, loyal and fun office companions; to Nathalie Schicktanz for helping me solve countless scientific and semi-scientific riddles and to Annette Milnik for all the fruitful discussions about methodology and statistical rigor.

I am thankful to all the students I had the pleasure to work with. It was their intelligent and persistent questions that kept my own critical thinking alive, and their enthusiasm which allowed my experiments to go the extra mile.

Thanks to the Swiss National Science Foundation, which financially supported my PhD with a research grant (Doc.CH: P0BSP1_168917).

Thanks to my family for keeping me both grounded and going. And finally, thank you, Marielle, for unconditionally standing by my side as a brilliant neuroscientist and partner.

Table of Content

Figure Index

Abbreviations

AOI(s)	Area(s) of interest
AR	Augmented reality
BOLD	Blood-oxygen-level-dependent
dlPFC	Dorsolateral prefrontal cortex
FEF	Frontal eye field
fMRI	Functional magnetic resonance imaging
IEBI	Inter-eye blink interval
MTL	Medial temporal lobe
PEF	Parietal eye field
PSA	Public speaking anxiety
PTSD	Posttraumatic stress disorder
SAD	Social anxiety disorder
SEF	Supplementary eye field
VOI(s)	Volume(s) of interest
VR	Virtual reality

‘The eyes are the window to the soul’
(Cicero, *Tusculan Disputations*, ca. 45 B.C.E)

1 Introduction

Vision is the primary perceptual system of humans (Meister & Buffalo, 2016). What we see defines to a major extent what can be processed by our brains.

As such, to objectively measure gaze characteristics by eye tracking has become one of the main research methods to study cognitive functions like spatial attention (Oakes, 2012), but also learning and memory (Hannula, 2010). The idea that eye movements are not only guided by physical properties of our world, but reflect cognitive states, may at first glance be counter-intuitive. It goes back to the 1960s, when Russian psychologist Alfred Yarbus initially showed that the eye movements of an observer of a painting were drastically different depending on the specific question that was asked shortly before (Yarbus, 1967). If the question was about the age of the depicted persons, the eyes systematically sampled the faces of the people, whereas a question about the wealth provoked eye movements to the furniture in the picture (see figure 1). This phenomenon is thought to reflect (pre-experimental) knowledge of the observer about task-relevant information and where to find it in such a scene (Hannula, 2010).

Ever since, a variety of eye tracking parameters have been proposed as markers that represent certain cognitive states, especially with regards to attention and memory. For example, it has been demonstrated that experts are able to scan expertise-related scenes faster and with more focus (Brams et al., 2019; Gegenfurtner et al., 2011), indicating long-term memory formation. On a much shorter timescale, studies have shown (1) increased visual exploration of manipulated regions and (2) decreased visual exploration of previously viewed scenes compared to new scenes, both indicating successful short-term memory formation (Lancry-Dayan et al., 2019; Ryan et al., 2000, 2007). The advance of neuroimaging techniques, most of all the introduction of functional magnetic resonance imaging (fMRI) 30 years ago, has added considerably to the understanding of the neuronal mechanisms underlying the relationship between attention, memory and viewing. Several findings have linked strategic visual exploration to activity in the medial temporal lobe (MTL) (Liu et al., 2017), in particular to the hippocampus. Based on these results, a new framework has emerged in which the hippocampus orchestrates overt attention allocation and thus visual exploration, guided by memory (Voss et al., 2017).

However, while it is widely acknowledged that visual exploration depends on the cognitive state of the viewer, the opposite has been far less investigated. Under the assumption that

the link between visual exploration and cognition is reciprocal (Hannula, 2010), visual exploration is not only a consequence of cognition, but also its prerequisite. One reason that some research has been blind to this fact may be that it becomes trivial in its extremes – if Yarbus' observer would have closed the eyes, questions about the picture would have become obviously unsolvable – but it is not otherwise. The question in how far cognition is modulated by viewing is of particular interest in neuropsychiatric conditions that are characterized by altered exploration patterns. Visual exploration deficits could partly cause cognitive deficits in such conditions, including memory-related impairments reported in autism spectrum disorders (Fedor et al., 2018), schizophrenia (Williams et al., 2010) and dementia (Shakespeare et al., 2015), but also attentional biases in depression (Elliott et al., 2011; Kellough et al., 2008), posttraumatic stress disorder (PSTD) (Armstrong et al., 2013; de Quervain, 2007) and anxiety disorders (de Quervain et al., 2017; LeMoult & Joormann, 2012).

Therefore, this thesis is aimed at presenting new scientific advances with regards to the relationship between visual exploration, memory and attention. Specifically, the aim was to document the potential to guide viewing behavior to (1) affect memory formation and (2) counteract an attentional bias to reduce anxiety. The effort to achieve this aim is documented in three studies that made use of eye tracking data generated at the Transfaculty Research Platform, Molecular and Cognitive Neurosciences at the University of Basel. The letters indicate the author's contribution to each study: **A** - designed the experiment; **B** - performed the experiment; **C** - analyzed the data; **D** - wrote the paper.

1) Fehlmann, B., Coynel, D., Schicktanz, N., Milnik, A., Gschwind, L., Hofmann, P., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Visual exploration at higher fixation frequency increases subsequent memory recall. *Cerebral Cortex Communications*, tgaa032.
   (A–D)
2) Fehlmann, B., Müller, F., Wang, N., Ibach, M. K., Schlitt, T., Bentz, D., Zimmer, A., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Effectiveness of a virtual reality-based eye contact training to reduce fear of public speaking: a randomized controlled trial. *Submitted for Publication.*
   (A–D)

3) Loos, E.*, Schicktanz, N.*, Fastenrath, M., Coynel, D., Milnik, A., Fehlmann, B., Egli, T., Ehrler, M., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Reducing Amygdala Activity and Phobic Fear through Cognitive Top-Down Regulation. *Journal of Cognitive Neuroscience,* 1–13.
(C–D)

In the first study, Fehlmann, Coynel et al. (2020), we combined an fMRI picture learning paradigm with eye tracking in a sample of 967 healthy subjects. This allowed us to replicate and extend earlier findings establishing a triadic correlation between individual visual exploration characteristics (i.e. frequency and location of fixations), MTL-activity and memory performance. In an additional experiment, we manipulated visual exploration patterns in an independent population of 64 subjects. The analysis revealed that both the fixation frequency and location can be altered by guided viewing conditions, which in turn affects episodic memory performance.

In the second study, Fehlmann, Müller, et al. (2020), we investigated the intervention potential of guided viewing in 89 participants suffering from public speaking anxiety (PSA). PSA can be viewed as a subclinical manifestation or a symptom of social anxiety disorder (SAD) and is known to be accompanied by gaze avoidance in social situations, based on negative experiences (i.e. memory representations). The repeated use of a stand-alone, smartphone- and virtual reality (VR)-based mutual gaze training was effective in reducing gaze avoidance as well as the fear of public speaking in real-life speech situations.

In the third study, Loos et al. (2020), we used eye tracking to monitor the eye movements of 43 participants with fear of snakes in an fMRI study. The results suggest that visual exploration of the feared stimuli is independent of cognitive load. This highlights the importance to assess visual exploration parameters to preclude that effects of cognition are confounded by biases in visual attention allocation. However, since the focus of this thesis is about the active manipulation rather than the passive control of visual exploration, it is only considered of secondary interest for the points to be made.

In the following chapters, an overview on the previous literature is provided regarding the interplay between visual exploration and cognition and how it can be used to assess – and potentially improve – cognitive functioning.

---

*these authors contributed equally

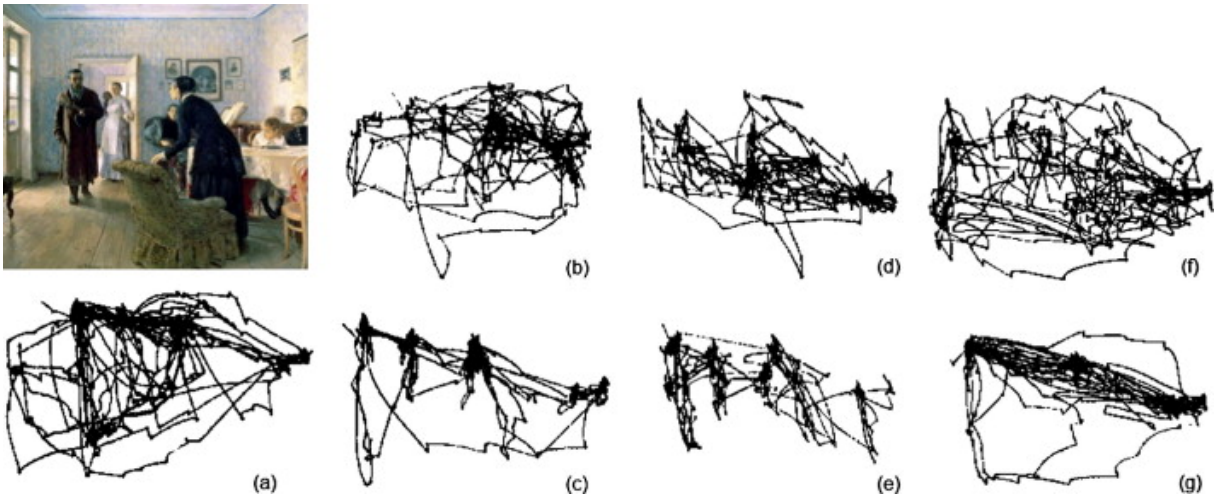


Figure 1. Visual exploration patterns of a subject, looking at a painting. The painting (*top left*; 'An Unexpected Visitor' by Ilya Repin, 1884–1888) was presented for 3 min in each task. Tasks were to *a)* freely examine the picture, *b)* estimate material circumstances of the family, *c)* give the ages of the people, *d)* surmise what the family had been doing before the arrival of the unexpected visitor, *e)* remember the clothes worn by the people, *f)* remember the positions of people and objects in the room and *g)* estimate how long the visitor had been away from the family. The different visual exploration patterns are thought to reflect the current goals and (pre-experimental) knowledge of the observer about the location of relevant information in such a scene. The figure is adapted from Yarbus (1967).

2 Theoretical Background

Despite the fact that visual exploration and cognition are tightly intertwined, eye movements have ‘mysteriously gone unmeasured in the vast majority of experiments on memory’ (Voss et al., 2017, p. 1). A similar claim can be made for neuroscientific experiments in general. This is problematic because visual exploration is (1) the dominant perceptual modality in humans and (2) very much incomplete, which is further elaborated in the following two chapters.

2.1 The Primacy of Visual Exploration

It has been known for decades that vision plays a dominant role in the humans’ and nonhuman primates’ perception (Pezdek et al., 1989; Shepard, 1967). The preference for visual processing was first illustrated in monkeys that naturally start looking at pictures without being trained or rewarded. They also gazed at pictures for longer when they were visually more entertaining, as opposed to homogenous color fields (Wilson & Goldman-Rakic, 1994).

It may not be surprising that cognitive processes, including attention allocation and memory formation, are particularly responsive to visual scenes in humans as well. However, the speed and ease with which scenic information is processed after it has been projected to the retina has fascinated researchers and become a subject of research in itself. It has been demonstrated in several studies that human observers can reliably recognize thousands of pictures that they have just seen for seconds, and that they can do so for days (Shepard, 1967; Standing, 1973). This capability is unmatched by other modalities, which was illustrated in experiments additionally presenting verbal (Standing et al., 1970) or auditory (M. A. Cohen et al., 2009) stimuli.

2.2 The Selectivity of Visual Exploration

Although we have the undeniable perceptual experience of a complete and detailed visual world, viewing is in fact a highly selective process (Henderson & Hollingworth, 2002).

During phases of steady gaze on a scene (fixations), only the small fraction of the visual field projected to the fovea can be seen in high spatial resolution. Rapid eye movements (saccades) serve to constantly relocate fixations in order to accumulate information from the environment (Henderson, 2003). They typically occur three to four times per second. For the time of a saccade, visual processing is selectively blocked, resulting in partial blindness throughout

its occurrence (Ross et al., 2001). Because saccades come at the cost of information interruption and fixations at the cost of information updating interruption, their interplay has to be tightly balanced to ensure a high quantity and quality of visual input to the brain. The functionality of the interplay between fixations and saccades has been demonstrated in experiments including fixations restricted to a given location (Henderson et al., 2005) and information provided outside their reach (Nelson & Loftus, 1980), both leading to impoverished perception. Noticeably, even during free viewing, only a minor fraction of the visual world is sampled and only this fraction can be the basis for subsequent cognitive processing (see figure 2). This makes it worthwhile to investigate the causes for individual differences in visual exploration patterns.

Figure 2. Viewing as a task of constant visual search.*Left:* Even under free viewing conditions, humans visually explore just a fraction of their environment. Only during fixation periods (*orange*), visual information reaches the brain. Fixations are quickly relocated across the scene by saccades in order to increase visual sampling. *Right:* Only the areas covered by fixations provide visual information in high spatial resolution, thereby limiting what is the basis for subsequent cognitive processes like memory formation. The figure is adapted from Voss et al. (2017).

2.3 Bottom-Up Control of Eye Movements

Before the ground-breaking experiments of Yarbus in 1967, it had been generally assumed that visual scenes rather than the viewers standing in front of them are the major driver of visual exploration (O'Regan, 1992). Newer image guidance theories still emphasize that attention can be directed to scene regions on the basis of salient image features, regardless of the semantic information they may or may not contain. On this view, attention is a reaction to image properties, 'pulled' by visual salience (Henderson & Hayes, 2017). The most influential image guidance theory proposes that features like color, color contrast, luminance contrast and edge orientation together form a saliency map for each scene, which eventually guides visual perception in a bottom-up manner (Itti et al., 1998; Itti & Koch, 2001) (see figure 3).

With regards to attention, the theory has been empirically supported by a study showing that salient, but task irrelevant objects diminished fixations on target objects (Pertzov et al., 2009), in line with the idea that elements of the visual world compete for limited attentional resources.

With regards to memory processing, the theory has been empirically supported by at least two studies combining eye tracking with a picture recognition paradigm (Kafkas & Montaldi, 2011; Sharot et al., 2008). Both found that pictures were more likely to be remembered if they were encoded with fixations focused around smaller regions. The authors concluded that smaller fixation clusters were triggered by salient details of some of the pictures. The attentional narrowing may in turn have led to a more focused encoding of distinctive details, facilitating subsequent recognition.

2.4 Top-Down Control of Eye Movements

More recently however, the emphasis on a top-down, meaning-based guidance of eye movements has emerged (Henderson & Hayes, 2017). In this framework, fixations are supposed to be preferentially allocated to regions of dense semantic information rather than to dense visual information.

What is semantically important and thus visually explored is partly based on the current goal of the observer, as shown by the experiment of Yarbus (1967). In one line of research, blinking has been proposed to be a marker for goals driving visual exploration. It was observed that subjects quickly and strategically adapted their blinking pattern to the task at hand. In environments dynamically changing over time, they were able to cognitively suppress blinks during important events and compensate them during unimportant events, minimizing the loss of visual information (Hoppe et al., 2018; Shin et al., 2015).

Yarbus (1967) also showed in his experiment that semantical importance is dependent on the observers' semantic knowledge about the world (e.g., 'furniture can be an indicator of wealth'), which can be accumulated over a lifetime. This was replicated in experts across domains (e.g., arts, chess, soccer, aviation), capable of optimizing the amount of information processed in their expertise-related environments. They did so by selectively allocating their attentional resources to task relevant stimuli and by ignoring irrelevant stimuli (Brams et al., 2019). Semantic knowledge can be acquired on shorter timescales as well. For example, in a study presenting pictures to the participants repeatedly, it was found that fixations got fewer and longer as the pictures got more familiar, along with shorter saccades and less exploration time

in semantically meaningful regions (Lancry-Dayan et al., 2019). Even within a picture, on the scale of milliseconds, previous experience can guide subsequent fixations. The inhibition of return is such a phenomenon, describing the observation that less important picture regions are less likely to be revisited under free viewing conditions, increasing the efficiency of visual exploration (see Hannula, 2010).

Together, these findings have contributed to the view that gaze behavior is guided to a substantial degree by an observer's goals and knowledge. Both are based on memory, formed over timescales from milliseconds to decades (Wolfe & Horowitz, 2017).

2.5 Eye Movements and fMRI

To study the connection between viewing, attention and memory on a neural level, the MTL has been the primary target of research. Extensive work suggests a key function of the MTL in binding fragmented information across cortical regions to permit episodic representations of scenes and events (see Cohen et al., 1999). Its connection to viewing is anatomically reflected by extensive pathways that connect the MTL to the oculomotor control regions, including the parietal eye field (PEF), the frontal eye field (FEF) and the supplementary eye field (SEF) (Liu et al., 2017).

It has been hypothesized that the MTL represents the visual space in a partly allocentric fashion (i.e. with reference to the real world as opposed to the retina), which in turn can be used to direct eye movements to relevant stimuli in the visual environment (Meister & Buffalo, 2016). The parieto-medial temporal pathway is a potential mediator of this process, linking the caudal intraparietal lobule with the MTL via place-selective regions such as the posterior cingulate and the retrospinal cortex (Kravitz et al., 2011). Since this finding originates from spatial navigation experiments in monkeys, the generalizability to human visual exploration has yet to be established (Meister & Buffalo, 2016). In humans, there is clinical evidence for the connection of the MTL to the dorsolateral prefrontal cortex (dlPFC), required for goal-dependent allocation of spatial attention. One study compared the visual exploration strategy of amnesic patients with hippocampal damage to healthy counterparts in a visual discrimination task (Voss, Gonsalves, Federmeier, Tranel, & Cohen, 2011). The authors reported that spontaneous re-visitations of recently viewed objects rarely occurred in amnesic patients. In healthy participants, they were related to enhanced hippocampal activity in conjunction with frontocerebellar circuits, as well as to superior subsequent recognition and object-location-memory performance. Further clinical work has shown that MTL damage leads to a lacking ability to increase fixations

on new stimuli while reducing them on already sampled ones (i.e. inhibition of return) (Chau et al., 2016; Zola et al., 2013). All these studies emphasize the top-down control of eye movements (see 'Top-Down Control of Eye Movements'). According to this framework that was later formalized as the active-memory hypothesis (Voss et al., 2017), the hippocampus uses memories, irrespective of the timescale, to guide future fixations (see figure 3).

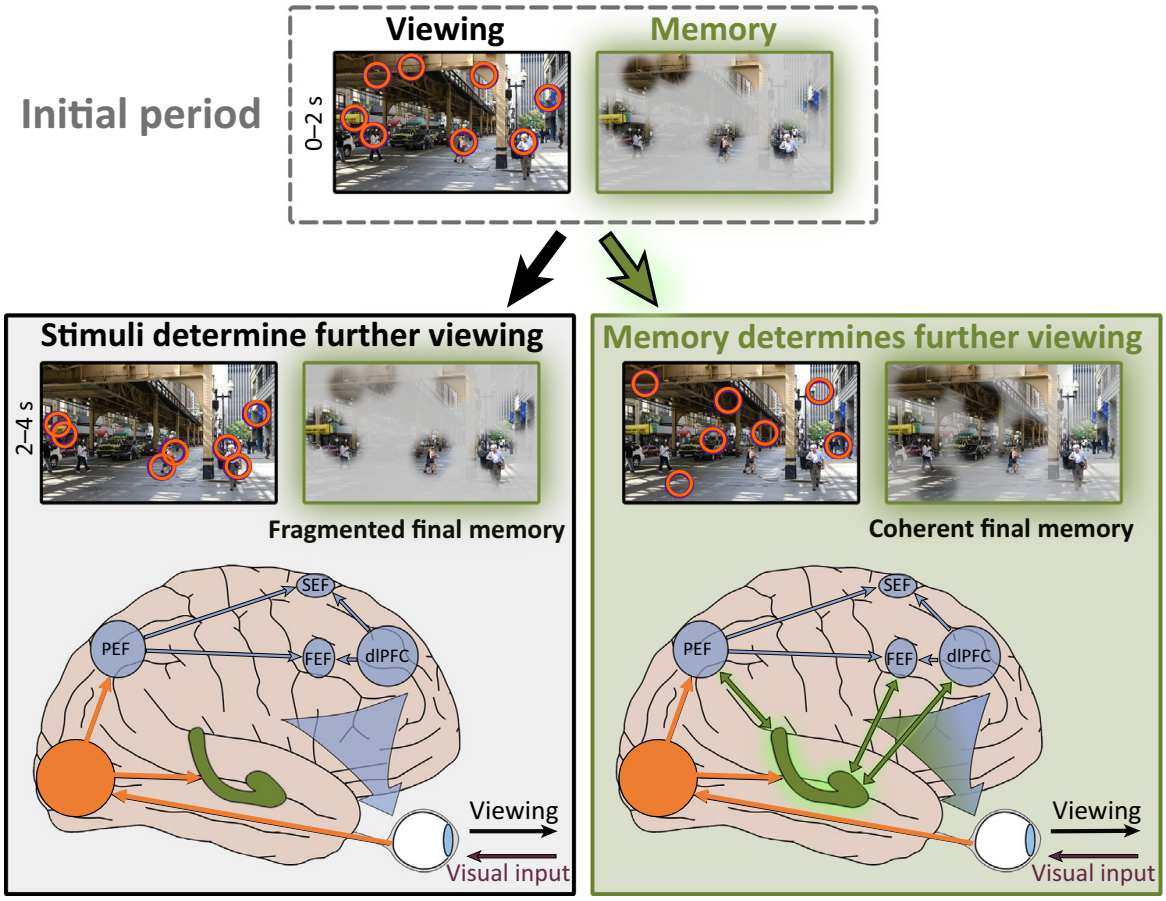


Figure 3. Illustration of the active-memory hypothesis. *Top:* Only a fraction of the environment is visually explored, and only this fraction can be the basis for subsequent memory formation. *Bottom left:* According to image guidance theories, visual exploration is guided by features of the scene like color, color contrast, luminance contrast and edge orientation. In this scenario, visual information is independently sent to the hippocampus for memory processing and to oculomotor control regions for eliciting subsequent fixations. *Bottom right:* By contrast, the active-memory hypothesis states that the memory is used to direct further viewing. Relevant memory representations can be built on timescales from milliseconds (e.g., by previous fixations on relevant locations within a scene) to decades (e.g., through expertise). In this scenario, the hippocampus interacts with cortical oculomotor control regions such as the parietal eye field (PEF), the frontal eye field (FEF) and the supplementary eye field (SEF) as well as the dorsolateral prefrontal cortex (dlPFC), driving visual exploration by online memory representations. Under this assumption, a reciprocal relationship between viewing and memory is plausible. However, in how far visual exploration can causally affect memory remains to be further investigated. The figure is adapted from Voss et al. (2017).

2.6 Eye Movements as a Tool for Diagnostics

In the clinical context, eye movements have been used as biobehavioral markers in many neuropsychiatric conditions that are characterized by memory impairments, including autism spectrum disorders (Fedor et al., 2018), schizophrenia (Williams et al., 2010) and dementia (Shakespeare et al., 2015). For example, patients suffering from Alzheimer's disease were unable to appropriately adjust their visual exploration pattern to different viewing tasks – unlike the observer in Yarbus' experiment. Instead, they inclined to fixate upon the low-level salient parts of the scenes, irrespective of their relevance (Shakespeare et al., 2015).

Further supporting evidence for a close relationship of eye movements and disease comes from clinical conditions that are accompanied by attentional biases, such as depression, PTSD and anxiety disorders. In all these conditions, altered visual exploration as well as memory anomalies are particularly selective to certain stimuli. Depressed patients are known to visually explore dysphoric images for a longer period of time at the cost of positive images (Kellough et al., 2008). Similarly, war veterans with PTSD typically maintained their overt attention longer on fearful and disgusted facial expressions (Armstrong et al., 2013). This attentional bias towards negative information has been discussed as a robust feature in both disorders, playing an essential role in the maintenance of negative beliefs and the (re-)consolidation of negative memories.

In anxiety disorders, early vigilance towards the feared stimulus has also been reported, but is typically followed by subsequent avoidance (Onnis et al., 2011). In one of the most well-described cases, SAD, anxious individuals extensively avoid mutual gaze in social situations, which can be quantified by eye tracking as a diagnostic marker (Chen et al., 2020). The causes of SAD are still being debated and include genetic susceptibility (Furmark, 2009) and environmental risk factors (Mathew & Ho, 2006). However, the interplay between visual exploration, attention and memory formation seems to be particularly important for the understanding of the disorder. It has been pointed out that negative experiences could provoke gaze avoidance as a submissive gesture to reduce the anticipated social threat (Chen et al., 2020; Horley et al., 2004; Weeks et al., 2013). Withdrawn eye contact could in turn trigger negative reinforcement learning, when socially anxious individuals perceive not being rejected as a consequence of not holding eye contact. Because mutual gaze avoidance prevents sufferers from disconfirming their negative beliefs about an interaction partner's attitude towards them, it could maintain social anxiety (Schulze et al., 2013).

Taken together, the role of visual exploration is still under debate in a lot of neuropsychiatric conditions that are accompanied by attentional and memory deficits. It is unclear if altered visual exploration is merely a consequence of those deficits, or if it represents a risk factor causing and/or maintaining them. Especially in the context of social anxiety, it is likely that gaze avoidance reflects an attentional bias, which is caused by past experiences (i.e. memory representations) and affects the formation of new ones.

If the link between viewing and cognition is indeed causal, the guidance of visual exploration could provide a new interventional tool, targeted at improving memory and reducing anxiety. In the next chapter, the first attempts to actively manipulate eye movements are summarized.

2.7 Eye Movement as a Tool for Intervention

Literature on manipulating eye movements during visual exploration as an intervention is scarce, with only two empirical studies investigating its effect on memory performance. In a first attempt combining fMRI with a recognition task, Voss and colleagues (2011) manipulated the control that subjects had over the position of a moving window through which they studied objects and their locations. For half of the trials, the window position was actively controlled by the subject with a joystick, while it was pre-determined for the other half. In this passive condition, the path was mimicking the active control of the previous subject, which ensured that the visual content was matched across conditions. The results showed that volitional control was beneficial for memory performance and associated with brain activity centered around the hippocampus. This supports the hypothesis that memory at any moment is used to direct further viewing, and that this process cannot be easily replaced by externally guided visual exploration. It has been implicated thereby that the mediating role of the hippocampus in relational processing can be entirely independent from long-term memory (see ‘Eye Movements and fMRI’).

The potential of guiding visual exploration was further explored by Chan and colleagues (2011). In a face recognition paradigm, visual exploration was restricted to a window, whose path was either actively controlled or pre-determined, much like in the study by Voss et al. (2011). Importantly however, the pre-determined exploration pattern was not mimicking the active control, but a third free viewing condition of another subject. The recognition performance was highest in the free viewing condition, but it was higher in the passive condition than in the active condition. In addition, increased fixation frequency on the faces during encoding was associated with subsequent recognition performance across conditions.

Three conclusions can be drawn from these findings: (1) Peripheral vision is critical for memory formation. Whenever it is restricted (e.g., by a search window), memory formation is attenuated due to its reliance on fragmented visual information. (2) Volitional control is not beneficial under all circumstances. If the passive visual exploration path is informed by earlier free viewing trials, memory formation can outperform conditions in which visual search is actively controlled but based on incomplete information. (3) Besides the location, the frequency of fixations is an interesting target to guide externally in order to affect memory performance.

In the first study, Fehlmann, Coynel et al. (2020), we further explored the potential of manipulating visual exploration to affect memory performance. In the second experiment of the study, we experimentally manipulated the frequency and location of fixations in 64 subjects. This was done by guided viewing conditions without restricting peripheral vision. The aim was to thereby increase subsequent episodic memory performance.

In the second study, Fehlmann, Müller, et al. (2020), we guided the gaze of 89 participants suffering from PSA to faces of a virtual audience by using a stand-alone, smartphone- and VR-based mutual gaze training without restricting peripheral vision. We expected to thereby correct an attentional bias (i.e. gaze avoidance) in socially anxious individuals and to reduce fear of public speaking by challenging their negative beliefs about real-life speaking situations.

The next chapter describes the relevant methods underlying these two studies.

3 Methods

The use of eye tracking in memory research presents some challenges, which may be part of the reason why it is still not extensively used. Especially the combination with fMRI goes along with several peculiarities, from study design to data analysis, that need to be considered to obtain reliable and valid results. Outside the scanner, the recent development of mobile eye trackers and VR devices with eye tracking has opened a new intriguing field for eye tracking studies. It offers the potential to measure visual exploration in real – or realistic – 3D environments while granting a high degree of control to the experimenter. However, the increased data complexity of this new technology also requires some aspects of data collection, analysis and interpretation to be reconsidered.

In the following chapters, the basics of eye tracking methodology are introduced. Further, they illustrate how the different challenges were approached in the two studies this thesis is focused on and provide some advice on how they can be addressed in future work.

3.1 Eye Tracking

Eye movements can be summarized at the level of an entire experimental display to quantify overall viewing behavior. Alternatively, they can be summarized at the level of specific elements of a visual stimulus that are defined as areas of interest (AOIs) to quantify directed viewing behavior (e.g., gaze to the eye and mouth region of a face; gaze to visually or emotionally salient features of a sunset) (Hannula, 2010).

3.1.1 Fixation Frequency

One of the most widely established and most straightforward visual exploration characteristics that is used in memory research is the fixations frequency as a proxy for sampling intensity. It has been consistently reported that there is a positive association between the number of fixations on simple objects and recognition memory strength for these objects (Kafkas & Montaldi, 2011; Pertzov et al., 2009; Tatler et al., 2005). Similar effects were found in face recognition, where the number of fixations at first encoding was related to memory performance (Heisz et al., 2013).

3.1.2 Fixation Location

With all the evidence for top-down guidance of eye movements, it is clear that not all the spots in a scene are equally important for memory formation. As a consequence, the most commonly used eye tracking analysis techniques rely on the definition of AOIs. AOI-based methods are implemented in almost every standard software package and straightforward in their application (Duchowski, 2017). In most cases, geometric forms are pre-selected by the experimenter and drawn around the visual area that is considered to be of interest. After eye tracking data have been collected, all the established measures like the number of fixations can still be analyzed, but are typically restricted to AOIs. In the context of directed viewing, the total duration of fixations in AOIs (i.e. total dwell time) is often indicated instead of or in addition to the number of fixations in AOIs. Given that the duration of fixations is more or less constant within stimuli and experimental conditions, both measures are usually highly intercorrelated and sometimes used redundantly in the literature.

Besides the more traditional measures, there is a multitude of additional ones that get available by defining AOIs, like the proportion of AOI covered by fixations, the number of transitions in an out of an AOI and the scan path between AOIs, just to name a few (Holmqvist, 2011). While it may be tempting to build all kinds of metrics from such an approach, there are at least three caveats that need to be considered:

1) Decisions on the number, size and placement of the AOIs are most often based on theoretical considerations. They are not always specified in the scientific literature and even when they are, can be hard to justify. A certain AOI may for example consist of several subareas that attract and repel fixations in an opposite manner. If this is not recognized and AOIs are therefore ill-defined, a significant amount of noise can be added to the data.
2) If characteristics of AOIs differ significantly from one stimulus to another, summary statistics across stimuli are severely biased. Pictures with bigger AOIs comprising more fixations gain for example much more weight in such an analysis by providing more data. Similarly, it usually only makes sense to compare the scan path between AOIs within a single stimulus, because it is highly dependent on picture properties.
3) The definition of AOIs can be associated with high amounts of data loss. Typically, eye tracking measures outside the AOIs are ignored. However, the assumption that no visual information of an AOI is processed because no fixations are identified within its borders may be wrong. The importance of peripheral vision to guide visual

exploration (Yamamoto & Philbeck, 2013), but also for the representation of objects (Chan et al., 2011; Fortenbaugh et al., 2007) has been repeatedly demonstrated. Fixations in the vicinity of an AOI may thus be sufficient for its encoding, leading to a focal vision bias when excluding them from the analysis. The bias can result in underpowered studies and false conclusion, especially in the case of AOIs only comprising a small proportion of the visually explored space.

All three caveats can have a major impact on data collection, analysis and interpretation and hamper comparisons across studies. One countermeasure is to define the AOIs in a data-driven fashion. In Fehlmann, Coynel et al. (2020), we therefore used empirical fixation data from 200 randomly chosen subjects per picture to define the AOIs and iterated the procedure to estimate its reliability. This was based on the idea of a meaning-based guidance of eye movements, whereby semantically important regions of a picture are simply those that are covered by most fixations (Henderson & Hayes, 2017). To prevent unequal weighing of the presented stimuli as well as data loss, we analyzed the data both with and without restriction to AOIs and defined the AOIs based on a relatively loose spatial threshold, which was constant across pictures (i.e. each AOI had to contain a minimum of 2.5% of the total fixations in order to be considered as such; all AOIs represented 50% of the total spatial variance of fixations).

To avoid focal vision bias in Fehlmann, Müller, et al. (2020), the AOIs were drawn such that they comprised the whole faces instead of only the eyes of the virtual and real-life audiences that participants had to hold eye contact with. The hypothesis of mutual gaze avoidance justified a pre-defined placement of the AOIs on the faces. Due to the fact that they were all almost identical in size, eye tracking data with regards to AOIs did not have to be weighted differently.

3.1.3 Blinks and Further Eye Tracking Measures

Besides the frequency and location of fixations that are most prominently reported in the neuroscientific research, other eye tracking parameters have been linked to cognitive states as well. Visual sampling continuity is sometimes quantified by the total blink duration or the inter-eye blink interval (IEBI) (Shin et al., 2015) and discussed as a an index of cognitive activity, or in its reverse, sleepiness. Furthermore, the pupil response and saccadic patterns have both been associated with cognitive effort and emotional arousal (van Steenbergen et al., 2011). However, because both measures are largely influenced by salient (bottom-up) features of the stimuli and come with specific methodological criteria that were hardly met in the experimental

conditions of the studies presented in this thesis (see ‘Eye Tracking and fMRI Scanner’ and ‘Eye Tracking in 3D environments‘), they are not further discussed.

3.2 Eye tracking and fMRI

The combination of eye tracking and fMRI offers the intriguing possibility to study neuronal underpinnings of the relationship between attention, memory and viewing. However, fMRI paradigms pose a number of limitations upon eye tracking that need to be addressed to ensure valid results. The most trivial limitation is of purely technical nature. Eye trackers need to be specifically equipped to withstand the strong magnetic fields in the MRI scanner, usually making them considerably more expensive than their counterparts without MRI-compatibility. As a result, the quality (e.g., time and spatial resolution) of eye trackers used in fMRI paradigms can be lower than elsewhere, precluding some of the more sophisticated analysis (e.g., of saccadic patterns).

Other limitations concern the experimental design. In fMRI studies, time spent in the scanner is reduced to a minimum to avoid that the participants start to feel uncomfortable or sleepy. This in itself puts some time constraints onto the time-consuming eye tracking calibration and recalibration procedures, negatively impacting data quality. During the restricted time in the scanner, many stimuli have to be presented to increase the signal to noise ratio (Maus et al., 2010). As a consequence, the visual exploration time for a single stimulus is often limited to seconds, which may artificially reduce the occurrence of certain eye movement characteristics (e.g., re-visitations of previously fixated AOIs) and preclude the analysis of others (e.g., the pupil response, not becoming fully apparent on this timescale).

The limitations highlight the need for well-designed and well-powered studies. While this is a prominent demand in fMRI research (Poldrack et al., 2017), it becomes even more essential in combination with eye tracking.

3.2.1 The Basel-Protocol

In 2008, a large-scale study was launched from two labs that are today an integral part of the Transfaculty Research Platform Molecular and Cognitive Neurosciences of the University of Basel. The aim was to investigate the neuronal correlates of memory formation with the use of fMRI. In order to additionally study the genetic underpinnings of memory formation with reasonable power, the target sample size was set to approximately 2’000 subjects. The

recruitment was stopped in 2016. By then, more than 1'800 subjects completed the study, making it the biggest single-center fMRI study worldwide. The gained insights have been reported in numerous publications throughout the years (e.g., Coynel et al., 2017; Egli et al., 2018; Heck et al., 2014; Loos et al., 2019; Petrovska et al., 2017; Spalek et al., 2015), but the eye tracking data have not been analyzed before. The fact that the fMRI paradigm was not mainly aimed at investigating visual exploration patterns made it necessary to carefully identify biases in the data and extensively filter trials to prevent data quality issues, as pointed out in the previous chapter. However, it allowed us to include a final subsample of 967 participants in the first experiment of the first study discussed in this thesis (Fehlmann, Coynel, et al., 2020), representing the world's largest sample of combined eye tracking and fMRI up to date. Because all the data was generated within the same paradigm without any relevant changes to the design, it is not only unique in terms of quantity, but also in terms of homogeneity. The number of subjects tested allowed us to focus on interindividual differences. This is an essential conceptual feature of the presented work that sets it apart from many other studies grouping trials based on pictures instead (e.g., Chipchase & Chapman, 2013; Kafkas & Montaldi, 2011; Pertzov et al., 2009; Sharot et al., 2008). In all these studies, salient visual stimulus features possibly contributed to systematic differences in visual exploration and thus to cognitive effects. While this was a deliberate choice in some cases, it is a difficult to control artefact in others. To take such bottom-up effects into account in our study, we z-standardized all eye tracking parameters within pictures. This is a rigorous method to correct for systematic variation between pictures and exclude it as a potential driver of the results.

To maximize statistical power for investigating the relationship between eye tracking and the fMRI signal, we conducted a parametric modulation analysis of fMRI data on the 1st level, which is described in the next chapter.

3.2.2 Parametric Modulation

The pictures presented in the Basel-protocol that was the basis for Fehlmann, Coynel et al. (2020) can be assigned to three valence categories, depending on their emotional content being rated as positive, neutral or negative. In a traditional model, the average effect of picture viewing on the expected blood-oxygen-level-dependent (BOLD)-contrast of the fMRI signal would be modelled based on one single value per participant, namely the average activation across pictures. Deviations from this average effect are considered to be measurement errors under such circumstances. However, it was evident that a given participant did not cover all

pictures with the same number of fixations, even when ruling out bottom-up effects of salient picture features. Due to the known variation, the contribution to the elicited fMRI signal may have differed considerably from one picture to another, which lowers the likelihood to detect true activation differences (i.e. causing type-II error inflation) (Wood et al., 2008).

To prevent this, we followed a parametric modulation analysis approach proposed by Büchel et al. (1998). This method allowed us to represent the different number of fixations per picture by a parametric regressor. The parametric regressor differentially modulates the average activation, which is still included in the model, on a single trial basis. Thereby, the regressor can absorb variation inherent to the number of fixations per picture, separating it from genuine measurement error and increasing overall statistical power (see figure 4).

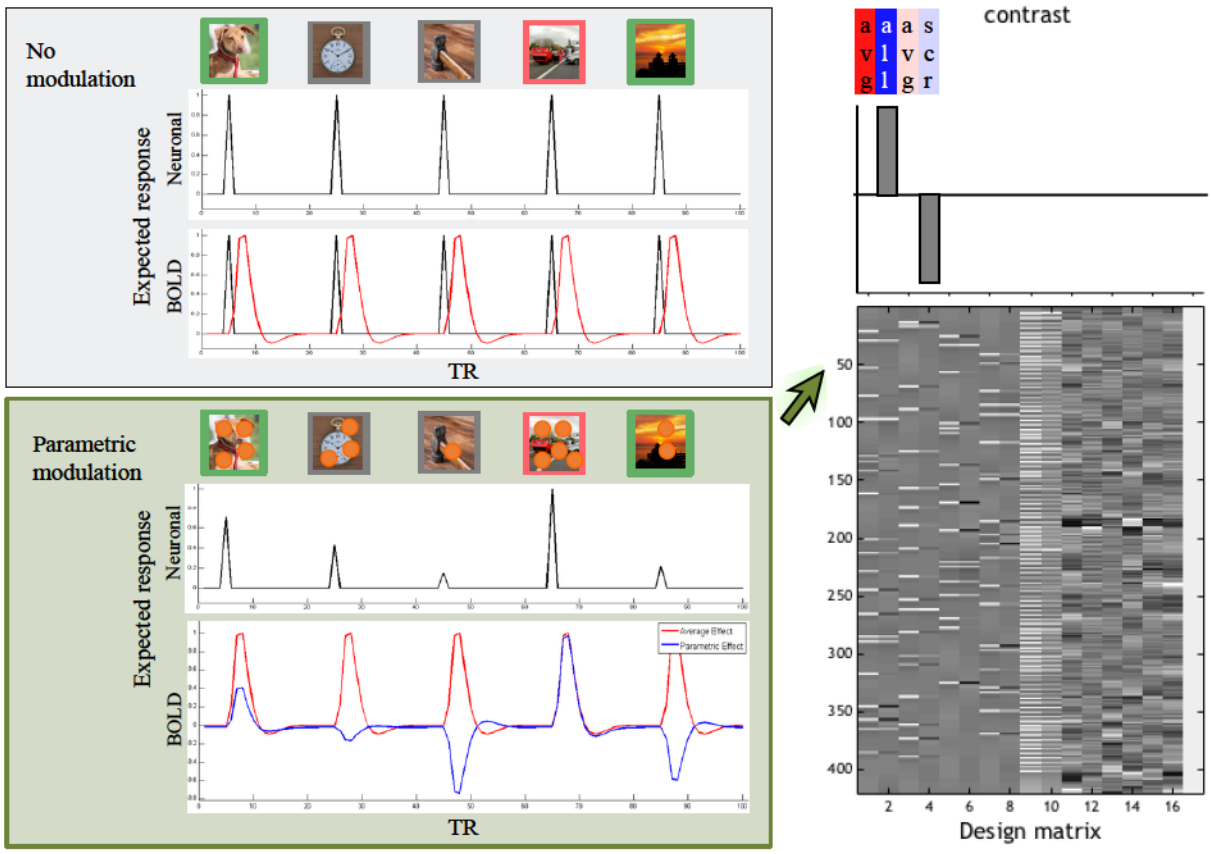


Figure 4. Classical vs. parametric modulation approach for 1st level fMRI analysis. *Top left:* classical 1st level approach, in which the expected blood-oxygen-level-dependent (BOLD) response is based on the average activation for viewing positive (*green*), neutral (*grey*) and negative (*red*) pictures per subject. The expected BOLD response is modelled by one single value, namely the average activation across the picture viewing task (*red line*), which can be insensitive to detect condition effects. *Bottom left:* The alternative parametric modulation approach allowed us to refine the expected BOLD response based on the number of fixations per picture (indicated by the *orange dots* and the *blue line*), modulating the average activation (*red line*) per subject on a single trial basis. This eventually decreased the error term of the model and increased statistical power. *Right:* The parametric regression estimates for all picture trials (*all*) were then further analyzed by contrasting them to the parametric regression estimates of a baseline, in this case trials with meaningless, scrambled (*scr*) pictures (not depicted on the *left*). The example is adapted from https://www.bobspunt.com/resources/teaching/single-subject-analysis/parametric-modulation/, August 2020.

3.3 Eye Tracking in 3D environments

Technical advances in mobile eye tracking and eye tracking in VR have made it possible to unobtrusively assess eye movements outside the lab. This has increased the popularity of the method in a broad range of fields, including sports (Kredel et al., 2017), aviation (Rudi et al., 2020), music (Fink et al., 2019) and retailing (Meißner et al., 2019), but also cognitive research (Kiefer et al., 2017). The main argument for using eye tracking in 3D environments to investigate cognitive processes is that they may differ considerably between the lab and real-world

settings (Foulsham et al., 2011). However, the increased ecological validity comes at the cost of increased data complexity and less experimental control.

In the next two chapters, some of the chances and challenges for mobile eye tracking and eye tracking in VR are illustrated, together with the implementation of both methods in Fehlmann, Müller, et al. (2020).

3.3.1 Mobile Eye Tracking

Mobile eye trackers are worn as glasses, which gives the wearer a high degree of flexibility and comfort. Experimental settings are thereby no longer bound to a computer screen – with the head often tightly fixed in front of it – but can be taken to virtually everywhere in the real world.

One of the challenges of the high degree of flexibility is with regards to the lighting conditions, which are hard to keep constant in these conditions. Especially for pupil measures that are highly susceptible to light, this should be taken into consideration. While the average pupil size over a larger period of time may still be informative, more fine-grained analyses of event-related pupil responses are often precluded (Palinko & Kun, 2011).

Another big challenge is the correct annotation and later interpretation of eye movements. Because the visual environment is no longer constant, it hardly ever makes sense to assess eye movement parameters without respect to pre-defined AOIs. However, the head movements of the wearer continuously change the position of the objects in the head-mounted scene camera recordings of the environment. In other words, the spatial coordinates of the AOIs are no longer in static relation to the observer, which necessitates new approaches to identify their position in each video frame. The most obvious way to achieve this is to manually analyze the video frame by frame, coding the parameters of interest like the number and duration of fixations in AOIs. Besides being extremely time-consuming, such an approach can be criticized for its lack of objectivity. The alternative is the automated detection of AOIs in the head-mounted scene camera recordings, allowing to compare AOI- and eye-movement-coordinates for each frame. Depending on the number of AOIs and their complexity, automated detection can be difficult and is often not implemented in the available software. However, in experimental settings such as in Fehlmann, Müller, et al. (2020), automated AOI detection provides a valid option for the fast and reliable annotation of eye movement parameters. Here, we used a mobile eye tracking system to quantify the eye contact that participants were able to hold with three experimenters while providing real-life speeches. The borders of

the AOIs were defined around the faces of the three experimenters. By automated face detection, we extracted the coordinates of these AOIs in each frame of all the recorded videos, allowing us to measure the dwell time as well as the number of fixations on faces. The approach turned out to be extremely reliable, with almost no faulty or missed annotations. The following factors are likely to have contributed to this outcome: (1) faces are known to be robustly detected by state-of-the-art algorithms (Tao et al., 2016), (2) to quantify the total dwell time, wrongly annotated face identities were irrelevant, as long as they were recognized as AOIs, (3) the identity of the three experimenters as well as their angle and distance from the participants was held constant, facilitating automated face detection.

Annotations may turn out to be much less reliable in other, more complex real-life experimental settings. They typically offer less control to the experimenter, which can be considered as one of the major drawbacks. The next chapter introduces eye tracking in VR, hereby offering a valid alternative.

3.3.2 Eye Tracking in VR

The benefits of using VR as opposed to real-life settings has already been recognized before World War II, when first flight simulators were introduced to train new pilots (Lele, 2013). In the research context, VR can combine the benefits of mobile eye tracking settings and lab experiments. It has the potential to provide fully immersive 3D environments without obvious movement restrictions for the user. At the same time, data analysis is not significantly more complicated than with 2D desktop eye trackers. Importantly, once the 3D models of the target stimuli have been created, they can be easily annotated to define areas – or better: volumes – of interest (VOIs) (Meißner et al., 2019). From the eyes of a user wearing the VR device, gaze-rays are cast into the VR environment, which are either based on the head position and orientation or on eye tracking. Whenever a VOI is hit by a gaze-ray, an event can be logged and be the basis for analysis. It is theoretically possible to annotate a VOI to every object in the VR and to analyze the complete eye tracking data as in 2D settings. In Fehlmann, Müller, et al. (2020), about half of the participants suffering from PSA were assigned to a stand-alone, smartphone- and VR-based mutual gaze training to reduce the subjectively perceived fear in real-life public speaking situations. In the VR environment that consisted of variously sized audiences, we used the gaze-tracking system of the VR software. As the goal was to keep eye contact with pre-specified members of the audience, VOIs were defined on their face regions. This allowed us to monitor at any point in time if the participants were

holding mutual gaze with a given target member of the audience and to adapt the further training procedure based on the success of doing so.

Summarized, VR experiments maintain a high level of reproducibility and control even in the most complex settings, while often providing greater ecological validity than their 2D counterparts. Some of the challenges irrespective of eye tracking, however, are the degree of immersion and interactivity, lack of embodiment, simulation sickness and ethics (e.g., 'nobody is *really* hurt as a result of participants' decisions'), which need to be further addressed in future work (Pan & Hamilton, 2018).

4 Original research papers

4.1 Visual Exploration at Higher Fixation Frequency Increases Subsequent Memory Recall

Title

Visual exploration at higher fixation frequency increases subsequent memory recall

Authors

Bernhard Fehlmann[a,e], David Coynel[a,e], Nathalie Schicktanz[a,e], Annette Milnik[b,e], Leo Gschwind[b,e], Pascal Hofmann[a,e], Andreas Papassotiropoulos[1,b,c,d,e] & Dominique J.-F. de Quervain[1,a,d,e]

Affiliations

[a]Division of Cognitive Neuroscience, Department of Psychology, University of Basel, 4055 Basel, Switzerland

[b]Division of Molecular Neuroscience, Department of Psychology, University of Basel, 4055 Basel, Switzerland

[c]Life Sciences Training Facility, Department Biozentrum, University of Basel, 4056 Basel, Switzerland

[d]University Psychiatric Clinics, University of Basel, 4002 Basel, Switzerland

[e]Transfaculty Research Platform, University of Basel, 4055 Basel, Switzerland

[1]These authors jointly supervised this work

Running Title

Higher fixation frequency increases memory recall

Abstract

Only a small proportion of what we see can later be recalled. Up to date it is unknown in how far differences in visual exploration during encoding affect the strength of episodic memories. Here, we identified individual gaze characteristics by analyzing eye tracking data in a picture encoding task performed by 967 healthy subjects during fMRI. We found a positive correlation between fixation frequency during visual exploration and subsequent free recall performance. Brain imaging results showed a positive correlation of fixation frequency with activations in regions related to vision and memory, including the medial temporal lobe. To investigate if higher fixation frequency is causally linked to better memory, we experimentally manipulated visual exploration patterns in an independent population of 64 subjects. Doubling the number of fixations within a given exploration time increased subsequent free recall performance by 19%. Our findings provide evidence for a causal relationship between fixation frequency and episodic memory for visual information.

32 **Keywords**

33 encoding, eye fixations, fMRI, memory, medial temporal lobe

Visual episodic memory consists of the voluntary recollection of previously encoded visual information along with contextual information. As such, visual episodic memory fundamentally depends on the visual sampling during encoding. Visual sampling is temporally restricted to phases of steady gaze referred to as fixations (Ross et al. 2001) and fine-grained information is spatially bound to the visual field projected to the fovea (Henderson 2003). As a consequence, only a minor fraction of the visual world sampled by an individual builds the basis for memory formation. It is thus crucial to consider visual exploration characteristics as an integrative part of memory processing and to ask how they affect memory encoding and later performance (Voss et al. 2017). Previous studies reported positive inter-individual correlations between the number of fixations and memory of objects (Pertzov et al. 2009; Kafkas and Montaldi 2011) and faces (Heisz et al. 2013; Olsen et al. 2016). It is unknown, however, if such a correlation also exists for memory of complex scenes and, importantly, if there is an underlying causal relationship.

In a first experiment, we explored how individual visual exploration characteristics, quantified by the number of fixations, blink duration and inter-fixation distance, are related to free recall memory performance across complex scenes. Based on previous findings (Pertzov et al. 2009; Kafkas and Montaldi 2011; Heisz et al. 2013; Olsen et al. 2016), we defined the number of fixations as the variable of primary interest. We analyzed eye tracking data of 967 subjects completing a memory paradigm, while controlling for inherent differences in stimulus properties. This enabled us to focus on interindividual exploration differences that are not attributable to the stimulus itself (e.g., saliency (Itti and Koch 2001), complexity (Voss et al. 2017), emotional valence (Sharot et al. 2008), semantic density (Henderson and Hayes 2017), memorability (Bylinskii et al. 2015)).

Using an fMRI paradigm allowed us to study the relationship between fixation frequency and memory on a neural level. In the context of face recognition, previous studies have linked visual sampling to activity in the medial temporal lobe (MTL; Liu et al. 2017) and to increased memory performance (Olsen et al. 2016). The aim of the first experiment was to combine these findings and to extend them from face recognition to episodic memory and a broad range of complex scenes. In a second experiment, we – for the first time to our knowledge – investigated causality of the found relationship between fixation frequency and memory by manipulating the scan paths of 64 additional subjects during memory encoding.

Materials and Methods Experiment 1

Participants

We analyzed data of 1485 subjects (917 females, mean age = 22.34, *SD* = 3.25, range 18–35) participating in a large-scale, simultaneous fMRI and eye tracking study conducted at the University Hospital of Basel, Switzerland (see Heck et al. 2014). Participants were free of any neurological or psychiatric conditions and did not take any medication at the time of testing (except hormonal contraceptives). Procedures were approved by the ethics committee of the Cantons of Basel-Stadt and Basel-Landschaft.

Experimental procedure

After participants received the general study information and provided written informed consent upon arrival, they were instructed and trained on a picture encoding task. Following training completion, participants were positioned in the scanner for the actual task, lasting for 20 min. Immediately afterwards, they performed an n-back working memory task for 10 additional min. Having left the scanner, participants were confronted with a surprise free recall memory test of the pictures without time limit. Participants were then repositioned in the scanner and performed a recognition task for 20 min, before structural MRI (T1) and diffusion MRI data were acquired for the remaining 20 min (see Fig. 1). The total length of the experimental procedure ranged from 3 to 4.5 hours per subject. Participants were rewarded with 25 CHF/h.

Stimuli

72 color images, divided into three valence groups, served as the main stimuli for the encoding task. On the basis of normative valence scores, 64 pictures from the International Affective Picture System (IAPS; Lang et al. 2008) were assigned to emotionally negative (2.3 ± 0.6), neutral (5.0 ± 0.3), and positive (7.6 ± 0.4) groups. In order to equate the picture set for visual complexity and content (e.g. human presence), 8 neutral pictures were selected from an in-house standardized picture set, resulting in a total of 24 pictures per valence category. Examples of pictures are as follows: erotica, sports, and appealing animals for the positive valence; bodily injury, snakes, and attack scenes for the negative valence; neutral faces, household objects, and buildings for the neutral condition. Furthermore, 24 scrambled pictures were included. Their

background contained the color information of the previously described stimuli, overlaid with a crystal and distortion filter (Adobe Photoshop CS3). In the foreground was one geometrical object, varying between pictures in terms of form, size, position and orientation. To control for primacy and recency effects in memory, two additional pictures of neutral objects were shown at the beginning and the end of the task respectively, but were discarded from further analyses.

Encoding task

The task was to visually explore 72 pictures under a free viewing condition, i.e. without any restrictions. Each of the 72 encoding trials started with a fixation cross, presented for 500 ms against a dark background, and was followed by the presentation of one picture for 2.5 s. The picture onset time was jittered within 3 s [1 repetition time (TR)] per valence category with regard to the scan onset. A blank, dark screen followed the offset of the picture for 1 s. Trials were separated by a variable period of 9–12 s (jitter). During this period, participants rated the emotional valence (negative, neutral, positive) and perceived arousal (low, middle, high) of the meaningful pictures on two separate three-point Likert scales by button press. The scrambled pictures were rated in terms of form (vertical, symmetric, horizontal) and size (small, medium, large) of the geometrical object (see Fig. 1). Across all trials, pictures were presented in a quasi-randomized order, allowing for a maximum of 4 consecutive pictures with identical valence categories.

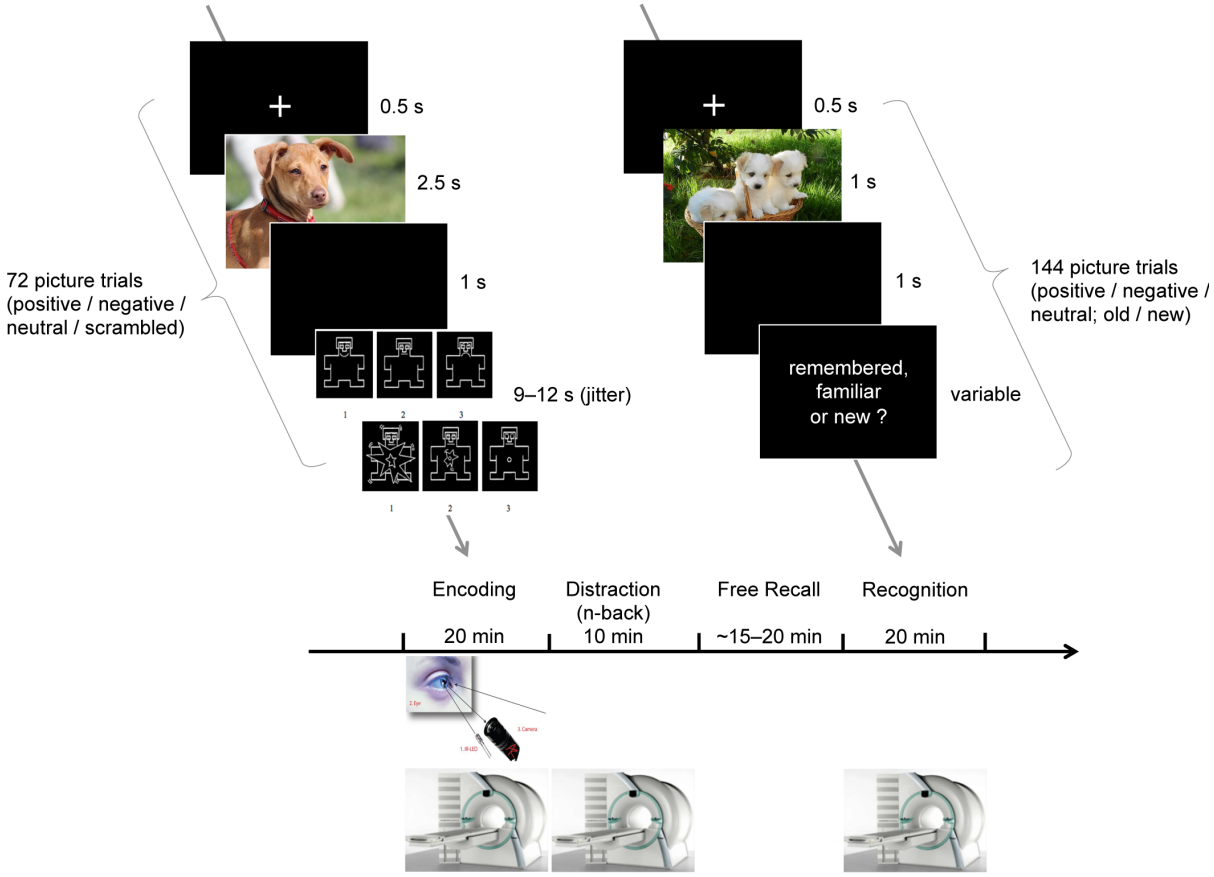


Fig. 1: Paradigm of experiment 1.

For 20 min., participants encoded 72 pictures in the fMRI scanner under free viewing conditions while their eyes were monitored by an eye tracker. The pictures were rated in terms of a) valence and b) arousal. Scrambled pictures were rated in terms of form (vertical, symmetric, horizontal) and size (small, medium, large) of the geometrical object instead. After 10 min. of an n-back working memory task that served as a distraction, they were confronted with a free recall memory test of the pictures outside the scanner and without time limit (approx. duration: 15–20 min.). Participants were then repositioned in the scanner and performed a recognition task for 20 min. Because of copyright restraints from the IAPS, the pictures in the figure are not the original IAPS pictures used in the study, but illustrative pictures that resemble them.

Free recall task

To document their free recall, subjects had to write a short description of each remembered picture. A picture was judged as correctly recalled if two trained investigators independently allocated the description to the same picture from the encoding set (inter-rater reliability > 98%,

reflecting the accordance rate between the two investigators across all trials). A third, blinded rater made a final decision for pictures that were rated differently. Free recall performance was assessed by the total number of correctly recalled pictures. We additionally assessed recognition performance, but focused on free recall rather than recognition, because for the latter we used subjective remember-know judgments (see Supplementary Material), a procedure that has been questioned as being suited to differentiate between episodic and semantic memory (Wixted 2009; McCabe et al. 2011).

Eye tracking data acquisition

During fMRI-sessions, the eye movements were recorded using an infrared camera integrated into the goggle system (Nordic NeuroLab, Bergen, Norway). The left eye position was sampled at 60 Hz and a spatial accuracy of about 1° (according to the manufacturer). The acquisition was controlled by ViewPoint eye tracker software (Arrington Research) and calibration was performed following the built-in 9-point procedure at the beginning of the experiments. In total, 967 subjects (596 females, mean age = 22.28, *SD* = 3.28, range 18–35) had eye tracking data before preprocessing.

Eye tracking data preprocessing

Collected raw data were pre-processed in R (v3.3.3; RRID:SCR_001905; R Core Team, 2015; http://www.r-project.org/). For each subject, fixation detection was done with an individual, velocity-based algorithm ('saccades' package, Malsburg 2015). Fixations with duration of less than 100 ms and saccades were discarded for further analyses. Slow, drift-like displacements of the recorded fixation coordinates were corrected as follows. The value of correction was calculated for each timepoint as its displacement relative to a baseline, represented by a moving median with a window size of 3301 sampling points (approximately 55 s). This procedure is roughly familiar to high-pass-filtering at 0.008 Hz, but more appropriate for time domain encoded signals (Smith 2003).

If not indicated otherwise, outlier detection of eye tracking data was based on boxplots. The first (q_1) and third (q_3) quartiles were estimated based on ideal fourths. After determining the interquartile range (*IQR*), a data point x was defined as an outlier if $x < q_1\text{-}1.5(IQR)$ or $x > q_3+1.5(IQR)$ (Wilcox 2012). Subjects were excluded if they were identified as outliers for

calibration data (total gaze deviation from expected grid, n = 80), or the eye movement velocity distribution (with respect to the x- and/or the y-axis, n = 87). Additionally, trials with the following characteristics were discarded: only one trial fixation (assuming that at least one saccade had to be made to ensure picture encoding), high signal loss (pupil aspect ratio < 0.5 in over 50% of the picture data samples) and/or pupil profile distortions (low correlation of the pupil response with the grand average profile, see Supplementary Materials and Methods Experiment 1). After removing data of 9 additional subjects with no valid trials, further analyses related to eye tracking were based on a total of 791 subjects (475 females, mean age = 22.35, *SD* = 3.39, range 18–35).

Eye tracking parameters

All eye movement measures were extracted per subject and picture trial. To quantify visual sampling intensity, the number of fixations at encoding was counted within the 2.5 s of picture presentation (N_{fix}) and restricted to areas of interest (N_{fix} in AOIs). For the definition of different AOIs, fixations of 200 random subjects were sampled per picture. Afterwards, a data-driven method to identify semantic areas of interest was applied (mean shift clustering, 'MeanShift' package, Ciollaro and Wang 2016; see Supplementary Materials and Methods Experiment 1). Visual sampling quality was specified by the number of unique AOIs covered by fixations (N_{AOIs}). This measure is perfectly inversely correlated with skipped AOIs and represents the completeness by which the distinct semantic aspects of a visual stimulus have been encoded (Holmqvist 2011). However, due to its high redundancy to the number of fixations in AOIs (N_{fix} in AOIs; r = .91; see Supplementary Table S2), the AOIs visited were not further investigated. As measures of secondary interest, visual sampling continuity was measured by the total time the eyes were closed during a complete picture trial (blink duration). Blink detection was based on the geometry of an ellipse fitted to the pupil, a default aspect ratio threshold of 0.6 (ViewPoint, Arrington Research) and a minimum duration of 83 ms (see Van Orden et al. 2000). Visual sampling dispersion was quantified by the average distance between two sequential fixation points across all picture fixations (interfixation distance), elsewhere described as an inverse index of attentional narrowing (Sharot et al. 2008). Importantly, to account for inherent differences in stimulus properties and their effect on visual exploration, all eye tracking parameters were z-standardized within pictures.

(f)MRI data acquisition

MRI imaging was performed using a 3 T Siemens Magnetom Verio whole-body MR unit with a 12-channel head coil. Head movements were minimized by using small cushions and the instruction to lie as still as possible. Blood oxygen level-dependent fMRI was obtained by a single-shot echo-planar sequence using parallel imaging (GRAPPA) with the following parameters: TE (echo time) = 35 ms, FOV (field of view) = 22 cm, acquisition matrix = 80 × 80, interpolated to 128 × 128, voxel size: 2.75 × 2.75 × 4 mm^3, GRAPPA acceleration factor r = 2.0. With a midsaggital scout image and an ascending interleaved sequence, we acquired 32 contiguous axial slices, placed along the anterior-posterior commissure plane and covering the entire brain with a TR = 3000 ms (α = 82°). The first two acquisitions were discarded due to T1 saturation effects. A high-resolution T1-weighted anatomical image was obtained by a magnetization prepared gradient echo sequence (MPRAGE) with the following parameters: TR = 2000 ms; TE = 3.37 ms; TI = 1000 ms; flip angle = 8°; 176 slices; FOV = 256 mm; voxel size = 1 × 1 × 1 mm^3 (see Heck et al. 2014). Stimuli were presented with Neurobs Presentation (Neurobehavioral Systems, Inc., Berkeley, CA, http://www.neurobs.com) and presented via an MR-compatible goggle system (VisualSystem; Nordic NeuroLab, Bergen, Norway). The system provided 800 × 600-pixel resolution with a field of view that nominally spans 23.5° in the vertical direction and 30° in the horizontal direction. Dioptric correction lenses were used when necessary. Responses were collected with an MR-compatible response box.

Construction of a population-average anatomical probabilistic atlas

The first 1000 participants that participated in the study and passed the T1 quality check were selected as a representative sample of our young and healthy population. Their T1-weighted images were used for the construction of a population-specific probabilistic anatomical atlas. The atlas consists of 35 cortical regions per hemisphere, as well as 17 subcortical regions (for details, see Supplementary Materials and Methods Experiment 1).

Preprocessing of (f)MRI data

If not indicated otherwise, brain-imaging data were processed in SPM 12 (Statistical Parametric Mapping; v6685; Wellcome Trust Centre for Neuroimaging, London; http://www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB R2016a (MathWorks). Volumes were

slice-time corrected to the first slice, realigned using the 'register to mean' option, and coregistered to the anatomical image by applying a normalized mutual information 3D rigid-body transformation. Successful coregistration was visually verified for each subject. Subject-to-template normalization was done using DARTEL (Ashburner 2007), which allows registration to both cortical and subcortical regions and has been shown to perform well in volume-based alignment (Klein et al. 2009). Normalization incorporated the following four steps: 1. The structural image of each subject was segmented using the 'Segment' procedure. 2. The resulting gray and white matter images were used to compute a subject-to-template transformation. The template employed here comes from a subgroup of 1000 subjects, part of which were included in the present experiment (Heck et al. 2014). 3. An affine transformation was applied to map the group template to MNI space. 4. Subject-to-template and template-to-MNI transformations were combined to map the functional images to MNI space. The functional images were smoothed with an isotropic 8 mm full-width at half-maximum (FWHM) Gaussian filter. Normalized functional images were masked using information from their respective T1 anatomical file as follows: At first, the three-tissue classification probability maps of the 'Segment' procedure (grey matter, white matter, and csf) were summed to define the mask. The mask was binarized, dilated and eroded with a 3 × 3 × 3 voxels kernel using fslmaths (FSL; v5.0.9; RRID: SCR_002823, Jenkinson et al. 2012) to fill in potential small holes in the mask. The previously computed DARTEL flowfield was used to normalize the brain mask to MNI space, at the spatial resolution of the functional images. The resulting non-binary mask was thresholded at 50% and applied to the normalized functional images. Consequently, the implicit intensity-based masking threshold usually employed to compute a brain mask from the functional data during the first level specification (spm_get_defaults('mask.thresh'), by default fixed at .8) was not needed any longer and set to a lower value of .05.

Fixations and memory

After preprocessing the eye tracking data, a subject-specific average value was computed for each eye tracking parameter. This was done separately for each of the 3 valence categories. If for a given subject and valence category, an average parameter value was based on less than 25% (= 6 out of 24 pictures) of the available trials, it was not further considered. For 709 subjects,

we had complete memory performance data in addition to the eye tracking data (433 females, mean age = 22.47, *SD* = 3.46, range 18–35). Data of these subjects entered the analyses, done in R. We applied linear mixed models ('nlme' package, Pinheiro et al. 2019) in combination with ANOVA (SS II). The participant-ID was included as the random effect in the mixed models. The dependent variable was the free recall performance. Independent variables were the specified eye movement parameters. Each independent variable was assessed in a separate model, together with the factor picture valence. Sex, age as well as the factors 'goggles' (accounting for a software update of the eye tracker) and 'recall room' (accounting for three changes of the room in which the free recall task took place during the course of the experiment) were included as covariates. We tested for main effects of eye movement measures and their interactions with picture valence. In a first step, we assessed significance of the interaction terms of the 4 full models using an FDR-correction over respective *p* values. In case of significant interactions, the main effect is still reported, but while accounting for the interaction term (SS III). In addition, post-hoc tests were applied to further investigate the interaction effect. In case of non-significant interactions, the model was recalculated without the interaction term. The reported *p* values of the main effects in the final models are again FDR-corrected for multiple comparisons (= 4, corresponding to N_{fix}, N_{fix} in AOIs, blink duration and interfixation distance). *P* values of post-hoc-tests are FDR-corrected by the number of conducted post-hoc tests (= 3, corresponding to the 3 valence categories) within each main model. Effect sizes for repeated measures are indicated by generalized semi-partial R^2 ($R^2\beta^*$; Jaeger et al. 2017), a generalization of the widely used marginal R^2-statistics (Nakagawa and Schielzeth 2013) which is comparable to effect size measures of between-subject designs. $R^2\beta^* > 0.01$, $R^2\beta^* > 0.09$ and $R^2\beta^* > 0.25$ are considered small, intermediate and large effects respectively.

Fixations and fMRI: first-level analyses

To investigate the relation between the number of fixations at encoding and functional brain activity, we conducted a parametric modulation analysis. Thereby, we only considered subjects with data for both fMRI and number of fixations (*N* = 775; 463 females, mean age = 22.37, *SD* = 3.40, range 18–35) and defined a general linear model (GLM) on the individual level. The following regressors were included: picture presentation, rating scale presentation (separately for

all 3 valence categories and scrambled pictures) and button presses. Picture and rating scale presentation were modeled by a boxcar function of constant duration, whereas button presses were modeled by a delta function at press onset. Mean-centered number of fixations per trial was included as a linear parametric modulator, for each picture category separately. The regressors were convolved with the canonical hemodynamic response function (HRF). Intrinsic autocorrelations were accounted for by AR(1), and low-frequency drifts were removed via high-pass filtering (time constant 128 s). The final design matrix was completed with six movement parameters obtained from spatial realignment. We aimed at identifying brain activity related to numbers of fixations, independently of valence. Therefore, we contrasted the parametric regressors of the three emotional valences versus the scrambled condition (referred to as parametric all vs. scrambled contrast), which served as a baseline [+1, +1, +1, -3].

Fixations and fMRI: second-level analyses

The subject-specific contrast estimates from the first-level analyses entered the second-level group analyses as dependent variables. Sex, age as well as the factor 'goggles' (accounting for a software update of the eye tracker) were included as covariates into the GLM. The main effect of the parametric modulator 'number of fixations' was assessed with a linear model. The statistical threshold of the two-sided hypothesis was set using family-wise error (FWE) correction for multiple comparisons across the whole brain (WB) at the voxel level at $p_{FWE\text{-}WB} < .05$ (corresponding to $t(771) \geq/\leq \pm 4.82$).

fMRI and memory: first-level analyses

We additionally analyzed whether brain activation in clusters associated with the number of fixations at encoding was related to memory performance. This was done on the basis of 1395 subjects with valid data for fMRI and the free recall task (854 females, mean age = 22.40, *SD* = 3.27, range 18–35). Unlike the parametric model, the underlying model does not contain any eye tracking regressors as parametric modulators, but is otherwise identical (the model contains the following regressors: picture presentation, rating scale presentation and button presses, convolved with the HRF, as well as 6 movement parameters). On the first level, the difference between the parameter estimates of the three emotional valences and the scrambled condition (referred to as all vs. scrambled contrast) were calculated for each subject and voxel [+1, +1, +1,

-3].

fMRI and memory: second-level analyses

The subject-specific contrast estimates from the first-level analyses entered the second-level group analyses as dependent variables. Sex, age as well as the factors ‘gradient’ (accounting for 2 changes of gradient coils), ‘software’ (accounting for a change in scanner software), and ‘recall room’ (accounting for three changes of the room in which the free recall task took place during the course to the experiment) were included as covariates. The free recall performance was entered as the independent variable of interest, and its association with brain activity was assessed with a linear model. Because we were only interested in voxels previously showing an association between the encoding signal and the number of fixations in AOIs, the statistical threshold of the two-sided hypothesis for the FWE-correction at voxel level was adjusted accordingly ($p_{FWE\text{-}SVC} < .05$, corresponding to $t(1388) \geq/\leq \pm 4.03$).

Materials and Methods Experiment 2

Participants

We collected data of 66 subjects (33 females; mean age = 23.30, *SD* = 3.89, range 18–32) participating in an eye tracking experiment conducted at the University of Basel, Switzerland. Participants did not participate in experiment 1, were free of any neurological or psychiatric conditions and did not take any medication at the time of testing (except hormonal contraceptives). The experiment was approved by the ethics committee of the Cantons of Basel-Stadt and Basel-Landschaft.

Experimental procedure

Participants received the general study information and gave their written informed consent upon arrival. They were then briefed and trained on a guided picture encoding task. The completion of the encoding task per se took 20 min. and was followed by 10 min. of an n-back working memory task that served as a distraction. Immediately afterwards, participants were confronted with a free recall memory test of the pictures with an upper time limit of 12 min. Finally, participants performed a recognition task for 10 min (see Fig. 2). The experiment took a total time between 1.25 and 2 hours per subject and was rewarded with 25 CHF/h.

Stimuli

From the 72 color images of experiment 1, a subset of 54 pictures was used, including their predefined AOIs (see Supplementary Materials and Methods Experiment 1). All the pictures from the original set with a maximum of 2 AOIs were thereby discarded. To arrive at an equal amount of 18 pictures per valence category, 1 negative picture with a high number of AOIs (= 9) and 4 negative pictures with the most redundant content were additionally excluded. Each of the 54 chosen pictures had a specific number of AOIs ($AOIs_{pic}$), ranging from 3 to 8. The pictures were divided into 3 sets, each containing 18 pictures matched in terms of valence category (6 negative, neutral and positive pictures respectively), number of AOIs and free recall memorability (i.e. the likelihood that a picture was recalled across participants) based on the data of experiment 1. The same 4 pictures as in experiment 1 served to control for primacy and recency effects, and were not further analyzed.

Encoding task

The task was to visually explore 54 pictures under guided viewing conditions. Each of the 54 encoding trials started with a fixation cross presented for 500 ms against dark background, and was followed by the presentation of one picture for an average of 8.2 s. Afterwards, the participants had to evaluate simple arithmetic operations that could either be correct (33%; i.e. 2 + 5 = 7) or wrong (67%; i.e. 3 × 2 = 8) for 9 s (see Fig. 2). This distractor task was chosen instead of the picture ratings in experiment 1. It was assumed that rating tasks related to the content of the picture could provoke scanning patterns deviating from the guided viewing in experiment 2 and would therefore be less suitable than in experiment 1. Across all trials, pictures were presented in a quasi-randomized order, allowing for a maximum of 4 consecutive pictures with identical valence categories and/or being of the identical set.

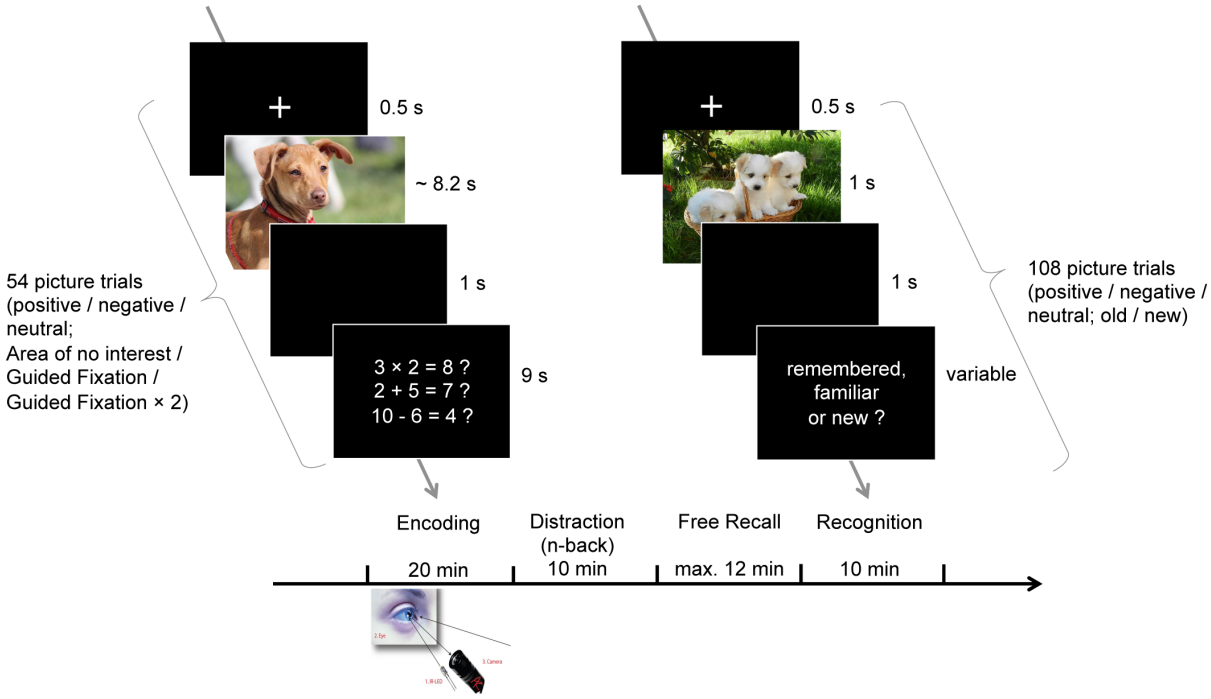


Fig. 2: Paradigm of experiment 2.

For 20 min., participants encoded 54 pictures under guided viewing conditions while their eyes were monitored by an eye tracker. Thereby, a quasi-random subset of 18 pictures had to be encoded under the 'Area of no Interest', the 'Guided Fixation' and the 'Guided Fixation × 2' condition, respectively. After 10 min. of an n-back working memory task that served as a distraction, they were confronted with a free recall memory test of the pictures with an upper time limit of 12 min., followed by a recognition task of 10 min. Because of copyright restraints from the IAPS, the pictures in the figure are not the original IAPS pictures used in the study, but illustrative pictures that resemble them.

Guided viewing conditions

Importantly, for each participant, the 3 picture sets were randomly assigned to three guided viewing conditions, referred to as 'Area of no Interest', 'Guided Fixation' and 'Guided Fixation × 2' (see Fig. 3). The instruction for all conditions was to follow the path of a moving circle and focus exclusively on the content lying within while ignoring the rest of the picture. The circle alternated between phases of steady state similar to fixations periods and fast movements imitating saccade-like movements. In the 'Area of no Interest' condition, the pathway of the circle included

n = $AOIs_{pic}$ fixation periods, but initially only covered one AOI in the center of the picture, while the subsequent fixation periods ($AOIs_{pic}$ - 1) were lying outside all AOIs. In the 'Guided Fixation' condition, the circle stopped at each AOI exactly once, corresponding to n = $AOIs_{pic}$ fixation periods. Finally, in the 'Guided Fixation × 2' condition, the pathway of the circle had n = $AOIs_{pic}$ × 2 fixation periods, allowing it to cover each AOI exactly twice.

The centroids of the AOIs were derived from experiment 1. Centroids of areas of no interest were calculated in a similar way (see Supplementary Materials and Methods Experiment 1), but based only on fixations that were not in AOIs. The size of the moving circle represents the average size of an AOI in experiment 1. The moving characteristics of the circle were also based on the data of experiment 1, with the intent of roughly mimicking plausible eye movement patterns. The velocity of the circle therefore alternates between 0°/s, imitating fixations, and 200°/s, imitating saccades. The latter is corresponding to an estimation of the peak saccadic velocity for the maximal possible angular distance of 30° (Boghen et al. 1974) between two AOIs in the current experiment. Depending on the number of AOIs ($AOIs_{pic}$) and the doubled fixation periods in the 'Guided Fixation × 2' condition, the amount of fixation periods per picture ranges from 3 to 16. The duration of these fixation periods is inversely related to their number, varying between 2333.33 ms for $AOIs_{pic}$ = 3 and 437.50 ms for $AOIs_{pic}$ = 16. Based on the distribution of the expected fixation frequencies (see Supplementary Fig. S1), the average fixation period (i.e. the duration the circle stayed at one point before moving to the next) was 1608.52 ms (SD = 475.54 ms) in the 'Area of no interest' condition, 1606.61 ms (SD = 476.43 ms) in the 'Guided Fixation' condition and 736.04 ms (SD = 238.50 ms) in the 'Guided Fixation × 2' condition'. The lower threshold was set based on the median fixation duration of 436.04 ms in experiment 1 and ensured that the encoding of any given AOI was still easily possible. This setting was chosen in order to allow all pictures to be fixated and thus encoded for the same amount of time (7000 ms) and thereby to avoid mere effects of exposure time. Depending on the different lengths of saccade paths, the actual picture presentation varied between 7712 ms and 9312 ms (M = 8153.63 ms, SD = 276.77 ms). The order of the AOIs was quasi-randomized for each subject and picture. The first restriction is with regards to the first fixation period always starting at the center of the picture, where the fixation cross had been previously located. The second restriction is with regards to the 'Guided Fixation × 2' condition, where the scan path first covered each AOI first

once in random order and then recapitulated itself to cover each AOI a second time. This procedure was chosen to keep the cognitive load and thus the potential risk of interference with the actual memory processes as low as possible, even at minimum fixation durations/maximum fixation periods as present in the ‘Guided Fixation × 2’ condition. The compliance of the subjects and the accuracy with which their gaze could be guided was estimated by correlating the scan path of the moving circle and the actual fixation pattern for each trial separately. In addition, it was characterized by the time lag between the moving circle arriving at a new picture region and the first fixation within this region. The fixations were thereby searched in a time window of 416.67 ms (50 sampling points) after the circle came to a halt.

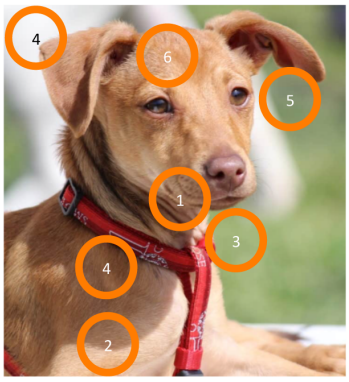
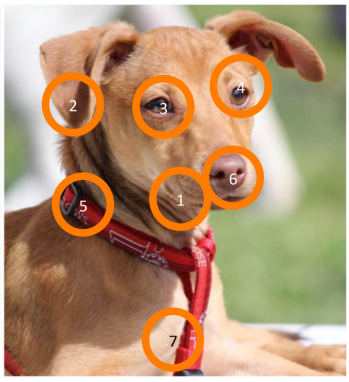
Condition	Area of no interest	Guided fixation	Guided fixation × 2
			
Fixation periods	7	7	14
AOIs	1	7	7
			
Fixation periods	5	5	10
AOIs	1	5	5
			
Fixation periods	6	6	12
AOIs	1	6	6

Fig. 3: Experimental manipulation of fixation frequency and location.

The task was to follow the path of a moving circle (1-2-3-...) within 54 presented pictures. Paths varied between conditions. Top left: Circle covering 1 AOI (1) and 6 areas of no interest (2–7), resulting in 7 fixation periods. Top center: 7 AOIs, each covered by the circle once, resulting in 7

412 fixations periods. Top right: 7 AOIs, each covered by the circle twice, resulting in 14 fixation
413 periods. Scanning time was constant across conditions. AOI 1 was related to the starting position
414 in the middle of the picture. AOIs and areas of no interest were derived from the empirical data of
415 experiment 1 (see Supplementary Materials and Methods Experiment 2). Middle and lower rows:
416 Further examples of neutral ('watch') and negative ('car accident') pictures, including their AOIs
417 and areas of no interest. Because of copyright restraints from the IAPS, the pictures in the figure
418 are not the original IAPS pictures used in the study, but illustrative pictures that resemble them.

419 **Working memory- and free recall task**

420 The working memory- and the free recall task are almost identical to experiment 1. The free recall
421 had a slightly different timing due to the reduced picture set (see Experimental procedure). The
422 inter-rater reliability in the free recall task was equally high (> 95%). We additionally assessed
423 recognition performance (see Supplementary Material). However, following the reasoning of
424 experiment 1, we focused on free recall rather than recognition.

425 **Eye tracking data acquisition**

426 To investigate how good subjects were able to comply with the 3 viewing conditions, their eye
427 movements were recorded with an SMI RED device. The gaze position accuracy of this system
428 was 0.4°, the spatial resolution 0.03° of visual angle. The eye tracker was controlled by the iView
429 X software (SMI iView X, SensoMotoric Instruments, Tetow, Germany) and fixated to the
430 presentation monitor with a display mode of 1680 × 1050 pixels. Subjects were placed
431 approximately 65 cm in front of the monitor, while the position of their left eye was sampled at
432 120 Hz. Calibration was performed following the built-in 9-point procedure at the beginning of the
433 experiments.

434 **Preprocessing of eye tracking data**

435 One subject reported to have misunderstood the instructions not to focus on the visual areas
436 outside the circle and was therefore excluded from any further analyses. For the remaining
437 subjects, fixation detection was done following the same pipeline as described in experiment 1.
438 Since there was no evidence for slow, drift-like displacements of the recorded fixation
439 coordinates, no correction was applied. No subjects (due to calibration outliers or no valid trials)

and no trials (due to only one fixation detected) were excluded using the outlier criteria of experiment 1. To identify possible deviations from the viewing instructions per trial, the theoretically expected fixation pattern, given by the coordinates of the moving circle at each time point, was correlated to the actual fixation pattern measured by eye tracking. One subject had a correlation below the outlier threshold (defined analogously to experiment 1) for more than 1/3 of all trials and was therefore excluded. Further analyses of the free recall performance were therefore based on a total of 64 subjects (32 females; mean age = 23.28, *SD* = 3.94, range 18–32).

Experimental manipulation of visual exploration

The influence of the three experimental conditions on memory performance was assessed in R with a linear mixed model ('nlme' package, Pinheiro et al. 2019) combined with ANOVA (SS II). The Participant-ID was included as the random effect in the model. The dependent variable was the free recall performance. The independent variable was the factor 'experimental condition'. We tested for the main effect of the experimental condition while controlling for sex and age. Post-hoc tests were applied to further investigate the effect, and *p* values were FDR-corrected by the number of post-hoc tests conducted (= 2, corresponding to the comparison of the 'Guided Fixation' condition with the 'Area of no Interest' condition and with the 'Guided Fixation × 2' condition, respectively). Effect sizes for repeated measures are indicated by generalized semi-partial $R^2\beta^*$(Jaeger et al. 2017).

Results

Experiment 1

In this experiment, subjects viewed 72 photographs of complex scenes with different emotional valences (i.e. neutral, positive and negative scenes) in the fMRI scanner, followed by a free recall test outside of the scanner.

Fixations – descriptive statistics

On average, subjects made 5.30 fixations per picture (*SD* = 1.01; mean duration 459.39 ms, *SD* = 126.01 ms) with a mean distance of 7.82° (visual angle; *SD* = 1.22) between two subsequent fixations. An average of 3.51 (*SD* = 0.94) fixations were lying inside any AOI of a picture, covering a total of 2.45 unique AOIs (54%, *SD* = 0.52). Blinks occurred in 26% of all trials and covered 8% (*M* = 206.85 ms, *SD* = 186.95 ms) of a picture trial on average.

Visual exploration and memory

On the behavioral level, we first investigated the association between several visual exploration characteristics and subsequent memory performance (for descriptive statistics, see Supplementary Results).

For the variable of primary interest (i.e. the number of fixations) we found a positive correlation with free recall performance ($t(1351) = 3.70$, $p = 3.1e\text{-}04$, $R^2\beta^* = .011$, 95% CI [.004, .021]). The interaction of the number of fixations with emotional picture valence was not significant ($F(2, 1351) = 1.36$, $p = .83$). Because memory performance might rely on the sampling of specific regions of each picture, fixations were subsequently restricted to areas of interest (AOIs), which covered semantically informative areas (Holmqvist 2011). In our experiment, the fixation frequency in AOIs was highly correlated with the fixation frequency in the entire picture ($r = .82$). There was a significant positive correlation of the number of fixations in AOIs with free recall performance ($t(1336) = 5.34$, $p = 4.4e\text{-}07$, $R^2\beta^* = .021$, 95% CI [.011, .035]). The interaction with valence was not significant ($F(2,1336) = 0.19$, $p = .83$). Within the variables of secondary interest, blink duration was negatively associated with free recall performance ($t(1303) = -4.46$, $p = 1.8e\text{-}05$, $R^2\beta^* = .014$, 95% CI [.006, .026]) without a significant valence interaction ($F(2,1303) = 0.36$, p

= .83). No association was found for the average distance between sequential fixations and free recall performance ($t(1351) = -1.43$, $p = .15$).

The positive link between sampling frequency in semantically informative regions and memory performance is in line with previous literature, and extends this finding with regard to the type of information processed (i.e. complex scenes) and type of memory tested (i.e. free recall, a purely episodic form of memory).

Fixations and fMRI

Next, we examined the relationship between fixation frequency in semantically informative regions and functional brain activation (independent of memory). The linear parametric modulation effect of the number of fixations in AOIs on the fMRI signal was investigated for the contrast 'viewing real pictures versus viewing scrambled pictures' throughout the brain (N_{voxels} = 57,032).

A positive modulation of brain activation by the fixation frequency was found in two large clusters, both being located predominantly in the medial left hemisphere. One was located in the left pericalcarine gyrus, extending to left lingual areas and bilaterally to precuneus and cingulum. The other was found in the left orbitofrontal cortex, including parts of the anterior cingulate and the bilateral superior frontal cortices. Additional clusters comprised bilateral parts of the MTL and thalamus as well as the left inferior parietal cortex and the right cerebellum. Negative modulation of brain activation by the fixation frequency comprised of a major cluster in the right cuneus, reaching into right lingual areas and bilateral parts of the superior parietal cortex (see Fig. 4 and Supplementary Table S1).

To summarize, a positive relationship between fixations and the fMRI signal was identified in early perceptual processing (e.g. left lingual) regions and regions known to be related to successful memory encoding, including the MTL.

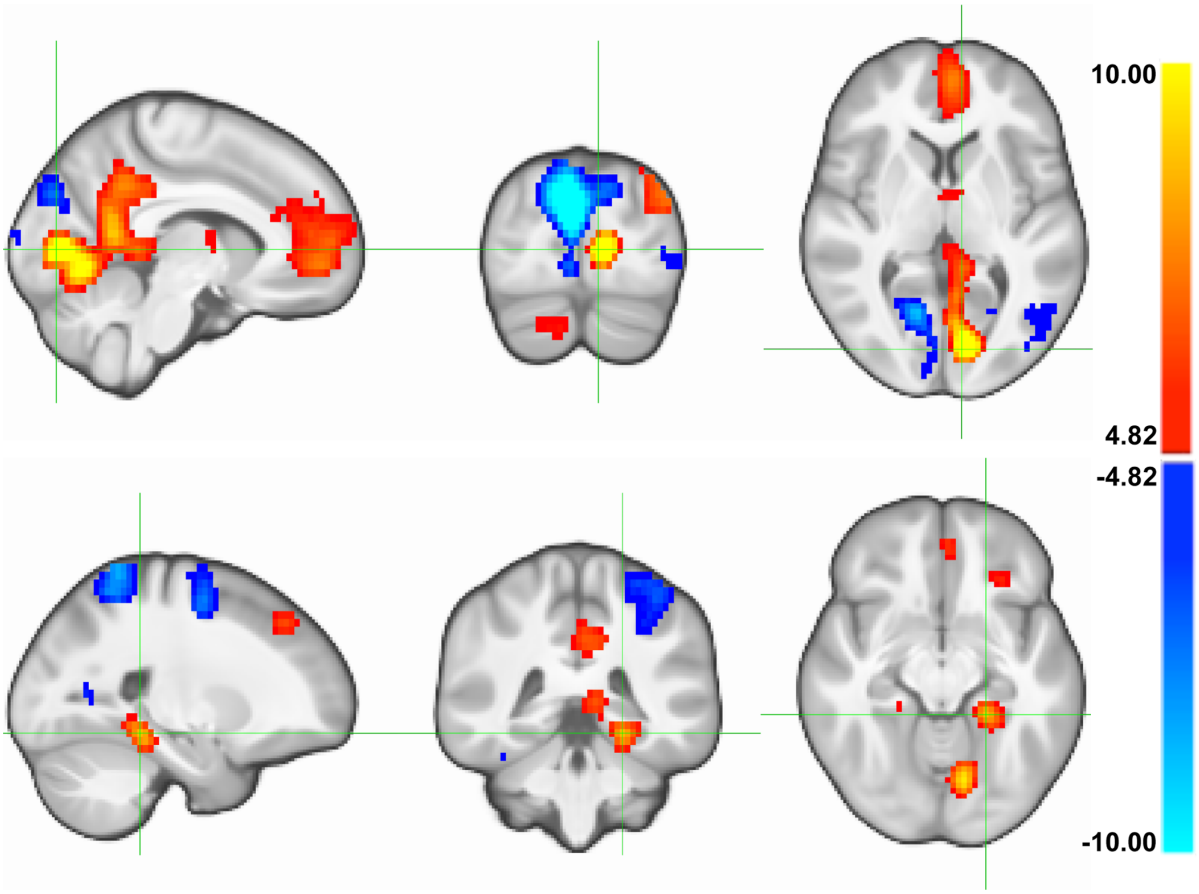


Fig. 4: Association between the number of fixations in AOIs and the fMRI encoding signal.

Parametric modulation effect of the number of fixations in AOIs on the fMRI encoding signal in 775 subjects, for every voxel of the brain (N_{voxels} = 57,032) for the contrast between viewing real pictures and scrambled pictures ($p_{FWE\text{-}WB}$ < .05 corresponding to $t(771) \geq/\leq \pm 4.82$). Top: Focus on the large cluster located in the left pericalcarine gyrus, extending to left lingual areas and bilaterally to precuneus and cingulum. Bottom: Focus on the cluster located in the left MTL. FWE: family-wise error; WB: whole brain correction for the investigated number of voxels in brain.

fMRI and memory

We then determined if brain activation in the previously detected fixations-related clusters is related to memory performance. Therefore, only voxels showing either a positive or a negative modulation of activity by the number of fixations in AOIs were considered. The associations between the fMRI-encoding signal and free recall were investigated in those voxels, revealing clusters of the right cerebellar cortex, the left superior frontal cortex and the left parahippocampal gyrus (see Fig. 5 and Supplementary Table S1). Importantly, the analyses revealed only positive associations, exclusively and consistently found in clusters with positive parametric modulation

effects. We thereby show that activity in the reported regions was associated to a great extent with free recall performance.

It is important to note that the behavioral and imaging findings reported so far are of correlational nature. Although our and previous results are in line with the idea that sampling intensity affects memory performance, they do not speak of causation. At this stage we cannot rule out that there is no causal relation, or that the causal direction is inversed, meaning that a good memory in general positively affects sampling frequency. In fact, there is evidence suggesting that previous experiences might influence visual exploration, including the number of fixations (Sharot et al. 2008; Hannula 2010; Wolfe and Horowitz 2017; Lancry-Dayan et al. 2019).

Furthermore, in the present experimental setting the number of fixations and the number of AOIs covered by them were highly correlated. It is therefore impossible to distinguish between the importance of the frequency and location of such fixations for subsequent memory performance. We cannot preclude that the increased amount of gathered semantic information, rather than the sampling intensity per se is related to successful memory processing.

To address these questions, we conducted an additional experiment pre-defining the visual scan paths for 64 subjects during picture encoding. The first aim was to examine the causal link between exploration characteristics and memory performance under this experimental manipulation. The second aim was to separately investigate the effects of the number of fixations and the number of AOIs covered by them on memory performance.

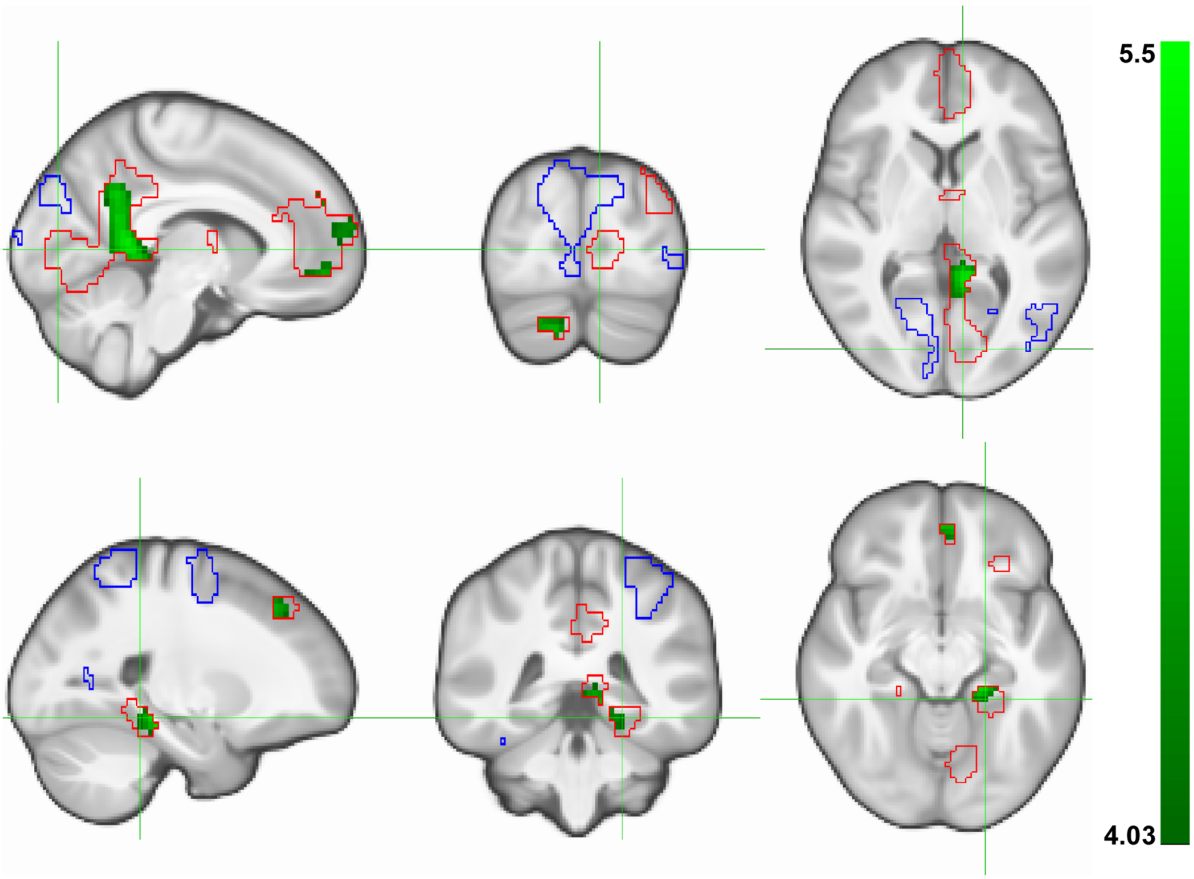


Fig. 5: Association between the fMRI encoding signal and episodic memory performance.

Association between the fMRI signal at encoding and subsequent free recall in 1395 subjects, restricted to voxels showing either a positive (red outline) or negative (blue outline) activation related to the number of fixations in AOIs ($p_{FWE\text{-}SVC} < .05$ corresponding to $t(1388) \geq/\leq \pm 4.03$). Top: Focus on the large cluster located in the left pericalcarine gyrus, extending to left lingual areas and bilaterally to precuneus and cingulum. Bottom: Focus on the cluster located in the left MTL.

FWE: family-wise error; SVC: small volume correction for the number of voxels in outlined regions.

Experiment 2

In this experiment, subjects viewed 54 photographs of complex scenes with different emotional valences (i.e. neutral, positive and negative scenes) under guided viewing conditions, followed by a free recall test.

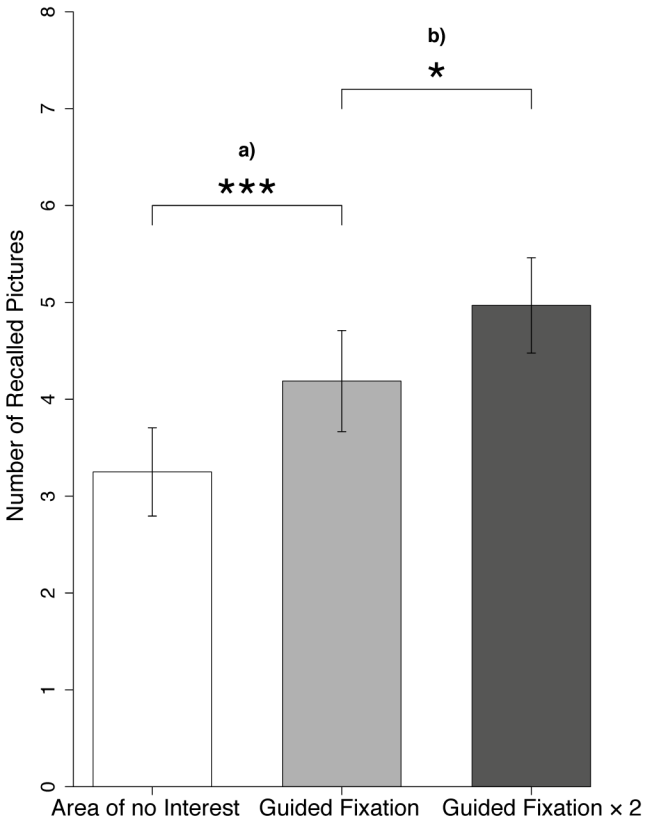
Scan path – descriptive statistics

The mean correlation between the scan path of the moving circle and the actual fixation pattern was high (r = .87), with only a small time lag between the moving circle arriving at a new picture region and the first fixation within this region (M = 74.05 ms, SD = 66.29), indicating good compliance and accuracy. Furthermore, the empirically measured fixations per picture indicated that the fixation frequencies in the 'Area of no Interest' (M = 5.81, SD = 0.92) and the 'Guided Fixation' condition were similar (M = 5.59, SD = 0.94), while increased in the 'Guided Fixation × 2' condition (M = 10.24, SD = 2.23; see Supplementary Fig. S1). The empirically measured durations of fixations per picture in the 'Area of no Interest' (M = 1047.18 ms, SD = 451.61 ms) and the 'Guided Fixation' condition were again similar (M = 1162.10 ms, SD = 479.67 ms), while decreased in the 'Guided Fixation × 2' condition (M = 692.54 ms, SD = 227.08 ms).

Visual exploration and memory

To study the effects of fixation frequency and location (i.e. the number of fixations in AOIs and the number of AOIs covered, respectively) on episodic memory performance in isolation, we introduced three within-subject experimental conditions as described in Fig. 3. The task was to encode the pictures following the path of a moving circle. The 'Area of no Interest' and 'Guided Fixation' condition only differed with regard to the number of AOIs (i.e. increased in the Guided Fixation' condition). The 'Guided Fixation' and the 'Guided Fixation × 2' conditions only differed in the number of fixations (i.e. doubled in the 'Guided Fixation × 2' condition) in the AOIs.

For free recall performance, there was a positive main effect of the factor 'experimental condition' (F(2,126) = 20.70, p = 1.7e-08). Post-hoc tests revealed an average decrease of freely recalled pictures in the 'Area of no Interest' condition compared to the 'Guided Fixation' condition by 22% (t(63) = -3.79, p = 2.9e-04, $R^2\beta^*$ = .063, 95% CI [.007, .165]), as well as an average increase of freely recalled pictures in the 'Guided Fixation × 2' condition compared to the 'Guided Fixation' condition by 19% (t(63) = 2.51, p = .015, $R^2\beta^*$ = .036, 95% CI [.001, .122]; Fig. 6).



584

585 **Fig. 6: Effect of scan path manipulation on free recall performance.**

586 Episodic memory effect in 64 subjects by a) decreasing the number of AOIs covered by fixations,
587 leading to a lower amount of freely recalled pictures in the 'Area of no Interest' condition (*M* =
588 3.25, *SE* = 0.46) compared to the 'Guided Fixation' condition (*M* = 4.19, *SE* = 0.52) as well as by
589 b) increasing fixation frequency, leading to a higher free recall performance in the 'Guided
590 Fixation × 2' condition (*M* = 4.97, *SE* = 0.49) compared to the 'Guided Fixation' condition.

Discussion

Experiment 1 revealed a link between visual exploration and memory. Most importantly, there was a positive correlation of the number of fixations in semantically informative picture areas and the performance in the subsequent episodic memory task. Fixations allow for the extraction and processing of detailed information like the position or orientation of objects (Pertzov et al. 2009). Episodic memory is typically characterized by the ability to recall such details.

Further, we found a negative correlation between blink duration and episodic memory performance. A possible explanation is that during blinks, which might reflect sleepiness, no visual sampling is taking place and hence, less information is encoded. Finally, we did not find a correlation between the interfixation distance and episodic memory performance, indicating that this exploration characteristic per se is not affecting memory strength.

The neuroimaging data of experiment 1 revealed that the number of fixations in semantically informative picture areas was correlated with brain activation in regions important for vision and memory. The activated brain regions overlap with the parietal medial temporal pathway, which is associated to target-directed fixations in animals (Kravitz et al. 2011) and includes the MTL. Activation of the MTL has been repeatedly and consistently related to successful memory encoding in human neuroimaging studies (Dickerson and Eichenbaum 2010). Growing evidence additionally attributes to the MTL a critical role in perceptual processing (Ringo et al. 1994; Hodgetts et al. 2017; Dalton et al. 2018; Ryan et al. 2020). Our findings further suggest the recruitment of frontocerebellar circuits. They are implicated in strategic planning of subsequent eye movements (Voss, Warren, et al. 2011; Voss, Gonsalves, et al. 2011), have been previously found to be activated in visual episodic memory tasks (Niu 2012), and could be functionally connected to the MTL via the thalamus (Ito et al. 2015). Interestingly with regards to the MTL, experiment 1 has predominantly shown an association between memory performance and activation in the parahippocampal gyrus, while an earlier study using a face recognition paradigm identified the hippocampus to be most critically involved in both in visual sampling and memory formation (Liu et al. 2017). This discrepancy could be interpreted in the light of the ‘binding of item and context’-model (Diana et al. 2007). The model proposes that both spatial and non-spatial contextual information is stored by the parahippocampal gyrus, while the hippocampus

binds this information, integrating additional item information provided by the perirhinal cortex (Graham et al. 2010). We argue that for complex scene viewing, as opposed to faces, the picture context (e.g., the number and strength of the contextual associations within a scene) and the temporal context (e.g., the order of scenes when trying to memorize them as a story) are particularly important (Bar et al. 2008). In addition, due to the heterogeneity of the scenes used, it might be sufficient to recall some of them based on a serial, single feature identification strategy without rich associations between them, which would require less hippocampus involvement (Graham et al. 2010). This might partly explain why the effect size for the association between fixation frequency and memory in experiment 1 ($R^2\beta^*$ = .011) was smaller than reported by Olsen at al. (2016) in a face recognition paradigm (R^2 = .11).

In experiment 2, we found that manipulating the scan path affects subsequent memory performance. As anticipated, visual exploration in semantically unimportant picture regions, as compared to regions with more semantic information, decreased subsequent free recall performance by 22%. This is in line with earlier findings indicating that restricted focused visual input decreases subsequent memory (Pertzov et al. 2009; Damiano and Walther 2019). More importantly, however, experiment 2 is the first to our knowledge to show a memory effect by repeating focused visual input that was otherwise held constant. Doubling the number of fixations in semantically important regions within a given exploration time increased subsequent free recall performance by 19%. Since guided viewing might per se interfere with memory encoding (Voss, Warren, et al. 2011; Voss, Gonsalves, et al. 2011), it remains to be investigated how this finding translates to free viewing. However, the evidence for causality and direction of the relationship between number of fixations and recall performance enabled further interpretation of the imaging results of experiment 1. Specifically, the findings are in line with the idea that visual sampling frequency triggers not only visual regions, but a larger brain circuitry relevant for memory processing, including the MTL. What caused individuals to scan scenes at varying sampling frequency in the first place has to be further examined. One explanation might be different levels of expertise. Several studies have associated expert knowledge to altered scanning of expertise-related objects or scenes. Interestingly, more fixations and more dwells in AOIs were specifically identified as key features in experts that could be responsible for their superior performance on perceptual-cognitive tasks (Gegenfurtner et al. 2011; Brams et al. 2019). Due to the variety of the

650 scenes used in both experiments, it seems implausible that some of the subjects had specific
651 expertise for the scenes presented. However, it has been argued that very short-term relational
652 memory signals, provided by the MTL, are needed for effective visual episodic memory formation,
653 for example to bind together important visual features over space and time (Liu et al. 2017; Voss
654 et al. 2017). As such, effective MTL signaling is likely to be not only a consequence of visual
655 scanning, but also its prerequisite (Ryan et al. 2020).
656 Our results have several important implications. First, they suggest the importance of measuring
657 eye movements in visual memory studies. A similar claim has already been put forward by Voss
658 et al. (2017), arguing that visual exploration systematically co-varies with cognitive variables of
659 interest such as attention, emotion and intentionality, thereby confounding their effect on memory
660 mechanisms. We add the notion that the frequency and location of fixations are cognitive
661 variables of interest by themselves. They vary across individuals, are associated with brain
662 activations in memory-related regions, affect episodic memory formation and should thus be
663 considered as an integrative part of memory processing.
664 Second, our findings may partly explain memory deficits in neuropsychiatric conditions and open
665 a new treatment approach. Both memory deficits and altered exploration patterns are often
666 observed in neuropsychiatric conditions, such as depression (Kellough et al. 2008; Elliott et al.
667 2011), dementia (Shakespeare et al. 2015), anxiety disorders (LeMoult and Joormann 2012),
668 autism spectrum disorders (Fedor et al. 2018), posttraumatic stress disorder (Armstrong et al.
669 2013; de Quervain et al. 2017) and schizophrenia (Williams et al. 2010). Patients with
670 schizophrenia, for example, have difficulties executing simple visual tasks like smooth pursuit
671 (O'Driscoll and Callahan 2008). They also show less fixations and visual exploration of
672 semantically complex pictures (Beedie 2011). Therefore, a training aimed at increasing fixations
673 during visual exploration might prove useful for enhancing memory in conditions of impaired
674 memory functions. In summary, our data indicate that higher fixation frequency improves visual
675 memory, a phenomenon with great therapeutic potential.

Acknowledgments

This work was supported by the Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, 4055 Basel, Switzerland and by the Swiss National Science Foundation (P0BSP1_168917 to B.F.). Correspondence to Bernhard Fehlmann, Spalenring 72, 4055 Basel, Switzerland, bernhard.fehlmann@unibas.ch.

Notes

Conflicts of Interest: None declared.

Data Availability

The data that support the findings of this study are available from the corresponding authors on request.

Code Availability

Custom code that supports the findings of this study is publicly available on OSF (https://osf.io/r9bxd/).

References

Armstrong T, Bilsky SA, Zhao M, Olatunji BO. 2013. Dwelling on potential threat cues: an eye movement marker for combat-related PTSD. Depress Anxiety. 30(5):497-502. doi:10.1002/da.22115.

Ashburner J. 2007. A fast diffeomorphic image registration algorithm. NeuroImage. 38(1):95-113. doi:10.1016/j.neuroimage.2007.07.007.

Bar M, Aminoff E, Schacter DL. 2008. Scenes Unseen: The Parahippocampal Cortex Intrinsically Subserves Contextual Associations, Not Scenes or Places Per Se. J Neurosci. 28(34):8539-8544. doi:10.1523/JNEUROSCI.0987-08.2008.

Beedie S. 2011. Atypical scanpaths in schizophrenia: Evidence of a trait- or state-dependent phenomenon? J Psychiatry Neurosci. 36(3):150-164. doi:10.1503/jpn.090169.

Boghen D, Troost BT, Daroff RB, Dell'Osso LF, Birkett JE. 1974. Velocity characteristics of normal human saccades. Invest Ophthalmol. 13(8):619-623.

Brams S, Ziv G, Levin O, Spitz J, Wagemans J, Williams AM, Helsen WF. 2019. The relationship between gaze behavior, expertise, and performance: A systematic review. Psychol Bull. 145(10):980-1027. doi:10.1037/bul0000207.

Bylinskii Z, Isola P, Bainbridge C, Torralba A, Oliva A. 2015. Intrinsic and extrinsic effects on image memorability. Vision Res. 116:165-178. doi:10.1016/j.visres.2015.03.005.

Ciollaro M, Wang D. 2016. MeanShift: Clustering via the Mean Shift Algorithm. https://CRAN.R-project.org/package=MeanShift.

Dalton MA, Zeidman P, McCormick C, Maguire EA. 2018. Differentiable Processing of Objects, Associations, and Scenes within the Hippocampus. J Neurosci. 38(38):8146-8159. doi:10.1523/JNEUROSCI.0263-18.2018.

Damiano C, Walther DB. 2019. Distinct roles of eye movements during memory encoding and retrieval. Cognition. 184:119-129. doi:10.1016/j.cognition.2018.12.014.

714 Diana RA, Yonelinas AP, Ranganath C. 2007. Imaging recollection and familiarity in the medial
715 temporal lobe: a three-component model. Trends Cogn Sci. 11(9):379-386.
716 doi:10.1016/j.tics.2007.08.001.

717 Dickerson BC, Eichenbaum H. 2010. The Episodic Memory System: Neurocircuitry and
718 Disorders. Neuropsychopharmacology. 35(1):86-104. doi:10.1038/npp.2009.126.

719 Elliott R, Zahn R, Deakin JFW, Anderson IM. 2011. Affective Cognition and its Disruption in Mood
720 Disorders. Neuropsychopharmacology. 36(1):153-182. doi:10.1038/npp.2010.77.

721 Fedor J, Lynn A, Foran W, DiCicco-Bloom J, Luna B, O'Hearn K. 2018. Patterns of fixation during
722 face recognition: Differences in autism across age. Autism. 22(7):866-880.
723 doi:10.1177/1362361317714989.

724 Gegenfurtner A, Lehtinen E, Säljö R. 2011. Expertise Differences in the Comprehension of
725 Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains. Educ
726 Psychol Rev. 23(4):523-552. doi:10.1007/s10648-011-9174-7.

727 Graham KS, Barense MD, Lee ACH. 2010. Going beyond LTM in the MTL: A synthesis of
728 neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory
729 and perception. Neuropsychologia. 48(4):831-853. doi:10.1016/j.neuropsychologia.2010.01.001.

730 Hannula DE. 2010. Worth a glance: using eye movements to investigate the cognitive
731 neuroscience of memory. Front Hum Neurosci. 4. doi:10.3389/fnhum.2010.00166. [accessed
732 2018 May 23]. http://journal.frontiersin.org/article/10.3389/fnhum.2010.00166/abstract.

733 Heck A, Fastenrath M, Ackermann S, Auschra B, Bickel H, Coynel D, Gschwind L, Jessen F,
734 Kaduszkiewicz H, Maier W, et al. 2014. Converging Genetic and Functional Brain Imaging
735 Evidence Links Neuronal Excitability to Working Memory, Psychiatric Disease, and Brain Activity.
736 Neuron. 81(5):1203-1213. doi:10.1016/j.neuron.2014.01.010.

737 Heisz JJ, Pottruff MM, Shore DI. 2013. Females Scan More Than Males: A Potential Mechanism
738 for Sex Differences in Recognition Memory. Psychol Sci. 24(7):1157-1163.
739 doi:10.1177/0956797612468281.

740 Henderson J. 2003. Human gaze control during real-world scene perception. Trends Cogn Sci.
741 7(11):498-504. doi:10.1016/j.tics.2003.09.006.

742 Henderson JM, Hayes TR. 2017. Meaning-based guidance of attention in scenes as revealed by
743 meaning maps. Nat Hum Behav. 1(10):743. doi:10.1038/s41562-017-0208-0.

744 Hodgetts CJ, Voets NL, Thomas AG, Clare S, Lawrence AD, Graham KS. 2017. Ultra-High-Field
745 fMRI Reveals a Role for the Subiculum in Scene Perceptual Discrimination. J Neurosci.
746 37(12):3150-3159. doi:10.1523/JNEUROSCI.3225-16.2017.

747 Holmqvist K, editor. 2011. Eye tracking: a comprehensive guide to methods and measures.
748 Oxford ; New York: Oxford University Press.

749 Ito HT, Zhang S-J, Witter MP, Moser EI, Moser M-B. 2015. A prefrontal-thalamo-hippocampal
750 circuit for goal-directed spatial navigation. Nature. 522(7554):50-55. doi:10.1038/nature14396.

751 Itti L, Koch C. 2001. Computational modelling of visual attention. Nat Rev Neurosci. 2(3):194-203.
752 doi:10.1038/35058500.

753 Jaeger BC, Edwards LJ, Das K, Sen PK. 2017. An R 2 statistic for fixed effects in the generalized
754 linear mixed model. J Appl Stat. 44(6):1086-1105. doi:10.1080/02664763.2016.1193725.

755 Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. 2012. FSL. NeuroImage.
756 62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015.

757 Kafkas A, Montaldi D. 2011. Recognition Memory Strength is Predicted by Pupillary Responses
758 at Encoding While Fixation Patterns Distinguish Recollection from Familiarity. Q J Exp Psychol.
759 64(10):1971-1989. doi:10.1080/17470218.2011.588335.

760 Kellough JL, Beevers CG, Ellis AJ, Wells TT. 2008. Time course of selective attention in clinically
761 depressed young adults: An eye tracking study. Behav Res Ther. 46(11):1238-1243.
762 doi:10.1016/j.brat.2008.07.004.

763 Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, Christensen GE, Collins
764 DL, Gee J, Hellier P, et al. 2009. Evaluation of 14 nonlinear deformation algorithms applied to

765 human brain MRI registration. NeuroImage. 46(3):786-802.
766 doi:10.1016/j.neuroimage.2008.12.037.

767 Kravitz DJ, Saleem KS, Baker CI, Mishkin M. 2011. A new neural framework for visuospatial
768 processing. Nat Rev Neurosci. 12(4):217-230. doi:10.1038/nrn3008.

769 Lancry-Dayan OC, Kupershmidt G, Pertzov Y. 2019. Been there, seen that, done that:
770 Modification of visual exploration across repeated exposures. J Vis. 19(12):2.
771 doi:10.1167/19.12.2.

772 Lang PJ, Bradley MM, Cuthbert BN. 2008. International Affective Picture System (IAPS):
773 Affective ratings of pictures and instruction manual.

774 LeMoult J, Joormann J. 2012. Attention and Memory Biases in Social Anxiety Disorder: The Role
775 of Comorbid Depression. Cogn Ther Res. 36(1):47-57. doi:10.1007/s10608-010-9322-2.

776 Liu Z-X, Shen K, Olsen RK, Ryan JD. 2017. Visual Sampling Predicts Hippocampal Activity. J
777 Neurosci. 37(3):599-609. doi:10.1523/JNEUROSCI.2610-16.2017.

778 Malsburg T von der. 2015. saccades: Detection of Fixations in Eye-Tracking Data.
779 https://CRAN.R-project.org/package=saccades.

780 McCabe DP, Geraci L, Boman JK, Sensenig AE, Rhodes MG. 2011. On the validity of
781 remember–know judgments: Evidence from think aloud protocols. Conscious Cogn. 20(4):1625-
782 1633. doi:10.1016/j.concog.2011.08.012.

783 Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R 2 from
784 generalized linear mixed-effects models. O'Hara RB, editor. Methods Ecol Evol. 4(2):133-142.
785 doi:10.1111/j.2041-210x.2012.00261.x.

786 Niu. 2012. Affective salience can reverse the effects of stimulus-driven salience on eye
787 movements in complex scenes. Front Psychol. doi:10.3389/fpsyg.2012.00336. [accessed 2018
788 May 31]. http://journal.frontiersin.org/article/10.3389/fpsyg.2012.00336/abstract.

O'Driscoll GA, Callahan BL. 2008. Smooth pursuit in schizophrenia: A meta-analytic review of research since 1993. Brain Cogn. 68(3):359-370. doi:10.1016/j.bandc.2008.08.023.

Olsen RK, Sebanayagam V, Lee Y, Moscovitch M, Grady CL, Rosenbaum RS, Ryan JD. 2016. The relationship between eye movements and subsequent recognition: Evidence from individual differences and amnesia. Cortex. 85:182-193. doi:10.1016/j.cortex.2016.10.007.

Pertzov Y, Avidan G, Zohary E. 2009. Accumulation of visual information across multiple fixations. J Vis. 9(10):2-2. doi:10.1167/9.10.2.

Pinheiro J, Bates D, R-core. 2019. nlme: Linear and Nonlinear Mixed Effects Models. https://CRAN.R-project.org/package=nlme.

de Quervain D, Schwabe L, Roozendaal B. 2017. Stress, glucocorticoids and memory: implications for treating fear-related disorders. Nat Rev Neurosci. 18(1):7-19. doi:10.1038/nrn.2016.155.

Ringo JL, Sobotka S, Diltz MD, Bunce CM. 1994. Eye movements modulate activity in hippocampal, parahippocampal, and inferotemporal neurons. J Neurophysiol. 71(3):1285-1288. doi:10.1152/jn.1994.71.3.1285.

Ross J, Morrone MC, Goldberg ME, Burr DC. 2001. Changes in visual perception at the time of saccades. Trends Neurosci. 24(2):113-121. doi:10.1016/S0166-2236(00)01685-4.

Ryan JD, Shen K, Liu Z-X. 2020. The intersection between the oculomotor and hippocampal memory systems: empirical developments and clinical implications. Ann N Y Acad Sci. 1464(1):115-141. doi:10.1111/nyas.14256.

Shakespeare TJ, Pertzov Y, Yong KXX, Nicholas J, Crutch SJ. 2015. Reduced modulation of scanpaths in response to task demands in posterior cortical atrophy. Neuropsychologia. 68:190-200. doi:10.1016/j.neuropsychologia.2015.01.020.

Sharot T, Davidson ML, Carson MM, Phelps EA. 2008. Eye Movements Predict Recollective Experience. Lauwereyns J, editor. PLoS ONE. 3(8):e2884. doi:10.1371/journal.pone.0002884.

814 Smith SW. 2003. Digital signal processing: a practical guide for engineers and scientists.
815 Amsterdam ; Boston: Newnes (Demystifying technology series).

816 Van Orden KF, Jung T-P, Makeig S. 2000. Combined eye activity measures accurately estimate
817 changes in sustained visual task performance. Biol Psychol. 52(3):221-240. doi:10.1016/S0301-
818 0511(99)00043-5.

819 Voss JL, Bridge DJ, Cohen NJ, Walker JA. 2017. A Closer Look at the Hippocampus and
820 Memory. Trends Cogn Sci. doi:10.1016/j.tics.2017.05.008. [accessed 2017 Jun 19].
821 http://linkinghub.elsevier.com/retrieve/pii/S1364661317301092.

822 Voss JL, Gonsalves BD, Federmeier KD, Tranel D, Cohen NJ. 2011. Hippocampal brain-network
823 coordination during volitional exploratory behavior enhances learning. Nat Neurosci. 14(1):115-
824 120. doi:10.1038/nn.2693.

825 Voss JL, Warren DE, Gonsalves BD, Federmeier KD, Tranel D, Cohen NJ. 2011. Spontaneous
826 revisitation during visual exploration as a link among strategic behavior, learning, and the
827 hippocampus. Proc Natl Acad Sci. 108(31):E402-E409. doi:10.1073/pnas.1100225108.

828 Wilcox RR. 2012. Modern statistics for the social and behavioral sciences: a practical
829 introduction. Boca Raton: Taylor & Francis.

830 Williams LE, Must A, Avery S, Woolard A, Woodward ND, Cohen NJ, Heckers S. 2010. Eye-
831 Movement Behavior Reveals Relational Memory Impairment in Schizophrenia. Biol Psychiatry.
832 68(7):617-624. doi:10.1016/j.biopsych.2010.05.035.

833 Wixted JT. 2009. Remember/Know judgments in cognitive neuroscience: An illustration of the
834 underrepresented point of view. Learn Mem. 16(7):406-412. doi:10.1101/lm.1312809.

835 Wolfe JM, Horowitz TS. 2017. Five factors that guide attention in visual search. Nat Hum Behav.
836 1(3):0058. doi:10.1038/s41562-017-0058.

Supplementary Materials and Methods Experiment 1

Recognition task

For the recognition task, an additional 72 pictures were taken from the same database and matched to the encoding set in terms of valence category and semantic content (24 pictures per category). This resulted in a total of 144 pictures for the recognition task, half of which were old (i.e. presented during the encoding task), the other half being new (i.e. not presented before). Each of the 144 recognition trials started with a fixation cross, presented for 500 ms against a dark background, and was followed by the presentation of one picture for 1 s. The stimulus onset time was jittered within 3 s (1 TR) per valence and old/new category with regard to the scan onset. A blank, dark screen followed the offset of the picture for 1 s. Afterwards, participants subjectively rated the picture as remembered, familiar, or new by button press. Picture rating was possible in a time window of 3 s. Across all trials, pictures were presented in a quasi-randomized order, allowing for a maximum of 4 consecutive pictures with identical valence categories. Recognition performance was assessed by the difference between subsequently remembered and subsequently not remembered pictures (Luksys et al. 2015).

We assessed the relationship between recognition performance and eye tracking parameters as well as the fMRI signal associated with the number of fixations at encoding. Therefore, the models specified for the free recall were used (see Materials and Methods Experiment 1), but with the recognition performance as the dependent variable. Results are summarized in the Supplementary Tables S1 and S3.

Construction of ET-AOIs

Fixations of 200 subjects were randomly chosen per picture. All fixations are iteratively and simultaneously moved towards locations of higher spatial density until points of eventual convergence. The points represent the modes of the distribution and indicate the distinct clusters identified. In order to build AOIs, large clusters (containing a minimum of 2.5% of the fixations) were given a parametrized representation as covariance ellipses. The ellipses are centered at the cluster mean (centroid) and represent 50% of the spatial variance of the original cluster. The procedure is based on previous work (Santella and DeCarlo 2004), with a modified threshold for spatial variance to prevent overlapping clusters. To estimate the reliability of this approach, the

same procedure was then repeated 100 times per picture. For each iteration, the spatial overlap of the resulting AOIs with the initial solution was calculated by the Sørensen-Dice coefficient, ranging from 0 to 1 for non-overlapping and perfectly identical AOIs, respectively (Dice 1945). The mean across all 72 pictures and 100 repetitions was .93 (*SD* = .03), suggesting a high reliability of the AOI estimation.

Population-average anatomical probabilistic atlas

The T1-weighted image of each subject of the 1000 subjects included in the construction of the study-specific template was automatically segmented into cortical and subcortical structures using FreeSurfer (v4.5, RRID:SCR_001847; http://surfer.nmr.mgh.harvard.edu/, Fischl et al. 2002). On the basis of the Desikan-Killiany-atlas (Desikan et al. 2006), 35 cortical gyri were labeled, as well as 17 subcortical regions (see Fischl et al. 2002). The segmented T1 images were then normalized to the study-specific anatomical template. Finally, the normalized segmentations were averaged across subjects, resulting in a population-average probabilistic atlas. Every voxel of the template could consequently be assigned a probability of belonging to a given anatomical structure.

Grand average pupil profile

The grand average pupil profile is based on 800 subjects (482 females; mean age = 22.34, *SD* = 3.38, range 18–35) with eye tracking data that were not identified as outliers for calibration data or the eye movement velocity distribution. Eye-blink related artifacts were replaced by linear interpolation throughout the dataset. Pupil data were then smoothed using a five-point unweighted average filter applied twice (Siegle et al. 2003) and segmented per trial. The segmentation window spanned a total of 3 seconds, including the 500 ms before picture onset that served as a baseline. For each trial with a minimum of 2 fixations during picture presentation, pupil data (recorded in arbitrary units) were baseline-corrected. The correction value was computed as a one-step M-estimator of location of the corresponding 30 pupil height samples ('WRS2' package, Mair and Wilcox 2019). This procedure is a proposed alternative to mean averaging that is more robust against outliers (Wilcox 2012). Valid trials, each consisting of a time series of 150 corrected pupil height samples that describe the pupil dynamics during the 2.5 s of picture presentation, formed the basis for further aggregation. This was done in two steps, again

using the one-step estimator. First, for each subject and time point separately, trials were averaged across valence categories. Second, the resulting pupil profiles were averaged across subjects, resulting in the grand average pupil profile. The pupil profile of each trial was separately correlated with this grand average. A low correlation led to trial exclusion, with the cutoff-threshold ($r < .75$) based on boxplots and ideal fourths used for quartile estimation (Wilcox 2012).

Supplementary Materials and Methods Experiment 2

Recognition task

The recognition task is almost identical to the one used in experiment 1, with the exception of the stimulus onset time not being jittered. The 54 pictures representing the semantically matched counterparts from experiment 1 were included, resulting in a total of 108 pictures, half of which were presented during the encoding task.

We assessed the influence of the three experimental conditions on recognition performance. Therefore, the same model specified for the free recall was used (see Materials and Methods Experiment 2), but with the recognition performance as the dependent variable. One subject did not fully complete the recognition task, leading to 63 subjects being considered for analyses (31 females; mean age = 23.19, *SD* = 3.90, range 18–32).

Supplementary Results

Effects of fixation frequency and location on recognition performance, Experiment 2

Regarding passive recognition performance, there was a positive main effect of the factor 'experimental condition' ($F(2,124)= 26.54$, $p = 2.5e\text{-}10$). Post-hoc tests revealed an average decrease of recognized pictures in the 'Area of no Interest' condition compared to the 'Guided Fixation' condition by 16% ($t(62)= -6.13$, $p = 1.4e\text{-}07$. $R^2\beta^* = .074$, 95% CI [.011, .180]). There was no evidence for a higher amount of recognized pictures in the 'Guided Fixation × 2' condition compared to the 'Guided Fixation' condition ($t(62)= -.18$, $p = .85$, $R^2\beta^* = .000$, 95% CI [.000, .040]) (see Supplementary Fig. S2).

Supplementary Discussion

In Experiment 1 the number of fixations in semantically informative picture areas were not only positively correlated to free recall, but also to recognition performance (see Supplementary Table S3). In experiment 2 we found that manipulating the scan path again affects both types of memories. However, a dissociation of the effects of fixation location and frequency was revealed. Both for free recall and recognition performance, it was beneficial to sample from semantically informative regions of the pictures during encoding. Only for free recall performance, however, it was beneficial if informative regions were sampled twice within a given time. A higher sampling frequency might promote associations between different regions of a picture (Wolfe and Horowitz 2017). We argue that free recall of episodic memories is likely to depend more on the association of visual areas of a picture than recognition, which can be achieved solely based on familiarity of isolated semantic areas of an image without requiring associations between them (Heisz et al. 2013). The finding of a positive correlation between fixation frequency and recognition memory in experiment 1 might be due to the high correlation of fixation frequency and number of semantic regions visited, with the latter being the actual driving factor.

Supplementary References

Boghen D, Troost BT, Daroff RB, Dell'Osso LF, Birkett JE. 1974. Velocity characteristics of normal human saccades. Invest Ophthalmol. 13(8):619-623.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, et al. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage. 31(3):968-980. doi:10.1016/j.neuroimage.2006.01.021.

Dice LR. 1945. Measures of the Amount of Ecologic Association Between Species. Ecology. 26(3):297-302. doi:10.2307/1932409.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, et al. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 33(3):341-355.

Heisz JJ, Pottruff MM, Shore DI. 2013. Females Scan More Than Males: A Potential Mechanism for Sex Differences in Recognition Memory. Psychol Sci. 24(7):1157-1163. doi:10.1177/0956797612468281.

Luksys G, Fastenrath M, Coynel D, Freytag V, Gschwind L, Heck A, Jessen F, Maier W, Milnik A, Riedel-Heller SG, et al. 2015. Computational dissection of human episodic memory reveals mental process-specific genetic profiles. Proc Natl Acad Sci. 112(35):E4939-E4948. doi:10.1073/pnas.1500860112.

Mair P, Wilcox R. 2019. WRS2: A Collection of Robust Statistical Methods. https://CRAN.R-project.org/package=WRS2.

Santella A, DeCarlo D. 2004. Robust clustering of eye movement recordings for quantification of visual interest. ACM Press. p. 27-34.

958 Siegle GJ, Steinhauer SR, Stenger VA, Konecky R, Carter CS. 2003. Use of concurrent pupil
959 dilation assessment to inform interpretation and analysis of fMRI data. NeuroImage. 20(1):114-
960 124. doi:10.1016/S1053-8119(03)00298-2.

961 Wilcox RR. 2012. Modern statistics for the social and behavioral sciences: a practical
962 introduction. Boca Raton: Taylor & Francis.

963 Wolfe JM, Horowitz TS. 2017. Five factors that guide attention in visual search. Nat Hum Behav.
964 1(3):0058. doi:10.1038/s41562-017-0058.

Supplementary Tables

Table S1.

Positive and negative modulation of activation by the number of fixations in AOIs during the encoding task in experiment 1.

	Cluster region names	Peak voxel region name	Cluster		Voxel				EFA	ERA
			P_{corr}	n	T_{peak}	MNI_{peak}			%n [k]	%n [k]
						X	Y	Z		
Positive Modulation	Precuneus (L) Lingual gyrus (L) Isthmus cingulum (L) Precuneus (R)	Pericalcarine gyrus (L)	< .001	995	13.40	-11	-82.5	4	28.6 [285]	19.7 [196]
	Superior frontal cortex (L) Rostral anterior cingulum (L) Medial orbitofrontal cortex (L) Superior frontal cortex (R) Caudal anterior cingulum (R)	Medial orbitofrontal cortex (L)	< .001	643	8.09	-2.75	55	-4	8.6 [81]	33.6 [216]
	Inferior parietal cortex (L)	Inferior parietal cortex (L)	< .001	331	8.73	-46.8	-77	32	9.7 [11]	0.0 0
	Parahippocampus (L) Fusiform gyrus (L) Lingual gyrus (L)	Parahippocampus (L)	< .001	97	8.56	-22	-38.5	-12	28.9 [28]	13.4 [13]
	Superior frontal cortex (L) Rostral middle frontal cortex (L)	Superior frontal cortex (L)	< .001	55	6.45	-19.2	33	44	50.9 [28]	21.8 [12]
	Cerebellar cortex (R)	Cerebellar cortex (R)	.001	43	5.32	11	-85.2	-36	79.1 [34]	9.3 [4]
	Thalamus (L) Thalamus(R)	Thalamus (L)	.006	25	5.78	0	-5.5	8	4.0 [1]	0.0 [0]
	Parahippocampus (R)	Parahippocampus (R)	.046	13	5.75	22	-33	-16	7.7 [1]	7.7 [1]

Negative Modulation	Cuneus (R) Superior parietal cortex (R) Lingual gyrus (R) Cuneus (L) Superior parietal cortex (L)	Cuneus (R)	< .001	964	15.6	5.5	-82.5	28	0.0 [0]	0.0 [0]
	Superior parietal cortex (L) Supramarginal gyrus (L) Postcentral gyrus (L)	Superior parietal cortex (L)	< .001	524	8.30	-22	-49.5	68	0.0 [0]	0.0 [0]
	Lateral occipital cortex (L) Inferior temporal cortex (L) Middle temporal cortex (L)	Inferior temporal cortex (L)	< .001	123	6.51	-49.5	-66	0	0.0 [0]	0.0 [0]
	Precentral gyrus (L) Superior frontal cortex (L) Caudal middle frontal cortex (L)	Superior frontal cortex (L)	< .001	108	7.34	-24.8	-5.5	56	0.0 [0]	0.0 [0]
	Superior parietal cortex (R)	Superior parietal cortex (R)	< .001	88	6.66	24.8	-49.5	64	0.0 [0]	0.0 [0]

Clusters with a voxel threshold of > 10 lying > 60% outside of cerebral white matter are reported. Cluster region name: Anatomical regions contributing a minimum of 5% to a given cluster are listed, ordered by the magnitude of the contribution. Regions are in accordance with the in-house atlas; Peak voxel region name: highest value in the cluster that is not lying in cerebral white matter is indicated; P_{corr} represents the whole-brain *FDR*-corrected cluster *p* value, n the number of voxels in the cluster, T_{peak} the *t* value of the peak voxel and [x,y,z] its coordinates in the MNI-space. EFA: Encoding-free-recall-association. Indicating in relative (%*n*) and absolute terms *[k]* how many voxels of a given cluster show a positive association between the encoding signal and free recall performance. ERA: Encoding-recognition-association. Indicating in relative (%*n*) and absolute terms *[k]* how many voxels of a given cluster show a positive association between the encoding signal and passive recognition performance.

Table S2.

Intercorrelations of z-standardized eye tracking parameters in experiment 1.

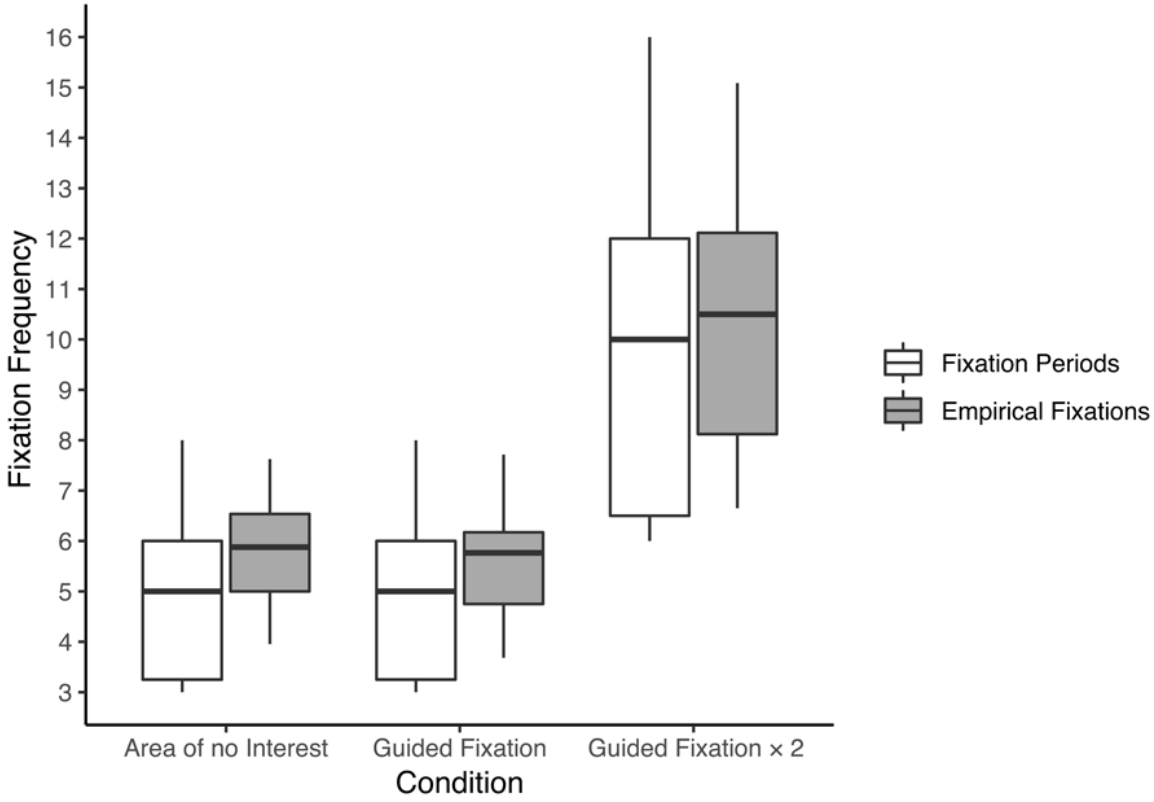
	N_{fix}	N_{fix} in AOIs	N_{AOIs}	Blink duration	Interfixation distance
N_{fix}	1				
N_{fix} in AOIs	.82	1			
N_{AOIs}	.73	.91	1		
Blink duration	-.22	-.29	-.30	1	
Interfixation distance	-.21	-.34	-.23	.10	1

Table S3.

Models regressing eye tracking parameters on recognition performance in experiment 1.

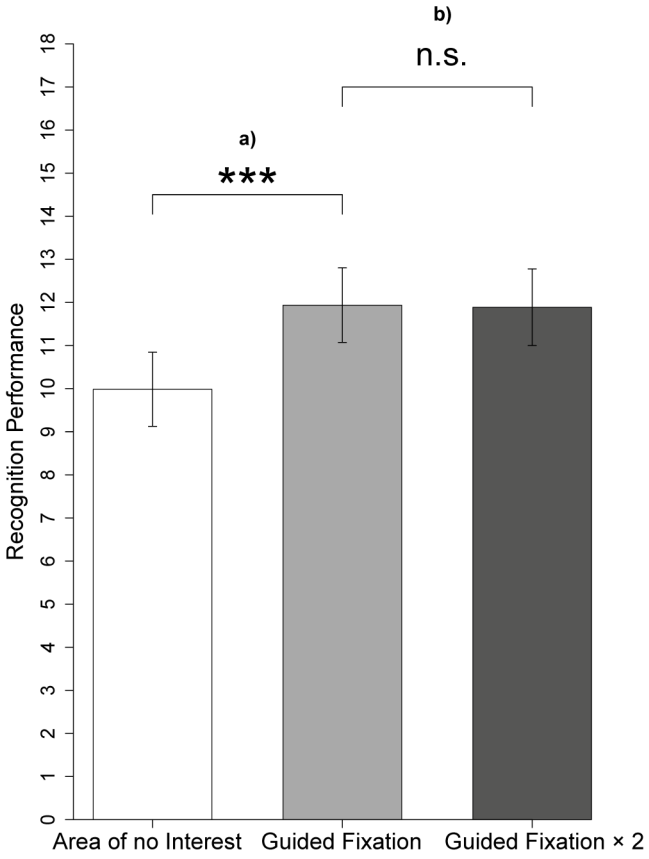
	Eye tracking parameter	Main effect			Valence interaction					
		t (*df*)	*p*	$R^2\beta$* 95% CI []	*F* (*df*)	*p*	Posthoc-test	*t* (*df*)	*p*	R^2
Passive recognition	N_{fix}	4.50 (1339)	2.9e-05 ***	.014 [.006, .026]	3.16 (2,1339)	.06				
	N_{fix} in AOIs	3.41 (1322)	.001	.004 [.000, .011]	5.67 (2,1322)	.014 *	negative	3.38 (676)	.001	.017
							neutral	3.47 (675)	.001	.018
							positive	1.64 (658)	.10	.004
	Blink duration	-0.57 (1291)	.57	.000 [.000, .006]	0.48 (2,1291)	.62				
	Interfixation distance	-1.72 (1337)	.11	.000 [.000, .005]	4.73 (2, 1337)	.018 *	negative	-1.33 (679)	.27	.003
							neutral	-0.12 (680)	.91	.000
							positive	1.59 (666)	.27	.004

Supplementary Figures



Supplementary Fig. 1: Comparison between fixation periods and empirical fixations

We compared the fixation periods, given by the moving circle, and the empirically measured fixation frequency for each of the 54 pictures in 64 subjects. The figure shows the general level of empirical fixations to be slightly higher, suggesting that one fixation period usually is accompanied by more than one fixation. While the frequency of fixations was expected to be doubled in the 'Guided Fixation × 2' compared to the 'Area of no Interest' and the 'Guided Fixation' condition, the actual difference was slightly lower. The average increase in the 'Guided Fixation × 2' condition is 1.75 (*SD* = 0.18) compared to the 'Area of no Interest' condition and 1.82 (*SD* = 0.16) compared to the 'Guided Fixation' condition.



Supplementary Fig. 2: Effect of scan path manipulation on recognition performance

Recognition memory effect in 63 subjects by a) decreasing the number of AOIs covered by fixations, leading to a lower amount of freely recalled pictures in the 'Area of no Interest' condition (*M* = 9.98, *SE* = 0.86) compared to the 'Guided Fixation' condition (*M* = 11.94, *SE* = 0.87) but not by b) increasing fixation frequency, leading to no significant differences in the 'Guided Fixation × 2' condition (*M* = 11.89, *SE* = 0.89) compared to the 'Guided Fixation' condition.

4.2 Effectiveness of a Virtual Reality-Based Eye Contact Training to Reduce Fear of Public Speaking: A Randomized Controlled Trial

Effectiveness of a virtual reality-based mutual gaze training to reduce fear of public speaking: a randomised controlled trial

Bernhard Fehlmann, Fabian Müller, Nan Wang, Merle K Ibach, Thomas Schlitt, Dorothée Bentz, Anja Zimmer, Andreas Papassotiropoulos & Dominique JF de Quervain

Division of Cognitive Neuroscience, Department of Psychology (B Fehlmann MA, F Müller BA, N Wang PhD, MK Ibach MA, T Schlitt PhD, D Bentz PhD, A Zimmer MA, Prof DJ de Quervain MD) **and Division of Molecular Neuroscience**, **Department of Psychology, and Life Sciences Training Facility, Department Biozentrum** (Prof A Papassotiropoulos MD)**, and University Psychiatric Clinics** (Prof A Papassotiropoulos MD, Prof DJ de Quervain MD), **and Transfaculty Research Platform** (B Fehlmann MA, F Müller BA, N Wang PhD, MK Ibach MA, T Schlitt PhD, D Bentz PhD, A Zimmer MA, Prof A Papassotiropoulos MD, Prof DJ de Quervain MD), **University of Basel, CH**

Correspondence to:

Prof Dominique de Quervain

Division of Cognitive Neuroscience

University of Basel, CH

dominique.dequervain@unibas.ch

+41 61 207 02 37

Summary (250 words)

Background

Public speaking anxiety (PSA) is the most widespread social fear, with prevalence estimates up to 30%. Anxious individuals typically avoid mutual gaze and report fears of being evaluated by others. We developed a virtual reality (VR) training application (app) aimed at increasing mutual gaze and tested its effectiveness to reduce fear of public speaking.

Methods

We used a single-blind, parallel-group, randomised controlled trial design. Between June 18, 2019 and September 19, 2019, 89 participants aged 18–40 years with PSA were randomly assigned to a gaze training (treatment) or a control group. During training, participants were exposed to social situations and instructed to maintain eye contact with people in the VR. Assessments were done at baseline, after a first intervention (acute intervention on a single day: three gaze training sessions of 20 min; control: *Google Street View*), and after a second intervention phase (repeated intervention over two weeks: nine gaze training sessions of 20 min; control: no further intervention) two months later. The primary outcome measure was subjective fear, rated on the Subjective Units of Distress Scale (SUDS) in a real-life public speaking test (PST). This trial was pre-registered at ClinicalTrials.gov (NCT03970187).

Findings

Repeated, but not acute, gaze training led to a significant reduction of subjective fear as compared to the control group (treatment group, baseline: 48·95 [SD 19·52], post intervention 2: 26·60 [SD 19·23]; control group, baseline: 48·92 [SD 18·43], post intervention 2: 56·34 [SD 28·15]; adjusted mean group difference: −29·82, 95% CI: −41·77 to −17·87; Cohen's d=−1·07, p<0·0001).

Interpretation

Repeated usage of our stand-alone, VR gaze training app reduces the subjectively perceived fear in individuals with PSA in real-life public speaking situations.

Funding

Transfaculty Research Platform, University of Basel; Swiss National Science Foundation (SNSF).

Research in Context (Box)

Evidence before this study

In January 2019 and in May 2020, we searched PubMed with no restrictions on date for all published work in English on stand-alone virtual reality (VR) applications to treat fear of public speaking using the search terms ('virtual reality') AND ('treatment' OR 'therapy' OR 'exposure') AND ('public speaking' OR 'public anxiety' OR 'social phobia' OR 'social fear' OR 'social anxiety') AND ('stand-alone' OR 'self-guided' OR 'self-training' OR 'automated').

We only retrieved one empirical study from 2016 by Kampmann and colleagues. The authors reported that a stand-alone, desktop-based VR exposure treatment using social situations with avatars reduces perceived stress in individuals with social anxiety disorder, with a medium effect size using a questionnaire-based assessment.

Added value of this study

We show that repeated use of our smartphone-based stand-alone VR app leads to a profound reduction in the fear of public speaking and a decrease of mutual gaze avoidance in a real-life speech situation.

Our study is the first to target mutual gaze with real people, projected to the VR as dynamic stimuli (i.e. movie clips). Changing the size, distance and facial expression of the virtual audience allowed us to adaptively modulate the perceived stress in the VR and to follow a gradual exposure scheme. Results show that mutual gaze training is sufficient to alleviate symptoms of PSA, without requiring verbal interactions. Our app does not rely on additional cognitive behavioural elements, assistance or therapeutic input. For sufferers of public speaking anxiety (PSA), the effectiveness of our 12 exposure sessions of approximately 20 min (total duration of 4 h) was comparable to real-life exposure therapy, the current gold standard to treat social anxiety.

Implications of all the available evidence

Stand-alone, smartphone-based VR exposure apps provide a cost-efficient and fully controllable treatment option for anxiety, together with a low threshold for initiation. This is particularly important in the context of PSA, where the fear of being evaluated by others is inherent to the therapeutic setting. Without requiring external input, such apps can serve as widely accessible training tools, countering the dissemination problem of traditional in-vivo treatment. In the clinical setting, stand-alone apps can provide a valuable add-on to guided standard exposure therapy – for example by making VR homework assessments more feasible.

Main article

Introduction

Disproportionate social fear reactions can be triggered by a wide range of situations with the potential of evaluation by others. The most frequently reported fears relate to public speaking, with prevalence estimates ranging from 6.5% to 30%.[1–5] Fear of public speaking is associated with impairments in daily life, including an increased need for medical care, decreased educational success, lower income and impaired personal relationships. Furthermore, socially anxious individuals suffer from a high degree of comorbidity with psychiatric disorders, such as the often more generalized and clinically relevant form of social anxiety disorder (SAD),[4] depression or substance abuse.[6]

Along with the extensive fear of evaluation, socially anxious individuals typically show altered gaze behaviour.[7] Early vigilance towards the eye region and subsequent avoidance of mutual gaze in social situations are thought to be submissive gestures to reduce the anticipated social threat.[8–11] Eye tracking studies indicate that socially anxious individuals exhibit fewer total fixations and total dwell time in the eye regions of faces in response to both negative and positive expressions,[11,12] suggesting that averting gaze due to fear of social evaluation is valence independent.[13,14] A recent study replicated and extended these findings in a computerized social simulation, specifying gaze avoidance as a biobehavioural marker of SAD.[15] It is currently unknown, however, if gaze avoidance is merely a sign of social anxiety or an essential risk and maintenance factor of it, which could be targeted by behavioural training in order to reduce fear.[12,16]

In the clinical context, options to treat SAD include medication, psychotherapy or the combination of both. While the efficacy of in vivo exposure therapy is well documented,[17] further evidence suggests comparable efficacy of virtual reality exposure therapy (VRET),[18,19] at increased acceptance levels in patients.[17] In the current study, we therefore investigated the potential of a virtual reality (VR)-based mutual gaze training to reduce fear of public speaking in individuals with public speaking anxiety (PSA), a symptom of SAD. Our study sets itself apart from studies investigating the efficacy of VRET in the context of SAD in the following aspects: 1) we used a mutual gaze training to increase face gaze, 2) we implemented movies of real people in VR rather than avatars, 3) we used a stand-alone smartphone application without relying on cognitive behavioural elements, assistance or therapeutic input and 4) we tested the effectiveness in a real-life public speaking situation.

The primary outcome measure is subjectively perceived fear in a real-life public speaking test (PST). Secondary outcome measures include the relative dwell time on faces as well as external assessments of the speech performance.

Methods

Study design and participants

In a single-blind, parallel-group, randomised controlled trial, we investigated the effectiveness of a smartphone-based VR app in reducing the fear of public speaking, compared to a VR condition without social exposure during intervention phase 1 and to no intervention in intervention phase 2. We recruited from the German speaking general population of Switzerland via print and online advertisements. We included physically healthy individuals, aged between 18 and 40 years that were fluent in German and indicated high fear in social situations with the potential of being evaluated by others. In addition, they had to affirm to endure those situations under high fear and/or to avoid them (for a detailed description of exclusion criteria see appendix). All participants gave written informed consent for trial participation. Participants received a compensation of CHF 25/h for their participation and CHF 50 for the successful completion of the study. The study protocol, including the definition of primary and secondary outcomes and statistical analysis plan, has been approved by the Independent Ethics Committee (IEC; Ethics Committee of North-West and Central Switzerland) before the start of the study (i.e. June 18, 2019, first subject in). There were no deviations from the protocol after trial start. The trial has been registered at ClinicalTrials.gov before start of the study with the identifier: NCT03970187 on May 31, 2019.

Randomisation and masking

Participants were randomly (matched for sex) allocated to the treatment group (VR gaze training in intervention phase 1 and 2) or the control group (a fear-unrelated VR task in intervention phase 1, no intervention in intervention phase 2). Each eligible participant was allocated to one of two randomisation lists (male/female). Within each list, groups were randomised according to a maximum tolerated imbalance (MTI-) procedure implemented in R[20] ('RandomizeR'-package;[21] MTI across lists=4). All experimenters who collected outcome measures in the real-life public speaking test were unaware of the group assignment of participants (single-blind).

Procedures

Inclusion and exclusion criteria were initially checked via an online questionnaire that was sent to all individuals interested in the study. Eligible individuals were scheduled for two visits at the Division of Cognitive Neuroscience at the University of Basel, Switzerland.

Before study enrolment, we rechecked inclusion and exclusion criteria and collected basic demographic data. If the enrolment criteria were still being met, participants were asked to fill in questionnaires to evaluate their fear of public speaking (for detailed information on implemented questionnaires and tests, see 'Outcomes' and

appendix) and were allocated to one of the two intervention groups. We then collected salivary cortisol with a saliva-sampling device (SalivetteSarstedt, Rommelsdorf, Germany).

Immediately afterwards, the baseline PST was conducted, representing real-life exposure to a socially threatening situation. The task was to give three short speeches in front of a committee, consisting of three experimenters that were trained to maintain neutral facial expression, eye contact and body posture. Participants were requested to choose three out of five predefined general topics and to prepare the speeches for 10 min. Before presenting the first topic, participants were asked to rate their general fear – our primary outcome measure – as well as their fear to hold eye contact on the Subjective Units of Distress Scale (SUDS). The maximum duration of each given topic was 3 min. As soon as the participant indicated feeling too uncomfortable to proceed or after the 9 min period (3 × 3 min) was over, the PST was terminated. After each speech, the participants were again asked to rate their general fear as well as their fear to hold eye contact using the SUDS. In addition, participants and committee members were asked to independently and covertly rate speech quality. To that end, visual analogue scales (VAS) were used with respect to the global performance as well as eight specific subscales covering separate aspects of the performance. Speeches were video and audio recorded, and the eyes were tracked using a mobile eye tracking system (see appendix).

Following the PST, participants provided a second cortisol sample, filled out the Simulator Sickness Questionnaire (SSQ)[22] and started the VR intervention.

For the exposure training in public speaking situations, we used a VR gaze training app for smartphones that we developed at the University of Basel. The app requires no assistance (stand-alone). It consists of an initial description of the app content in 2D and three subsequent VR scenarios, comprising of different audiences (i.e. close proximity, classroom, lecture hall scenario). At the beginning of each scenario, users are standing in an empty room (level 0) without time limit. Afterwards, the room is filled with a virtual audience (level 1–6). Users are asked to briefly introduce themselves, which is played back via the microphone of the smartphone, and then to maintain eye contact with specific audience members during 3 min. Individuals of the virtual audience direct their gaze towards the user of the VR app in all levels. Advancing to further levels follows a predefined exposure scheme based on VAS-ratings indicating the perceived fear at a given level (0: no fear, 10: maximum fear) as well as the maintenance of mutual gaze with the virtual audience, measured by gaze tracking. Each level is repeated until the VAS-rating indicates low subjective fear (≤3) and the face gaze maintained exceeds a predefined time threshold. VAS-ratings are provided by gaze selection throughout the three exposure sessions. Each scenario contains six levels with a net duration of 3 min each (i.e. without introductory slides and VAS-ratings). The levels are gradually increasing in difficulty, which is operationalized by the emotional valence of

the audience's facial expression (level 1–2: positive, level 3–4: neutral, level 5–6: negative), the size of the audience (close proximity scenario: 2–12, classroom scenario: 2–21, lecture hall scenario: 2–100) as well as the time required to keep mutual gaze (see figure 1 and appendix table 3). Each scenario ends after 20 min, irrespective of the achieved level. The training is based on a behavioural exposure approach and includes no psycho-educative elements or specific cognitive interventions (e.g. challenging of cognitive distortions).[23] The exposure paradigm contains a minimal number of gamified reinforcement elements (i.e. a clapping sound after successful level completion).

Participants of both groups received smartphones with a noise-cancelling headset, a head mounted display (HMD) to enable stereoscopic view and a controller for the HMD. For the treatment group, the gaze training app was pre-installed, while participants of the control group received devices with Google Street View (v2.0.0.25751656). The task given to the control group was to explore three pre-defined virtual scenarios in VR (Nautilus House, Mexico; Montpelier, Orange County; Wethersfield, Connecticut) for 20 min each and to answer several basic questions (e.g., 'describe the weather in the scenario'). The three scenarios were selected because they did not contain any stimuli of potential social threat. Participants were not prompted to indicate their fear level and were allowed to explore each scenario at their own pace by teleporting themselves using the controller.

Participants in both groups were instructed to stand still while in VR. Each scenario was followed by a 5 min break, and the first training session ended after completion of all three scenarios.

After completion of the VR training session, symptoms of cybersickness were reassessed by the SSQ, followed by the Igroup Presence Questionnaire (IPQ)[24] assessing presence in VR. Subsequently, participants completed a second PST following a procedure identical to the first, but with five new topics to choose from. At the end of study phase 1, participants rated the VR app acceptability and usability by questionnaire, as well as the subjectively perceived improvement regarding their fear, eye contact and speech performance. Participants of the treatment group then received smartphones with detailed instructions about the home training.

For the training at home, the treatment group was requested to complete 9×20 min sessions with the gaze training app (three sessions per scenario, irrespective of achieved level), while the control group did not receive any task. Each training session was following the procedure of the trainings on site. The training was requested to start one to five weeks after study phase 1, spanning over a maximum duration of two weeks. Participants were allowed to freely schedule the nine sessions, but to only train once a day (i.e. 20 min). The completion of a minimum of two-thirds of the planned VR home training sessions (with regard to the amount of sessions and the completeness of single sessions) was defined per protocol as mandatory to be further included in the study. The

second visit took place approximately two months after study phase 1 (mean number of days, treatment condition: 57·46 [SD 3·46], control condition: 56·55 [SD 6·41]) and one month after training completion of the treatment group (mean number of days: 34·00 [SD 4·44]). We again checked inclusion and exclusion criteria and reassessed participants' fear of public speaking by questionnaires (see appendix). A third PST was then conducted, accompanied by cortisol sampling shortly before and afterwards. Subsequently, participants indicated their subjectively perceived improvement regarding fear, eye contact and speech performance as well as the amount of self-exposure to social situations between the two visits. The treatment group additionally filled in the VR app acceptability and usability scale and a questionnaire concerning their feeling of immersion.[25]

Figure 1: **Examples of the lecture hall scenario used in the gaze training app**
Top left: Introductory level, used for displaying general instructions about the task and for the subjective fear ratings between levels. *Top right*: Lecture hall, level 1; 4 people with positive emotional expression. *Bottom left*: Lecture hall, level 3; 16 people with neutral emotional expression. *Bottom right:* Lecture hall, level 6; 100 people with negative emotional expression; the blue circles indicate the location and order of the targets with whom the participants had to keep mutual face gaze (not shown in the app). The green arrow indicated the current target. As soon as a participant gazed towards the correct target area, the green area disappeared and an invisible timer was started. After a predefined time threshold, the arrow switched to the next target to prompt a change of face gaze (randomised order indicated by the white numbers). Whenever a participant exited the target area before the predefined time threshold, the timer stopped (but was not set back) and the green arrow reappeared above the current target.

Outcomes

The primary outcome measure was subjective fear during the PST (SUDS; 0: no fear, 100: maximum fear), averaged across four time points (i.e. before the first speech and after each speech, respectively).

The main secondary outcomes were (1) the average relative dwell time on faces of any of the committee members during the PST, assessed by eye tracking as an indicator of fear-related gaze behaviour (see appendix) and (2) the global external assessment of performance in the PST (VAS-ratings; 0: very bad, 100: very good), assessed three times (i.e. after each speech) by each of the three committee members and averaged across time points and committee members. Measures were taken at baseline, as well as after the first and after the second intervention phase (figure 2). All further outcome measures are described in the appendix, together with their statistical analysis.

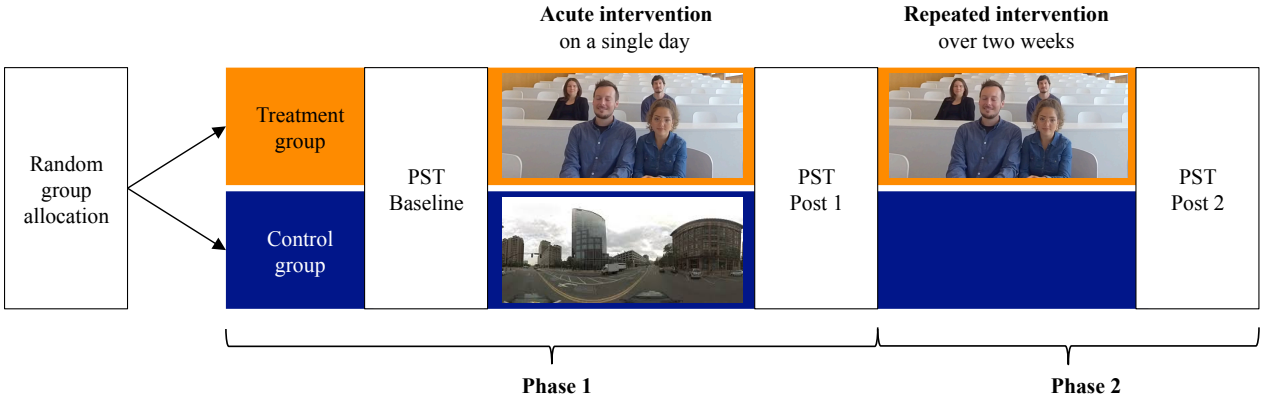


Figure 2: **Schematic study procedure**
Left to right: Individuals were randomly allocated to either the treatment or the control group. Afterwards, the baseline public speaking test (PST) was conducted, which comprised a preparation time of 10 min and three speeches with a maximum duration of 3 min each. Before the first and after every speech, participants were asked to indicate their general fear, which was the primary outcome measure. During the speeches, eye tracking allowed us to quantify the relative dwell time on faces of any of the committee members during the PST as a main secondary outcome measure. We asked the three members of the committee rate the global speech performance after each speech, which was another main secondary outcome measure. After the first PST, the participants of the treatment group underwent the mutual gaze training in the VR app for 3×20 min, while the participants of the control group explored virtual scenarios without social threat for the same amount of time (acute intervention). The second PST followed a procedure identical to the first. In study phase 2, the treatment group was requested to complete 9×20 min home training sessions with the gaze training app, while the control group did not receive any task (repeated intervention). Approximately two months after study phase 1 and one month after training completion of the treatment group, a third PST following the same procedure was conducted. This enabled us to compare the two groups at baseline, after the first intervention phase and after the second intervention phase.

Statistical analyses

We performed a per-protocol analysis, done in R (v3.6.1; RRID:SCR_001905; R Core Team, 2019; http://www.r-project.org/) and applied linear mixed models ('nlme' package)[26] in combination with ANOVA (SS II).

The Participant-ID was included as the random effect in the mixed models. To test our main hypothesis, the SUDS fear rating during the PST was included as the dependent variable. Group (treatment vs. control group; between-subject factor), time point (t0: baseline, t1: after intervention phase 1, t2: after intervention phase 2; within-subject factor) as well as their interaction were included as independent variables. In case of significant interactions, post-hoc tests were applied to further characterize the effects of group. Age and sex were included as covariates. To account for baseline differences, we also included the measurement of the SUDS fear rating at baseline testing in case of post-hoc analyses. Potential significant interactions of covariates with either of the independent variables were further described by post-hoc tests. Non-significant interactions were removed from the statistical model. To quantify a potential attrition bias, we performed an intention to treat (Last-Observation-Carried-Forward, LOCF) analysis for the primary outcome measure in addition to the analyses defined in the protocol. We present results as means (SD) for the treatment and control group, and associated two-sided p values, as well as adjusted group difference with 95% CIs.

For the analysis of the main secondary outcomes, we replaced the primary outcome measure as the independent variable by them. This was done for both main secondary outcomes in a separate model that was otherwise identical. They were statistically treated the same way as the primary outcome measures, but corrected for multiple comparisons ($p<0{\cdot}025$, corresponding to Bonferroni correction for two independent tests).

We estimated Cohen's d as effect size measurement. The estimate of d was based on the t value of the linear mixed models. Therefore, d is corrected for the effects of all included confounding variables. By convention, $d=0{\cdot}2$ is considered to be a small, $d=0{\cdot}5$ to be a medium and $d=0{\cdot}8$ to be a large effect.[27] Based on studies regarding brief as well as continuous VR interventions using concepts of exposure therapy and including participants with fear of public speaking and public performance, we expected medium to large effect sizes.[28]

We estimated a required minimum sample size of N=80 based on a power analysis assuming two-sample t tests, equivalent to the least complex post-hoc test performed within the linear mixed model analyses (with a power of 80% at $\alpha=0{\cdot}05$, two-tailed; Cohen's $d=0{\cdot}65$; software: G*power 3.1).

A clinical trial monitor oversaw data collection and data entry according to a written monitoring plan, approved by the IEC before trial conduction.

Role of the funding source

The study was funded by the Transfaculty Research Platform of the University of Basel. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. BF was supported by a grant from the Swiss National Science Foundation (doc.CH: P0BSP1_168917). This funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Results

A total of 221 individuals were screened for trial participation (figure 3). The inclusion criteria were met by 89 individuals, of which 43 were randomly allocated to the treatment group and 46 to the control group. Study phase 1 was completed as planned by 86 participants (treatment group: 41, control group: 45).

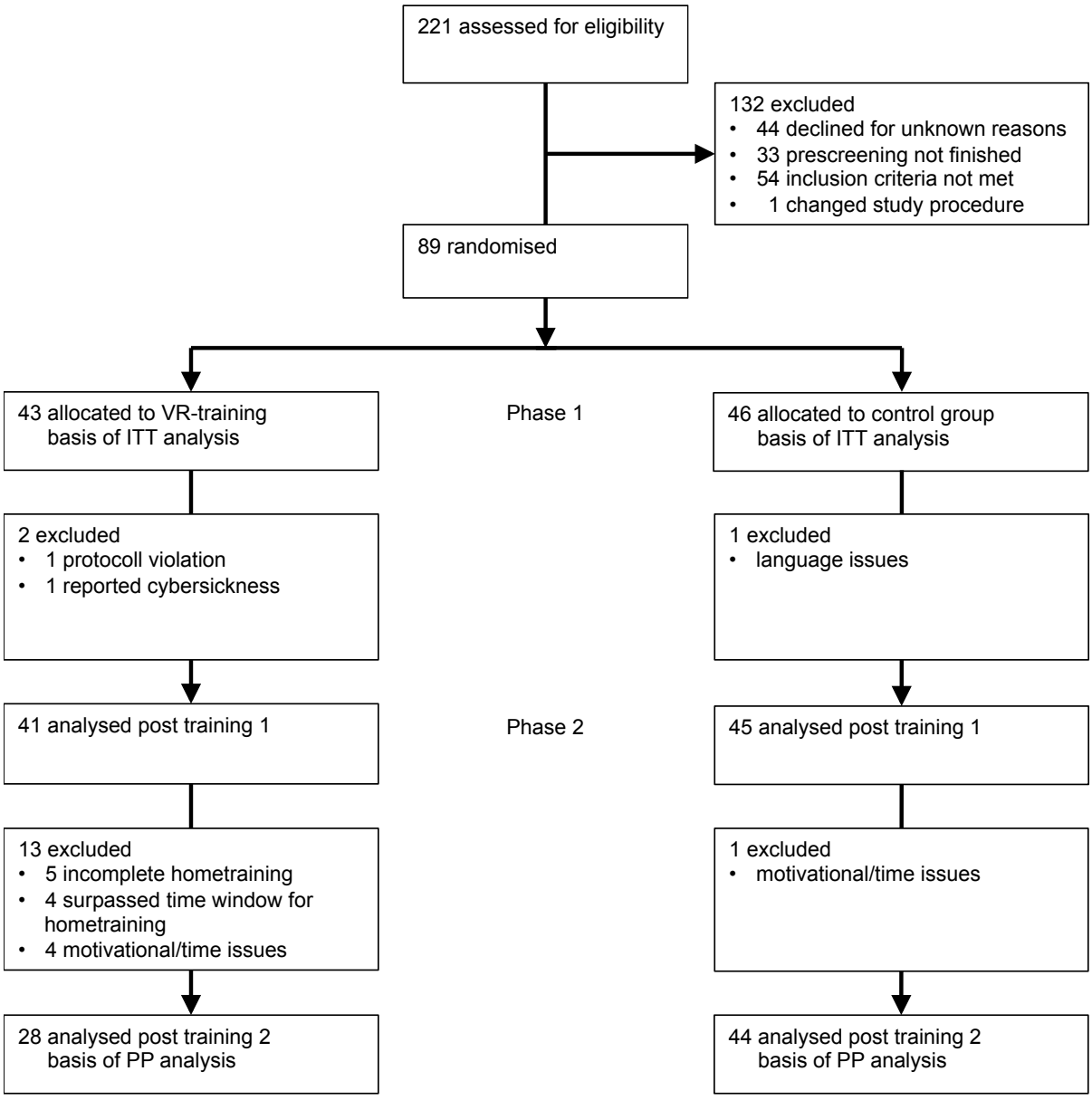


Figure 3: **Study profile**
VR, virtual reality; ITT, intention to treat; PP, per protocol. In study phase 2, more participants were excluded in the treatment group compared to the control group. In order to quantify a potential attrition bias, we therefore performed an intention to treat (Last-Observation-Carried-Forward, LOCF) analysis for the primary outcome measure in addition to the analyses defined in the protocol.

Participants' baseline characteristics were balanced across groups (table 1). Final data was collected on September 19, 2019. In study phase 1, one participant of the treatment group dropped out due to reported VR

side effects. For all participants in study phase 1, the time spent in the VR app was 60 min. 72 participants (treatment group: 28, control group: 44) completed study phase 2. In the treatment group, 65% correctly completed the full intervention course with three VR exposure trainings at visit 1 and at least six VR exposure trainings at home with a minimum training duration of 14 min. Uptake of the VR exposure to social threat was high in study phase 1 (100%) and study phase 2 (90%). Outcome data was missing for two participants in study phase 1 (1 of the control group and 1 of the treatment group) with regards to global improvement and the usability of the app (see appendix). For two participants (1 of the control group and 1 of the treatment group) eye tracking data was missing from study phase 1 due to a technical error.

***Table 1*: Demographic characteristics of study participants**

Data are numbers of participants or means (SDs).

	Treatment group			Control group		
Intervention phase	**Baseline**	**Post 1**	**Post 2**	**Baseline**	**Post 1**	**Post 2**
Participants included	43	41	28	46	45	44
Age (years) (SD)	26·7 (5·5)	26·7 (5·6)	26·2 (5·1)	28·2 (6·2)	28·1 (6·2)	28·1 (6·2)
Men	14	14	9	17	17	17
Women	29	27	19	29	28	27
Education						
Master degree, equivalent or higher	11	11	9	9	8	8
Bachelor degree or equivalent	9	8	4	13	13	12
Vocational education	7	6	3	10	10	10
High school education	13	13	9	10	10	10
Compulsory education	1	1	1	3	3	3
Other	2	2	2	1	1	1

For the primary outcome measure, the perceived fear during the PST, we found a significant interaction between group and time point (F[2,154]=23·32, p<0·0001). Contrary to our hypothesis, participants undergoing the mutual gaze training did not show a reduction in the subjectively perceived fear in PST immediately after one hour of acute VR exposure in study phase 1, as compared to the control group (treatment group, post intervention phase 1: 34·64 [SD 21·55]; control group, post intervention phase 1: 32·69 [SD 21·95]; p=0·20; adjusted group difference=3·63, 95% CI: −2·05 to 9·31, Cohen's d=0·27).

The repeated home training with the gaze training app (mean total min spent in the app: 168·84, SD 9·36) in study phase 2, spanning over an average of 11·96 days (SD 2·30), led to a reduction of subjective fear during PST (treatment group, post intervention phase 2: 26·60 [SD 19·23]; control group, post intervention phase 2: 56·34 [SD 28·15]; p<0·0001; adjusted group difference=−29·82, 95% CI: −41·77 to −17·87; Cohen's d=−1·07). The analysis with intention to treat confirmed these results, with only marginally smaller effect size estimates at post intervention phase 2 (p<0·0001; Cohen's d=−1·04; see figure 4 and appendix table 1).

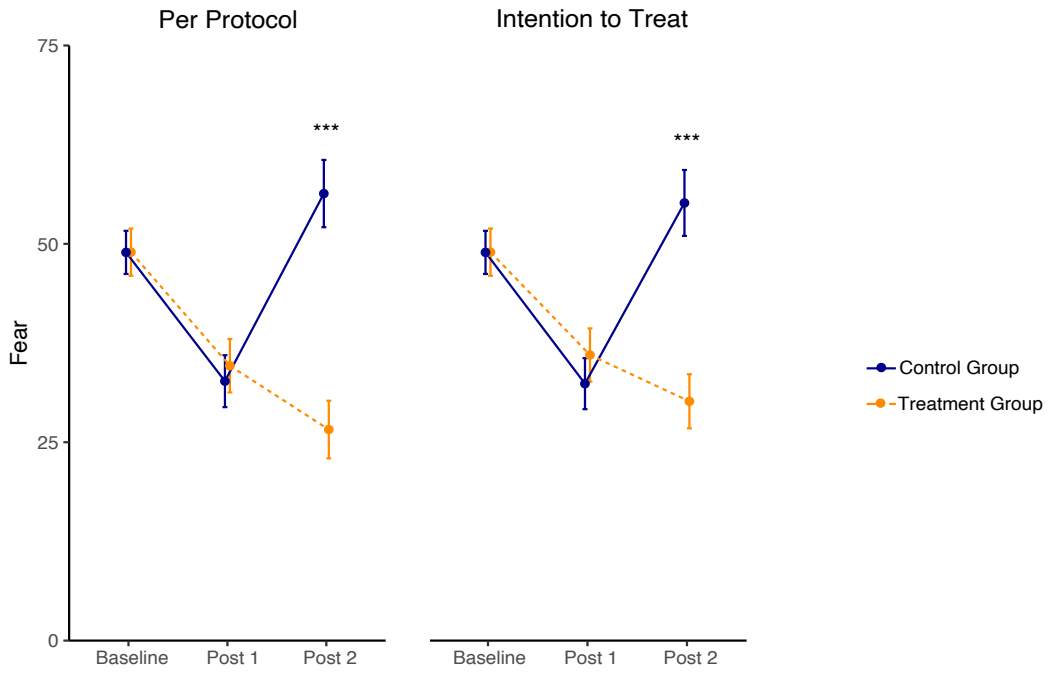


Figure 4: **Effect of the mutual gaze training on perceived fear in the subsequent PST**
Left: The per-protocol analysis suggested no reduction regarding the subjectively perceived fear in PST in the treatment group immediately after acute VR exposure in study phase 1 (Post 1) as compared to the control group. There was a beneficial effect on fear reduction for the additional home training with the gaze training app, as evident by the difference in fear ratings between the groups after intervention phase 2 (Post 2). *Right:* The intention-to-treat analysis confirmed the results, with only marginally smaller effect size estimates at post intervention phase 2.
*** p<0·001. Error bars indicate standard errors of the mean.

With regards to the relative dwell time on faces during the PST, we again found a significant interaction between group and time point (F[2,151]=8·71, p=0·00026). Users of the gaze training app, compared to the control group, did not show an increase in relative dwell time on faces immediately after acute VR exposure in study phase 1 (treatment group, post intervention phase 1: 0·22 [SD 0·12]; control group, post intervention phase 1: 0·19 [SD 0·12]; p=0·20; adjusted group difference=0·02, 95% CI: −0·01 to 0·06, Cohen's d=0·28). However, the home training with the gaze training app was associated with an increase in the relative dwell time on faces (treatment group, post intervention phase 2: 0·30 [SD 0·09]; control group, post intervention phase 2: 0·19 [SD 0·12]; p<0·0001; adjusted group difference=0·09, 95% CI: 0·05 to 0·13; Cohen's d=0·97). The analysis with intention to treat confirmed these results, with slightly smaller effect size estimates at post intervention phase 2 (p=0·00012; Cohen's d=0·83; see figure 5 and appendix table 1).

For the global performance, assessed by the committee, there was no significant interaction between group and time point (F[2,154]=024, p=0·79), indicating neither training effects of the gaze training at post VR assessment in study phase 1 nor in study phase 2. Mean scores as well as acute and repeated training effects at of all further outcome measures are summarized in appendix tables 1 and 2.

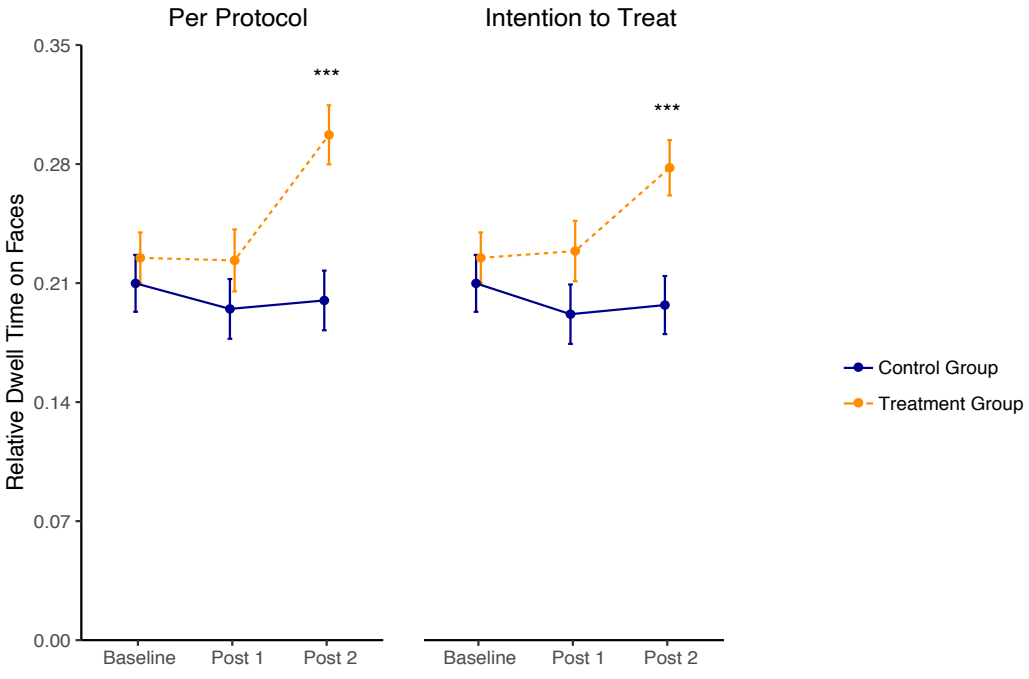


Figure 5: **Effect of the mutual gaze training on relative dwell time on faces in the subsequent PST**
Left: The per-protocol analysis suggested no reduction regarding the relative dwell time on faces in the PST in the treatment group immediately after acute VR exposure in study phase 1 (Post 1) as compared to the control group. There was a beneficial effect on reducing gaze avoidance for the additional home training with the gaze training app, as evident by the difference in the relative dwell time on faces between the groups after intervention phase 2 (Post 2). *Right:* The intention-to-treat analysis confirmed the results, with slightly smaller effect size estimates at post intervention phase 2.
*** p<0·001. Error bars indicate standard errors of the mean.

Discussion

Repeated use of our stand-alone, smartphone-based VR gaze training app reduced subjective fear of public speaking in a real-life speech situation at a large effect size (d=−1·07). Importantly, the effect size reported here is comparable to the effect size of the intention to treat analysis and to effect sizes in established, therapist-guided in-vivo[29] and in-virtuo[28] exposure studies. Furthermore, the repeated use of the app led to an increase of face gaze in real-life speech situations (d=0·97), measured objectively with eye tracking. The uptake of the VR exposure to social threat was high in both study phases, indicating that the app was generally well accepted.

The present findings are in line with the reported gaze avoidance in SAD.[13–15] However, from these studies it is unknown if there is a causal relationship between gaze avoidance and social anxiety. Here, we show that mutual gaze training with a virtual audience is sufficient to reduce fear of public speaking, without relying on verbal interactions, added cognitive behavioural elements, further assistance or input from a therapist. This sets the current study apart from the classical VRET approach, where elements of the VR (e.g., dialogues, number and gestures of avatars in the VR etc.) are externally controlled to enable interactions between the participant and its environment. Not relying on these interactions allowed us to project 360° videos of real-life audiences to the VR instead of programmed avatars and to capitalize not only on realistic, but on real facial expressions. Their complex and dynamic features might be a critical component in the context of mutual gaze training to treat PSA.[30,31]

In addition, fully automated treatment options allow home training without exposure to real social situations, which can be less predictable, and without a therapist. This might be particularly important in the context of PSA, where therapeutic settings can provoke overwhelming fears that prevent many sufferers from seeking professional help or discontinuation of treatment. Most importantly, we show that face gaze training is sufficient to reduce the fear of public speaking. For people with PSA who find it already too frightening to verbally perform in front of a VR audience, this could further lower the threshold to overcome initial fears.

Our study has the following limitations. First, although the task during mutual gaze training was to hold eye contact, we cannot exclude that participants also directed their gaze towards other salient facial features (e.g. the mouth). However, gaze towards these features might be equally important in evaluating an individual's emotional state.[31] Second, the intervention did not significantly reduce traits of general social anxiety or the fear of negative evaluation by others, as measured by the German version of the Social Phobia Inventory (SPIN)[32] or the German version of the Brief Fear of Negative Evaluation Scale–Revised (FNE-K),[33] respectively (see appendix table 1). Third, the mutual gaze training increased no external measure of speech performance except for eye contact. Arguably, the addition of cognitive or interactive elements to the VR training sessions could still

increase the feeling of immersion and thereby the training effects in terms of the externally perceived speech performance.[30] Finally, we have tested the training in a population with PSA that did not have the full clinical presentation of SAD. We therefore have no data on how well our findings translate to the clinical practice in the treatment of SAD.

In conclusion, we present evidence for the high effectiveness of an easy-to-use, stand-alone, smartphone- and VR-based mutual gaze training in reducing the fear of public speaking as well as gaze avoidance behaviour in real-life speech situations. Based on the promising results in a population with subclinical SAD, its full clinical potential remains to be investigated.

References

1. Pollard CA, Henderson JG. Four Types of Social Phobia in a Community Sample: J Nerv Ment Dis. 1988;176(7):440–5.

2. Ruscio AM, Brown TA, Chiu WT, Sareen J, Stein MB, Kessler RC. Social fears and social phobia in the USA: results from the National Comorbidity Survey Replication. Psychol Med. 2008;38(1):15–28.

3. Seim RW, Spates CR. The Prevalence and Comorbidity of Specific Phobias in College Students and Their Interest in Receiving Treatment. J Coll Stud Psychother. 2009;24(1):49–58.

4. Pull CB. Current status of knowledge on public-speaking anxiety: Curr Opin Psychiatry. 2012;25(1):32–8.

5. Wittchen HU, Stein MB, Kessler RC. Social fears and social phobia in a community sample of adolescents and young adults: prevalence, risk factors and co-morbidity. Psychol Med. 1999;29(2):309–23.

6. Canton J, Scott KM, Glue P. Optimal treatment of social phobia: systematic review and meta-analysis. Neuropsychiatr Dis Treat. 2012;8:203–15.

7. Langer JK, Rodebaugh TL. Social Anxiety and Gaze Avoidance: Averting Gaze but not Anxiety. Cogn Ther Res. 2013;37(6):1110–20.

8. Chen NTM, Clarke PJF. Gaze-Based Assessments of Vigilance and Avoidance in Social Anxiety: a Review. Curr Psychiatry Rep. 2017;19(9).

9. Weeks JW, Howell AN, Goldin PR. GAZE AVOIDANCE IN SOCIAL ANXIETY DISORDER: Research Article: Social Anxiety Disorder and Gaze Avoidance. Depress Anxiety. 2013;30(8):749–56.

10. Wieser MJ, Pauli P, Alpers GW, Mühlberger A. Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. J Anxiety Disord. 2009;23(1):93–103.

11. Horley K, Williams LM, Gonsalvez C, Gordon E. Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia. Psychiatry Res. 2004;127(1–2):43–53.

12. Moukheiber A, Rautureau G, Perez-Diaz F, Soussignan R, Dubal S, Jouvent R, et al. Gaze avoidance in social phobia: Objective measure and correlates. Behav Res Ther. 2010;48(2):147–51.

13. Weeks JW, Heimberg RG, Rodebaugh TL, Norton PJ. Exploring the relationship between fear of positive evaluation and social anxiety. J Anxiety Disord. 2008;22(3):386–400.

14. Weeks JW, Howell AN. The Bivalent Fear of Evaluation Model of Social Anxiety: Further Integrating Findings on Fears of Positive and Negative Evaluation. Cogn Behav Ther. 2012;41(2):83–95.

15. Weeks JW, Howell AN, Srivastav A, Goldin PR. “Fear guides the eyes of the beholder”: Assessing gaze avoidance in social anxiety disorder via covert eye tracking of dynamic social stimuli. J Anxiety Disord.

2019;65:56–63.
16. Chen J, van den Bos E, Westenberg PM. A systematic review of visual avoidance of faces in socially anxious individuals: Influence of severity, type of social situation, and development. J Anxiety Disord. 2020;70:102193.
17. Garcia-Palacios A, Botella C, Hoffman H, Fabregat S. Comparing Acceptance and Refusal Rates of Virtual Reality Exposure vs. In Vivo Exposure by Patients with Specific Phobias. Cyberpsychol Behav. 2007;10(5):722–4.
18. Kampmann IL, Emmelkamp PMG, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. J Anxiety Disord. 2016;42:71–84.
19. Chesham RK, Malouff JM, Schutte NS. Meta-Analysis of the Efficacy of Virtual Reality Exposure Therapy for Social Anxiety. Behav Change. 2018;35(3):152–66.
20. Berger VW, Antsygina O. A review of randomization methods in clinical trials. Clin Investig. 2015;5(12):847–53.
21. Uschner D, Schindler D, Hilgers R-D, Heussen N. randomizeR: An R Package for the Assessment and Implementation of Randomization in Clinical Trials. J Stat Softw. 2018;85(8).
22. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. Int J Aviat Psychol. 1993;3(3):203–20.
23. Feske U, Chambless DL. Cognitive behavioral versus exposure only treatment for social phobia: A meta-analysis. Behav Ther. 1995;26(4):695–720.
24. Schubert TW. The sense of presence in virtual environments:: A three-component scale measuring spatial presence, involvement, and realness. Z Für Medien. 2003;15(2):69–71.
25. Borkovec TD, Nau SD. Credibility of analogue therapy rationales. J Behav Ther Exp Psychiatry. 1972;3(4):257–60.
26. Pinheiro J, Bates D, R-core. nlme: Linear and Nonlinear Mixed Effects Models. 2019.
27. Cohen J. A power primer. Psychol Bull. 1992;112(1):155–9.
28. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. J Anxiety Disord. 2019;61:27–36.
29. Powers MB, Sigmarsson SR, Emmelkamp PMG. A Meta–Analytic Review of Psychological Treatments for Social Anxiety Disorder. Int J Cogn Ther. 2008;1(2):94–113.
30. Kampmann IL, Emmelkamp PMG, Hartanto D, Brinkman W-P, Zijlstra BJH, Morina N. Exposure to

virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. Behav Res Ther. 2016;77:147–56.

31. Kegel LC, Brugger P, Frühholz S, Grunwald T, Hilfiker P, Kohnen O, et al. Dynamic human and avatar facial expressions elicit differential brain responses. Soc Cogn Affect Neurosci. 2020;nsaa039.

32. Connor KM, Davidson JR, Churchill LE, Sherwood A, Foa E, Weisler RH. Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. Br J Psychiatry J Ment Sci. 2000;176:379–86.

33. Reichenberger J, Schwarz M, K nig D, Wilhelm FH, Voderholzer U, Hillert A, et al. Angst vor negativer sozialer Bewertung: Übersetzung und Validierung der Furcht vor negativer Evaluation–Kurzskala (FNE-K). Diagnostica. 2016;62(3):169–81.

Acknowledgments

We thank Gabriel Guerra, Galya Iseli and Theresa Müggler for their work as experimenters in the PST. We thank Daria Bühler, Milena Cavegn, Tatjana Fuchs, Christiane Gerhards, Irena Kovacic, Cécile Longoni, Anna Magos, Johanna Otte, Jule Schr der, Nina Waldthaler and Johanna Weibel for helping in trial conduction as well as Amanda Aerni for monitoring the study. This study was funded by the Transfaculty Research Platform of the University of Basel and the Swiss National Science Foundation (SNSF). The views expressed are those of the author(s) and not necessarily those of the SNSF.

Author Contributions

FM, BF, DQ and AP conceived and designed the study. NW and MKI programmed and visually designed the VR app. BF, FM and AZ collected and analysed the data. TS supervised data collection, storage and processing. DB provided clinical advice. BF, DQ and FM wrote the manuscript, with substantial input from the other authors. DQ and AP provided critical oversight and feedback of the work.

Competing Interests Statement

Dominique de Quervain (DQ) and Andreas Papassotiropoulos (AP) are co-founders of GeneGuide AG, a spin-off company of the University of Basel. DQ and AP have acquired a license from the University of Basel to use the developed technology for commercial purposes. DQ nor AP have been involved in data acquisition or data analysis. BF reports a grant from the Swiss National Science Foundation during the conduct of the study. All other authors declare no competing interests.

Data sharing

Pseudonymised data will be made available upon reasonable request, which must include an approved ethics protocol and statistical analysis plan. R code for the analyses is freely available from the corresponding author.

Appendix

Methods

Exclusion Criteria

We excluded individuals suffering from clinically relevant social anxiety, based on the criteria of the corresponding section of the structured Diagnostic Interview for Mental Disorders for DSM-5.[1] Further, individuals were not allowed to participate if they were ever treated against fear of public speaking, participated in a parallel study, showed signs of depression (Beck Depression Inventory II; BDI-II;[2] total score≥20) or suicidal ideation (BDI-II item 9>0), received concurrent psychotherapy, had other serious psychological or medical conditions (including epilepsy and migraine), had restricted 3D sight or chronic drug or medication intake (except intake of oral contraceptives), as well as females if they were pregnant. Participants were instructed to abstain from the intake of alcohol (for 12 h), medication (for 24 h), as well as psychoactive substances (including benzodiazepines; for 5 d) before days of testing.

Further information on questionnaires, scales and tests

Depressive symptomatology and suicidal ideation were assessed by the BDI-II. Alcohol consumption and intake of prescribed or illicit drugs was queried. The 3D sight was assessed with the standard for stereo depth perception test (Precision Vision ®). To collect baseline measures for the fear of public speaking, participants filled in the SPIN and the FNE-K. The severity of PSA was further investigated by the corresponding section for social anxiety of a structured clinical interview of the diagnostic interview for mental disorders for DSM-5. Additionally, all participants received some general psycho-educative material to explain the rationale of exposure to the feared public speaking situations.

Further outcome measures

Further secondary outcomes measures with regards to behaviour were (1) fear of eye contact during the PST (SUDS; 0: no fear of eye contact, 100: maximum fear of eye contact), measured analogously to the primary outcome measure, (2) the global self-assessment of performance in the PST, measured analogously to the global external assessment, but with regard to the participants' own rating, (3–5) the global subjectively perceived improvement of fear, eye contact and performance by the VR app (VAS-ratings; −100: much worse than before, 100: much better than before), (6) social anxiety, as measured by the SPIN (0–68, with higher scores corresponding to a higher burden) and (7) fear of negative evaluation, as measured by the FNE-K (12–60, with higher scores corresponding to greater distress).

Further secondary outcomes with regards to eye tracking were (1) the average relative fixation frequency on faces as well as (2) the average pupil size during the PST (see 'Eye tracking').

The further secondary outcome with regards to physiology was (1) the difference between salivary cortisol concentrations before and after performing the PST.

Outcomes of further interest were a total of 16 measures (1–16) with regards to the externally (averaged across committee members) and self-assessed performance in the PST, characterized by 8 VAS-ratings of detailed performance aspects (i.e. verbal fluency, verbal expression, vocal modulation, tempo, posture, facial expression, eye contact, nervousness; 0: very bad, 100: very good) and averaged within each PST, (17) the speech duration in seconds, averaged within each PST, (18) the feeling of presence in the VR-session, assessed by the IPQ (−42–42, with higher scores corresponding to stronger feeling of presence), (19) the usability (0–90, with higher scores indicating higher usability) and (20) social immersion (0–500, with higher scores indicating a higher degree of social immersion) of the app, assessed by questions created in-house as well as (21) the minutes of self-exposure to social situations since phase 1.

Measures were taken at baseline, as well as after the first and after the second intervention phase, with the following exceptions: Measures directly addressing improvement (i.e. of fear, eye contact and performance in the VR app) were assessed after intervention phase 1 and at intervention phase 2. Scores for SPIN, FNE-K and the cortisol difference were assessed at baseline and intervention phase 2.

The IPQ and usability of the app were assessed after intervention phase 1 and the amount of self-exposure only at intervention phase 2. Additionally, only the treatment group provided fear ratings during VR exposure and ratings of the usability and social immersion of the app, rated after the home training at intervention phase 2.

Statistical analysis of further outcome measures

For the analysis of the further secondary outcomes and the outcomes of further interest, we replaced the primary outcome measure as the independent variable by them. This was done for all of them in a separate model that was otherwise identical. For measures directly addressing improvement (i.e. of fear, eye contact and performance in the VR app), no baseline differences were accounted for in case of post-hoc analyses. For variables that were only assessed at one time point (i.e. IPQ and usability of the app for both groups only after intervention phase 1, amount of self-exposure to social situations only after intervention phase 2), analyses were reduced to linear models. Variables only assessed for one group (i.e. the app usability and social immersion after intervention phase 2) are reported in a descriptive way. Further secondary outcome measures were grouped according to behaviour (7, $p<0{\cdot}007$), eye tracking (2, $p<0{\cdot}025$) and physiology (1, $p<0{\cdot}05$) and corrected for

multiple comparisons within those categories. Outcomes of further interest were analysed in an explorative way. Therefore, statistical significance is reported based on nominal p values.

Gaze training app

The gaze training app was developed using Unity3D (v2018.3.11f1; Unity Technologies, San Francisco, CA, USA) under MacOS Mojave (v10.14.6) and compiled into standard Android Package file (.apk). All visual material is based on 360° panoramic video clips taken by a 360°-camera (Insta360° Pro; Insta360, Shenzhen, GD, China) and stitched with Insta360Sticher (v3.0.0). Videos are used as skybox textures and all panels were created in the world coordinate system in the Unity3D game engine. All audio material (such as the logo and effects sound) was produced using Ableton Live 10 Suite (v10.0.1). The VR environment consists of the 360° video clips, accompanied by sounds characteristic to each VR scenario and level (e.g. sounds of the audience rustling or coughing). In appendix table 3, the different settings of each level are provided, together with the minimal requirements to proceed to the next level.

Eye tracking

We used eye tracking to characterize participants' social gaze behaviour during the PST. Both eyes were tracked at 120 hz by a mobile, head mounted eye tracking system (Pupil Labs, Berlin, Germany; field of view: 100°, world-camera: 30 hz, 720 p). The eye tracking was controlled by the software Pupil Capture (v1.12.17). Gaze detection and mapping was based on a 2D model. At the beginning of the first PST, a built-in 10-point calibration procedure was run and repeated until an accuracy of at least 0·5° and a precision of at least 0·2° was achieved. Therefore, calibration targets were used, which were projected at a distance of 2·5 m. In order to keep the distance to the calibration plane constant, participants were instructed to stay within an area marked on the ground for the duration of their speech. After each speech, the predefined 5-point accuracy test was run, allowing for offline recalibration in case of shift/slippage. Speeches where the predefined minimum accuracy and precision was neither reached by online calibration nor offline recalibration were excluded from further analyses (<5%). Blinks (filter length: 0·2 s, onset and offset confidence: 0·5) and fixations (maximum dispersion: 3°, minimum duration: 0·3 s, confidence threshold: 0·75) were detected by the default settings of the software. To quantify pupil size, we used the 2D pupil detection algorithm with default settings (pupil range: 10–100, pupil size and diameter indicated in image pixels as observed in the eye image frame, confidence threshold: 0·6). We indirectly controlled for artefacts (e.g., with regards to blinks or movement) by excluding pupil diameter outliers. A data point x was defined as an outlier if $x<M-3\times SD$ or $x>M+3\times SD$ or if it was labeled as a blink and replaced by linear interpolation. Subsequently, pupil recordings were smoothed by using a sliding average (83 ms time

window, ten samples; see 28). The light settings and exact position of the participants in the room were held constant throughout the entire experiment to control for confounding effects of the lighting conditions. All recorded data were visualized and exported by pupil player (v1.14.9), with the default minimum data confidence of 0·6.

Because we used a mobile eye tracker, the coordinate system of the exported gaze data was fixed to the head of the participant, not to the real world. However, automated face detection allowed us to detect the location of the faces of the committee.

To this end, we implemented a Convolutional Neural Network (CNN), which was based on a face detector available in the dlib library (Python3.7). Supported by GPU (NVidia Tesla v100) and based on the video recorded by the world camera (exported mp4-file), we extracted the faces of the committee in each frame of the video. The total area within the bounding boxes of all recognized faces per video frame was defined as the area of interest (AOI; see appendix figure 1).

To quantify fear-related gaze behaviour as one of our main secondary outcome measures, we calculated the duration of all valid fixations that were lying in the AOI (i.e. one of the three faces of the PST committee). We then divided this value by the total duration of all valid fixations in order to account for different speech durations. The resulting measure is the relative dwell time on faces. As a further secondary outcome measure and another marker of attention towards the committee, we counted the relative number of fixations on faces, irrespective of their length.[4] Finally, we calculated the mean pupil diameter during the course of each speech as a potential marker of general task difficulty and engaged attention.[3] All eye tracking measures were averaged within participants and PSTs.

Figure A1: **Example of face detection**
A video frame is shown, which was recorded from the world camera of the mobile eye tracking system, thus representing a participant's perspective during the PST. The green boxes mark the bounding area of the faces of the three committee members that were recognised by our deep learning approach. The total area comprised by the boxes was defined as the area of interest (AOI), and fixations within (as indicated by the green dot) were considered to reflect mutual face gaze.

546

Results

***Table A1*: Results with regards to primary and secondary outcome measures**
If not indicated otherwise, the results are based on the per-protocol analysis.
Post-hoc analyses for each time point separately were only conducted in case of significant time point × group interactions. The adjusted group differences are only indicated if significant. Descriptive values are means (SDs).

		Treatment group		Control group		p value (F-value)	Adjusted group difference (95% CI)	Cohen's d
		Mean (SD)	n	Mean (SD)	n			
Primary outcome	**1.0 PST fear**							
	Time point × group					<0·0001 (F[2, 154]=23·32)		
	Baseline	49·0 (19·5)	43	48·9 (18·4)	46	0·81		
	Post training 1	34·6 (21·5)	41	32·7 (21·9)	45	0·20		
	Post training 2	26·6 (19·2)	28	56·3 (28·1)	44	<0·0001	−29·8 (−41·8 to −17·9)	−1·07
	1.1 PST fear ITT analysis							
	Time point × group					<0·0001 (F[2,174]=23·15)		
	Baseline	49·0 (19·5)	43	48·9 (18·4)	46	0·15	..	..
	Post training 1	36·0 (22·0)	43	32·4 (21·8)	46	0·55	..	..
	Post training 2	30·2 (22·3)	43	55·2 (28·2)	46	<0·0001	−24·6 (−34·7 to −14·6)	−1·04
Main secondary outcomes	**1.0 PST relative dwell time on faces**							
	Time point × group					0·00026 (F[2, 151]=8·71)		
	Baseline	0·22 (0·10)	42	0·21 (0·11)	45	0·56	..	..
	Post training 1	0·22 (0·12)	40	0·19 (0·12)	44	0·20	..	..
	Post training 2	0·30 (0·09)	28	0·19 (0·12)	44	<0·0001	0·09 (0·05 to 0·13)	0·97
	1.1 PST relative dwell time on faces ITT analysis							
	Time point × group					0·0030 (F[2,170]=6·07)		
	Baseline	0·22 (0·10)	43	0·21 (0·11)	46	0·56	..	..
	Post training 1	0·22 (0·12)	43	0·19 (0·12)	46	0·16	..	..
	Post training 2	0·28 (0·11)	43	0·20 (0·12)	46	0·00012	−0·07 (0·03 to 0·11)	0·83
	2.0 PST global external assessment of performance							
	Time point × group					0·79 (F[2,154]=0·24)		
	Baseline	47·5 (11·9)	43	45·2 (11·7)	46	..	..	..
	Post training 1	52·3 (11·1)	41	48·9 (10·9)	45	..	..	..
	Post training 2	55·1 (7·5)	28	52·2 (8·7)	44	..	..	..
	2.1 PST global external assessment of performance ITT analysis							
	Time point × group					0·37 (F[2,174]=1·01)		
	Baseline	47·5 (11·9)	43	45·2 (11·7)	46	..	..	..
	Post training 1	51·2 (12·6)	43	48·1 (11·9)	46	..	..	..
	Post training 2	52·1 (11·3)	43	51·0 (10·5)	46	..	..	..

Further secondary outcomes	**1. PST fear of eye contact**							
	Time point × group					<0·0001 (F[2, 154]=18·58)		
	Baseline	37·6 (21·0)	43	37·4 (22·6)	46	0·75	··	··
	Post training 1	25·0 (20·1)	41	26·0 (23·3)	45	0·79	··	··
	Post training 2	18·2 (14·5)	28	49·1 (31·1)	44	<0·0001	−29·6 (−42·0 to −17·3)	−1·03
	2. PST global self−assessment of performance							
	Time point × group					<0·0001 (F[2, 154]=25·10)		
	Baseline	35·3 (18·3)	43	33·4 (15·4)	46	0·48	··	··
	Post training 1	47·0 (17·5)	41	43·8 (19·7)	45	0·68	··	··
	Post training 2	61·9 (21·1)	28	34·6 (20·0)	44	<0·0001	27·6 (18·1 to 37·1)	1·24
	3. Perceived improvement of fear							
	Time point × group					<0·0001 (F[1, 68]=21·95)		
	Baseline	··	··	··	··	··	··	··
	Post training 1	66·3 (15·5)	40	66·9 (16·9)	44	0·67	··	··
	Post training 2	76·3 (14·0)	28	49·2 (23·8)	44	<0·0001	26·1 (16·1 to 36·1)	1·27
Behaviour	**4. Perceived improvement of eye contact**							
	Time point × group					<0·0001 (F[1, 68]=27·44)		
	Baseline	··	··	··	··	··	··	··
	Post training 1	63·2 (14·0)	40	60·9 (15·6)	44	0·60	··	··
	Post training 2	75·8 (16·3)	28	45·3 (21·5)	44	<0·0001	30·1 (20·4 to 39·8)	1·51
	5. Perceived improvement of performance							
	Time point × group					<0·0001 (F[1, 68]=24·94)		
	Baseline	··	··	··	··	··	··	··
	Post training 1	60·8 (15·2)	40	62·5 (16·6)	44	0·51	··	··
	Post training 2	73·3 (17·3)	28	46·6 (22·4)	44	<0·0001	25·9 (15·8 to 36·0)	1·25
	6. SPIN score							
	Time point × group					0·62 (F[1, 70]=0·24)		
	Baseline	21·7 (10·9)	43	22·7 (8·0)	46	··	··	··
	Post training 1	··	··	··	··	··	··	··
	Post training 2	17·6 (9·9)	28	17·8 (8·9)	44	··	··	··
	7. FNE-K score							
	Time point × group					0·78 (F[1, 70]=0·07)		
	Baseline	40·2 (10·7)	43	38·5 (9·2)	46	··	··	··
	Post training 1	··	··	··	··	··	··	··
	Post training 2	37·5 (10·2)	28	34·7 (10·6)	44	··	··	··
Eye-tracking	**1. PST relative fixation frequency on faces**							
	Time point × group					<0·0001 (F[2, 151]=8·26)		
	Baseline	0·21 (0·09)	42	0·20 (0·11)	45	0·57	··	··
	Post training 1	0·22 (0·11)	40	0·19 (0·11)	44	0·24	··	··
	Post training 2	0·29 (0·09)	28	0·19 (0·11)	44	<0·0001	0·09 (0·05 to 0·13)	0·94
	2. PST Pupil size							
	Time point × group					0·34 (F[2, 151]=1·09)		
	Baseline	48·6 (10·3)	42	47·7 (9·7)	45	··	··	··
	Post training 1	49·2 (10·9)	40	49·6 (10·7)	44	··	··	··
	Post training 2	53·1 (10·0)	28	49·3 (7·6)	44	··	··	··
Physiology	**1. PST Cortisol (Δ)**							
	Time point × group					0·74 (F[1, 70]=0·11)		
	Baseline	1·3 (5·7)	43	2·0 (6·8)	46	··	··	··
	Post training 1	··	··	··	··	··	··	··
	Post training 2	0·4 (4·7)	28	0·7 (4·7)	44	··	··	··

PST, public speech test; ITT, intention to treat; Δ=Cortisol levels after the PST−Cortisol levels before the PST.

***Table A2*: Results with regards to outcomes of further interest**

If not indicated otherwise, the results are based on the per-protocol analysis.
Post-hoc analyses for each time point separately were only conducted in case of significant time point × group interactions. The adjusted group differences are only indicated if significant. Descriptive values are means (SDs).

	Treatment group		Control group		p value (F-value)	Adjusted group difference (95% CI)	Cohen's d
	Mean (SD)	n	Mean (SD)	n			
Outcomes of further interest							
1. PST self-assessment of verbal fluency							
Time point × group					<0·0001 (F[2, 154]=15·99)		
Baseline	40·3 (17·6)	43	39·3 (16·1)	46	0·15	··	··
Post training 1	52·3 (17·5)	41	52·1 (18·8)	45	0·090	··	··
Post training 2	61·4 (18·7)	28	39·3 (21·5)	44	<0·0001	22·6 (13·0 to 32·2)	1·01
2. PST self-assessment of verbal expression							
Time point × group					<0·0001 (F[2, 154]=16·37)		
Baseline	37·8 (17·0)	43	37·0 (15·4)	46	0·60		
Post training 1	48·8 (19·2)	41	48·5 (19·6)	45	0·84		
Post training 2	58·5 (17·9)	28	36·7 (20·5)	44	<0·0001	22·7 (13·7 to 32·0)	1·08
3. PST self-assessment of vocal modulation							
Time point × group					<0·0001 (F[2, 154]=21·66)		
Baseline	45·0 (19·5)	43	45·3 (17·5)	46	0·98	··	··
Post training 1	50·5 (19·5)	41	48·4 (19·5)	45	0·57	··	··
Post training 2	65·6 (20·3)	28	38·3 (20·4)	44	<0·0001	27·0 (18·0 to 36·0)	1·29
4. PST self-assessment of tempo							
Time point × group					<0·0001 (F[2, 154]=26·86)		
Baseline	37·9 (15·7)	43	37·5 (16·2)	46	0·75	··	··
Post training 1	49·3 (18·6)	41	48·3 (18·9)	45	0·96	··	··
Post training 2	60·9 (20·2)	28	35·2 (21·7)	44	<0·0001	26·2 (17·1 to 35·4)	1·22
5. PST self-assessment of posture							
Time point × group					<0·0001 (F[2, 154]=23·14)		
Baseline	34·0 (13·7)	43	30·7 (13·6)	46	0·26	··	··
Post training 1	43·8 (17·8)	41	41·7 (18·0)	45	0·83	··	··
Post training 2	57·9 (17·6)	28	31·9 (17·2)	44	<0·0001	24·0 (15·9 to 32·1)	1·27
6. PST self-assessment of facial expression							
Time point × group					<0·0001 (F(2, 154)=24·25)		
Baseline	36·7 (14·7)	43	35·6 (14·0)	46	0·65	··	··
Post training 1	45·0 (18·8)	41	42·2 (17·1)	45	0·53	··	··
Post training 2	61·0 (17·5)	28	34·3 (19·2)	44	<0·0001	25·9 (17·5 to 34·3)	1·32
7. PST self-assessment of eye contact							
Time point × group					<0·0001 (F[2, 154]=22·38)		
Baseline	39·8 (18·7)	43	39·9 (20·2)	46	0·89	··	··
Post training 1	51·7 (21·5)	41	51·8 (19·4)	45	0·89	··	··
Post training 2	66·5 (18·5)	28	39·5 (22·8)	44	<0·0001	27·6 (17·6 to 37·6)	1·18
8. PST self-assessment of nervousness							
Time point × group					<0·0001 (F[2, 154]=19·24)		
Baseline	33·0 (15·8)	43	31·1 (15·9)	46	0·42	··	··
Post training 1	40·7 (20·3)	41	47·6 (21·0)	45	0·019	−8·6 (−16·0 to −1·3)	−0.50
Post training 2	57·5 (20·9)	28	33·7 (22·6)	44	<0·0001	23·0 (12·5 to 33·4)	0·94

9. PST external assessment of verbal fluency							
Time point × group					0·57 (F[2, 154]=0·55)		
Baseline	60·4 (12·5)	43	58·2 (12·2)	46	..	..	..
Post training 1	63·6 (12·0)	41	62·8 (11·6)	45	..	..	..
Post training 2	66·8 (7·7)	28	64·9 (10·1)	44	..	..	..
10. PST external assessment of verbal expression							
Time point × group					0·41 (F[2, 154]=0·90)		
Baseline	58·1 (13·6)	43	54·1 (13·9)	46	..	..	..
Post training 1	60·7 (13·3)	41	58·4 (12·8)	45	..	..	..
Post training 2	62·2 (9·2)	28	59·3 (9·6)	44	..	..	..
11. PST external assessment of vocal modulation							
Time point × group					0·88 (F[2, 154]=0·12)		
Baseline	51·0 (11·1)	43	48·5 (11·7)	46	..	..	..
Post training 1	54·0 (12·3)	41	50·8 (11·0)	45	..	..	..
Post training 2	55·5 (7·7)	28	52·7 (9·6)	44	..	..	..
12. PST external assessment of tempo							
Time point × group					0·77 (F[2, 154]=0·26)		
Baseline	43·6 (10·2)	43	41·0 (10·3)	46	..	..	..
Post training 1	48·0 (10·6)	41	45·1 (10·0)	45	..	..	..
Post training 2	53·6 (6·4)	28	52·2 (8·5)	44	..	..	..
13. PST external assessment of posture							
Time point × group					0·68 (F[2, 154]=0·39)		
Baseline	41·6 (10·3)	43	40·2 (12·0)	46	..	..	..
Post training 1	44·1 (10·5)	41	44·1 (10·8)	45	..	..	..
Post training 2	48·0 (8·9)	28	45·5 (8·6)	44	..	..	..
14. PST external assessment of facial expression							
Time point × group					0·95 (F(2, 154)=0·05)		
Baseline	47·0 (8·3)	43	46·4 (9·3)	46	..	..	..
Post training 1	51·0 (8·5)	41	50·0 (9·6)	45	..	..	..
Post training 2	53·5 (7·8)	28	52·5 (8·0)	44	..	..	..
15. PST external assessment of eye contact							
Time point × group					0·0060 (F[2, 154]=5·23)		
Baseline	47·7 (14·1)	43	47·1 (16·2)	46	0·79	..	..
Post training 1	54·2 (14·8)	41	49·0 (16·7)	45	0·016	4·9 (0·8 to 8·9)	0·51
Post training 2	55·3 (7·5)	28	48·4 (10·1)	44	<0·0001	7·3 (4·2 to 10·5)	0·99
16. PST external assessment of nervousness							
Time point × group					0·59 (F[2, 154]=0·54)		
Baseline	36·1 (9·5)	43	32·9 (11·2)	46	..	..	..
Post training 1	46·6 (8·1)	41	43·7 (10·8)	45	..	..	..
Post training 2	50·2 (8·2)	28	46·7 (10·2)	44	..	..	..
17. PST speech duration (s)							
Time point × group					0·40 (F[2, 154]=0·93)		
Baseline	171·3 (16·9)	43	173·9 (14·9)	46	..	..	..
Post training 1	171·6 (18·4)	41	169·9 (21·0)	45	..	..	..
Post training 2	177·9 (8·9)	28	176·5 (11·1)	44	..	..	..
18. IPQ score							
Baseline	..	..	..	..	..	..	..
Post training 1	0·7 (11·8)	41	−2·4 (12·4)	45	0·19	..	..
Post training 2	..	..	..	..	..	..	..
19. App usability							
Baseline	..	..	..	..	..	..	..
Post training 1	50·4 (12·8)	40	34·5 (16·9)	44	<0·0001	15·7 (9·1 to 22·4)	1·03
Post training 2	49·3 (13·2)	28	..	..	..	..	..
20. App social immersion							
Baseline	..	..	..	..	..	..	..
Post training 1	..	..	..	..	..	..	..
Post training 2	283·1 (105·3)	28	..	..	..	..	..

21. Self-exposure to social situations (min)							
Baseline	..	..	..	..	..	..	..
Post training 1	..	..	..	..	..	..	..
Post training 2	87·2 (223·7)	28	105·6 (159·0)	44	0·68	..	..

PST, public speech test.

566 ***Table A3:*** **Level settings of the gaze training app**
567 The levels were gradually increasing in difficulty, which was operationalized by the emotional valence of the
568 audience's facial expressions, the size of the audience as well as the time required to maintain face gaze. While
569 the absolute number of faces required to gaze at remained constant, the dwell time required on each face was
570 increased in higher levels, resulting in a longer total dwell time on faces needed for successful level completion.
571

Scenario	Level	Emotional valence of audience	Size of audience	Number of target faces	Number of faces required to gaze at	Dwell time required per face (s)	Total dwell time required on faces (s)
Close proximity	1	Positive	2	2	8	7	56
Close proximity	2	Positive	4	4	8	8	64
Close proximity	3	Neutral	6	2	8	9	72
Close proximity	4	Neutral	6[a]	2	8	10	80
Close proximity	5	Negative	9[a]	3	8	11	88
Close proximity	6	Negative	12[a]	4	8	12	96
Classroom	1	Positive	2	2	8	7	56
Classroom	2	Positive	4	3	8	8	64
Classroom	3	Neutral	8	4	8	9	72
Classroom	4	Neutral	12	4	8	10	80
Classroom	5	Negative	16	5	8	11	88
Classroom	6	Negative	21	6	8	12	96
Lecture hall	1	Positive	4	2	8	7	56
Lecture hall	2	Positive	8	4	8	8	64
Lecture hall	3	Neutral	16	6	8	9	72
Lecture hall	4	Neutral	32	6	8	10	80
Lecture hall	5	Negative	64	6	8	11	88
Lecture hall	6	Negative	100	7	8	12	96

572
573 [a] levels with<1 m of personal distance

References

1. Margraf J, Cwik JC, Suppiger A, Schneider S. DIPS Open Access: Diagnostisches Interview bei psychischen St rungen. Ruhr-Univ Boch RUB. 2017.

2. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. J Pers Assess. 1996;67(3):588–97.

3. Ajasse S, Benosman RB, Lorenceau J. Effects of pupillary responses to luminance and attention on visual spatial discrimination. J Vis. 2018;18(11):6.

4. Moukheiber A, Rautureau G, Perez-Diaz F, Soussignan R, Dubal S, Jouvent R, et al. Gaze avoidance in social phobia: Objective measure and correlates. Behav Res Ther. 2010;48(2):147–51.

4.3 Reducing Amygdala Activity and Phobic Fear through Cognitive Top-Down Regulation.

Reducing Amygdala Activity and Phobic Fear through Cognitive Top–Down Regulation

Eva Loos*, Nathalie Schicktanz*, Matthias Fastenrath, David Coynel, Annette Milnik, Bernhard Fehlmann, Tobias Egli, Melanie Ehrler, Andreas Papassotiropoulos, and Dominique J.-F. de Quervain

Abstract

■ The amygdala is critically involved in emotional processing, including fear responses, and shows hyperactivity in anxiety disorders. Previous research in healthy participants has indicated that amygdala activity is down-regulated by cognitively demanding tasks that engage the PFC. It is unknown, however, if such an acute down-regulation of amygdala activity might correlate with reduced fear in anxious participants. In an fMRI study of 43 participants (11 men) with fear of snakes, we found reduced amygdala activity when visual stimuli were processed under high cognitive load, irrespective of whether the stimuli were of neutral or phobic content. Furthermore, dynamic causal modeling revealed that this general reduction in amygdala activity was partially mediated by a load-dependent increase in dorsolateral PFC activity. Importantly, high cognitive load also resulted in an acute decrease in perceived phobic fear while viewing the fearful stimuli. In conclusion, our data indicate that a cognitively demanding task results in a top–down regulation of amygdala activity and an acute reduction of fear in phobic participants. These findings may inspire the development of novel psychological intervention approaches aimed at reducing fear in anxiety disorders. ■

INTRODUCTION

The amygdala is fundamentally involved in processing emotional stimuli of positive and negative valence (Janak & Tye, 2015), including fear in animals (Fanselow & Gale, 2003; Davis & Whalen, 2001) and humans (Shin & Liberzon, 2010; LeDoux, 2007; Adolphs, Tranel, Damasio, & Damasio, 1995). Furthermore, amygdala hyperactivity has been associated with many anxiety disorders including phobias. Specifically, phobic participants show higher amygdala activation compared with healthy participants when confronted with phobic stimuli (Ipser, Singh, & Stein, 2013; Straube, Mentzel, & Miltner, 2006; Schienle, Schäfer, Walter, Stark, & Vaitl, 2005; Dilger et al., 2003). Proper regulation of emotional reactions, including a down-regulation of fear, is thought to rely on the successful interplay between prefrontal and limbic regions (Okon-Singer, Hendler, Pessoa, & Shackman, 2015; Dolcos & Denkova, 2014; Pessoa, 2013). Within the prefrontal network, the dorsolateral PFC (dlPFC) is critically involved in higher cognitive processes like working memory and executive control (Kohn et al., 2014; Barbey, Koenigs, & Grafman, 2013; Owen, McMillan, Laird, & Bullmore, 2005; Curtis & D'Esposito, 2003) and has been reported to interact with regions engaged in emotion processing and emotion regulation (Dolcos, Iordan, & Dolcos, 2011; Van Dillen, Heslenfeld, & Koole, 2009; Ochsner & Gross, 2005; Phillips, Drevets, Rauch, & Lane, 2003). fMRI studies have consistently shown that cognitively demanding tasks are associated with increased dlPFC activity and decreased amygdala activity (de Voogd et al., 2018; Straube, Lipka, Sauer, Mothes-Lasch, & Miltner, 2011; Erk, Kleczar, & Walter, 2007; Mitchell et al., 2007). In situations that demand high cognitive functioning, the dlPFC is assumed to inhibit limbic regions, including the amygdala through top–down control mechanisms, to ensure that emotional reactions do not interfere with goal-directed behavior (Okon-Singer et al., 2015; Clarke & Johnstone, 2013; Iordan, Dolcos, & Dolcos, 2013). On a behavioral level, increased cognitive load has been associated with reduced state anxiety and startle response (Balderston et al., 2016; Vytal, Arkin, Overstreet, Lieberman, & Grillon, 2016; Vytal, Cornwell, Arkin, & Grillon, 2012; King & Schaefer, 2011) and with reduced subjectively experienced negative emotion in response to negative stimuli (Van Dillen et al., 2009). Additionally, performing a cognitively demanding task over several weeks resulted in better cognitive control in healthy (Cohen et al., 2016; Schweizer, Grahn, Hampshire, Mobbs, & Dalgleish, 2013; Schweizer, Hampshire, & Dalgleish, 2011) as well as in anxious individuals (Sari, Koster, Pourtois, & Derakshan, 2016).

To our knowledge, it has not yet been investigated whether a cognitively demanding task known to engage the dlPFC could be used to acutely decrease amygdala activity and reduce subjectively felt fear in anxious participants.

University of Basel
*These authors contributed equally to this work.

Journal of Cognitive Neuroscience 32:6, pp. 1117–1129
https://doi.org/10.1162/jocn_a_01537

To address this question, we designed a pictorial *n*-back task and measured amygdala activity during the viewing of snake pictures and neutral pictures in participants with fear of snakes under different cognitive load conditions. The task included a high cognitive load condition (2-back) and a low cognitive load condition (0-back), whereby the snake pictures and neutral pictures served as targets in the different conditions (Figure 1). This design ensured that the visual input during the *n*-back task was identical across load conditions. We hypothesized reduced amygdala activity and reduced subjective fear ratings during the high load condition, as compared with the low load condition. Additionally, we applied dynamic causal modeling (DCM) to investigate a possible load-dependent change in effective connectivity between the dlPFC and the amygdala.

METHODS

Participants

Forty-three participants (11 men; mean age = 23.12 years, *SD* = 3.37 years) were included in the final analysis after removing four participants because of corrupted fMRI data (in three participants, there was a mistake in the scanning procedure, and one participant displayed excessive head motions, which resulted in poor quality of the imaging data). Another participant was excluded because of low fear ratings of the snake pictures during the pictorial *n*-back task (> 2.5 *SD* from sample mean).

Participants were recruited from the Basel and Zurich area in Switzerland through advertisements on the Internet, trams, and buses as well as through distribution of flyers. Participants had to meet the following inclusion criteria: (1) age between 18 and 35 years, (2) body mass index between 18 (women)/19 (men) and 35 kg/m^2, (3) native or fluent German-speaking, (4) capable of viewing pictures of snakes without turning the head away, and (5) a score of 12 or higher in a snake anxiety questionnaire (Schlangenangst Screening [fear of snakes screening]; SCANS questionnaire; Reinecke, Hoyer, Rinck, & Becker, 2009). The SCANS is a short and time-efficient self-report questionnaire consisting of four items pertaining to the four relevant *Diagnostic and Statistical Manual of Mental Disorders IV* (DSM-IV) diagnosis criteria of snake phobia (persistent fear, anxiety response, avoidance, distress). Participants judge each statement on a 7-point Likert scale (0–6). Reinecke et al. (2009) reported that the SCANS discriminated well between participants fulfilling DSM-IV criteria for specific phobia and healthy controls. Although controls showed a mean SCANS score of 1.5 (*SD* = 1.8), participants with fear of snakes showed a mean score of 18.3 (*SD* = 2.3). In our sample, the mean SCANS score was 17.51 (*Mdn* = 18, *SD* = 3.04, range = 12–23), indicating medium to high fear of snakes. With regard to psychometric properties, the SCANS shows a good test–retest reliability (r = .84) as well as high correlations with other snake questionnaires (convergent validity) like the Snake Anxiety Questionnaire (Klorman, Weerts, Hastings, Melamed, & Lang, 1974; r = .76) or the Aspects of Snake Fear Questionnaire (Suedfeld & Hare, 1977; r = .87).

Participants were free of any neurological or psychiatric illness (except the fear of snakes), did not take any medication at the time of the experiment (oral contraception was allowed), and had normal or corrected-to-normal vision. Participants gave their written informed consent to participate in the study, which was approved by the Ethics Committee of Northwest/Central Switzerland (Registration Number BASEC 2016-01330). All participants received 25 CHF/hr as compensation for their participation.

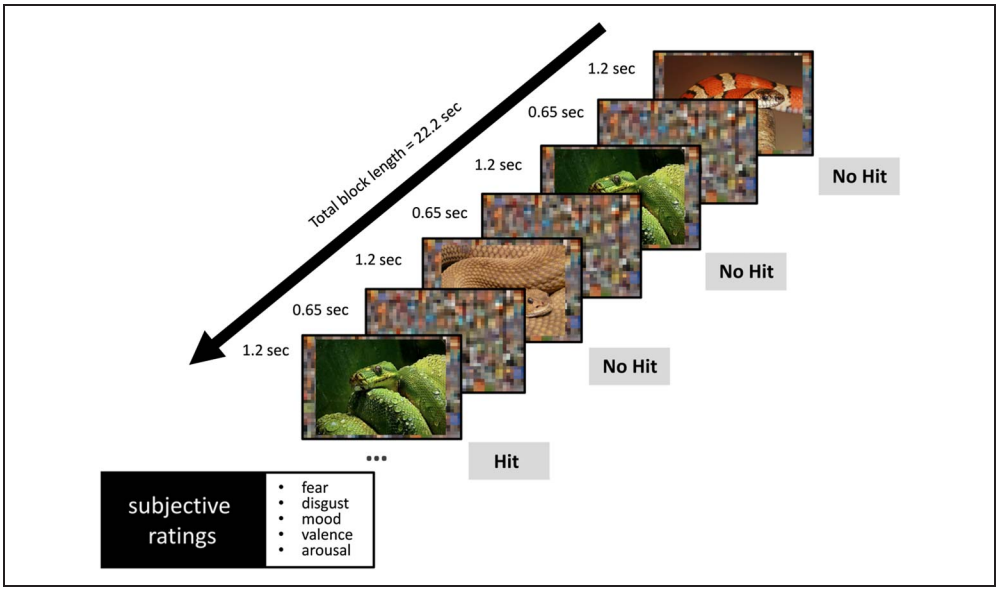


Figure 1. *n*-back task performed during fMRI. The figure illustrates a 2-back task block with snake pictures. Participants had to remember the snake picture presented two positions before and indicate if the currently presented picture was the same (Hit) or a different one (No Hit). During 0-back blocks, a target picture was presented at the beginning of each block, and participants had to respond each time it was presented during the block. For the 2-back and 0-back neutral condition, snake pictures were replaced by pictures of neutral objects. In total, each participant completed 32 *n*-back task blocks (eight blocks of each condition: 0-back/snake, 2-back/snake, 0-back/neutral, and 2-back/neutral).

Stimuli and Task Description

Description of the Pictorial n-back Task

The task designed for this study was a pictorial *n*-back task consisting of two levels of cognitive load (low load: 0-back; high load: 2-back) and two types of pictures (snake pictures, neutral pictures), resulting in four conditions for each participant (within-participant design): (1) 0-back/snake, (2) 2-back/snake, (3) 0-back/neutral, and (4) 2-back/neutral (see Figure 1).

During the 0-back conditions, participants needed to respond as quickly as possible to the occurrence of a target picture (snake or neutral). The 0-back mainly requires general attention processes (Owen et al., 2005) and is thus considered to induce only low task load. In the 2-back condition, participants had to judge whether the currently presented picture was identical to the one presented two positions before. The 2-back condition served as a high load condition because it requires online monitoring, updating, and manipulation of remembered information (Owen et al., 2005).

In total, the task comprised 32 blocks (eight blocks per condition), presented in a quasi-randomized order. In each block, four different pictures of the same picture type were quasi-randomly presented three times, resulting in a total of 12 presented pictures per block. Each block contained three target stimuli and nine nontarget stimuli, resulting in a target rate of 25%. Participants had to react to these targets as quickly as possible. At the beginning of each block, an introduction was displayed for 5 sec to introduce the next task (0-back or 2-back). In case of a 0-back condition, the instruction also comprised a randomly selected picture that served as a target in the following block. After the instruction, a black screen appeared for 1 sec before the block started. Pictures were presented on a scrambled background for 1.2 sec, with only scrambled background between pictures (0.65 sec). Every block lasted for 22.2 sec.

After each block, participants had to rate how they had felt during the last block on five separate visual analog scales via button presses (see Emotion Ratings during *n*-back Task section). The ratings lasted for a total of 40 sec. After a break (random duration, min = 1 sec and max = 8 sec; 20 sec in total over four consecutive blocks), an empty screen was presented for 1 sec until the instructions of the next task block appeared.

Conditions were presented in a quasi-randomized order, that is, each of the four conditions was presented once before being presented again. Furthermore, snake and neutral pictures were assigned to 0-back and 2-back blocks in a counterbalanced fashion. As blocks of snake pictures always depicted the same animal (snake), we took care that neutral pictures also depicted the same type of object within one task block (e.g., chairs) to control blocks for task difficulty. However, the type of neutral object changed for each new task block.

Picture Selection

In total, 64 pictures of snakes and 64 pictures depicting neutral objects were used in this study. All snake pictures and 21 neutral pictures were selected from the Geneva Affective Picture System (Dan-Glauser & Scherer, 2011), and 24 neutral pictures were from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008). Because these standard picture systems did not provide us with a sufficient number of neutral pictures in accordance with our selection criteria, we selected 19 additional neutral pictures from in-house standardized picture sets. Neutral pictures comprised single inanimate objects like chairs, clocks, cups, or shoes. With regard to visual complexity, these pictures were comparable to each other as well as to the selected snake pictures.

Emotion Ratings during n-back Task

To measure participants' emotional reaction to the pictures presented during the task blocks, five separate visual rating scales (11-point Likert scales) were presented after each block. An instruction slide appeared for 5 sec, announcing the first three ratings. Afterward, participants had to indicate how much fear (none–maximal) and how much disgust (none–maximal) they had felt during the last task block as well as the state of their mood (very good–very bad). In total, the participants were given 18 sec to indicate their ratings (on average 6 sec per rating) by moving the curser stepwise to the according scale position on an fMRI-compatible finger-controlled button box. If participants were faster than 18 sec, a cross appeared in the middle of the screen for the remaining time. Next, a second instruction slide appeared for 5 sec, announcing the last two ratings. Participants rated the pictures of the last block according to overall valence (positive–negative) and arousal (low–high). Participants were given a total of 12 sec to indicate their ratings (6 sec per rating).

Experimental Procedure

Before the day of the experiment, participants received general information about the study and filled out an online questionnaire to assess study eligibility. The software SoSci Survey was used for online assessments (Leiner, 2014).

The experiment took place at the University Hospital of Basel. Upon arrival, participants gave written informed consent. They were then trained on the *n*-back task. Only neutral pictures were used for training. The training was repeated if the number of correct responses was lower than 90% in the 0-back or lower than 70% in the 2-back. Afterward, participants entered the scanner. All participants received earplugs and headphones during MR scans to reduce scanner noise and were instructed not to move during the scans. Small foam pads were used for additional head fixation. We used MR-compatible

LCD goggles (VisualSystem, NordicNeuroLab) to present the *n*-back task inside the scanner and to track eye movements during the task. Eye-tracking data were acquired with the ViewPoint eyetracker software (Arrington Research), and calibration was done at the beginning of the experiment. Vision correction was used if necessary. Participants gave their responses via a button box placed on their lower abdomen using the index, middle, and ring finger of their dominant hand. The *n*-back task lasted 40 min and was followed by 10 min of magnetization-prepared rapid gradient-echo and B0 field map acquisition.

We used Presentation software (Version 14.5; Neurobehavioral Systems, Inc., www.neurobs.com) to present the tasks inside the scanner.

Statistical Analysis of the Behavioral Data

All statistical analyses of behavioral data were performed in R (Version 3.3.2; RRID:SCR_001905) by using linear mixed models in combination with ANOVA. The two within-participant factors Cognitive Load (0-back, 2-back) and Picture Type (snake, neutral), as well as the interaction between Cognitive Load and Picture Type, were entered into the model. In case of a significant interaction, post hoc tests were applied separately for each picture type. Participant-IDs were included as random effect. In case that model assumptions were not met (visual inspection of normal distribution and random intercept, Shapiro–Wilk normality test), we used nonparametric two-way repeated ANOVA by means of ANOVA-type statistic (ATS) as provided in the R package *nparLD* (Noguchi, Gel, Brunner, & Konietschke, 2012). The ATS rank-based method tests the hypothesis of equality of distributions rather than the equality of means (Shah & Madden, 2004).

Assessment of n-back Performance

To assess whether 2-back blocks were more demanding and therefore induced a higher cognitive load than 0-back blocks, we measured participants' task performance, that is, accuracy (hits plus correct rejections divided by the total number of pictures shown) and d' (Stanislaw & Todorov, 1999) measures. Two separate statistical models were calculated, with accuracy and d' serving as dependent variables.

Preprocessing and Analysis of Eye-tracking Data

For each participant, fixation detection was performed with an individual, velocity-based algorithm using the *saccades* package (von der Malsburg, 2015) and a lower fixation duration threshold of 100 msec. The analysis was restricted to ROIs, which were manually defined for each picture as the region covered by the main object in the picture (e.g., a snake or a chair; opposed to the background of the picture). The average dwell time in a ROI was calculated for each task condition to quantify the overt attention drawn to these regions.

Analysis of Emotion Ratings during n-back Task

As an experimental manipulation check, we first investigated for each emotion rating whether snake pictures, on average, induced more negative emotions than neutral pictures, irrespective of cognitive load. Here, we calculated separate dependent two-sided *t* tests for each rating of the *n*-back task.

Our main analyses of interest (interaction of Cognitive Load × Picture Type and effect of load) were performed by using separate statistical models for each of the emotion ratings.

As the rating of fear during the *n*-back task constituted our primary variable of interest, we set the significance threshold to $p < .05$ for this rating. Bonferroni correction was implemented to account for multiple testing for all remaining ratings (disgust, mood, valence, and arousal; Bonferroni correction for four independent tests).

fMRI Data Acquisition

Measurements were performed on a Siemens Magnetom SkyraFit 3 T whole-body MR unit equipped with a 32-channel head coil. Functional series were acquired using a single-shot echo-planar sequence using parallel imaging (Generalized Autocalibrating Partial Parallel Acquisition; GRAPPA). The following acquisition parameters were applied: echo time (TE) = 30 msec, field of view = 24 cm, acquisition matrix = 96 × 96, voxel size = 2.5 × 2.5 × 3 mm^3, GRAPPA acceleration factor $R = 2.0$. Using a mid-sagittal scout image, 42 contiguous axial slices placed along the AC–PC plane covering the entire brain with a repetition time (TR) of 2600 msec ($\alpha = 82°$) were sampled with an ascending interleaved sequence. A high-resolution T1-weighted anatomical image was acquired for each participant using a magnetization-prepared rapid gradient echo (TR = 2000 msec, TE = 2.26 msec, inversion time = 1000 msec, flip angle = 8°, 176 slices, field of view = 256 mm, voxel size = 1.5 × 1.5 × 1.5 mm^3).

To correct the fMRI data for geometric distortions caused by magnetic field inhomogeneities, B0 field-map scans were collected as well (TR = 550 msec, TE = 4.92 msec/7.38 msec, flip angle = 60°, voxel size = 2.5 × 2.5 × 3.0 mm^3).

Processing of Structural MRI Data and Construction of Probabilistic Atlas

Each participant's anatomical image was automatically segmented into cortical and subcortical structures using FreeSurfer v5.3.0 (Fischl et al., 2002). Labeling of the cortical gyri was based on the Desikan–Killiany atlas (Desikan et al., 2006), yielding 35 cortical and 7 subcortical regions per hemisphere.

The segmentations were used to build a population-averaged probabilistic anatomical atlas. Individual segmented anatomical images were subsequently normalized to the study-specific anatomical template space using the participant's previously computed warp field and affine-registered to the Montreal Neurological Institute (MNI) space. Nearest-neighbor interpolation was applied to preserve labeling of the different structures. The normalized segmentations were finally averaged across all 43 participants to create a population-averaged probabilistic atlas. Each voxel of the template could consequently be assigned a probability of belonging to a given anatomical structure.

fMRI Data Analysis

Preprocessing and First-level Analysis

Analyses were performed using SPM12 (Version 6470; Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging; www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB R2014b (The Mathworks, Inc.).

To account for magnetization effects, the first four volumes were discarded from further analyses. The remaining volumes were slice-time-corrected to the first slice, realigned and unwarped with the field maps, and coregistred to the anatomical image by applying a normalized mutual information 3-D rigid body transformation. Successful coregistration was visually verified for every participant. Each volume was masked with the participant's T1 anatomical image to exclude voxels outside the brain. The EPI volumes were normalized to MNI space by applying DARTEL, which leads to an improved registration between participants (Klein et al., 2009; Ashburner, 2007). Normalization incorporated the following steps: (1) structural images of each participant were segmented using the "Segment" procedure in SPM12. (2) The resulting gray and white matter images were used to derive a study-specific group template. The template was computed from all participants included in this study ($n = 43$). (3) An affine transformation was applied to map the group template to MNI space. (4) Participant-to-template and template-to-MNI transformations were combined to map the functional images to MNI space. The functional images were smoothed with an isotropic 5-mm FWHM Gaussian filter.

Normalized functional images were masked using information from their respective T1 anatomical file as follows: At first, the three-tissue classification probability maps of the "Segment" procedure (gray matter, white matter, and CSF) were summed to define a brain mask. The mask was binarized, dilated, and eroded with a $3 \times 3 \times 3$ voxels kernel using fslmaths (FSL) to fill in potential small holes in the mask. The previously computed DARTEL flowfield was used to normalize the brain mask to MNI space, at the spatial resolution of the functional images. The resulting nonbinary mask was thresholded at 50% and applied to the normalized functional images. Consequently, the implicit intensity-based masking threshold usually employed to compute a brain mask from the functional data during the first-level specification (*spm_get_defaults('mask.thresh')*, by default fixed at .8) was not required anymore and therefore set to a lower value of .05.

Analyses were conducted in the framework of the general linear model. Intrinsic autocorrelations were accounted for by AR(1), and low-frequency drifts were removed via a high-pass filter (time constant, 128 sec). Regressors, which modeled the onset and duration of each block, were convolved with a canonical hemodynamic response function. Separate regressors were constructed for each of the four *n*-back conditions: (1) 0-back neutral, (2) 0-back snake, (3) 2-back neutral, and (4) 2-back snake. Events between blocks, that is, task instructions, ratings, and breaks, were modeled as separate regressors. Additionally, six movement regressors from spatial realignment were included as regressors of no interest.

The resulting parameter estimates were used to specify contrasts using fixed effects models (first-level analysis). The following contrasts were specified: (1) "picture contrast": brain activity related to presentation of snake pictures compared with neutral pictures (snake pictures–neutral pictures), independently of whether the picture was shown under 0-back or 2-back; (2) "load contrast": brain activity related to pictures presented under 0-back or 2-back (0-back–2-back), irrespective of whether the picture depicted a snake or a neutral object; and (3) "interaction contrast": brain activity related to the interaction of load and picture type ([0-back snake–2-back snake]–[0-back neutral–2-back neutral]).

Group-level Analysis

The single-participant contrast maps of the first-level analysis were entered in a random effects model to make inferences on group level. We controlled for sex and age by including them as covariates. As the amygdala and the dlPFC served as ROIs in this analysis, we applied a small volume correction (SVC) for these regions. We first created one probabilistic mask for the amygdala and one for the dlPFC by combining the respective masks from both hemispheres. These masks were taken from the population-specific atlas (corresponding Freesurfer labels for dlPFC mask: ctx-lh-rostralmiddlefrontal/ctx-rh-rostralmiddlefrontal). The probabilistic masks were consequently thresholded at 50%, binarized and applied to the group-level contrast maps ($p < .05$, family-wise error [FWE]-corrected for multiple comparisons within the mask [$p_{\text{FWE-SVC}}$]).

DCM: Extracting Time Courses from Volumes of Interest

We used DCM to investigate a possible inhibition of amygdala activity through top–down control of prefrontal regions when cognitive load was high. Volumes of interest

(VOIs) were defined as the amygdala and the dlPFC. Time courses were extracted separately per hemisphere.

The applied approach was similar to the one used by Fastenrath et al. (2014). First, we identified local maxima at the group level for each of the four anatomical masks (amygdala and dlPFC from both hemispheres). Local maxima were based on the load contrast (0-back–2-back). Second, group-level coordinates in MNI space were mapped to native participant space. Based on these participant space coordinates, participant-specific local maxima were identified within a distance of 10 mm. Time courses were extracted by computing the principal eigenvariate of the data across all significant voxels ($p <$ .05 uncorrected, minimum cluster size 3) within a 10-mm sphere around the participant-specific local maxima and within the participant-specific anatomical mask (masks were retrieved from the FreeSurfer segmentations). The application of the aforementioned p value threshold of .05 with a minimal cluster size of 3 allowed us to separate voxels with task-related signal from voxels with noisy signal (see, e.g., Stephan et al., 2010). This procedure implies that time series are extracted only from those voxels reaching this threshold. As data from all VOIs in all participants are a prerequisite to run DCM (Stephan et al., 2010; Friston, Harrison, & Penny, 2003), participants who did not show sufficient activation in line with these criteria were excluded from further DCM analysis. Consequently, 9 of 43 participants had to be excluded per hemisphere. The extracted time courses were adjusted to the F contrast (i.e., effects of interest) of each participant and entered into the DCM models.

DCM: Defining Model Space and Model Comparison

We applied bilinear, deterministic DCM with two states (Version DCM12 r6432 in SPM 12 r6470; Marreiros, Kiebel, & Friston, 2008). We ran DCM for the left and right hemispheres separately. In each hemisphere, models were set up consisting of two nodes, corresponding to the amygdala and the dlPFC, respectively. We allowed full bidirectional connectivity between the two nodes. The two load conditions 0-back and 2-back, as well as instructions and emotion ratings, served as driving input to either one of the regions or to both regions. This resulted in three input possibilities to the network, that is, three different models per hemisphere. Within each model, the connections between both regions could be modulated by either the 0-back or the 2-back condition, irrespective of picture type. We focused on the difference in task load because we did not find any significant interaction effect between task load and picture type in the fMRI group-level analysis, which would have been a prerequisite to extract peak coordinates for the DCM analysis.

DCM is based on Bayesian statistics. The model evidence denotes the probability of the data given the model while adjusting for model complexity and dependencies among parameters (Penny et al., 2010; Stephan, Penny, Daunizeau, Moran, & Friston, 2009). Models were compared by conducting random-effects Bayesian model selection (BMS; Penny et al., 2010; Stephan et al., 2009) and differed only in the location of the driving input. This allowed us to test whether one of the models was more likely than any of the other two, which is expressed as exceedance probability.

DCM: Bayesian Model Averaging and Parameter Analysis

Bayesian model averaging (BMA) was applied to obtain a summary measure of likely connectivity values (Penny et al., 2010). The connectivity parameters of each model were weighted by the posterior model probability and subsequently averaged within each participant. Overall, the BMA weighting procedure resulted in participant-specific connectivity estimates that were independent of a particular model while ensuring that models with a high probability contributed more than models with a lower probability.

The BMA modulatory parameter estimates of each participant were further analyzed using R (www.r-project.org). We checked for possible sex and age effects by calculating linear mixed models. As we did not find any significant effects of sex or age ($p > .05$), these variables were not considered in the following analyses.

To test for differences in connectivity strength, we applied the same statistical approach as for the analysis of behavioral data (see Statistical Analysis of the Behavioral Data section). Bonferroni correction was implemented to account for multiple testing (left and right hemisphere; $p < .025$).

RESULTS

Behavioral Results

n-back Performance

Performance in the 0-back condition was significantly better than in the 2-back condition (accuracy: ATS(1) = 261.47, $p < 8.2 \times 10^{-59}$; d': $t(126) = 17.42$, $p < 2.4 \times 10^{-35}$; see also Table 1), indicating effective manipulation of load. There was no significant interaction between Cognitive Load and Picture Type or main effect of Picture Type on n-back performance (accuracy: $p > .07$; d': $p > .19$).

Eye-tracking

We used eye-tracking to investigate if the average dwell time in informative picture regions, as an index for overt attention, varied depending on cognitive load, picture type, or their interaction. Such a difference in overt attention allocation might have affected fear ratings by altering the visual input of fear-inducing information.

For the average dwell time in ROI of pictures, there was neither an effect of Cognitive Load ($p = .35$) nor of Picture Type ($p = .14$) nor an interaction effect ($p = .42$). This

Table 1. *n*-back Performance (Accuracy, d')

	Neutral Pictures		*Snake Pictures*	
	0-back	*2-back*	*0-back*	*2-back*
Accuracy	0.97 (0.03)	0.86 (0.07)	0.97 (0.04)	0.88 (0.07)
d'	3.57 (0.4)	2.39 (0.58)	3.62 (0.48)	2.51 (0.73)

Depicted are mean and standard deviation (in parentheses).

finding indicates that participants spent an equal amount of time looking at relevant regions of the picture, irrespective of valence and load.

Emotion Ratings

As an experimental manipulation check, we first investigated whether snake pictures induced more negative emotions than neutral pictures, irrespective of cognitive load. Participants reported significantly more fear, $t(42) = 17.16$, $p < 1.4 \times 10^{-20}$; disgust, $t(42) = 20.52$, $p < 1.7 \times 10^{-23}$; negative mood, $t(42) = 10.42$, $p < 3.3 \times 10^{-13}$; negative valence, $t(42) = 11.95$, $p < 4.3 \times 10^{-15}$; and arousal, $t(42) = 12.41$, $p < 1.3 \times 10^{-15}$. after blocks depicting snake pictures compared with neutral pictures, indicating that snake pictures evoked more negative emotions than neutral pictures.

Subsequently, we investigated whether there were any interaction effects between Cognitive Load and Picture Type on emotion ratings and whether emotion ratings differed between levels of cognitive load. We found a significant interaction between Cognitive Load and Picture Type for ratings of fear, ATS(1) = 10.5, $p = .001$, and arousal, ATS(1) = 8.92, $p_{\text{Bonferroni corrected}} = .011$, and a trend for disgust, ATS(1) = 6.0, $p_{\text{Bonferroni corrected}} = .057$, and valence, ATS(1) = 5.54, $p_{\text{Bonferroni corrected}} = .074$, but not for mood, ATS(1) = 4.42, $p_{\text{Bonferroni corrected}} = .14$. Consequently, we analyzed ratings of fear and arousal separately for snake pictures and neutral pictures. We also ran post hoc tests for the remaining emotion ratings. As fear rating served as the primary variable of interest, the respective p value threshold was set to $<.05$ in all analyses, whereas p values for all other ratings (disgust, mood, arousal, and valence) were Bonferroni-corrected to account for multiple testing (see Methods section).

Post hoc tests revealed that snake pictures presented during 2-back blocks evoked less fear, ATS(1) = 8.1, $p = .004$, and less disgust, ATS(1) = 9.13, $p_{\text{Bonferroni corrected}} = .01$, than snake pictures presented during 0-back blocks. No significant effects of Load on mood, valence, or arousal were found (all $p_{\text{Bonferroni corrected}} > .45$; see Table 2).

In contrast, neutral pictures presented during 2-back blocks resulted in higher fear ratings, ATS(1) = 4.84, $p = .028$, than neutral pictures presented during 0-back blocks. This finding could reflect an increased fear of failure during the cognitively demanding task blocks compared with low demanding ones. For all other emotion ratings on neutral pictures, no significant effects of Load were found (all $p_{\text{Bonferroni corrected}} > .17$).

fMRI Results for ROIs

Main Effect of Picture Type

Voxel-wise analysis revealed that the amygdala was bilaterally activated during blocks of snake pictures compared

Table 2. Effect of Cognitive Load on Emotional Ratings of Snake and Neutral Pictures during the *n*-back Task

Emotion Ratings	*0-back*	*2-back*	*2-back–0-back*	*p*
Snake pictures				
Fear	62.14 (17.13)	60.47 (16.52)	−1.68 (3.95)	.004*
Disgust	69.09 (17.59)	66.95 (17.05)	−2.14 (4.39)	.003*
Mood	62.78 (12.09)	62.18 (12.43)	−0.59 (4.52)	.53
Valence	65.93 (11.74)	65.03 (11.00)	−0.90 (3.71)	.36
Arousal	64.68 (15.75)	63.46 (15.46)	−1.22 (5.03)	.11
Neutral pictures				
Fear	8.87 (8.92)	11.42 (10.46)	2.56 (6.67)	.03*
Disgust	7.94 (7.7)	9.35 (8.78)	1.41 (5.62)	.13
Mood	29.87 (14.8)	31.84 (15.55)	1.97 (7.72)	.1
Valence	30.87 (12.69)	32.33 (14.49)	1.45 (6.64)	.06
Arousal	20.98 (14.66)	23.33 (14.5)	2.36 (6.84)	.04

Depicted are mean and standard deviation (in parentheses) of each condition, as well as the difference score and nominal p values for the effect of cognitive load derived by nonparametric ATS.

*For fear ratings, $p < .05$; for all other ratings $p < .125$ (Bonferroni correction for four comparisons).

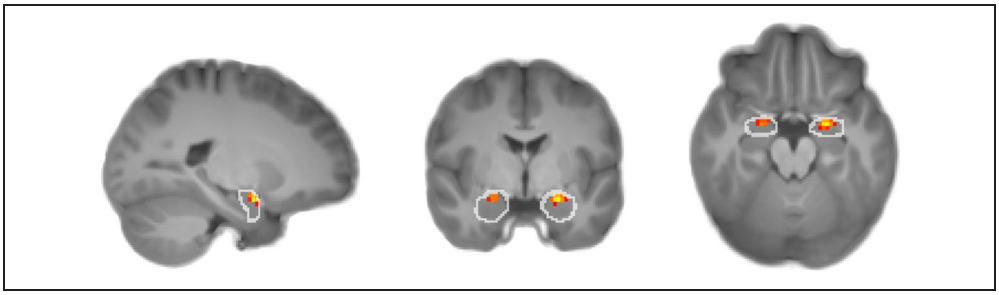


Figure 2. Amygdala activity for the picture contrast (snake pictures–neutral pictures) on group level. Increased amygdala activity during snake pictures as compared with neutral pictures. Depicted are only significant voxels (yellow to red) within the probabilistic amygdala mask (white circles) used for SVC ($p_{FWE\text{-}SVC} < .05$).

with blocks of neutral pictures (peak voxel: left: MNI −22, −2, −18, $t = 6.23$, $k = 23$; right: MNI 22, −2, −18, $t = 4.95$, $k = 10$; $p_{FWE\text{-}SVC} < .05$, corresponding to $t_{SVC} = 3.56$; see Figure 2). The reported activation in the left amygdala also survived whole-brain correction ($t = 5.75$, $p_{FWE} < .05$; see Supplementary Table 1 on whole-brain corrected results[1]). Activation in the same direction was also found in a small cluster in the right dlPFC (peak voxel: MNI −25, 50, 33, $t = 4.72$, $k = 3$), but not in the left dlPFC. We did not observe any effects of sex or age with regard to this activation.

Main Effect of Cognitive Load

Amygdala activity was reduced during 2-back blocks compared with 0-back blocks in both the left hemisphere (peak voxel: MNI −22.5, −10, −15, $t = 8.06$, $k = 73$) and right hemisphere (peak voxel: MNI 22.5, −7.5, −15, $t = 6.71$, $k = 66$; $p_{FWE\text{-}SVC} < .05$, corresponding to $t_{SVC} = 3.5$; see Figure 3A). Furthermore, 2-back blocks resulted in higher bilateral dlPFC activity compared with 0-back blocks (peak voxel: left: MNI −47.5, 25, 33, $t = -10.64$, $k = 640$; right: MNI 45, 30, 39, $t = -11.05$, $k = 357$; $p_{FWE\text{-}SVC} < .05$, corresponding to $t_{SVC} = 4.42$; see Figure 3B). All reported peak activations also survived whole-brain correction ($t = 5.75$, $p_{FWE} < .05$; see Supplementary Tables 2 and 3 on whole-brain corrected results). No significant sex or age effects were found.

Interaction between Cognitive Load and Picture Type

There was no significant interaction effect between cognitive load and picture type on amygdala activity when applying SVC ($p_{FWE\text{-}SVC} > .05$), indicating that there was a similar decrease in amygdala activity during 2-back compared with 0-back for snake pictures and for neutral pictures. No interaction effects were observed for the dlPFC either ($p_{FWE\text{-}SVC} > .05$). Furthermore, we did not observe any whole-brain corrected interaction effects (Cognitive Load × Picture Type).

DCM Results

Extraction of Time Courses in VOIs

Activity in the respective peak voxel within the amygdala and the dlPFC were obtained from the load contrast (0-back–2-back) of the group-level analysis (for peak coordinates, see Table 3). We focused only on the contrast of load as we did not find any interaction

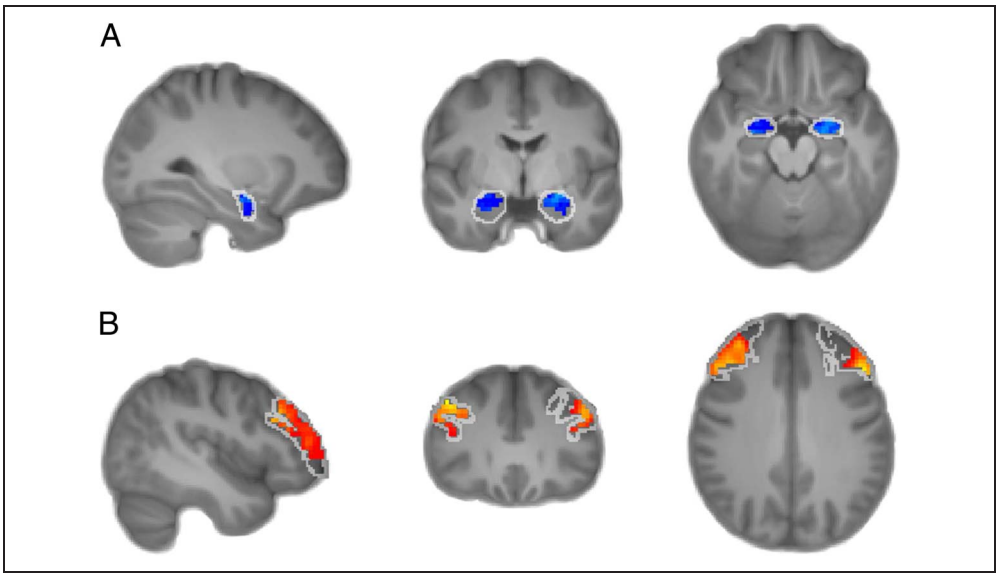


Figure 3. Amygdala and dlPFC activity for the load contrast (0-back–2-back) on group level. Decreased amygdala activity during 2-back blocks as compared with 0-back blocks (A, blue color). Increased dlPFC activity during 2-back blocks as compared with the 0-back blocks (B, yellow to red color). Depicted are significant voxels within the respective probabilistic masks used for SVC ($p_{FWE\text{-}SVC} < .05$).

Table 3. Peak Activation in VOIs Extracted for DCM from the Load Contrast (0-back–2-back) of the Group-level Analysis

Cluster No.	*Max. t Value within Cluster*	*Regional Correspondence of Maximum*	*MNI Coordinates at Maximum*			*No. of Voxels*
			x	*y*	*z*	
1	−8.06	Left amygdala (53%)	−22.5	−10	−15	66
2	−6.71	Right amygdala (86%)	22.5	−7.5	−15	44
3	−10.64	ctx-lh-rostralmiddlefrontal (57%)	−47.5	25	33	640
4	−11.05	ctx-rh-rostralmiddlefrontal (54%)	45	30	39	357

Regions and probabilities are in accordance with population-specific atlas. ctx = cortex; lh = left hemisphere; rh = right hemisphere.

between cognitive load and picture type in the fMRI group-level analysis.

Robust task-related activation in all regions in all participants is a prerequisite to run DCM (see Methods section). Five participants did not show task-related activation, neither in the left nor in the right amygdala. Another four participants did not show task-related activation in the left amygdala and another four participants in the right amygdala. Per hemisphere, time courses were consequently available for 34 of 43 participants.

Model Comparison Using BMS

We constructed three different models based on the site of input to the network. BMS was used to identify the most plausible model given the data. Comparison between the three models indicated that the model where the input entered both regions (amygdala and dlPFC) was most likely (left hemisphere: model exceedance probability = .995; right hemisphere: model exceedance probability = .605) even though there was a considerably high exceedance probability for input only to the dlPFC in the right hemisphere as well (exceedance probability = .395).

Modulatory Influence of Load on Connectivity Parameters

As there was no clearly superior input model for the right hemisphere, we used BMA to calculate connectivity parameters. Modulators of effective connectivity describe whether the strength of the connection (i.e., the influence that one region exerts upon another) increases or decreases under the influence of experimental manipulations (Friston et al., 2003). Modulator estimates were calculated for the 0-back and for the 2-back condition. In addition to modulators of connection strength, DCM also estimates intrinsic connectivity parameters. They represent the connectivity in the absence of experimental perturbations. The effective connection strength associated with the experimental condition can be obtained by adding the value of the intrinsic connection and the value of the modulator. A negative sum indicates that the activity in the source region decreases (inhibits) activity in the target region (Sokolov et al., 2018).

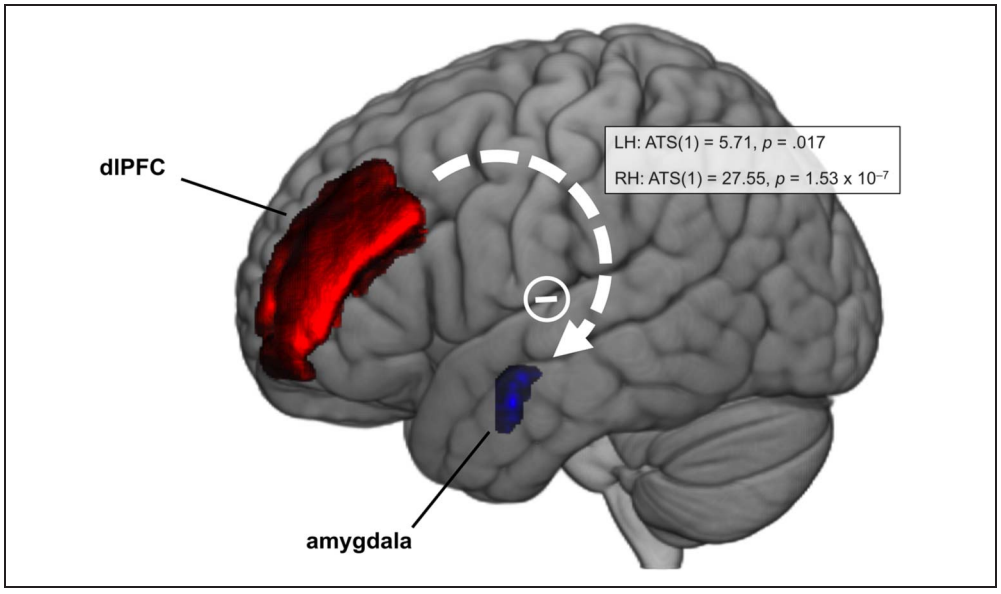


Figure 4. Change in connectivity between the dlPFC (red) and the amygdala (blue) during 2-back blocks compared with 0-back blocks. The white arrow indicates the influence from the dlPFC to the amygdala. The ATS values in the box indicate a stronger decrease in connectivity strength during the 2-back than during the 0-back in the left hemisphere (LH) and the right hemisphere (RH). For parameter values per condition, see Tables 4 and 5.

Table 4. BMA Parameter Values across All 34 Participants in the Left Hemisphere

	Parameter Values (Left Hemisphere)			
Direction of connection	dlPFC to amygdala		Amygdala to dlPFC	
Strength of intrinsic connectivity	−0.019 (0.019)		0.020 (0.012)	
	0-back	*2-back*	*0-back*	*2-back*
Change in connection strength (modulators) per condition	−0.002 (0.021)	−0.200 (0.066)	−0.015 (0.012)	0.003 (0.012)
Total connectivity	−0.021 (0.035)	−0.219 (0.067)	0.005 (0.020)	0.023 (0.019)

Depicted are mean and *SEM* (in parentheses).

The modulation of connectivity parameters was analyzed on group level. Visual inspection of the data distribution and subsequent statistical testing indicated a deviation from normality in both hemispheres (Shapiro–Wilk test: $p < .05$). For this reason, we report p values of the nonparametric statistical model to ensure robustness of the results.

Modulators of connection strength from the dlPFC to the amygdala differed significantly between the two load conditions in both hemispheres (left: ATS(1) = 5.71, $p = .017$; right: ATS(1) = 27.55, $p = 1.53 \times 10^{-7}$; see Figure 4 and Tables 4 and 5), showing an inhibitory influence of the dlPFC on the amygdala during the 2-back condition. For the direction from the amygdala to the dlPFC, no significant differences in connectivity strength were found between 2-back and 0-back (both hemispheres: $p > .26$).

We performed an additional DCM analysis (based on the same VOIs) that also accounted for possible effects of picture type on functional connectivity between dlPFC and amygdala. This analysis neither revealed a significant interaction effect (both hemispheres: $p > .21$) nor a main effect of picture type (both hemispheres: $p > .11$). The main effect of load remained significant (left: ATS(1) = 28.27, $p = 1.08 \times 10^{-7}$; right: ATS(1) = 51.24, $p = 8.16 \times 10^{-13}$). For the direction from the amygdala to the dlPFC, there was neither a significant interaction effect (both hemispheres: $p > .38$) nor a main effect of cognitive load (both hemispheres: $p > .22$). There was, however, a significant main effect of picture type in the left hemisphere, ATS(1) = 6.92, $p = .009$ ($M_{snake} = 0.015$, $SD = 0.06$; $M_{neutral} = -0.013$, $SD = 0.12$), but not in the right hemisphere ($p = .75$).

DISCUSSION

The confrontation with a feared object or situation typically results in an acute activation of limbic brain regions, including the amygdala, and in perceived fear. In the present fMRI study, we investigated if high cognitive load, induced via a demanding task, would lead to reduced activity of the amygdala as well as to attenuated phobic fear in participants with fear of snakes.

The findings revealed a decrease in amygdala activity during blocks of high cognitive load compared with blocks of low cognitive load. Importantly, we also observed acute effects of cognitive load on subjective fear ratings of snake pictures. When task load was high, participants reported less phobic fear in the presence of fearful pictures. To our knowledge, these results are the first to indicate that engaging in a highly demanding cognitive task can acutely decrease perceived fear toward phobic stimuli.

Applying DCM, we additionally investigated load-depended changes in functional connectivity between the dlPFC and the amygdala. We focused on the effect of task load in the DCM analysis as we did not detect any significant interaction effects between task load and picture type in the initial group-level analysis. This suggests

Table 5. BMA Parameter Values across All 34 Participants in the Right Hemisphere

	Parameter Values (Right Hemisphere)			
Direction of connection	dlPFC to amygdala		Amygdala to dlPFC	
Strength of intrinsic connectivity	0.014 (0.047)		0.075 (0.050)	
	0-back	*2-back*	*0-back*	*2-back*
Change in connection strength (modulators) per condition	0.056 (0.031)	−0.316 (0.078)	0.006 (0.011)	0.005 (0.024)
Total connectivity	0.071 (0.064)	−0.301 (0.110)	0.081 (0.053)	0.080 (0.040)

Depicted are mean and *SEM* (in parentheses).

a load-dependent decrease in amygdala activation irrespective of whether the presented pictures depicted fearful or neutral objects. This parallels findings from previous experiments in which solving a difficult task has been associated with a decrease in amygdala activity, independently of whether neutral or emotional stimuli were used (Cohen et al., 2016; Straube et al., 2011; Silvert et al., 2007). Our DCM analysis revealed a bilateral decrease in connectivity strength from the dlPFC to the amygdala during high cognitive load compared with low task load, indicating that the dlPFC exerted a stronger inhibitory influence on the amygdala when task demand was high.

What might be the mechanism(s) of the observed reduction of phobic fear under high cognitive load? One potential explanation for reduced amygdala activity and fear is that the high-load condition caused a visual distraction from the emotional pictures. However, as the pictures were used as targets in the *n*-back task, participants were "forced" to process the pictorial information in both load conditions. Our eye-tracking results indicate that, in the current design, participants spent a similar amount of time in the emotional ROIs of the snake pictures in both load conditions. Thus, reduced amygdala activity and fear as a consequence of visual distraction from emotional hotspots appears unlikely. This is in line with a study indicating that cognitive load does not alter dwell time on anxiety-related stimuli (Berggren, Koster, & Derakshan, 2012; see also MacNamara & Proudfit, 2014; Berggren, Richards, Taylor, & Derakshan, 2013). Instead, we argue that the observed fear reduction may have been a result of a load-induced top–down control mechanism, that is, an increase in dlPFC activity when participants engage in a cognitively demanding task induces a decrease in amygdala activity (Okon-Singer et al., 2015). In line with this idea, the current study found that reduced amygdala activation during high load could be explained by a change in effective connectivity between lateral prefrontal regions and the amygdala. The purpose for such a top–down regulation of amygdala activity might be the reallocation of resources from emotional processes to the cognitive task. It has been shown that the presentation of a distractor during a highly demanding task results in a competition for limited cognitive resources (Lavie, 2010; Lavie, Hirst, De Fockert, & Viding, 2004; Desimone & Duncan, 1995). Attention is directed toward accomplishing the cognitively demanding task, leaving little capacity to process the emotional stimulus, which in turn results in attenuated neuronal and behavioral reactions toward emotional stimuli (Balderston et al., 2016; Vytal et al., 2012; Mitchell et al., 2007; Bishop, Jenkins, & Lawrence, 2006). Even though it has been shown that the dlPFC does not have strong direct anatomical connections to the entire amygdala (Ray & Zald, 2012), studies suggest that the dlPFC might functionally interact with the amygdala either through direct projections to the basal nucleus of the amygdala (Birn et al., 2014) or over indirect pathways via the subgenual cingulate gyrus, dorsal ACC, OFC, or ventrolateral PFC (Clarke & Johnstone, 2013; Sladky et al., 2013; Ray & Zald, 2012). The investigation of potential mediating brain regions by load-dependent functional connectivity is interesting and should be considered in future studies examining top–down regulation effects on amygdala activity. Also, the use of TMS could reveal new insights about the involvement of specific cortical regions in processing stimuli under high cognitive load (see, e.g., Schicktanz et al., 2015). Taken together, the mechanism of reduced amygdala activity and reduced fear under high cognitive load may involve a top–down regulation and a reallocation of cognitive resources.

In summary, the findings of this study suggest that high cognitive load, induced by an *n*-back task comprising fearful stimuli, not only decreases amygdala activity but also reduces perceived phobic fear toward fearful stimuli. Future studies may investigate the acute effects of even higher load conditions on amygdala activity and fear. Furthermore, it would be interesting to look into the therapeutic potential of cognitive tasks in anxiety. It has been reported that distraction from anxiogenic stimuli might be a promising approach to treat anxiety-related disorders (see, e.g., Price, Paul, Schneider, & Siegle, 2013; Vytal et al., 2012; Oliver & Page, 2003). Our study now points to an additional beneficial effect of tasks involving high cognitive load, leading to a top–down regulation of amygdala activity and fear. Repeated exposure to phobic stimuli paired with reduced amygdala activity and reduced fear might facilitate fear extinction. In support of this idea, it has recently been shown that extinction learning can be improved in states of reduced amygdala activity (de Voogd et al., 2018). Thus, our findings might contribute to the development of novel psychological treatment approaches aimed at reducing fear in anxiety disorders.

Acknowledgments

The study was funded by the Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, Switzerland.

Reprint requests should be sent to Dominique de Quervain, Division of Cognitive Neuroscience, University of Basel, Birmannsgasse 8, CH-4055 Basel, or via e-mail: dominique.dequervain@unibas.ch.

Note

1. All three Supplementary Tables, as well as most relevant scripts for first and second level analysis, and DCM analysis can be accessed via this link: https://osf.io/rtz3d/?view_only=218b422676334ef6b4cb13024afc5b53. Furthermore, second level contrasts for analysis of cognitive load, picture type and their interaction can be found on Neurovault: https://neurovault.org/collections/FFXEDLNE/.

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. R. (1995). Fear and the human amygdala. *Journal of Neuroscience*, *15*, 5879–5891.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*, 95–113.
Balderston, N. L., Quispe-Escudero, D., Hale, E., Davis, A., O'Connell, K., Ernst, M., et al. (2016). Working memory maintenance is sufficient to reduce state anxiety. *Psychophysiology*, *53*, 1660–1668.
Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex*, *49*, 1195–1205.
Berggren, N., Koster, E. H., & Derakshan, N. (2012). The effect of cognitive load in emotional attention and trait anxiety: An eye movement study. *Journal of Cognitive Psychology*, *24*, 79–91.
Berggren, N., Richards, A., Taylor, J., & Derakshan, N. (2013). Affective attention under cognitive load: Reduced emotional biases but emergent anxiety-related costs to inhibitory control. *Frontiers in Human Neuroscience*, *7*, 188.
Birn, R. M., Shackman, A. J., Oler, J. A., Williams, L. E., McFarlin, D. R., Rogers, G. M., et al. (2014). Evolutionarily conserved prefrontal-amygdalar dysfunction in early-life anxiety. *Molecular Psychiatry*, *19*, 915–922.
Bishop, S. J., Jenkins, R., & Lawrence, A. D. (2006). Neural processing of fearful faces: Effects of anxiety are gated by perceptual capacity limitations. *Cerebral Cortex*, *17*, 1595–1603.
Clarke, R. J., & Johnstone, T. (2013). Prefrontal inhibition of threat processing reduces working memory interference. *Frontiers in Human Neuroscience*, *7*, 228.
Cohen, N., Margulies, D. S., Ashkenazi, S., Schäfer, A., Taubert, M., Henik, A., et al. (2016). Using executive control training to suppress amygdala reactivity to aversive information. *Neuroimage*, *125*, 1022–1031.
Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*, 415–423.
Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva Affective Picture Database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*, 468–477.
Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*, 13–34.
de Voogd, L. D., Kanen, J. W., Neville, D. A., Roelofs, K., Fernández, G., & Hermans, E. J. (2018). Eye-movement intervention enhances extinction via amygdala deactivation. *Journal of Neuroscience*, *38*, 8694–8706.
Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*, 968–980.
Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
Dilger, S., Straube, T., Mentzel, H.-J., Fitzek, C., Reichenbach, J. R., Hecht, H., et al. (2003). Brain activation to phobia-related pictures in spider phobic humans: An event-related functional magnetic resonance imaging study. *Neuroscience Letters*, *348*, 29–32.
Dolcos, F., & Denkova, E. (2014). Current emotion research in cognitive neuroscience: Linking enhancing and impairing effects of emotion on cognition. *Emotion Review*, *6*, 362–375.
Dolcos, F., Iordan, A. D., & Dolcos, S. (2011). Neural correlates of emotion–cognition interactions: A review of evidence from brain imaging investigations. *Journal of Cognitive Psychology*, *23*, 669–694.
Erk, S., Kleczar, A., & Walter, H. (2007). Valence-specific regulation effects in a working memory task with emotional context. *Neuroimage*, *37*, 623–632.
Fanselow, M. S., & Gale, G. D. (2003). The amygdala, fear, and memory. *Annals of the New York Academy of Sciences*, *985*, 125–134.
Fastenrath, M., Coynel, D., Spalek, K., Milnik, A., Gschwind, L., Roozendaal, B., et al. (2014). Dynamic modulation of amygdala–hippocampal connectivity by emotional arousal. *Journal of Neuroscience*, *34*, 13935–13947.
Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*, 341–355.
Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, *19*, 1273–1302.
Iordan, A. D., Dolcos, S., & Dolcos, F. (2013). Neural signatures of the response to emotional distraction: A review of evidence from brain imaging investigations. *Frontiers in Human Neuroscience*, *7*, 200.
Ipser, J. C., Singh, L., & Stein, D. J. (2013). Meta-analysis of functional brain imaging in specific phobia. *Psychiatry and Clinical Neurosciences*, *67*, 311–322.
Janak, P. H., & Tye, K. M. (2015). From circuits to behaviour in the amygdala. *Nature*, *517*, 284–292.
King, R., & Schaefer, A. (2011). The emotional startle effect is disrupted by a concurrent working memory task. *Psychophysiology*, *48*, 269–272.
Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M. C., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*, *46*, 786–802.
Klorman, R., Weerts, T. C., Hastings, J. E., Melamed, B. G., & Lang, P. J. (1974). Psychometric description of some specific-fear questionnaires. *Behavior Therapy*, *5*, 401–409.
Kohn, N., Eickhoff, S. B., Scheller, M., Laird, A. R., Fox, P. T., & Habel, U. (2014). Neural network of cognitive emotion regulation—An ALE meta-analysis and MACM analysis. *Neuroimage*, *87*, 345–355.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual*. Gainesville, FL: University of Florida.
Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*, 143–148.
Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*, 339–354.
LeDoux, J. (2007). The amygdala. *Current Biology*, *17*, R868–R874.
Leiner, D. J. (2014). SoSci Survey (Version 2.5.00-i) [Computer software]. https://www.soscisurvey.de/.
MacNamara, A., & Proudfit, G. H. (2014). Cognitive load and emotional processing in generalized anxiety disorder: Electrocortical evidence for increased distractibility. *Journal of Abnormal Psychology*, *123*, 557–565.
Marreiros, A. C., Kiebel, S. J., & Friston, K. J. (2008). Dynamic causal modelling for fMRI: A two-state model. *Neuroimage*, *39*, 269–278.
Mitchell, D. G., Nakic, M., Fridberg, D., Kamel, N., Pine, D., & Blair, R. (2007). The impact of processing load on emotion. *Neuroimage*, *34*, 1299–1309.
Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, *50*, 1–23.
Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*, 242–249.
Okon-Singer, H., Hendler, T., Pessoa, L., & Shackman, A. J. (2015). The neurobiology of emotion–cognition interactions: Fundamental questions and strategies for future research. *Frontiers in Human Neuroscience*, *9*, 58.
Oliver, N. S., & Page, A. C. (2003). Fear reduction during in vivo exposure to blood-injection stimuli: Distraction vs. attentional focus. *British Journal of Clinical Psychology*, *42*, 13–25.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). *n*-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*, 46–59.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., et al. (2010). Comparing families of dynamic causal models. *PLoS Computational Biology*, *6*, e1000709.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*. Cambridge, MA: MIT Press.

Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry*, *54*, 504–514.

Price, R. B., Paul, B., Schneider, W., & Siegle, G. J. (2013). Neural correlates of three neurocognitive intervention strategies: A preliminary step towards personalized treatment for psychological disorders. *Cognitive Therapy and Research*, *37*, 657–672.

Ray, R. D., & Zald, D. H. (2012). Anatomical insights into the interaction of emotion and cognition in the prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, *36*, 479–501.

Reinecke, A., Hoyer, J., Rinck, M., & Becker, E. S. (2009). Zwei kurzscreenings zur messung von angst vor schlangen: Reliabilität und validität im vergleich zum SNAQ [Two short-screenings measuring fear of snakes: Reliability and validity by contrast with the SNAQ]. *Klinische Diagnostik und Evaluation*, *2*, 221–239.

Sari, B. A., Koster, E. H., Pourtois, G., & Derakshan, N. (2016). Training working memory to improve attentional control in anxiety: A proof-of-principle study using behavioral and electrophysiological measures. *Biological Psychology*, *121*, 203–212.

Schicktanz, N., Fastenrath, M., Milnik, A., Spalek, K., Auschra, B., Nyffeler, T., et al. (2015). Continuous theta burst stimulation over the left dorsolateral prefrontal cortex decreases medium load working memory performance in healthy humans. *PLoS One*, *10*, e0120640.

Schienle, A., Schäfer, A., Walter, B., Stark, R., & Vaitl, D. (2005). Brain activation of spider phobics towards disorder-relevant, generally disgust- and fear-inducing pictures. *Neuroscience Letters*, *388*, 1–6.

Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D., & Dalgleish, T. (2013). Training the emotional brain: Improving affective control through emotional working memory training. *Journal of Neuroscience*, *33*, 5301–5311.

Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training. *PLoS One*, *6*, e24372.

Shah, D. A., & Madden, L. (2004). Nonparametric analysis of ordinal data in designed factorial experiments. *Phytopathology*, *94*, 33–43.

Shin, L. M., & Liberzon, I. (2010). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology*, *35*, 169–191.

Silvert, L., Lepsien, J., Fragopanagos, N., Goolsby, B., Kiss, M., Taylor, J. G., et al. (2007). Influence of attentional demands on the processing of emotional facial expressions in the amygdala. *Neuroimage*, *38*, 357–366.

Sladky, R., Höflich, A., Küblböck, M., Kraus, C., Baldinger, P., Moser, E., et al. (2013). Disrupted effective connectivity between the amygdala and orbitofrontal cortex in social anxiety disorder during emotion discrimination revealed by dynamic causal modeling for fMRI. *Cerebral Cortex*, *25*, 895–903.

Sokolov, A. A., Zeidman, P., Erb, M., Ryvlin, P., Friston, K. J., & Pavlova, M. A. (2018). Structural and effective brain connectivity underlying biological motion detection. *Proceedings of the National Academy of Sciences, U.S.A.*, *115*, E12034–E12042.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage*, *46*, 1004–1017.

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage*, *49*, 3099–3109.

Straube, T., Lipka, J., Sauer, A., Mothes-Lasch, M., & Miltner, W. H. (2011). Amygdala activation to threat under attentional load in individuals with anxiety disorder. *Biology of Mood & Anxiety Disorders*, *1*, 12.

Straube, T., Mentzel, H. J., & Miltner, W. H. (2006). Neural mechanisms of automatic and direct processing of phobogenic stimuli in specific phobia. *Biological Psychiatry*, *59*, 162–170.

Suedfeld, P., & Hare, R. D. (1977). Sensory deprivation in the treatment of snake phobia: Behavioral, self-report, and physiological effects. *Behavior Therapy*, *8*, 240–250.

Van Dillen, L. F., Heslenfeld, D. J., & Koole, S. L. (2009). Tuning down the emotional brain: An fMRI study of the effects of cognitive load on the processing of affective images. *Neuroimage*, *45*, 1212–1219.

von der Malsburg, T. (2015). Saccades: Detection of fixations in eye-tracking data. R package version 0.1-1. https://CRAN.R-project.org/package=saccades.

Vytal, K., Arkin, N., Overstreet, C., Lieberman, L., & Grillon, C. (2016). Induced-anxiety differentially disrupts working memory in generalized anxiety disorder. *BMC Psychiatry*, *16*, 62.

Vytal, K., Cornwell, B., Arkin, N., & Grillon, C. (2012). Describing the interplay between anxiety and cognition: From impaired performance under low cognitive load to reduced anxiety under high load. *Psychophysiology*, *49*, 842–852.

‘Our brain accepts what the eyes see and our eye looks for whatever our brain wants.’
(Daniel Gilbert, *Stumbling on Happiness*, 2006)

5 Discussion

Visual exploration is substantially affected by cognitive states. The aim of this thesis was to investigate the extent to which cognitive states can in turn be affected by visual exploration. This was tested by experimental manipulation. In the first study, ‘Visual exploration at higher fixation frequency increases subsequent memory recall’ (Fehlmann, Coynel, et al., 2020), we investigated the potential of manipulating the frequency and location of fixations to increase subsequent memory performance. The second study, ‘Effectiveness of a Virtual Reality-Based Eye Contact Training to Reduce Fear of Public Speaking: A Randomized Controlled Trial’ (Fehlmann, Müller, et al., 2020), was aimed at investigating the intervention potential of guiding fixation locations in order to reduce fear in a socially anxious population. In the following, the insights of both studies are discussed and integrated into already existing frameworks on the interplay between viewing and cognition.

In the first study, Fehlmann, Coynel et al. (2020), we replicated earlier findings by establishing a triadic correlation between individual visual exploration characteristics (i.e. frequency and location of fixations), MTL-activity and memory performance. We brought together results that separately linked visual sampling either to activity in the MTL (Liu et al., 2017) or to increased memory performance (Olsen et al., 2016) and replicated them in a large sample. Furthermore, we generalized findings with regards to the type of stimuli and memory tested. We showed that increased fixation frequency is not only related to increased memory performance for faces and objects, but for a broad range of complex, naturalistic scenes with varying emotional valence. We also extended results from passive recognition to active free recall, a purely episodic form of memory (Ferbinteanu et al., 2006; Tulving, 1993) with an arguably higher ecological validity. Based on these results, we conducted a second experiment, in which we experimentally manipulated visual exploration patterns in an independent sample of 64 subjects. We demonstrated that guided viewing is an effective way to manipulate the fixation frequency and location and to affect episodic memory performance. The idea that memory could be increased by guiding fixation locations was only discussed by one earlier study by Chan and colleagues (2011). We added to this the important notion that fixation frequency may be equally effective.

A limitation of our second experiment is that it lacked a free viewing condition. Although it is tempting to assume that memory performance can be elevated above baseline performance by a pre-determined visual exploration pattern, the demonstration of such an effect lies beyond the scope of the presented work. As reported in two studies, recognition performance seems to be extremely susceptible to any restrictions of volitional eye movement control (Chan et al., 2011; Voss et al., 2011). This is in line with the assumption that the MTL plays an active role in guiding fixations based on online memory representations (Voss et al., 2017). Every form of forced viewing is unable to profit from memory-based fixation guidance and potentially interferes with it, which poses a severe challenge to memory formation. We could therefore not preclude that our healthy participants had optimal or near-optimal individual visual exploration strategies under free viewing conditions and did not further profit from externally guided viewing. However, the hippocampus is discussed to be particularly involved in guiding the location rather than the frequency of fixations (Voss et al., 2011, 2017). Increasing fixation frequency arguably causes much less interference and is much easier to train than where to look, two crucial advantages that should be considered in future attempts to increase memory performance.

In the second study, Fehlmann, Müller, et al. (2020), we further investigated the intervention potential of guided viewing. In 89 participants suffering from PSA, we showed that the repeated use of a stand-alone, smartphone- and VR-based mutual gaze training effectively reduced gaze avoidance as well as the fear of public speaking in real-life speech situations. By focusing on PSA that is well-known for its association with dysfunctional viewing patterns, it could be assumed that visual exploration was not optimal at free viewing baseline (i.e. before treatment). This may have allowed for more room for improvement than in healthy participants, highlighting the clinical potential of guided viewing interventions under such conditions. To cause as little interference as possible and to increase compliance, the eye movements were not guided actively (as opposed to Fehlmann, Coynel et al. (2020)). Instead, we instructed participants in the treatment condition to hold eye contact in the VR, let them train, monitored their progress and gave them feedback about it. This relatively simple, passive intervention proved to be sufficient to correct an attentional bias (i.e. gaze avoidance) during socially stressful real-life speeches. Mutual gaze during the speeches was monitored by eye tracking, as well as by asking the participants and their interaction partners (i.e. the experimenters) about it. Although the increase of mutual gaze after repeated training was documented by all three measures independently, the effect size was larger when based on the subjective rating of participants and experimenters. This could have been caused by outcome expectancies of the participants and

by the experimenters misinterpreting dwell close to their eyes as actual eye contact, respectively. While the source of the discrepancy remains a matter of speculation, it highlights the use of eye tracking as an objective and reliable tool to measure eye movements, even in complex 3D environments.

A limitation of the second study is with regards to the relatively short time span from the completion of the training to the assessment of its effect (approximately four weeks). As gaze avoidance is a stable marker of PSA, long-term training effects need to be further investigated. An important aim thereby is to disconfirm negative beliefs of the sufferers about the interaction partner's attitude towards them (Schulze et al., 2013). However, it remains to be demonstrated if the increase of mutual gaze by repeated training can positively affect the cognitive evaluation of the situations and with it the formation of new memories.

Summarized, both studies illustrate from different angles the causal effect of viewing on cognition, with two major conclusions drawn.

First, the importance of assessing visual exploration in cognitive research was demonstrated. In a very basic sense, what is seen by a given participant should be taken into account in every experiment using visual stimuli, because it is the fundamental basis of what is processed upstream of the retina. If individual differences in visual exploration are not at least statistically controlled for in experiments, data analysis and interpretation can become severely biased. The often implicit assumption that every aspect of a visual scene is explored, and that it is equally explored within and between participants, is most obviously violated in anxious individuals that show clear gaze avoidance towards the feared stimuli. When studying human subjects in general and anxious individuals in particular, their different visual exploration patterns needs to be accounted for to avoid erroneous presumptions about their different cognitive states (Loos et al., 2020).

Second, the two studies that this thesis was focused on have documented the potential to experimentally guide viewing in order to increase memory and reduce anxiety. The reported findings provide guidance for future work that is dedicated to fully exploit this potential.

For memory improvement, it has been argued before that it is crucial not to interfere with the volitional control where to look. Because the locations of fixations are normally guided by memories, interferences can be caused whenever they are controlled externally. Further studies should therefore be focusing on increasing fixation frequency by active guidance, passive guidance or a combination of both. However, the interference with and replacement of old memories and beliefs could become a future goal of guided viewing interventions on its own,

and fixation locations could thereby be a very promising target for manipulation. Potential fields of clinical application include PTSD and anxiety disorders beyond PSA.

In the context of PSTD, which is characterized by a visual exploration bias towards negative stimuli together with strong memory traces for the traumatic events (Armstrong et al., 2013), manipulating fixations could be used as a novel approach to attenuate traumatic memories. This aim has already been pursued by a method called retrieval-induced forgetting, which has greatly impacted the therapeutic landscape. According to the underlying theory, stable memory representations of traumatic events can be destabilized by active retrieval (de Quervain, 2007). During this phase, they become sensitive to disruption, so that interrupting the reconsolidation process can lead to retrospective memory impairment for the event (Murayama et al., 2014). One way to impair certain aspects of an event is to selectively retrieve others that are related, but less negatively connotated (Brown et al., 2012). Given this mechanism, it would be interesting to see if guided viewing to parts of previously experienced negative scenes that are deliberately chosen to be uninformative (i.e. misguided viewing) leads to an overall memory impairment for those scenes, even when compared to conditions with no re-exposure.

For the treatment of specific phobias, an even more straightforward approach can be used to interfere with old memories and beliefs – or at least to correct attentional biases caused by them. Many phobias are characterized by avoidance of the feared stimuli. In the simplest case, the instruction to explore the feared stimuli and the participants' awareness that compliance is being monitored by eye tracking is sufficient to counteract stimulus avoidance. In the study by Loos el al. (2020), this arguably has been illustrated by the fact that participants with fear of snakes did not explore snake pictures differently compared to neutral objects, which would not be expected in unsupervised conditions. To further reduce avoidance behavior and improve exposure efficiency, systematic and repeated visual guidance towards the feared stimuli could be used, as reported in the context of PSA by Fehlmann, Müller, et al. (2020). This could be extended to other phobic conditions. With the integration of eye tracking into mobile devices that can produce VR or augmented reality (AR) environments, treatment settings are no longer bound to restrictions of the real world, offering the possibility for new, innovative interventions. For example, two recently submitted studies have been successful in reducing phobic fears by repeated exposure to heights in VR (Bentz et al., 2020) and to spiders in AR (Zimmer et al., 2020). It is an interesting possibility that the exposure efficiency could be further increased by guided viewing, which was not the interventional tool in either of the studies.

Irrespective of the clinical target, apps integrating guided viewing by eye- or head-tracking have the advantage that they do not longer need external input by an experimenter or a therapist. As such, they have the potential to be used as widely accessible training tools, countering the dissemination problem of many traditional treatment forms. In the clinical setting, such apps could provide an add-on to guided standard exposure therapy – for example by making VR homework assessments more feasible.

In overall conclusion, it is argued that viewing behavior is not just a confounding artefact that needs to be controlled to understand cognitive processes. Instead, visual exploration is a fundamental part of cognition. It varies between individuals, is affected by other cognitive domains like attention and memory and shares functional brain networks with them. Atypical visual exploration patterns can be assessed by eye tracking as biobehavioral markers in many neuropsychiatric conditions. Even more importantly however, visual exploration affects other cognitive states. This can be exploited by guided viewing, an eye tracking approach to increase memory and decrease anxiety. While advancements in eye tracking technology and new insights from empirical studies will have to reveal the full clinical potential of the described phenomenon, it is of great relevance and interest for neuroscientific research.

6 References

Armstrong, T., Bilsky, S. A., Zhao, M., & Olatunji, B. O. (2013). Dwelling on potential threat cues: An eye movement marker for combat-related PTSD. *Depression and Anxiety*, *30*(5), 497–502. https://doi.org/10.1002/da.22115

Bentz, D., Wang, N., Ibach, M. K., Schicktanz, N., Zimmer, A., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Effectiveness of a stand-alone, smartphone-based virtual reality exposure app to reduce fear of heights in real-life: A randomized controlled trial. *Submitted for Publication*.

Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A. M., & Helsen, W. F. (2019). The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological Bulletin*, *145*(10), 980–1027. https://doi.org/10.1037/bul0000207

Brown, A. D., Kramer, M. E., Romano, T. A., & Hirst, W. (2012). Forgetting Trauma: Socially Shared Retrieval-induced Forgetting and Post-traumatic Stress Disorder: Forgetting trauma. *Applied Cognitive Psychology*, *26*(1), 24–34. https://doi.org/10.1002/acp.1791

Büchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing Stimulus–Response Functions Using Nonlinear Regressors in Parametric fMRI Experiments. *NeuroImage*, *8*(2), 140–148. https://doi.org/10.1006/nimg.1998.0351

Chan, J. P. K., Kamino, D., Binns, M. A., & Ryan, J. D. (2011). Can Changes in Eye Movement Scanning Alter the Age-Related Deficit in Recognition Memory? *Frontiers in Psychology*, *2*, 92. https://doi.org/10.3389/fpsyg.2011.00092

Chau, S. A., Herrmann, N., Sherman, C., Chung, J., Eizenman, M., Kiss, A., & Lanctôt, K. L. (2016). Visual Selective Attention Toward Novel Stimuli Predicts Cognitive Decline in Alzheimer's Disease Patients. *Journal of Alzheimer's Disease*, *55*(4), 1339–1349. https://doi.org/10.3233/JAD-160641

Chen, J., van den Bos, E., & Westenberg, P. M. (2020). A systematic review of visual avoidance of faces in socially anxious individuals: Influence of severity, type of social situation, and development. *Journal of Anxiety Disorders*, *70*, 102193. https://doi.org/10.1016/j.janxdis.2020.102193

Chipchase, S. Y., & Chapman, P. (2013). Trade-offs in visual attention and the enhancement of memory specificity for positive and negative emotional stimuli. *Quarterly Journal of Experimental Psychology*, *66*(2), 277–298. https://doi.org/10.1080/17470218.2012.707664

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences*, *106*(14), 6008–6010. https://doi.org/10.1073/pnas.0811884106

Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., & Nash, C. (1999). Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus*, *9*(1), 83–98. https://doi.org/10.1002/(SICI)1098-1063(1999)9:1<83::AID-HIPO9>3.0.CO;2-7

Coynel, D., Gschwind, L., Fastenrath, M., Freytag, V., Milnik, A., Spalek, K., Papassotiropoulos, A., & de Quervain, D. J.-F. (2017). Picture free recall performance linked to the brain's structural connectome. *Brain and Behavior*, *7*(7), e00721. https://doi.org/10.1002/brb3.721

de Quervain, D. J.-F. (2007). Glucocorticoid-induced reduction of traumatic memories: Implications for the treatment of PTSD. *Progress in Brain Research*, *167*, 239–247. https://doi.org/10.1016/S0079-6123(07)67017-4

de Quervain, D. J.-F., Schwabe, L., & Roozendaal, B. (2017). Stress, glucocorticoids and memory: Implications for treating fear-related disorders. *Nature Reviews Neuroscience*, *18*(1), 7–19. https://doi.org/10.1038/nrn.2016.155

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice* (Third edition). Springer.

Egli, T., Coynel, D., Spalek, K., Fastenrath, M., Freytag, V., Heck, A., Loos, E., Auschra, B., Papassotiropoulos, A., de Quervain, D. J.-F., & Milnik, A. (2018). Identification of Two Distinct Working Memory-Related Brain Networks in Healthy Young Adults. *ENeuro*, *5*(1), ENEURO.0222-17.2018. https://doi.org/10.1523/ENEURO.0222-17.2018

Elliott, R., Zahn, R., Deakin, J. F. W., & Anderson, I. M. (2011). Affective Cognition and its Disruption in Mood Disorders. *Neuropsychopharmacology*, *36*(1), 153–182. https://doi.org/10.1038/npp.2010.77

Fedor, J., Lynn, A., Foran, W., DiCicco-Bloom, J., Luna, B., & O'Hearn, K. (2018). Patterns of fixation during face recognition: Differences in autism across age. *Autism*, *22*(7), 866–880. https://doi.org/10.1177/1362361317714989

Fehlmann, B., Coynel, D., Schicktanz, N., Milnik, A., Gschwind, L., Hofmann, P., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Visual exploration at higher fixation frequency increases subsequent memory recall. *Cerebral Cortex Communications*, tgaa032. https://doi.org/10.1093/texcom/tgaa032

Fehlmann, B., Müller, F., Wang, N., Ibach, M. K., Schlitt, T., Bentz, D., Zimmer, A., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Effectiveness of a virtual reality-based eye contact training to reduce fear of public speaking: A randomized controlled trial. *Submitted for Publication*.

Ferbinteanu, J., Kennedy, P. J., & Shapiro, M. L. (2006). Episodic memory—From brain to mind. *Hippocampus*, *16*(9), 691–703. https://doi.org/10.1002/hipo.20204

Fink, L. K., Lange, E. B., & Groner, R. (2019). The application of eye-tracking in music research. *Journal of Eye Movement Research*, *2*(1), 1–4. https://doi.org/10.16910/JEMR.11.2.1

Fortenbaugh, F. C., Hicks, J. C., Hao, L., & Turano, K. A. (2007). Losing sight of the bigger picture: Peripheral field loss compresses representations of space. *Vision Research*, *47*(19), 2506–2520. https://doi.org/10.1016/j.visres.2007.06.012

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Furmark, T. (2009). Neurobiological aspects of social anxiety disorder. *The Israel Journal of Psychiatry and Related Sciences*, *46*(1), 5–12.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Hannula, D. E. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00166

Heck, A., Fastenrath, M., Ackermann, S., Auschra, B., Bickel, H., Coynel, D., Gschwind, L., Jessen, F., Kaduszkiewicz, H., Maier, W., Milnik, A., Pentzek, M., Riedel-Heller, S. G., Ripke, S., Spalek, K., Sullivan, P., Vogler, C., Wagner, M., Weyerer, S., … Papassotiropoulos, A. (2014). Converging Genetic and Functional Brain Imaging Evidence Links Neuronal Excitability to Working Memory, Psychiatric Disease, and Brain Activity. *Neuron*, *81*(5), 1203–1213. https://doi.org/10.1016/j.neuron.2014.01.010

Heisz, J. J., Pottruff, M. M., & Shore, D. I. (2013). Females Scan More Than Males: A Potential Mechanism for Sex Differences in Recognition Memory. *Psychological Science*, *24*(7), 1157–1163. https://doi.org/10.1177/0956797612468281

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*(10), 743. https://doi.org/10.1038/s41562-017-0208-0

Henderson, J. M., & Hollingworth, A. (2002). *Eye movements, visual memory, and scene representation*. In MA Peterson & G. Rhodes (Eds.), Analytic and holistic processes.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106.

Holmqvist, K. (Ed.). (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Hoppe, D., Helfmann, S., & Rothkopf, C. A. (2018). Humans quickly learn to blink strategically in response to environmental task demands. *Proceedings of the National Academy of Sciences*, 201714220. https://doi.org/10.1073/pnas.1714220115

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2004). Face to face: Visual scanpath evidence for abnormal processing of facial expressions in social phobia. *Psychiatry Research*, *127*(1–2), 43–53. https://doi.org/10.1016/j.psychres.2004.02.016

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

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*(11), 1254–1259. https://doi.org/10.1109/34.730558

Kafkas, A., & Montaldi, D. (2011). Recognition Memory Strength is Predicted by Pupillary Responses at Encoding While Fixation Patterns Distinguish Recollection from Familiarity. *Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989. https://doi.org/10.1080/17470218.2011.588335

Kellough, J. L., Beevers, C. G., Ellis, A. J., & Wells, T. T. (2008). Time course of selective attention in clinically depressed young adults: An eye tracking study. *Behaviour Research and Therapy*, *46*(11), 1238–1243. https://doi.org/10.1016/j.brat.2008.07.004

Kiefer, P., Giannopoulos, I., Raubal, M., & Duchowski, A. (2017). Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation*, *17*(1–2), 1–19. https://doi.org/10.1080/13875868.2016.1254634

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews. Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008

Kredel, R., Vater, C., Klostermann, A., & Hossner, E.-J. (2017). Eye-Tracking Technology and the Dynamics of Natural Gaze Behavior in Sports: A Systematic Review of 40 Years of Research. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01845

Lancry-Dayan, O. C., Kupershmidt, G., & Pertzov, Y. (2019). Been there, seen that, done that: Modification of visual exploration across repeated exposures. *Journal of Vision*, *19*(12), 1–16. https://doi.org/10.1167/19.12.2

Lele, A. (2013). Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, *4*(1), 17–26. https://doi.org/10.1007/s12652-011-0052-4

LeMoult, J., & Joormann, J. (2012). Attention and Memory Biases in Social Anxiety Disorder: The Role of Comorbid Depression. *Cognitive Therapy and Research*, *36*(1), 47–57. https://doi.org/10.1007/s10608-010-9322-2

Liu, Z.-X., Shen, K., Olsen, R. K., & Ryan, J. D. (2017). Visual Sampling Predicts Hippocampal Activity. *The Journal of Neuroscience*, *37*(3), 599–609. https://doi.org/10.1523/JNEUROSCI.2610-16.2017

Loos, E., Egli, T., Coynel, D., Fastenrath, M., Freytag, V., Papassotiropoulos, A., de Quervain, D. J.-F., & Milnik, A. (2019). Predicting emotional arousal and emotional memory performance from an identical brain network. *NeuroImage*, *189*, 459–467. https://doi.org/10.1016/j.neuroimage.2019.01.028

Loos, E., Schicktanz, N., Fastenrath, M., Coynel, D., Milnik, A., Fehlmann, B., Egli, T., Ehrler, M., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Reducing Amygdala Activity and Phobic Fear through Cognitive Top–Down Regulation. *Journal of Cognitive Neuroscience*, *32*(6), 1117–1129. https://doi.org/10.1162/jocn_a_01537

Mathew, S. J., & Ho, S. (2006). Etiology and neurobiology of social anxiety disorder. *The Journal of Clinical Psychiatry*, *67 Suppl 12*, 9–13.

Maus, B., van Breukelen, G. J. P., Goebel, R., & Berger, M. P. F. (2010). Optimization of Blocked Designs in fMRI Studies. *Psychometrika*, *75*(2), 373–390. https://doi.org/10.1007/s11336-010-9159-3

Meißner, M., Pfeiffer, J., Pfeiffer, T., & Oppewal, H. (2019). Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *Journal of Business Research*, *100*, 445–458. https://doi.org/10.1016/j.jbusres.2017.09.028

Meister, M. L. R., & Buffalo, E. A. (2016). Getting directions from the hippocampus: The neural connection between looking and memory. *Neurobiology of Learning and Memory*, *134*, 135–144. https://doi.org/10.1016/j.nlm.2015.12.004

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *140*(5), 1383–1409. https://doi.org/10.1037/a0037505

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 391–399. https://doi.org/10.1037/0278-7393.6.4.391

Oakes, L. M. (2012). Advances in Eye Tracking in Infancy Research. *Infancy*, *17*(1), 1–8. https://doi.org/10.1111/j.1532-7078.2011.00101.x

Olsen, R. K., Sebanayagam, V., Lee, Y., Moscovitch, M., Grady, C. L., Rosenbaum, R. S., & Ryan, J. D. (2016). The relationship between eye movements and subsequent recognition: Evidence from individual differences and amnesia. *Cortex*, *85*, 182–193. https://doi.org/10.1016/j.cortex.2016.10.007

Onnis, R., Dadds, M. R., & Bryant, R. A. (2011). Is there a mutual relationship between opposite attentional biases underlying anxiety? *Emotion*, *11*(3), 582–594. https://doi.org/10.1037/a0022019

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461–488. https://doi.org/10.1037/h0084327

Palinko, O., & Kun, A. (2011). Exploring the Influence of Light and Cognitive Load on Pupil Diameter in Driving Simulator Studies. *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 329–336. https://doi.org/10.17077/drivingassessment.1416

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10), 1–12. https://doi.org/10.1167/9.10.2

Petrovska, J., Coynel, D., Fastenrath, M., Milnik, A., Auschra, B., Egli, T., Gschwind, L., Hartmann, F., Loos, E., Sifalakis, K., Vogler, C., de Quervain, D. J.-F., Papassotiropoulos, A., & Heck, A. (2017). The NCAM1 gene set is linked to depressive symptoms and their brain structural correlates in healthy individuals. *Journal of Psychiatric Research*, *91*, 116–123. https://doi.org/10.1016/j.jpsychires.2017.03.007

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*(2), 113–121. https://doi.org/10.1016/S0166-2236(00)01685-4

Rudi, D., Kiefer, P., & Raubal, M. (2020). The instructor assistant system (iASSYST)—Utilizing eye tracking for commercial aviation training purposes. *Ergonomics*, *63*(1), 61–79. https://doi.org/10.1080/00140139.2019.1685132

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a Deficit in Relational Memory. *Psychological Science*, *11*(6), 454–461. https://doi.org/10.1111/1467-9280.00288

Ryan, J. D., Hannula, D. E., & Cohen, N. J. (2007). The obligatory effects of memory on eye movements. *Memory*, *15*(5), 508–525. https://doi.org/10.1080/09658210701391022

Schulze, L., Renneberg, B., & Lobmaier, J. S. (2013). Gaze perception in social anxiety and social anxiety disorder. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00872

Shakespeare, T. J., Pertzov, Y., Yong, K. X. X., Nicholas, J., & Crutch, S. J. (2015). Reduced modulation of scanpaths in response to task demands in posterior cortical atrophy. *Neuropsychologia*, *68*, 190–200. https://doi.org/10.1016/j.neuropsychologia.2015.01.020

Sharot, T., Davidson, M. L., Carson, M. M., & Phelps, E. A. (2008). Eye Movements Predict Recollective Experience. *PLoS ONE*, *3*(8), e2884. https://doi.org/10.1371/journal.pone.0002884

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7

Shin, Y. S., Chang, W., Park, J., Im, C.-H., Lee, S. I., Kim, I. Y., & Jang, D. P. (2015). Correlation between Inter-Blink Interval and Episodic Encoding during Movie Watching. *PLOS ONE*, *10*(11), e0141242. https://doi.org/10.1371/journal.pone.0141242

Spalek, K., Fastenrath, M., Ackermann, S., Auschra, B., Coynel, D., Frey, J., Gschwind, L., Hartmann, F., van der Maarel, N., Papassotiropoulos, A., de Quervain, D., & Milnik, A. (2015). Sex-Dependent Dissociation between Emotional Appraisal and Memory: A Large-Scale Behavioral and fMRI Study. *Journal of Neuroscience*, *35*(3), 920–935. https://doi.org/10.1523/JNEUROSCI.2384-14.2015

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. https://doi.org/10.1080/14640747308400340

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74. https://doi.org/10.3758/BF03337426

Tao, Q.-Q., Zhan, S., Li, X.-H., & Kurihara, T. (2016). Robust face detection using local CNN and SVM based on kernel combination. *Neurocomputing*, *211*, 98–105. https://doi.org/10.1016/j.neucom.2015.10.139

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual Memory for Objects in Natural Scenes: From Fixations to Object Files. *The Quarterly Journal of Experimental Psychology Section A*, *58*(5), 931–960. https://doi.org/10.1080/02724980443000430

Tulving, E. (1993). What Is Episodic Memory? *Current Directions in Psychological Science*, *2*(3), 67–70. https://doi.org/10.1111/1467-8721.ep10770899

van Steenbergen, H., Band, G. P. H., & Hommel, B. (2011). Threat But Not Arousal Narrows Attention: Evidence from Pupil Dilation and Saccade Control. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00281

Voss, J. L., Bridge, D. J., Cohen, N. J., & Walker, J. A. (2017). A Closer Look at the Hippocampus and Memory. *Trends in Cognitive Sciences*, *21*(8), 577–588. https://doi.org/10.1016/j.tics.2017.05.008

Voss, J. L., Gonsalves, B. D., Federmeier, K. D., Tranel, D., & Cohen, N. J. (2011). Hippocampal brain-network coordination during volitional exploratory behavior enhances learning. *Nature Neuroscience*, *14*(1), 115–120. https://doi.org/10.1038/nn.2693

Weeks, J. W., Howell, A. N., & Goldin, P. R. (2013). Gaze Avoidance in social anxiety disorder. *Depression and Anxiety*, *30*(8), 749–756. https://doi.org/10.1002/da.22146

Williams, L. E., Must, A., Avery, S., Woolard, A., Woodward, N. D., Cohen, N. J., & Heckers, S. (2010). Eye-Movement Behavior Reveals Relational Memory Impairment in Schizophrenia. *Biological Psychiatry*, *68*(7), 617–624. https://doi.org/10.1016/j.biopsych.2010.05.035

Wilson, F. A. W., & Goldman-Rakic, P. S. (1994). Viewing preferences of rhesus monkeys related to memory for complex pictures, colours and faces. *Behavioural Brain Research*, *60*(1), 79–89. https://doi.org/10.1016/0166-4328(94)90066-3

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1–8. https://doi.org/10.1038/s41562-017-0058

Wood, G., Nuerk, H.-C., Sturm, D., & Willmes, K. (2008). Using parametric regressors to disentangle properties of multi-feature processes. *Behavioral and Brain Functions*, *4*(1). https://doi.org/10.1186/1744-9081-4-38

Yamamoto, N., & Philbeck, J. W. (2013). Peripheral vision benefits spatial learning by guiding eye movements. *Memory & Cognition*, *41*(1), 109–121. https://doi.org/10.3758/s13421-012-0240-2

Yarbus, A. (1967). Eye Movements and Vision (B. Haigh, Trans.) Plenum Press. *New York*.

Zimmer, A., Wang, N., Ibach, M. K., Fehlmann, B., Schicktanz, N., Bentz, D., Michael, T., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Effectiveness of a stand-alone, smartphone-based, gamified augmented reality exposure app to reduce fear of spiders in real-life: A randomized controlled. *Submitted for Publication*.

Zola, S. M., Manzanares, C. M., Clopton, P., Lah, J. J., & Levey, A. I. (2013). A Behavioral Task Predicts Conversion to Mild Cognitive Impairment and Alzheimer's Disease. *American Journal of Alzheimer's Disease & Other Dementiasr*, *28*(2), 179–184. https://doi.org/10.1177/1533317512470484

7 Declaration by candidate

I declare herewith that I have independently carried out the PhD thesis entitled ‘Guided Viewing: An Eye Tracking Approach to Increase Memory and Reduce Anxiety’. This thesis consists of original research articles that have been written in cooperation with the enlisted co-authors and have been published in peer-reviewed scientific journals or are in preparation for publication / submitted for publication. Only allowed resources were used and all references used were cited accordingly.

Date: _____

Signature: _____