The role of the frontopolar cortex in the exploratory redistribution of attentional resources

Thesis

for the degree of

doctor rerum naturalium

approved by the Faculty of Natural Science of Otto-von-Guericke University Magdeburg

by M.Sc. Lasse Güldener

born on 24.12.1989 in Westerstede

Examiner: Prof. Stefan Pollmann

Prof. Martin Geier

Submitted on: 18.08.2023

Defended on: 20.03.2024

Abstract

The successful navigation in rapidly changing environments heavily relies on focusing on what is important while ignoring the irrelevant. Attention helps us doing this by supporting the facilitated processing of relevant information, often automatically and governed by implicit motives. At the same time, changing circumstances leave us being confronted with the need to adjust our behavior. Such adaptation requires a redistribution of attentional resources towards a new strategy or goal. Although neuroimaging evidence could link the anterior prefrontal, specifically the frontopolar cortex (FPC), to exploratory re-weighting of attentional weights, we only have started to unravel the neurocognitive mechanisms that help us to find a balance between continuing to do what we know works for us and trying something new. The series of experiments presented in this dissertation was designed to investigate the role of FPC in exploratory shifts of attention. In the first experiment, a novel behavioral masking paradigm was designed to test if shifts of feature-based attention can occur in response to fully invisible stimulus changes in young healthy human participants. The second experiment followed up the first experiment using functional magnetic resonance imaging (fMRI) to study if and how the FPC supports shifts of visual attention in the absence of visual awareness. The third experiment, again carried out using fMRI, used a novel virtual foraging task to examine FPC's implication in so-called patch-leaving behavior, that is, to what extend FPC supports participants' decision to initiate behavioral exploration. Combining signal detection theory and subjective measures of awareness in experiment 1 and 2, we showed that performance on unaware trials was consistent with visual selection being weighted towards repeated orientations of Gabor patches and reallocated in response to a novel unconsciously processed orientation. This was particularly present in trials in which the prior feature was strongly weighted and only if the novel feature was invisible. The fMRI data revealed that the ventral attention network responded to invisible feature changes whereas activity patterns in FPC conveyed the feature information of the novel stimulus attentional resources needed to be redirected to. Together these results foster the notion that FPC, not specifically implicated in the detection of invisible stimulus changes, supports shifts of visual attention by representing information about alternative goals. Building on these findings, the third experiment showed that inter-individual differences in the propensity to pursue an either exploitative or exploratory foraging strategy, indexed by participants' giving-up times (GUT), modulated the signal strength in both medial, and lateral FPC time-locked to the onset of exploratory behavior. Those participants who showed a behavioral bias towards exploration, showed stronger signaling bilaterally in the medial and lateral FPC in the moment of exploration. Altogether, the experiments provide important new insights into FPC's functioning that extend existing findings by showing that FPC encodes information of an unconsciously perceived stimulus attention needs to be directed to as well as that it takes a specific role in exploratory decision-making by supporting the shifting from an exploitative towards an exploratory mode of cognitive control. At the same time, they highlight that further research benefits from taking differences in behavioral strategies into account when making predictions about FPC's involvement in exploratory decision-making.

Contents

List of Figures

List of Tables

List of Abbreviations

ADHD ..... attention deficit hyperactivity disorder

AG ..... angular gyrus

AIC ..... anterior insular

AL ..... awareness level

ANOVA ..... analysis of variance

BA ..... Brodmann area

BBR ..... boundary based registration

BF ..... bayes factor

bil ..... bilateral

BOLD ..... blood oxygenation level-dependent

CBF ..... cerebral blood flow

CI ..... confidence intervall

CMRO2 ..... cerebral metabolic rate of oxygen

COPE ..... contrast of parameter estimates

CR ..... correct rejection

dACC ..... dorsal anterior cingulate cortex

EEG ..... electroencephalography

EPI ..... echo planar imaging

EV ..... explaining variable

FA ..... false alarm

FEF ..... frontal eye field

fMRI ..... Functional magnet resonance imaging

FMS ..... frontomarginal sulcus

FOV ..... field of view

FPC ..... frontopolar cortex

FPR ..... false positive rate

FWE ..... family-wise error rate

FWHM ..... full width at half maximum

GLM ..... generalized linear model

GUT ..... giving-up time

H ..... Hit

HRF ..... hemodynamic response function

ICR ..... instantaneous collection rate

IFG ..... inferior frontal gyrus

IFJ ..... inferior frontal junction

IPS ..... inferior prarietal sulcus

ITI ..... inter-trial-interval

l ..... left

LFP ..... local field potential

LG ..... lingual gyrus

LMM ..... linear mixed model

LOC ..... lateral occipital cortex

m ..... medial

M ..... miss

MCR ..... mean collection rate

MFG ..... medial frontal gyrus

MTG ..... middle temporal gyrus

MVPA ..... multi voxel pattern analysis

MVT ..... marginal value theorem

OFC ..... orbitofronal cortex

OPC ..... opercular cortex

PAS ..... perceptual awareness scale

PCC ..... posterior cingulate cortex

PCG ..... paracingulate gyrus

Pcu ..... precuneus

PET ..... positronen emission tomographie

PFC ..... prefrontal cortex

PreCG ..... precentral gyrus

r ..... right

RCZ ..... rostral cingulate zone

REML ..... restricted maximum likelihood

rm ..... repeated measure

ROI ..... region of interest

RSVP ..... rapid serial visual presentation

RT ..... reaction time

SD ..... standard deviation

SDT ..... signal detection theory

SE ..... standard error

SFG ..... superior frontal gyrus

SLF ..... superior longitudinal fasciculus

SLR ..... searchlight radius

SMC ..... supplementary motor cortex

SMG ..... supramarginal gyrus

SOA ..... stimulus onset asynchrony

SPL ..... superior parietal lobe

TE ..... time to echo

TOFC ..... temporo-occipital fusiform cortex

TPJ ..... temporoparietal junction

TPR ..... true positive rate

TR ..... time to repeat

TVA ..... theory of visual attention

VFC ..... ventral frontal corte

VIF ..... variance inflation factor

WCST ..... wisconsin card sorting task

1 General Introduction

1.1 Motivation and structure of this thesis

Often we find ourselves in situations overly rich in perceptual information that is far beyond what we can effectively process. Thus, to make sense of the environment, and to be able to execute goal-directed behavior, we need to choose the relevant information and ignore the irrelevant. Let us picture ourselves on our way back home during rush hour. Here, safe navigation critically depends on our ability to detect and monitor signs and traffic lights as well as to observe and respond to the actions of others, no matter if we walk down the sidewalk, ride our bike, or drive a car. At the same time, we manage to ignore billboards, the user interface in the car, and bleeping phones. Similarly, a nurse rushing into a patient's room following an emergency call, needs to quickly scan the ill person and read the blinking lights around the patient's bed to understand what kind of emergency he is facing and what kind of actions he needs to take. These examples illustrate that, very often, our environment bears a high complexity and an overload of information. Attention as a cognitive function helps us to reduce this informational overload through *visual selection*. Put simply, attention as a process of visual selection helps us to focus on what is currently important and to ignore what is currently less relevant. However, sudden events in constantly changing environments can render unimportant what was important a moment ago, while a previously unnoticed or ignored aspect may take on the highest urgency. To adapt to such situations, attentional selection must be flexible, allowing a quick redistribution of attentional resources according to changing circumstances and priorities. While a solid framework exists that describes the general interplay of different frontoparietal brain regions supporting attentional control in general as well as attentional reorienting (e.g., Corbetta & Shulman, 2002; Corbetta et al., 2008), this dissertation is dedicated to examine the scope of the FPC in the execution of exploratory redistribution of attentional resources. Considered as the largest higher-order association area in the human brain, uniquely large in humans compared to other primates (Bludau et al., 2014), this most rostral part of the prefrontal cortex (PFC) has been consistently linked to shifts of visual attention (Lepsien & Pollmann, 2002; Pollmann et al., 2000; Weidner et al., 2002), but it is not well integrated in the most prominent frameworks describing the neural correlates of visual attention (Corbetta & Shulman, 2002; Corbetta et al., 2008). Key aim of this dissertation was thus to provide new evidence to promote and further elaborate our understanding of the FPC's role in the redistribution of attentional resources.

The remainder of this first chapter briefly outlines central models in cognitive neuroscience and psychology of visual attention and review empirical findings on the neural correlates that support attentional control. Here I will also address the anatomical and functional specificities of the FPC. Lastly, I will dwell briefly on the usage of fMRI and multi-voxel pattern analysis (MVPA) to analyze neural activity and close the chapter with an outlook on the experimental work. Together with Chapter 2, where I detail common methods of experiment 1 and 2, this chapter thus provides the background for the experimental work reported in Chapters 3 to 5. Finally, in Chapter 6, I will summarize the results of the experiments and discuss

their implications.

1.1.1 The cognitive neuroscience of visual attention

As this dissertation exclusively examined attention in the visual domain, the following chapter focuses solely on visual attention. Attention does however extend to other sensory domains (e.g. Carlson et al., 2018; Lukas et al., 2010; Shomstein & Yantis, 2004), and basic principles of attention as a cognitive function outlined in the following section may also apply to attention studied in other sensory domains.

1.1.2 Visual attention as a cognitive process

As illustrated in the beginning of this chapter, our day-to-day life often takes place in situations with a high load of information so that goal-directed behavior requires a prioritization of what is important to be processed first. Consistent with this idea, the *biased-competition* model (Desimone & Duncan, 1995) assumes that objects in a visual scene are in a constant competition for access to visual short-term memory of limited capacity. The competition is thought to be biased by top-down signals (e.g., reflecting action goals) that add higher so-called *selection weights* to behaviorally relevant objects (e.g., Roelfsema et al., 2002), object features (e.g., Maunsell & Treue, 2006; Treue & Trujillo, 1999) or locations (e.g., Posner, 1980) to promote their access to further processing resources. To enable adaptive and goal-directed behavior in changing environments, the top-down signals representing behavioral goals (Craighero et al., 2002; Rosenbaum et al., 1991) must interact with sensory (bottom-up) information about objects in the visual scene to facilitate the selective perception and further processing of what is relevant given the situation, at the cost of what is irrelevant (Bundesen, 1990; Wolfe, 1994).

Itti and Koch (2000, 2001) elaborated their saliency model with the aim to explain visual selection that is driven by differences in the physical strength of low-level stimulus features. In an early visual, pre-attentive perceptual process, the different components of a visual scene compete on some feature dimension such as color, contrast, or spatial orientation, whereby the more salient component wins - e.g., the location within a scene that contrasts more strongly - and consequently attracts our gaze and becomes the object of a now attention-binding perceptual process ("pop-up", Itti & Koch, 2000). However, as outlined above, such a computational model of attention assuming merely bottom-up factors fails to exhaustively predict attentional selection (e.g., Betz et al., 2010). Based on our everyday experience, we know the salience of what is surrounding us is also influenced by the behavioral relevance we attribute to it. If we are waiting for our best friend at the train station, we may not know what the friend is currently wearing, but we know that he has blonde hair. Consequently, we are more likely to notice people passing by with blonde hair. This illustrates that attentional control can act in line with our internal objectives supporting goal-directed behavior (e.g., Bourgeois et al., 2016; Folk et al., 1992; Raymond & O'Brien, 2009). Here the idea is that we maintain information about what we are looking for, e.g., we know the friend is blonde, in form of a so-called *search template*, i.e., a 'perceptual set' (Corbetta & Shulman,

2002), and use this template to bias the processing of incoming visual information. Whereas acknowledging top-down biases driving attentional selection is essential to develop a holistic understanding of the mechanisms that determine what we attend to (Theeuwes, 2010), the dichotomy of a bottom-up versus top-down control of visual attention appears to be insufficient as more factors such as prior selection and reward history have been shown to impact visual selection as well (e.g., Awh et al., 2012; Failing & Theeuwes, 2018). To integrate bottom-up, top-down biases, as well as influences based on reward history on visual attention, the idea of priority maps as a unifying concept was introduced (e.g., Bisley & Goldberg, 2010; Klink et al., 2014; Theeuwes, 2018). Such a map is best understood as a spatially organized representation of the stimuli within the visual field, where each stimulus is represented by a combination of internally attributed characteristics (such as the expected reward, or task- or goal relevance that may be associated with a stimulus), and external features such as the physical salience (Bisley, et al., 2009; Fecteau & Munoz, 2006). Consequently, such a representational map allows to distinguish between stimuli associated with more or less priority and guides shifts of visual attention accordingly so that the stimulus associated with the highest priority is selected (Bisley & Goldberg, 2010; Serences & Yantis, 2006). At the same time, attention itself contributes to the formation of such a priority map by providing top-down signals representing e.g., goals or expectations that are consequently integrated in the map. The notion that such priority map is modulated by both exogenous and endogenous biases is consistent with neuroimaging studies that show that both types of attentional control are linked to activations a range of brain regions that all belong to an overlapping frontoparietal network (Corbetta & Shulman, 2002; Corbetta et al., 2008; Katsuki & Constantinidis, 2012; Katsuki & Constantinidis, 2014).

So far I have briefly outlined computational accounts of visual attention that assume spatially organized saliency, and later priority maps, guiding visual selection in the face of limited processing resources. Crucial for goal-directed behavior are shifts of visual attention. They enable us to respond to sudden changes in the current environment as well as to attend to novel information that may be relevant for achieving a goal. In theory, e.g., the sudden appearance of a highly relevant stimulus within the visual field changes the configuration of the priority map and attentional selection is consequently shifted towards that new stimulus. Yet, how does the brain execute such changes? The remainder of this chapter will deal with the neural correlates assumed to constitute the so-called frontoparietal attention network and point out their interplay that support shifts of visual attention.

1.1.3 The frontoparietal network of visual attention

One of the most renowned frameworks about the neural foundations of attention was developed by Corbetta and Shulman in 2002 and later extended in 2008 and distinguishes between two networks orchestrating attentional control (see Figure 1). The *dorsal* network of attention comprises parietal regions including the intraparietal sulcus (IPS) and the superior parietal lobe (SPL). In the frontal lobe, it is the dorsal part along the precentral sulcus (PCS) that is considered as

part of the network, including the frontal eye field (FEF). Typical for these regions are consistent preactivation in various cognitive tasks, e.g., in the expectancy of the upcoming stimulus at a specific spatial location (Corbetta et al., 2000). Such preactivations are interpreted as internally generated and maintained signals reflecting current expectations or relevant behavioral goals. Consistent with the model of biased competition (Desimone & Duncan, 1995; see above), these internal representations are thought to be subsequently used as top-down signals, biasing the processing of the incoming sensory information, e.g., in the primary visual cortex (Corbetta & Shulman, 2002; Corbetta et al., 2008). Considering that both IPS and FEF possess subregions with retinotopic organization (Silver & Kastner, 2009), both regions show properties that enable the maintenance of spatially organized priority maps (section 1.1.2), presumably guiding covert shifts of spatial attention, saccade planning, as well as visual working memory (Jerde et al., 2012; see Vossel et al., 2014).

The second, *ventral* network of attention is suggested to predominantly enable stimulus-driven (bottom-up) shifts of attention, typically showing activity in response to relevant but unexpected stimulus onsets (Corbetta et al., 2008). It comprises the ventral frontal cortex (VFC), that includes parts of middle frontal gyrus (MFG), the ventral inferior frontal gyrus (IFG), the frontal operculum (OPC), and the anterior insular cortex (AIC). Posteriorly, the network includes the temporoparietal junction (TPJ) which is composed by the posterior part of the superior temporal sulcus (STS) as well as the ventral part of the supramarginal gyrus (SMG) were included (Corbetta et al., 2008). (Although note that different studies reporting TPJ activations also report differences in the exact spatial location (see Schurz et al., 2017). During viusal search TPJ, rMFG, and rIFG reportedly show sustained deactivation, whereas the same regions respond with increased activity to unattended targets (Shulman et al., 2003). This pattern of results promoted the hypothesis that the ventral network disrupts the currently maintained selection in the dorsal network, resulting in a shift of attention to a previously unattended stimulus (Corbetta et al., 2008).

Whereas the suppression of the ventral network during ongoing search presumably supports the prevention of responses to irrelevant stimuli, relevant targets still elicit positive responses suggesting that the activity in the ventral network is filtered by task relevance (Shulman et al., 2007). In particular, for the rTPJ, implicated in stimulus-driven spatial attentional reorienting (Chang et al., 2013), it was shown that the degree to which it was deactivated correlated positively with participants' detection performance in a rapid serial presentation task (RVSP), suggesting that a stronger filtering in the ventral network promotes a better task performance (Corbetta et al., 2008; Shulman et al., 2007). Consistent with the idea of a filtering by task relevance, in the domain of contextual cueing where unattended stimuli can carry predictive information for the target location, it was shown that TPJ is not suppressed if distractors are predictive for the the target location. Moreover, the activity in both TPJ and IFG is positively modulated by the contextual relevance of the distractors (i.e., in repeated displays in which the spatial configuration of distractors predicts the target location). Only if the distractors have no predictive value (novel displays), the FEF inhibits TPJ. These observations suggest that the

IFG together with TPJ constitute a network driven by sensory information that integrates knowledge about the context with steadily incoming sensory information to provide an attentional control signal to FEF (DiQuattro & Geng, 2011). At the same time, this shows that FEF, as part of the dorsal, and TPJ, as part of the ventral attention network, closely interact to allow flexible attentional control that enables us to pursue a goal but also to adapt to our environment. Using transcranial magnetic stimulation (TMS) to inhibit the right IPS and TPJ in a spatial cueing task showed that TPJ inhibition impaired only exogenously cued attentional orienting, i.e., if the the spatial cue had no predictive value of the upcoming target location (50% cue validity), whereas the inhibition of the rIPS hampered both exogenous as well as endogenous orienting of attention, when the spatial cues predicted the upcoming target location with 67% probability (Chica et al., 2011). Clearly, this evidence challenges the straightforward attribution of either top-down or bottom-up processing to one of the two networks. At the same time, and similar to DiQuattro and Geng (2011), it strongly suggests that both networks interact closely to promote flexible attentional control. On the functional level this interaction is possibly supported by the rMFG which shows spontaneous fMRI activity that correlates with both networks (Fox et al., 2006; He et al., 2007). On the anatomical level, the link between both attention networks is provided by the middle superior longitudinal fasciculus (SLF II) connecting the parietal nodes of the ventral with the frontal nodes of the dorsal network (de Schotten et al., 2011; see Vossel et al., 2014).

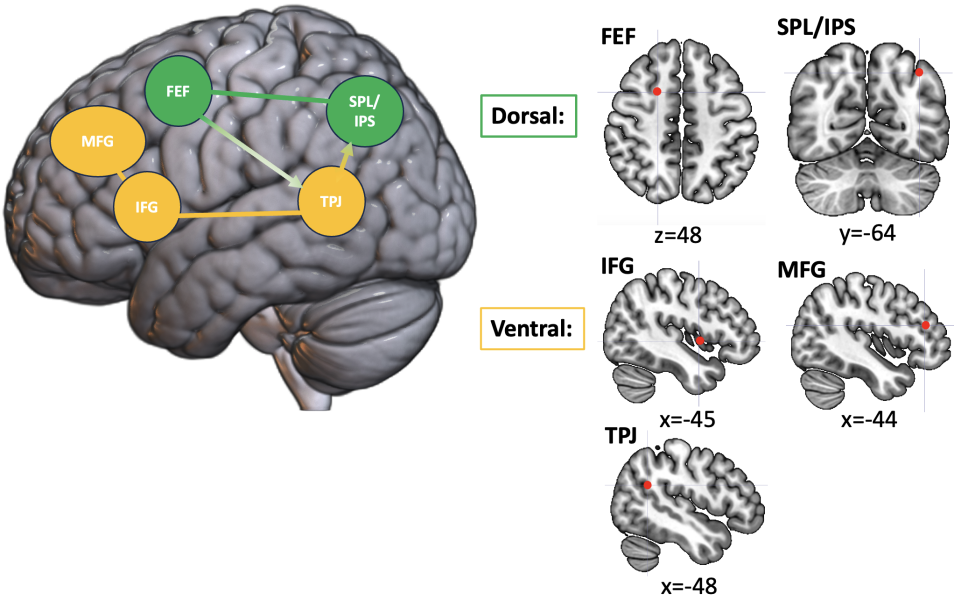


Figure 1: The network model of brain structures underlying attentional control adopted from Corbetta and Shulman, 2002. The FEF and IPs support the top-down regulation of visual processing whereas TPJ together with IFG and MFG are thought to serve stimulus-driven attentional control. Via connections between IPs and TPJ the current attentional focus is interrupted if novel stimuli outside of the attentional focus are detected. Suppressions of the stimulus-driven attentional control, on the other hand, rely on connections between the IPs and TPJ, as well as between the TPJ and FEF (DiQuattro & Geng, 2011).

Despite evidence demonstrating the role of the FPC in the execution of attentional shifting (Daw et al., 2006; Gramann et al., 2010; Ort et al., 2019; Pollmann, 2001; Pollmann, 2004; Pollmann et al., 2007; Pollmann et al., 2000), the model by Corbetta and Shulman (2002) does not consider the FPC as part of neither the dorsal nor the ventral attention network. A reason for this might be the diverse contexts in which the FPC has been implicated, such as mnemonic function (e.g., Dobbins et al., 2002), reasoning (e.g., Green et al., 2006), action planning (Dagher et al., 1999), cognitive branching (Koechlin et al., 1999), or reward-based learning (Herrojo Ruiz et al., 2021), as well as inconsistent evidence linking FPC to attentional weighting (Mansouri et al., 2020; Pollmann et al., 2007). These findings promote the idea that FPC capitalizes a more general role in cognitive control, exceeding the mere redistribution of attentional selection weights. In the remaining chapter I will briefly outline the anatomical distinctiveness of the FPC and review its putative functional role in managing attentional resources facilitating cognitive control.

1.1.4 Anatomical features of the FPC

The FPC, also referred to as the frontal pole or BA10, is located at the most rostral part of the prefrontal cortex in the human brain. It is positioned at the superior frontal gyrus (SFG), occupying the area at the front of the brain, just behind the forehead. From an evolutionary perspective, it is the neocortical brain region

that has extended the most in the course of human phylogenesis compared to other non-human primates (Semendeferi et al., 2001). It comprises Brodmann Area 10, (BA10) which is determined by its unique cytoarchitecture, distinct from adjacent prefrontal regions (Brodmann, 1909). The originally described BA10 encompasses the FPC in its center, which includes the frontomarginal sulcus (FMS) as well as the rostral SFG, and to some extent also the MFG (see Figure 2), (Bludau et al., 2014). In Brodmann's description, BA10 borders ventrally with the orbitofrontal cortex (OFC, BA11), dorsally with BA9 which contributes to the dorsolateral and medial prefrontal cortex (dlPFC, mPFC). In caudal direction BA10 adjoins BA46, the middle frontal area, which roughly corresponds with the dlPFC (Bludau et al., 2014; Brodmann, 1909). Later anatomical descriptions of BA10 mostly align with the originally proposed map (Bludau et al., 2014; von Economo & Koskinas, 1925; Sarkisov et al., 1949). Given the heterogeneous cytoarchitecture even within the BA10, Öngür and colleagues (Öngür et al., 2003) proposed a map of greater extent that comprises three subdivisions of BA10, 10m, 10r, and 10p, where only 10p contains the FPC and 10m and 10r reach further into the mesial surface of the brain reaching into the most ventral part of the anterior cingulate gyrus (Bludau et al., 2014; Öngür et al., 2003). Typically, all neocortical areas comprise six cell layers, the molecular layer (layer I), the outer granular layer (layer II), the outer pyramidal layer (layer III), the inner granular layer (layer IV), the inner pyramidal layer (layer V) and the polymorphic layer (layer VI), (Bludau et al., 2014; Brodmann, 1909; Strotzer, 2009). In the map of BA10 from Öngür an colleagues, 10p, the frontal pole, is described as having the most pronounced layer III. This layer contains projecting pyramidal cells (e.g., Bannister, 2005) with short- and long-range connections within and outside of BA10. In a more recent account of FPC's anatomical subdivisions, Bludau et al. (2014) propose an anatomical distinction between a lateral (Fp1) and medial (Fp2) FPC, stemming from a growing body of research stressing a functional divergence between the medial and lateral FPC. Comparing the cytoarchitecture of these two subregions revealed similarities in that both areas showed sharp borders between layer I, layer II, layer III, and layer IV with densely packed cells. At the same time, the lateral FPC (Fp1) could be distinguished by its higher cell density in the outer pyramidal layer (III) and it also appeared to have a greater layer II with densely packed cells and a more pronounced layer IV. Fp2's lower parts of layer V, in turn, possessed a small population of pyramid cells that was not found in area Fp1 (Bludau et al., 2014). In the human FPC pyramidal neurons have a higher number and density of dendritic spines compared with neurons in the orbitofrontal cortex (Jacobs et al., 2001), and pyramidal cells bodies of layer III display a greater horizontal distance between each other compared to the monkeys' FPC, suggesting that the human FPC may possess an enhanced capacity for neural integration crucial for cognition (Tsujimoto et al., 2011). The spatial expansion of the FPC with its lateral and medial subdivisions according to Bludau et al. (2014) is shown in Figure 2.

1.1.5 The functional role of FPC in attentional control

As mentioned above, a large and rather heterogeneous body of research has linked the FPC to different forms of intelligent human behavior. With an anatomical volume larger compared to other non-human primates (Semendeferi et al., 2001), it

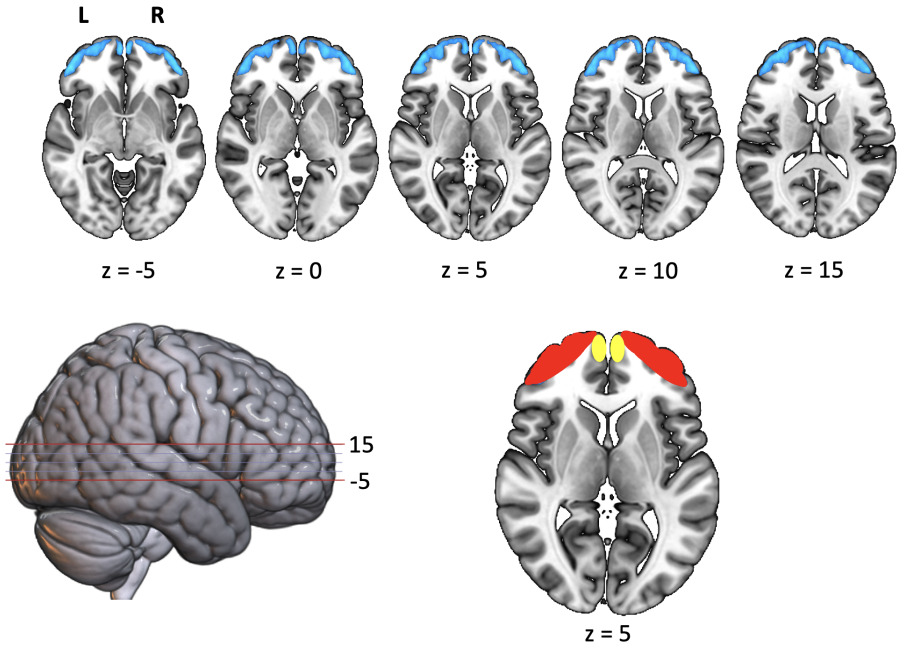


Figure 2: based on the Harvard-Oxford cortical atlas, the region colored in blue shows bilaterally the lateral (Fp1, area in red) and medial (Fp2, area in yellow) FPC (> 65% probability) plotted on an MNI152 standard brain template. Transverse slices z = -5 to z = 15 encompass the region with the highest probability of containing the FPC along the vertical axis (Bludau et al., 2014).

likely supports cognitive abilities unique to humans. According to an early account of FPC's function, it serves the evaluation of internally generated information (Christoff & Gabrieli, 2000). A more recent hypothesis focuses on the role of the PFC in coordinating concurrent goals (Hogeveen et al., 2022; Mansouri et al., 2017). Goal-directed behavior heavily relies on a constant balance between a tendency to either exploit or explore as we constantly decide to keep doing what we know works for us or to take the risk trying something new as it could work even better. Behavioral exploitation, on the one hand, is thought to be originating from posterior parts of the prefrontal cortex (Donoso et al., 2014; Koechlin et al., 1999), presumably allocating cognitive resources to an ongoing task. Exploratory tendencies, on the other hand, are thought to be driven by the FPC that orchestrates the redistribution of cognitive resources away from the current task whenever alternative goals receive a higher behavioral relevance (e.g., due to changes in the environment). Importantly, the successful balancing between exploitative and exploratory urges is crucial for the adequate execution of natural behaviors such as foraging which directly impacts the individual's and species's survival and fitness (Mansouri et al., 2017).

Weighing between exploitation and exploration. First evidence for FPC's role in resolving the exploration-exploitation dilemma was reported in a virtual gambling task originating from the principles of a casino's slot-machine (e.g., Daw et al., 2006; see Jones, 1975). Replications and extensions based on similar n-armed bandit tasks followed, all together implicating the FPC more generally in such higher-order executive control processes. In particular, the FPC together with the IPS was found

to show increased activity during exploratory gambling choices. The striatum and ventromedial prefrontal cortex (vmPFC), on the other hand, were linked to exploitative choice behavior (Daw et al., 2006; Laureiro-Martínez et al., 2015). Later, using TMS right lateral FPC was demonstrated to govern directed but not random exploration in a version of bandit-task that allowed the distinction between both types of exploration (Zajkowski et al., 2017). This finding was later confirmed by yet another bandit-task study showing that the relative decision uncertainty is represented in the right lateral FPC (Fp1) facilitating directed exploration (Tomov et al., 2020). Further confirmation of the idea that FPC takes a central role in the exploratory reallocation of cognitive resources is provided by evidence demonstrating that FPC represents the value of switching to a foregone alternative behavioral choice during uninstructed decision-making (Boorman et al., 2009). The same study also showed that switch decisions towards the alternative choice were associated with changes in the functional connectivity of FPC, and that inter-individual differences in the FPC signal strength predicted differences in how effectively the behavior was adapted (Boorman et al., 2009). This suggests that FPC keeps track of the potential value of not-selected choices and supports the redistribution of attentional resources if an alternative choice seems favorable.

Important convergent evidence in support of this idea is provided by lesion studies in apes. Focal FPC lesions impair rapid one-trial learning (Boschin et al., 2015), but do not generally hamper the performance in a cognitively demanding task (Mansouri et al., 2015). Intriguingly, control monkeys with an intact FPC are more easily distracted from exploiting the current task when confronted with unexpected interruptions by a secondary task or by free rewards. Lesion monkeys' performance in the primary task during the first trials after revisiting, on the other hand, is less affected by such unforeseen interruptions. This suggests that the lesion monkeys show a decreased tendency to explore the potential value of the new task or reward for the overall performance, which is consistent with the evidence from human gambling studies that implicate the FPC in exploratory choice making. Together these findings foster the notion that FPC is involved in the redistribution of cognitive resources away from the current task or goal towards an alternative opportunity (Hogeveen et al., 2022; Mansouri et al., 2015; Mansouri et al., 2017). Such shifting towards a behavioral alternative also requires the assessment of the value of the currently selected choice. Consistent with this idea, single-cell recordings in monkeys' FPC, carried out temporally within between action choices, that the animals had made, and before learning about the positive outcome related to their choices, show that FPC tracks the importance of the currently selected action (Tsujimoto et al., 2011, 2012).

The evidence reviewed above underscores the importance of FPC in serving the switching away from the currently pursued towards an alternative behavioral choice. This process certainly entails the shifting of attentional selection weights, but also requires the monitoring of competing values assigned to each choice. Thus, FPC's role appears to be tied to a broader cognitive ability crucial to a whole range of different behaviors that all entail the management of competing goals and an according adaptation of the current behavioral strategy (Hogeveen et al., 2022;

Mansouri et al., 2017).

In contrast to the monkey FPC, the human FPC is comprised by a lateral (Fp1) and a medial subdivision (Fp2) (see Figure 2 in section 1.1.4), (Bludau et al., 2014). This anatomical distinction based on the cytoarchitecture is supported by differences in co-activations of Fp1 and Fp2 as well as in their contribution to different cognitive function and behaviors (reviewed by Bludau et al., 2014). One hypothesis is that Fp1 supports *directed* whereas Fp2 drives *random* exploration (Mansouri et al., 2017). Directed exploration is considered to be based on the intent to seek further information, i.e., choosing information over immediate reward. Random exploration, on the other hand, (e.g., restless bandit task, Daw et al., 2006) is thought to rely on decision noise (Wilson et al., 2014). In favor of the hypothesis that Fp1 supports directed and Fp2 random exploration, is evidence that causally links the human Fp1 to directed exploration using TMS (Zajkowski et al., 2017) as well as the findings showing that the monkey FPC, corresponding to the human Fp2 (Carmichael & Price, 1996), supports the monitoring of the value of the currently selected behavior choice, and, further, weighs between maintaining it and the random exploration of alternative options (Mansouri et al., 2017; Tsujimoto, et al., 2010; 2012). Less consistent with the assumed functional distinction are findings linking the human right dlPFC but not Fp2 to random exploration (Tomov et al., 2020). Moreover, lacking a sub-region equivalent to the human Fp1 does not mean that monkeys are not able to engage in directed exploration, in which case it may be supported by meso-corticolimbic regions (Costa et al., 2019). Thus, directed exploration neither seems to be unique to human cognition, nor does it seem to depend on a Fp1-like structure, at least in monkeys. Moreover, it was shown that random exploration is linked to increases in neural variability and a loss of choice tuning in the FEF in the monkey brain (Ebitz et al., 2018) as well as to increased neural variability in motor related circuits in humans (Tomov et al., 2020). This evidence does not rule out that Fp2 serves random exploration, but it is less in keeping with the proposed functional distinction between Fp1 and Fp2 serving directed versus random exploration, respectively. Thus, albeit structural sub-regions with a distinct cytoarchitecture and differing anatomical as well as functional connections constitute the human FPC, and consistent evidence implicates Fp1 in directed exploration, the evidence is rather ambiguous regarding the putative link between Fp2 and random exploration.

Together, the existing body of research underscores the importance of FPC in serving the shifting away from the currently pursued towards an alternative behavioral choice. This process not only entails the redistribution of attentional selection weights, but also requires the monitoring of competing values assigned to each choice. Thus, FPC's role seems to be tied to a broader cognitive control mechanism which is crucial to a whole range of cognitive abilities facilitating goal-directed behavior that all share the need for managing competing goals and the according adaptation of the current behavioral strategy (Hogeveen et al., 2022; Mansouri et al., 2017).

1.1.6 Thematic focus and experimental framework

In real life, environmental changes that require behavioral adaptation are often subtle and may sometimes even occur outside of our awareness. This raises the question if adaptive attentional control necessitates visual awareness? Evidence has shown that FPC detects violations of implicitly learned co-variations between visible stimuli (Pollmann & Manginelli, 2009a). However, it is not clear if shifts of visual attention can occur in response to changes of fully invisible stimuli and if the redistribution of attentional resources, necessitated by such invisible changes, are supported by the FPC, similar to its role in shifting attention between feature dimensions of visible stimuli (Pollmann, 2001; Pollmann et al., 2007; Pollmann et al., 2006; Pollmann et al., 2000; although see Mansouri et al., 2020). Motivated by this questions, the first and second experiments were carried out to test if adaptive attentional control (i.e., exploratory shifts of attentional selection weights) necessitates visual awareness (experiment 1), and if the FPC supports these shifts outside of visual awareness (experiment 2). To this end, a novel masking paradigm is presented in Chapter 3, in which I tested if invisible orientation changes, requiring shifts of visual attention, lead to behavioral switch costs in response times, indexing the redistribution of attentional weights towards the novel stimulus orientation. In Chapter 4, I report fMRI data obtained using the same visual masking paradigm in a new group of participants that provides novel insights into the role of the FPC during exploratory shifts of visual attention in response to invisible feature changes.

Although implicated before in the redistribution of attentional selection weights (e.g., Pollmann, 2004), the reviewed literature in the previous section strongly suggests that FPC likely plays are broader role in managing cognitive resources in accordance with forgone behavioral choice options (Boorman et al., 2009; Mansouri et al., 2017). Consistent with this, decision-making research implicates the FPC in exploratory choices (e.g., Daw et al., 2006). Yet, economical choice studies are typically based on bandit-like gambling tasks or other decision-making tasks in which individuals choose from a predefined set of *simultaneously* presented choice options (e.g., Daw et al., 2006). Alternative behavioral options in many day-to-day decisions are however often encountered *serially* and cannot directly be compared to one another (Garrett & Daw, 2020; Kolling et al., 2012). Therefore, in Chapter 5, I present a third experiment examining FPC's involvement in supporting switches from a cognitive control mode of exploitation towards exploration when choice options are encountered serially. For this purpose, a novel patch foraging paradigm was designed that required participants to constantly weigh between behavioral exploitation and exploration. Natural foraging in both humans and animals, typically involves the question of how long a currently used source of food or energy resource should continue to be used, with the remaining resources steadily decreasing, or at what point it is better to leave, seek and explore potential alternative sources (Charnov, 1976). Foraging behavior is thus an ideal means to study the switch from an exploitative towards an exploratory mode of cognitive control (Cohen et al., 2007). Taken together, the first experimental series focused on FPC's role in the redistribution of attentional resources in response to low-level stimulus changes, examining its independence from visual awareness, whereas the third

experiment (Chapter 5) served to study FPC's involvement when a redistribution of attentional resources occurs in response to self-initiated changes in behavioral strategies. Importantly, both novel paradigms (visual masking and foraging) require uninstructed exploratory shifts of attentional resources not guided by task instruction, but more by situational changes that require a behavioral adaptation for optimal task performance (Boorman et al., 2009; Konishi et al., 2005; Mansouri et al., 2015).

Since fMRI was chosen to study FPC function in the second and third experiment, the next sections will briefly outline the principles of this imaging technique and highlight some caveats that come with using it.

1.1.7 What the fMRI signal can tell us and what not

First introduced in 1992 (Bandettini et al., 1992; Frahm et al., 1992; Kwong et al., 1992; Ogawa et al., 1992) functional MRI has become one of the most widely used methods to study various aspects of brain function, including perception, emotion, memory, and decision-making. It has a relatively high spatial resolution within millimeters but a rather low temporal resolution in the scale of seconds. It allows us to study the whole brain while participants perform a task, so that multiple brain regions can be studied simultaneously and brain activity and behavior can be directly linked.

The experiments reported in chapter 2 and 3 used the fMRI signal to study the neural correlates of exploratory shifts of visual attention. Therefore, this chapter will briefly address the underlying principles of the fMRI signal. I will also give a short overview on multivariate pattern analysis (MVPA) used in fMRI and discuss the caveats that come with studying the prefrontal cortex using fMRI and multivariate decoding technique. This chapter will not provide a full introduction to the fMRI method itself, but interested readers can find comprehensive introductions to the topic in the existing literature (e.g., Poldrack et al., 2011).

1.1.8 The BOLD contrast mechanism in fMRI

FMRI constitutes an indirect measure of neural activity, and, it may surprise, the relationship between the blood-oxygen-level-dependent (BOLD) fMRI signal and the underlying neural firing, that is supposedly reflected in the fMRI signal, is not yet fully understood. It is important to understand that the fMRI signal does not pick up the electrical activity of neurons directly, but that it heavily relies on the metabolic changes following neuronal activity. The ferromagnetic characteristic of the oxygen carrying molecule hemoglobin, that occurs in red blood cells, changes depending on its oxygenation. The molecule becomes paramagnetic following its deoxygenation (Logothetis, 2003). This magnetization increases local inhomogeneities of the magnetic field impacting the MR signal. In principle, the MR signal measures how hydro nuclei move between different energy states. FMRI relies heavily on the transverse relaxation of these nuclei (T2*) and it is exactly this transverse magnetisation that is reduced by increasing inhomogeneities of the magnetic

field (e.g., Logothetis & Wandell, 2004). Thus, in response to a relative increase of the proportion of paramagnetic deoxyhemoglobin directly after neural activity, the BOLD-signal decreases. However, we typically observe an increase in the cerebral blood flow (CBF) and an influx of oxygenated blood into the neighboring capillary system following neural activity. Importantly, the amount of oxygenated blood coming in is higher than the amount of oxygen that was consumed (Fox & Raichle, 1986; Fox, et al., 1988; see Logothetis & Wandell, 2004). This 'overcompensation' leads to an increase of diamagnetic blood cells stabilizing the magnetic field and thereby enhancing the BOLD contrast.

This dynamic stresses the temporally delayed nature of the BOLD signal as well as that it is not the neural activity itself, but the CBF and the cerebral metabolic rate of oxygen (CMRO2) that impact the BOLD signal. CBF increases the blood oxygenation while CMRO2 decreases it, and it is the relative balance of the changes in CBF and CMRO2 that is crucial for the BOLD signal (see Buxton, 2010). This circumstance introduces further complications: both CBF and CMRO2 are controlled through neural activity in parallel but not serially and the nature of that neural activity is not yet fully identified (Buxton, 2010; Logothetis & Wandell, 2004). Moreover, a significant positive signal change of the BOLD response, that may occur in response to a stimulus, does not necessarily index an increase in the firing of the neurons selective for that stimulus (Goense & Logothetis, 2008). Instead, there is good evidence showing that it is the local field potentials (LFP) but not the spiking activity that correlate more strongly with the BOLD response (Berens et al., 2010; Goense & Logothetis, 2008; Logothetis & Wandell, 2004). Both types of neural activity can but do not have to be correlated with each other. In the later case, it is the LFPs that remains correlated with the BOLD response but never the spiking activity (Logothetis et al., 2001; Mathiesen, et al., 1998; Mathiesen et al., 2000). In contrast to spiking activity, LFPs presumably reflect input signals from rather than output signals to a specific brain region, mostly driven by synaptic and dendritic activity rather than axonal action potentials. They emerge from complementary excitatory and inhibitory neural activity which are tied not only to processing of sensory signals but also, more importantly, to neuromodulatory processes (Logothetis, 2003; Logothetis & Wandell, 2004).

This short outline highlights the indirect nature of the BOLD signal. It also shows that, although we do know which type of neural activity is more likely to be contributing to broader signal changes of the BOLD response, the exact mechanisms are still debated. Both caveats should find consideration whenever we try to link empirical results based on the BOLD signal with neural activity and brain function.

1.1.9 Multi-voxel pattern analysis of fMRI data

Almost one decade after the propagation of fMRI as one of the key methods in cognitive neuroscience (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992), in 2001, Haxby and colleagues published a methodological prototype based on multivariate pattern analysis (MVPA) of fMRI data to study the functional structure for face and object recognition in the ventral temporal cortex (Haxby et

al., 2001). Instead of looking for univariate signal changes to map certain sensory or cognitive abilities to specific brain regions, MVPA allows to decode unique brain states from specific patterns of activation in an array of brain voxels (Haxby, 2012). By analyzing the neural responses as an entire ensemble of voxels, the method allows to relate specific activity patterns to specific experimental conditions. Thereby, it became possible to decode from the multivariate response patterns what kind of object subjects saw (e.g., Haynes & Rees, 2005; Kamitani & Tong, 2005), whether that object was animated or not (Kiani et al., 2007; Kriegeskorte et al., 2008), whether that object was the target of attention (Kamitani & Tong, 2005), or which task subjects intended to perform next (Haynes et al., 2007). MVPA thus became a powerful complementary means to analyze fMRI data. In the tradition of functional brain mapping, cognitive neuroscientists typically have tried to link specific brain regions to specific cognitive functions. The MVPA opened a new avenue for qualitatively different questions about how differences in cognitive states are represented in the brain and how certain brain regions encode information of different qualities (e.g., abstract versus sensory).

The first fMRI experiment reported in this thesis (Chapter 4), I carried out to firstly identify key regions underlying exploratory shifts of visual attention, and, secondly, to test if these regions also encode the relevant, yet unconscious stimulus information driving the shifts of attention. To this end, I used the mass-univariate analyses based on the GLM as well as MVPA to analyse the BOLD signals related to attention shifts. Thus, I will briefly outline the principles of MVPA and stress some caveats inherent to the method when used to study brain representations.

Typically, MVPA comprises four steps (Norman et al., 2006): firstly the array of voxels needs to be determined (i.e., *feature selection*). This is realized by either narrowing the analysis down to a region of interest (ROI), or by looking at the entire brain. In the latter case a searchlight analysis is often used (Kriegeskorte et al., 2006), where the analysis is carried out in so-called searchlight spheres of a specific size (e.g., 3 voxel diameter). The classification analysis is consequently performed for the vector of voxels contained in that sphere, and the procedure is repeated until each brain voxel has served once as the center of a searchlight sphere. The second step involves sorting the data so that the patterns of activity across the selected voxels are assigned to a particular time in the experiment (i.e., *pattern* assembly). The voxel patterns are then labeled in accordance with the experimental conditions in which they were recorded. In the third step (*classifier training*), parts of the labeled data are handed over to a multivariate pattern classification algorithm and the algorithm then learns to link the activity patterns to the experimental conditions resulting in a mapping function. One type of pattern classification algorithm that is widely used, are support vector machines (SVM). It is able to learn training examples that contain information on the basis of which two ore more categories (i.e., features or stimulus types) can be distinguished. Technically, the SVM classifier determines a boundary between the categories it is exposed to during training in such a way that this boundary maximizes the separation (margin) between the categories. The training examples, usually a subset of the data, are called support vectors. In the fourth step the classifier is tested whether it can apply the learned mapping function

to novel data. To do so, a new multi voxel pattern of brain activity is fed to the algorithm and tested whether it can be assigned correctly to the corresponding experimental condition (*generalization testing*), (Norman et al., 2006). For a deeper understanding of the mathematical foundations of the MVPA approach, the reader might be referred to e.g., Formisano et al., (2008).

By simultaneously taking into account the measured activity of multiple voxels, MVPA is able to pick up on combinatorial effects across voxels. Therefore, it often appears to have a higher sensitivity compared to mass-univariate analyses based on the GLM (e.g., Davis & Poldrack, 2013), meaning that even in the absence of statistically significant overall changes in the recorded BOLD signal, classifier algorithms may be able to decode information from sub-threshold variations in multi voxel activity patterns (Etzel et al., 2013). One crucial caveat, however, regards the representational ambiguity inherent to MVPA. Stimuli usually encompass various features, and likewise, task paradigms may encompass distinct sources of variation, each corresponding to different stimulus features or cognitive states that are reflected in the recorded brain activity. A classification algorithm may pick up on any consistencies in these variations, is thus able to find distinguishable activity patterns, and would map them successfully to the respective stimulus classes or task conditions. This then would indicate that the brain area conveying the informative activity patterns indeed encodes something about the stimuli or task conditions. However, as long as there are multiple alternative features correlating with each stimulus class or task condition, successful classification does not tell what exactly is encoded (Naselaris & Kay, 2015). Thus, on the one hand, MVPA does provide a powerful tool if the aim is to predict stimuli or cognitive states (e.g., does a brain region of interest represent distinguishable information about different attentional states?). On the other hand, due to the representational ambiguity, the method is limited in showing how the brain actually represents those stimuli or cognitive states (e.g., does the nature of brain representations of different attentional states differ, or how does the early visual cortex represent different visual stimulus features?).

Summarizing the above, although FMRI fails to assess the exact neuronal mechanisms that may underlie the task behavior studied in this thesis, it is well-suited to measure signal changes which reflect local changes in overall neural activity. Importantly, it enables measuring these changes with a high spatiotemporal resolution across the entire brain. Combining fMRI with novel forms of multivariate analyses simultaneously examining local activity patterns of a whole array of voxels (MVPA) additionally offers new ways to decode cognitive states that are specific to different experimental conditions. Thus, in combination with its non-invasive nature, fMRI appears to be an optimal tool to study the functional role of the FPC in the redistribution of attentional resources in healthy human participants.

2 Common Methods

This chapter covers those parts of the methodology overlapping in experiments 1 and 2 in Chapters 3-4 to avoid repetitive reporting. The methodological aspects specific to each experiment are detailed in the respective methods section.

2.1 Participants

In all three experiments young healthy adults of both genders took part. They were either recruited from the pool of psychology students of the Otto-von-Guericke University or from the participant pool of healthy adults of the University Klinik Magdeburg. Only those volunteers with no report of psychiatric or neurological illness were included. Handedness was determined by self-report and both left- and right-handed participants took part in the experiments. All participants reported a normal or corrected to normal vision. All gave written informed consent consistent with the protocols approved by the local ethics committee of the Otto-von-Guericke University prior to the experiments.

2.2 The behavioral tasks

2.2.1 Experiment 1 and 2

Parts of the following methods have been published in: Güldener, L., Jüllig, A., Soto, D., & Pollmann, S. (2021). Feature-Based Attentional Weighting and Re-weighting in the Absence of Visual Awareness. Frontiers in Human Neuroscience, 15, 610347. and in: Güldener, L., Jüllig, A., Soto, D., & Pollmann, S. (2022). Frontopolar activity carries feature information of novel stimuli during unconscious re-weighting of selective attention. Cortex, 153, 146-165.

2.2.2 The Masking paradigm

For the main experiment of Chapter 3 and 4 we developed a novel orientation categorization task based on masked Gabor patches presented in the center of the screen. Participants had to quickly make a forced-choice response with two custom buttons, deciding if a vertical or non-vertical grating had been presented. In each trial, after the categorization response participants rated how well they perceived the orientation of the masked grating using the four-point perceptual awareness scale (PAS), (Ramsøy & Overgaard, 2004). As we were particularly interested in examining differences between unaware and aware trial conditions, we aimed to achieve a maximum number of trials with unaware 1-ratings respectively aware 3 or 4-ratings. Thus, the luminance contrast was adjusted on every trial after the participants rated their subjective awareness: the luminance contrast value was further decreased following trials rated as fully aware (AL4), almost fully aware (AL3), and residually aware (AL2), and increased if the participant reported being fully unaware of the orientation (AL1). A trial started with the brief presentation of a central fixation cross for 500 ms, followed by a blank screen for another 500 ms. Next, the target Gabor occurred at the center of the screen for 33 ms. The mask

followed immediately for 350 ms. Subjects were trained to give their categorization response during the next 1500 ms following the onset of the Gabor. Following this, they had another 2.5 sec to rate the subjective visibility of the Gabor using a keyboard with four keys. Figure 3 shows an example of a trial sequence. All trials were separated by inter-trial-intervals (ITI) with varying durations (1.5-3.5 sec) following a logarithmic distribution. All volunteers that passed the calibration process (see next chapter) completed 10 runs of the main experiment (360 trials).

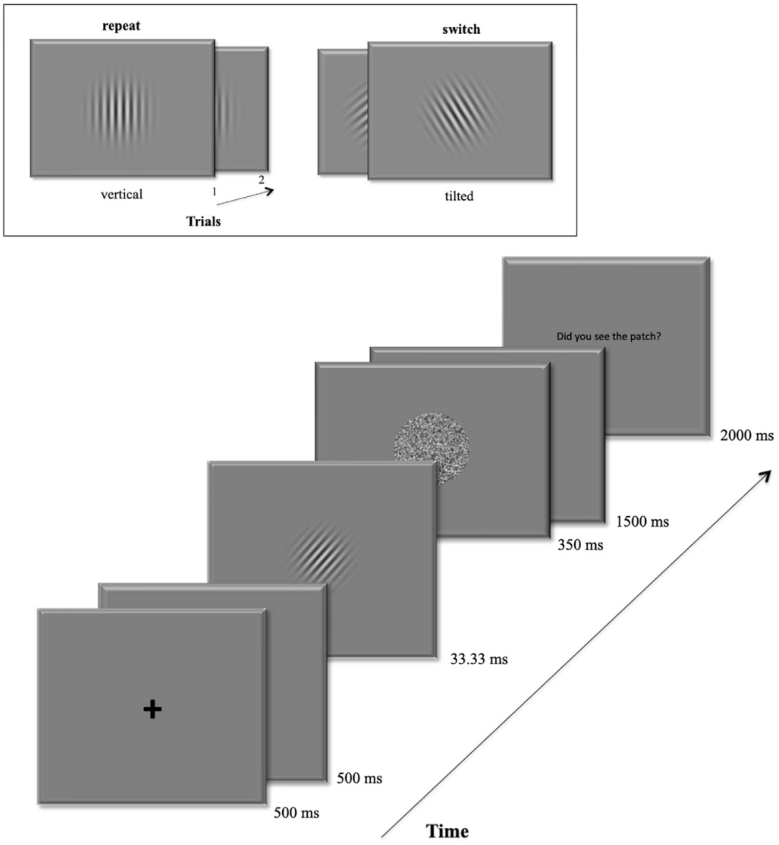


Figure 3: Example of a trial sequence. The box at the top shows an example of the repeat condition (left): a vertical target grating in the first trial is followed by another vertical grating in the second trial. On the right it shows an example for the switch condition: the left-tilted target grating is followed by a right-tilted grating in the next trial.

2.2.3 Threshold determination

Experiment 1 and the consecutive fMRI follow-up study (experiment 2) targeted the question if and how attentional resource allocation occurs in response to invisible feature changes. Therefore, it was crucial to determine participants' perceptual threshold for the luminance contrast of the experimental stimuli at which stimuli would still escape visual awareness.

For this purpose each experimental session started with a 1-up:1-down adaptive staircase procedure (adopted from Jachs et al., 2015) to determine the stimulus' luminance contrast for the first trial of the main experiment. The task was performed at the PC in the behavioral experiment (experiment 1). In the fMRI follow-up participants underwent the staircase procedure inside the scanner.

Gabor patches occurred centrally on the screen for 33 ms directly followed by a random-dot mask for 350 ms. If participants saw the grating's orientations they were to respond by pressing the "2" button, while the "1" button was to be pressed if they did not see anything at all. In the following main experiment participants would rate the subjective visibility of the target at the end of each trial using the four-point perceptual awareness scale (PAS): 1: "did not see anything at all", 2: "saw a brief glimpse without seeing the orientation", 3: "had an almost clear image of the stimulus", 4: "saw the stimulus and its orientation" (Ramsøy & Overgaard, 2004). During initial calibration, participants were thus instructed to give an unaware response only if they did not see anything at all which corresponded to the "1" rating of the PAS. Conversely, they were to give an aware response, corresponding to the remaining three points of the PAS whenever a brief glimpse or a more stable percept of the Gabor grating was experienced. The Gabor's luminance contrast was increased following an unaware response and further decreased following an aware response. Each participant completed 90 trials (30 trials for each of the three orientations) and the percentage of aware responses was calculated on a trial-by-trial basis. The subjective awareness threshold was reached when the percentage of aware responses was about 50% over the last ten trials. Then, the final threshold luminance contrast was defined as the mean luminance contrast across the last ten trials of the staircase. If the individual threshold was not reached within the 90 trials, the staircase was repeated. Next, participants performed one block of training under experimental conditions consisting of 36 practice trials. Here, the luminance contrast obtained with the first staircase procedure was used for the contrast value of the training stimuli. The practice phase was followed by a second calibration conducted according to the same protocol as the first staircase procedure. This recalibration provided the threshold value for the luminance contrast used in the first trial of the main experiment. On later trials, this value was further adjusted. If the grating's visibility was rated with a 2, 3 or 4 on the PAS, the contrast value was decreased in the next trial. If the grating was rated as being invisible (AL1), the contrast value was increased instead.

2.3 Neuroimaging - fMRI data acquisition in experiment 2 and 3

To obtain the functional magnetic resonance imaging data (fMRI), all participants were scanned on a 3 Tesla MAGNETOM Prisma (Siemens). fMRI data was sampled using a standard head coil and EPI-sequence (TR, 2000 ms; TE, 30 ms; flip angle, 90°; epi factor, 80; echo time, 0.49 ms; matrix size, 80 × 80; FOV, 240 mm; 36 slices with interleaved acquisition; 3 mm isotropic voxels; 0.3 mm interslice gap). The first three dummy scans were excluded prior to analysis. T1-weighted MPRAGE scans (TR, 2500 ms; TE, 2.82 ms; flip angle, 7°; matrix size, 256 x 256; FOV, 256 mm; 192 slices, 1 mm isotropic resolution) were additionally sampled for each participant.

3 Experiment 1: Feature-based attentional weighting and re-weighting in the absence of visual awareness

The results of the following behavioral experiment were first published in: Güldener, L., Jüllig, A., Soto, D., & Pollmann, S. (2021). Feature-Based Attentional Weighting and Re-weighting in the Absence of Visual Awareness. Frontiers in Human Neuroscience, 15, 610347.

3.1 Introduction

For survival in an unstable and uncertain world, it is crucial to detect contextual regularities, but also to adapt quickly when they change. Since such contextual changes may be complex and occur very rapidly, the question arises as to whether attention shifts in response to environmental changes are contingent on visual awareness. Previous studies examined the effect of exogenous invisible cues on the deployment of external visual selective attention, suggesting that subliminal spatial cues can capture attention and facilitate task performance at the cued location (McCormick, 1997; Mulckhuyse et al., 2007; for a review see Mulckhuyse & Theeuwes, 2010), that the association between a subliminal cue and a visible target can be learned implicitly (Lambert et al., 1999) and that subliminal stimulus can even induce cognitive control processes like response inhibition or task-switching effects (Farooqui & Manly, 2015; Lau & Passingham, 2007; Van Gaal et al., 2008; Van Gaal et al., 2010). This notion is further supported by evidence from clinical studies in "blindsight" patients, which indicate that visual cues presented in the patient's blind field are still capable of directing spatial attention (Kentridge et al., 1999).

It is, however, less clear whether feature-based attention can be redirected towards a novel feature (feature-based attentional re-weighting) in response to changes in unconsciously processed targets: according to Bundesen's theory of visual attention (TVA), (Bundesen, 1990), the attentional selection is a mechanism that operates in the service of perceptual categorization, i.e., by aiding the selection of a potential target item within a distractor display (*"filtering"*), or the discrimination of features in single items (selection of categories, *"pigeonholing"*). The processing speed for this visual selection depends on both the attentional weight and the perceptual decision bias. In theory, the attentional weight relies on the sensory evidence indicating the category a certain stimulus belongs to ("bottom-up"), and the goal-relevance of that category, i.e. the importance of attending to a certain stimulus category ("top-down"; Bundesen, 1990). Thus, the weaker the sensory evidence is, the more the attentional weighting should rely on the "top-down" mechanism (the importance to attend to this category). Based on the TVA's assumption a "top-down" driven attentional bias (i.e., the goal-relevance) on visual selection is predicted especially for invisible non-consciously processed visual stimuli because the sensory evidence that could support visual selection in a bottom-up fashion (i.e., the saliency of the stimulus) is very limited if the stimulus is only unconsciously perceived. Importantly, evidence is still missing as to whether such a feature-based selection bias can be elicited for subliminal, unconsciously

processed stimuli and whether it can be reweighted flexibly in response to feature changes of the unconsciously processed stimulus.

Later accounts of visual attention criticize the dichotomy of bottom-up vs. top-down attentional weighting and propose to include a history-driven weighting of attentional selection (e.g., Awh et al., 2012; Theeuwes, 2018; 2019) to better incorporate empirical evidence showing that not only can stimulus saliency and internal goals (volitional control) bias attentional selection but the "history" of former attention deployments driven by e.g., reward, intertrial priming, or statistical learning (Awh et al., 2012) can also have an influence. For consciously perceived visual stimuli, such history-driven attention weighting effects have been observed in singleton search tasks. For instance, repeated presentation of the same target-defining dimension leads to response time benefits and associated activation changes in dimension-specific visual processing areas (Pollmann et al., 2006) that were interpreted as evidence for an attentional weighting of the target-defining dimension (Liesefeld et al., 2019; Müller et al., 1995). In contrast, when the target-defining dimension changes, e.g., when the target was defined by a singleton color in recent trials and then is defined by a singleton motion direction, response time costs are observed, as would be expected when attention needs to be reweighted to the new target-defining dimension. These re-weighting processes occur incidentally, in the absence of an explicit instruction to attend to the new target-defining dimension (Müller et al., 2004). Furthermore, a comparable spatial attention weighting pattern is observed when implicitly learned target-distractor configurations change in the contextual cueing paradigm (Manginelli & Pollmann, 2009; Pollmann & Manginelli, 2009a). When attention-weighting processes occur in the absence of explicit task demand and even after changes of implicitly learned configurations, the next question would be whether attentional re-weighting can also occur as an adaptive adjustment to unconsciously perceived stimulus changes.

Therefore, this study addressed two key questions. First, we asked whether the repeated presentation of an invisible target feature can lead to a temporally persisting attentional selection bias. The second question was how flexible this attentional bias is, i.e., whether a novel invisible target can trigger the re-weighting of visual attention to the new target feature in the absence of awareness. Peremen et al. (2013) studied the relation of intertrial feature priming and visual awareness during a letter search task. They reported that the repetition of the target shape speeded visual search only when the target in the prime display had been consciously perceived. Yet, it remains unknown whether unconscious re-weighting of visual selection can occur for simpler orientation stimuli such as Gabor patches (Rajimehr, 2004). We also considered a different task setting in which the selection task occurred at a fixed attended location throughout the trials. In all previous studies, attention-weighting effects were examined in multi-item displays and search tasks for a singleton target. Our paradigm does not involve spatial shifts of attention but rather a process of visual selection in which the same spatial location is always attended.

Specifically, our paradigm involved an orientation discrimination task based on a central masked bar stimulus. Volunteers were instructed to discriminate whether the target stimulus was vertical or tilted irrespective of the specific direction of tilt. They had to make no further distinction between the two tilted orientations. Yet, to introduce the tilt-based attentional selection bias, we manipulated the likelihood of the two non-vertical gratings (left vs. right) so that one tilt would occur twice as often as the other. Consistent with the proportion congruency effect during priming (Blais et al., 2016; Bodner & Lee, 2014), and feature-based statistical learning (Chetverikov et al., 2017; Turk-Browne & Scholl, 2009), an increase of the frequency at which a right or left-tilted grating appeared should result in a high selection weight for the frequent orientation indicating the importance to attend to this category. This prediction is based on the idea that the relevant feature information (e.g., the spatial orientation) of the most likely target gets represented in a form of a short-term description - the attentional template (Desimone & Duncan, 1995), to control the sensory processing so that stimuli matching the description are favored, i.e., are more readily processed in the visual system. The degree to which a stimulus matches the attentional template defines its attentional weight. Thus, Gabor patches that fit the information stored in the template receive a high selection weight, e.g., 1, while mismatching Gabor patches (infrequent and vertical) have reduced selection weights as the whole weight is thought to be a constant value: if the weight increases for one feature it decreases for another (Duncan & Humphreys, 1989). Now, concerning behavior, a switch from the heavily weighted orientation to a target with a vertical or the infrequent spatial orientation should require a shift of selection weights due to the mismatch between the sensory input and the attentional template. This shift of attentional selection weights was expected to lead to slowing stimulus processing and response initiation eventually resulting in increased response latencies in such switch trials. The higher the selection bias for the Gabor patch's orientation in the preceding trial, the more re-weighting should be necessary to process and respond to a novel grating in the subsequent trial. Thus, particularly switch trials in which the prior orientation was the highly frequent tilt should show prolonged response latencies on the behavioral level, given that the increased likelihood of one orientation over the others was sufficient to induce a prior selection bias (e.g., Chetverikov et al., 2017; Leber et al., 2009). Importantly, a combination of signal detection theoretic measures (Stanislaw & Todorov, 1999) and subjective perceptual ratings (Ramsøy & Overgaard, 2004) was used to assess participants' awareness of the stimulus to avoid potential confounds due to criterion biases in reporting (un)awareness, e.g., reports of no experience for the knowledge held with low confidence (Soto et al., 2019; Wiens, 2007). Therefore the unconscious re-weighting of selection hypothesis was eventually tested by maintaining a clear separation between the measures of selective attention weighting, inferred by the pattern of response latencies, and the measures that we used to probe (un)awareness of the stimulus (objective orientation discrimination task and subjective reports). We predicted decision reaction time (RT) costs due to a change of the tilt direction. Costs should be highest if the prior orientation was the highly biased tilt, i.e., a switch from the frequent to the infrequent tilt or a vertical target, and they should occur even if the novel target is non-consciously perceived.

3.2 Materials and Methods

3.2.1 Participants

In total 21 native German students (three male) from the University of Magdeburg, Germany took part in the experiment. All volunteers were between 19 and 34 years old (M = 24.90 years). They were either monetarily reimbursed (8 euros per hour) or received course credits for the 2 h of participation. In two sessions an error in the response collection occurred and the respective participants were removed from the analysis. Another volunteer interrupted the session at an early stage and was thus excluded. During data analysis, five other participants were identified to have more than 40% missing responses during the 1.5 s response deadline (see below) and were thus excluded from analysis of RT data.

3.2.2 Apparatus and Stimuli

The stimulus display and responses were controlled with the Python toolbox "Psychopy" (Peirce, 2007; Peirce et al., 2019). The stimuli were presented on a 24" Samsung monitor (1920:1080 resolution, 60Hz refresh rate). All participants were placed 50 cm away from the screen. Stimuli were Gabor gratings with an individually calibrated contrast (see next section) centrally presented on a grey background subtending 3.4° visual angle. Its spatial frequency was 3.7° cycles per degree. The patch´s orientation was either vertical (180°), 165°, 150° or 135° if it was a left-tilted, non-vertical Gabor patch, and 195°, 210° or 225° if it was a non-vertical patch tilted to the right. To further reduce the visibility of the Gabor patch we used a circular backward mask of black and white random dots (3.4° visual angle).

3.2.3 Experimental Task and Procedures

3.2.4 Threshold Determination

A session started with a staircase procedure that is described in detail in the Chapter covering common methods in the section 2.2.1.2. The aim of the calibration was to calibrate the stimulus's luminance contrast so that its orientation was rendered invisible. The final threshold luminance was defined as the mean luminance contrast across the last 10 trials of the staircase. Next, participants performed one block of training under experimental conditions consisting of 36 practice trials. Here the luminance contrast obtained after the first staircase procedure was used for the contrast value of the training stimuli. The practice unit was followed by a second calibration conducted according to the same protocol as the first staircase procedure (see 2.2.1.2). Eventually, this recalibration provided the threshold value for the luminance contrast used in the main task.

3.2.5 Behavioral Task

In the main experiment, volunteers were asked to perform an orientation categorization task based on masked Gabor patches. The details of this task were already outlined in 2.2.1.1.

3.3 Design

To facilitate the occurrence of a tilt-based attentional selection bias, we introduced uneven proportions of the two non-vertical gratings (left vs. right). Consistent with feature-based statistical learning (Chetverikov et al., 2017; Turk-Browne & Scholl, 2009), the relative increase of the frequency at which a right or left-tilted grating appeared was expected to strongly weight attentional selection for this orientation. Its higher likelihood should increase the importance of attending to this feature, resulting in a high selection weight. At the same time, the selection weight for the other two orientations (vertical and the infrequent tilt) should be reduced (Bundesen, 1990). Eventually, switches away from the heavily weighted tilt were expected to result in a significant increase in volunteers' response times. Therefore, for a block of 36 trials, we chose 12 vertical targets ($\sim$33%), and used uneven proportions of the two tilts, with 18 trials (50%) and six trials ($\sim$16%), respectively. This way, each block was either left- (75% of all non-vertical trials were left-tilted) or right-weighted (75% of all tilt trials were right-tilted). The actual presentation of the three orientations was randomized within a single block.

The first 11 participants performed 14 blocks in the main experiment (504 trials). The eighth subject, however, interrupted the session after 12 blocks were completed. Subjects 12–21 completed 10 blocks (360 trials) as this amount of trials turned out to be sufficient to obtain enough trials for each awareness level (AL) while avoiding growing weariness that was reported by subjects completing 14 blocks.

3.4 Statistical Analysis

Sensitivity and response bias measures were calculated using custom-made Python code (Version 2.7). All statistical analyses were carried out with R (Version 3.5, R Core Team, 2014). For the Bayes factor (BF) analysis (Rouder et al., 2009), we used JASP (JASP Team, 2019).

3.4.1 Subjective awareness

To determine the level of subjective awareness (AL) in experiment 1 and 2, the number of trials for each subjective AL was counted using the trial-by-trial PAS-rating for each participant.

3.4.2 Discrimination performance

To examine whether the participant's ability to correctly discriminate between vertical and non-vertical gratings depended on the level of subjective awareness, the individual response bias and perceptual sensitivities was determined using signal detection theory (Macmillan & Creelman, 2004; Stanislaw & Todorov, 1999). Group-effects were subsequently assessed for each level of subjective awareness using BFs since it was required to prove the absence of sensitivity (H0), (Dienes & Mclatchie, 2018; Gallistel, 2009). A BF_{10} provides moderate evidence for H0 (e.g., A' = 0.5) if it stands between 0 and 0.33, anecdotal evidence if it stands between 1/3 and 1, and evidence for H1 (A' $>$ 0.5) if it exceeds 1 (Dienes & Mclatchie, 2018),

with a BF_{10} between 1 and 3, 3 and 10, 10 and 30, 30 and 100 and >100 providing anecdotal, moderate, strong, very strong, and extreme evidence, respectively, for H1 (Jeffreys, 1998; Quintana & Williams, 2018).

Under Yes/No-conditions A′ and the criterion location (C) were calculated to determine perceptual sensitivity and bias: we calculated false-positive rates [FPR = False alarms/(False Alarms + Correct Rejections)] and hit rates [TPR = Hits/(Hits + Misses)] defining a hit as the correct report of a non-vertical orientation when the Gabor's orientation truly was tilted; false alarms were defined as tilt response for vertical gratings. We used the following formulas to calculate the non-parametric response bias and sensitivity (Stanislaw & Todorov, 1999):

$$C = -(Z(TPR) + Z(FPR))/2$$

$$A' = 0.5 + |sign(TPR - FPR)((TPR - FPR)2 + |TPR - FPR|/(4max(TPR, FPR) - 4 * TPR * FPR))|$$

Values of C around 0 indicate unbiased discrimination performance. A liberal decision criterion favoring yes-responses (i.e., reporting a non-vertical grating) leads to values of $C < 0$, while positive values occur if participants are biased to report a vertical target. If volunteers possess perfect sensitivity at discriminating the target orientations, A′ appears to be equal to 1 and it decreases to 0.5 if the sensitivity diminishes (Stanislaw & Todorov, 1999).

3.4.3 Analysis of RT Data

We used the packages *lme4* (Bates et al., 2015) as well as *lmerTest* to make use of a linear mixed model (LMM) analysis. As the data was unbalanced due to the variations in the subjective awareness ratings (PAS) that lead to uneven numbers of trials across the four levels of visual awareness, LMMs were chosen over custom repeated measures ANOVAs to analyze the RT (e.g., Avneon & Lamy, 2018). Since all cases with missing data would be excluded in a repeated-measures ANOVA, the LMM approach is the better means to make use of all available data in the face of an unbalanced design (Magezi, 2015). Only RTs of trials with correct responses entered the analysis after each participant's individual outliers (mean RTs ± 3 SD) were removed.

Before assessing the significance of the fixed effects, we determined the random effect structure of the final model with likelihood ratio tests (i.e., comparisons of models differing in their random effect structure). Importantly, we did not use likelihood ratio tests to compare models with differences in their fixed effects as these were already determined by the design (see below). Once the final model for analysis was fully defined, we fitted this model with the RT data using a restricted maximum likelihood estimation (REML) and tested the statistical significance of the fixed effect predictors with a type III ANOVA with F-statistics as implemented in the *lmer* function of the *lme4* package (Version 1.1–23), (Bolker et al., 2009; Luke,

2017; McNeish, 2017; Richardson & Welsh, 1995). The p-values were calculated using Satterthwaite approximations to degrees of freedom with the *anova* function of the package *lmerTest* (Version 3.1-2), (Kuznetsova et al., 2017). We chose the ANOVA approach to test the statistical significance of the fixed effects as this approximation is thought to be producing acceptable Type 1 error rates even for small samples while the use of model comparisons (likelihood ratio tests) is not recommended to test fixed effects because they appear to be anti-conservative (Bolker et al., 2009; Luke, 2017; Pinheiro & Bates, 2000). Post-hoc tests (least squared means of the contrasts with Bonferroni correction) were performed using the R package *emmeans* (Version 1.4.7). Finally, we used the R function *r.squaredGLMM* as implemented in the R package *MuMin* to calculate the marginal R squared (R^2m) and conditional R squared (R^2c) to obtain standardized effect sizes. R^2m is interpreted as the variance explained by the fixed effects of awareness and switch and R^2c gives the variance explained by all fixed and random effects (Johnson, 2014).

The main goal we pursued in the study was the examination of whether a changing orientation from one trial to another (switch) affected participants' responses: we predicted a switch-related slowing of RTs compared to trials in which the orientation remained unchanged (repeat). Thus, the *switch* of orientations (switch vs. repeat) constituted the first fixed effect predictor in the LMM. RTs were also expected to decrease with increasing visual awareness: the more the participants saw, and the more confidently they should perform at categorizing the stimulus orientation, the faster they should be at responding to the grating's orientation. Therefore, visual *awareness* was defined as the second fixed effect predictor of the basic model. Finally, to make allowance for a possible interaction between the two fixed effects we included the interaction term of *switch* and *awareness* into the final LMM. Regarding inter-individual baseline differences in response latencies, we also defined a by-subject random intercept accounting for non-independency of single subjects' data. Thus, the basic model was formalized as RT $\sim$ *switch* + *awareness* + *switch:awareness* + (1 | *subject*).

In this model, however, the full random effect structure still needed to be determined. Therefore, we next used model comparisons based on likelihood ratio tests (χ^2) with the *anova* function of the *lme4* package (Baayen et al., 2008) to assign the full random effect structure (Barr, 2013) of this basic model. Defining the random effect structure is important to balance between the type I error rate that inflates if the random effect structure of an LMM is underspecified (Barr, 2013), and the model power that suffers if the random effect structure is more complex than the given data (Matuschek et al., 2017). The method of model comparisons based on likelihood ratio tests compares to the procedure of a hierarchical regression in which relevant predictors are added to the regression model and kept if they significantly improve the model fit (changes in R^2). Likelihood ratio tests are deemed to be appropriate to formally define the random effect structure of an LMM even if the sample size is small (Baayen et al., 2008; Bolker et al., 2009). Using this method, we tested the basic model containing only a by-subject intercept against alternative models containing an additional by-subject random slope for awareness and/or a by-subject random slope for the switch. The details of this analysis are reported in

the Appendix A.1. Importantly, we used the likelihood ratio tests only to determine the random effect structure of the final model that we used to fit the RT data with, while the significance of the fixed effects (i.e., the hypotheses testing) was assessed using the type III ANOVA with Satterthwaite approximations to degrees of freedom (Luke, 2017). Based on the model comparisons we included a by-subject random slope for *awareness* to model potential by-subject heteroscedasticity concerning visual awareness (i.e., allowing uneven variances across the levels of the fixed effect *awareness),* (Baayen et al., 2008). Eventually, the final model for significance testing was defined as RT $\sim$ *switch* + *awareness* + *switch:awareness* + (1 + *awareness* | *subject*).

The final LMM with the structure outlined above was applied in two RT models: In the first model (average RT model) we included all possible orientation changes in the switch condition. In the second model (weighted RT model) the switch condition contained only those switch trials in which we expected the highest RT costs to occur: the frequency differences between the three orientations were expected to boost the selection weight for the highly frequent non-vertical orientation (either left or right). Consequently, re-weighting to the infrequent non-vertical orientation should be associated with more pronounced switch costs than vice versa. The same was predicted for changes away from the heavily weighted to the vertical orientation requiring stronger attentional re-weighting. However, switches away from the low-frequent tilted orientation to vertical should lead to less prominent RT costs because the attentional selection weight for this tilted orientation is weaker, facilitating the shift of attentional resources towards the novel target orientation. Hence, these trials were not included in the weighted RT model. We separately report the results for the LMM analyses for the *average* and the *weighted* RT model.

3.5 Results

3.5.1 Subjective Awareness

In the majority of trials, participants' subjective awareness of the to-be-discriminated orientation was low (AL2, 25.94%), or reported experience was fully absent (AL1, 37.70%). In about 26.80% of all trials, subjects reported an almost clear perception of the grating (AL3) and its orientation. In only 9.54% of all trials did they clearly see the grating and its orientation (AL4). Due to the low number of these AL4 trials, we excluded them from the following analyses. A detailed summary of the number of trials for switch and repeat trials for each level of subjective awareness is reported in Appendix A.2. After the experimental session, each volunteer was asked to report whether any differences in the frequencies of the stimulus orientation had been noticed. The majority of subjects reported that more tilted than vertical orientations had been presented, but none of the participants noticed the block-wise changing frequency difference for the tilted orientations (left vs. right).

3.5.2 Objective Discrimination Ability and Subjective Awareness Concordantly Diminish

Signal detection analyses revealed that on trials with almost full (AL3) and partial awareness (AL2) participants' perceptual sensitivity was significantly above chance. Bayes-factors provided extreme evidence that sensitivity (A') was greater than 0.5 in AL3, $BF_{10} > 100$, 95% CI (0.733, 0.845), and AL2 trials, $BF_{10} > 100$, 95% CI (0.672, 0.750), (Quintana & Williams, 2018). The mean A' of 0.789 ± 0.026 (SE) in AL3 trials was 7.6 times more likely to be greater than the mean A' of 0.711 ± 0.018 in AL2 trials, BF_{10} AL3 $>$ AL2 = 7.608. On unaware trials (AL1), however, we observed a mean A' of 0.516 ± 0.030 that was more likely to be equal to 0.5 with moderate evidence for the H0, $BF_{10} = 0.317$, 95% CI (0.451, 0.582), indicating the absence of perceptual discriminability of the gratings' orientation.

In contrast to the perceptual sensitivity analyses, individual response biases remained unaffected by changes in subjective awareness. In none of the four ALs did we find clear evidence for a more liberal or more conservative response criterion to report a non-vertical orientation, than a C around 0. BFs were rather in favor of the null hypothesis indicating that the mean C of –0.188 ± 0.173 in AL1 trials was more likely not different from zero, yet with only anecdotal evidence for the H0, $BF_{10} = 0.457$, 95% CI (–0.560, 0.188), (Quintana & Williams, 2018). The same was true for the mean C of –0.073 ± 0.106 in AL2 trials, $BF_{10} = 0.342$, 95% CI (–0.303, 0.157), and for a mean C of 0.089 ± 0.123 in AL3 trials, $BF_{10} = 0.349$, 95% CI (–0.179, 0.357). To examine variations in the response criterion location (C) across the three levels of subjective visual awareness, we made use of LMM to optimally deal with the unbalanced data set (Magezi, 2015). Since we aimed to assess the absence of variations in C across the three levels of subjective awareness, we conducted a Bayesian-based LMM analysis using the R package *BayesFactor* to obtain a Bayes factor (BF_{10}) directly proving the null hypothesis (Morey et al., 2018): first we constructed a null model in which only a by-subject random intercept was included assuming that variations in C relied on inter-individual differences only, [$C \sim 0 + (1 \mid subject)$]. Next, we constructed an alternative model in which the subjective awareness reports (awareness ratings 1–3) served as a single fixed effect explaining variance in C in addition to the by-subject random intercept, [$C \sim awareness + (1 \mid subject)$]. Using the *lmBF* function, we then calculated BFs for each model and compared the two models by dividing the BF of the model that included awareness as a fixed effect by the BF of the null model. The analysis resulted in an inconclusive BF_{10} of 0.58 providing weak evidence for the absence of variation in C across the three levels of subjective awareness (Quintana & Williams, 2018).

Descriptive data of sensitivity and bias measures are reported in Table 1. The data showed that the performance of at least two subjects was highly biased in trials rated as fully unaware with a shift in C of +1 SD and –1 SD, respectively. A graphic illustration of the relation between the objective measures of awareness and the subjective measure is shown in Figure 4 depicting violin plots of A' and C for each level of subjective awareness. In sum, these results show that the ability to distinguish the two types of orientation (non-vertical vs. vertical) strongly

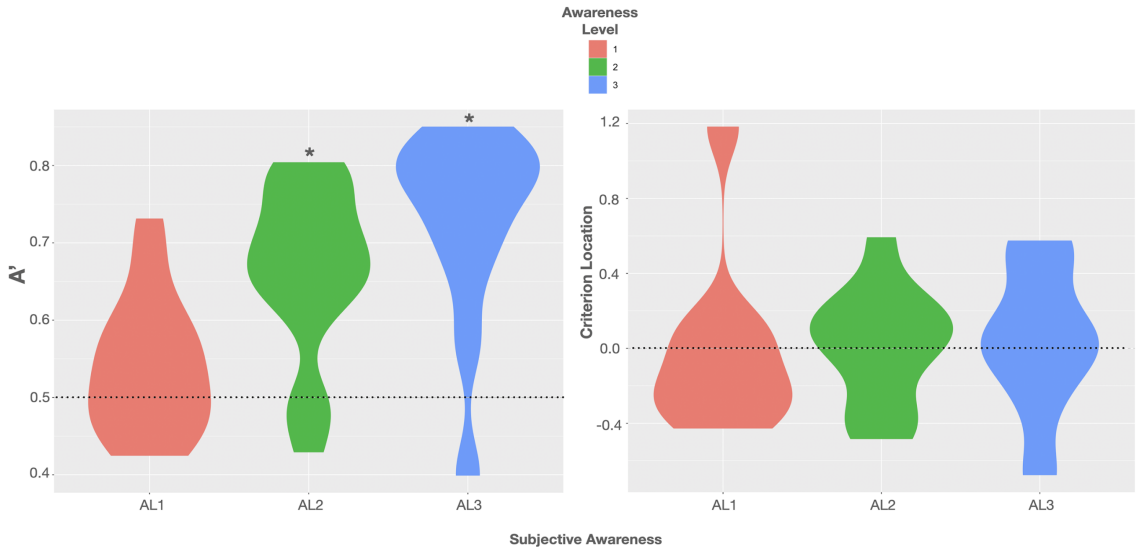


Figure 4: Violin plots of group distributions of sensitivity A′ (left) and the criterion location C (response bias; right) for each level of subjective awareness (perceptual awareness scale (PAS) ratings AL1–AL3). Left: the agreement of the objective and subjective measure of visual awareness is indicated by moderate evidence for the absence of sensitivity on trials rated as subjectively unaware; $BF_{10} < 0.33$. Black asterisks = $BF_{10} > 100$ indicating extreme evidence for A′ being truly > 0.5. Violin plots use density curves to depict distributions of numeric data. The width corresponds with the approximate frequency of data points in each region. The lower and upper limits of each plot are determined by the minimum and maximum values.

depended on the subjective visibility and fully diminished on subjectively unaware trials. In contrast, there was no clear evidence of a response bias, regardless of the level of subjective awareness. Importantly the absence of variation in volunteers' response bias likely suggested that perceptual decision criteria were not dependent on the awareness reports and that variations in participants' perceptual sensitivity regarding the stimulus orientation could thus not be caused by variations in the response bias.

Table 1: *Perceptual sensitivity and response bias*

	Level of awarness (PAS)					
	AL1		AL2		AL3	
	A′	C	A′	C	A′	C
M	0.515	0.048	0.670	−0.093	0.723	−0.152
SD	0.091	0.438	0.110	0.402	0.135	0.527

In the signal detection analysis outlined above, we assumed a binary yes/no task setup. However, as we deployed left and right-tilted gratings next to vertical ones, subjects were to sort three possible stimulus types into two categories. Moreover, we presented left- and right-tilted Gabors with varying angles so that subjects needed to map different stimuli to the same response. Hence, a classification scenario may better fit the scenario (Snodgrass et al., 2004). Importantly, such a setup requires the implementation of two rather than one decision criterion increasing the decision uncertainty, and the proportion of correct responses (i.e., proportion correct, p(c)) is then used to measure volunteers' classification sensitivity (Macmillan & Creelman, 2004, pp. 190–191). Hence, our sensitivity measure may not be exhaustive of all the information that the subject could hold, meaning that actual sensitivity on unaware trials could be higher than we measured.

Thus, we additionally calculated p(c) for each level of subjective awareness (AL1–AL3): p(c) can be defined as the prior probability of a positive stimulus (i.e., non-vertical grating) times the conditional probability of a positive response given a positive stimulus (i.e., a non-vertical response for a non-vertical target) added to the product of the prior probability of the negative stimulus (i.e., vertical) times the conditional probability of a negative response given a negative stimulus (Swets, 2014, p. 4). In other words, p(c) is found by using the presentation probability of the two non-vertical targets as weights for the hit rate and adding this to the product of the 1-False alarm rate (i.e., correct rejection rate) and the presentation probability of the vertical target (i.e., p(c) = (8/36)*H + (16/36)*H + (12/36)*(1-F); Macmillan & Creelman, 2004, p. 89)). Using this formula, we observed a mean p(c) of 54 ± 4.2% (SE) in trials rated as fully unaware. Here the BF was rather inconclusive as to whether p(c) was different from the 50% chance level with anecdotal evidence for the H0, BF_{10} = 0.409, 95% CI (44.9, 63.1), (Quintana & Williams, 2018). In AL2 trials the mean p(c) on group level was 77.1 ± 2% associated with a BF providing extreme evidence that p(c) was truly above chance, BF_{10} > 100, 95% CI (72.6, 81.5). In trials rated as almost fully aware (AL3) we observed a mean p(c) of 85.4 ± 3.2%. Here the BF again provided extreme evidence for p(c) to be greater than 50%, BF_{10} > 100, 95% CI (78.4, 92.4). Violin plots show the observed p(c) as a function of subjective awareness in Figure 5. For more transparency, we additionally included accuracy data obtained in the experimental task in Appendix A.3, as well as the average rates of hits (H), false alarms (FA), correct rejections (CR), and misses (M), and the mean number of hit, false alarm, miss, correct rejection trials (Appendix A, Table 14 and 15).

Taken together, using p(c) as a measure of volunteers' perceptual sensitivity did not change the conclusion that participants' classification ability was at chance in trials rated as subjectively fully unaware, while they showed considerable classification sensitivity in trials with residual and almost full subjective awareness. Importantly, the above measures are representative and exhaustive of the critical target feature that is relevant for the task (i.e., orientation), (Snodgrass et al., 2004). However, additional experimentation could be performed employing an even more stringent detection threshold in which one's sensitivity to detect the presence of any grating is null.

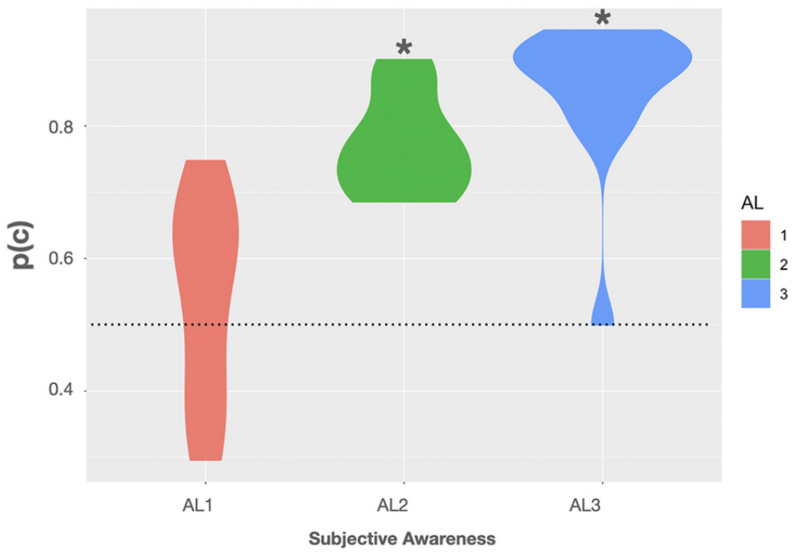


Figure 5: Violin plots show the observed proportions of correct responses (p(c)) as a function of subjective awareness (AL1–AL3). Black asterisks indicate that testing p(c) on group level against a theoretical chance level of 0.5 (dotted line) resulted in a BF_{10} providing extreme evidence for p(c) being greater than 0.5 ($BF_{10} > 100$). Violin plots use density curves to depict distributions of numeric data. The width corresponds with the approximate frequency of data points in each region. The lower and upper limits of each plot are determined by the minimum and maximum values.

3.5.3 RT data

We analyzed volunteers' RT data to test whether the latency of the manual responses slowed down during (unconscious) changes in the target orientation compared to repeating target orientations which would suggest a re-weighting of attentional selection weights. Individual outliers (M ± 3 SD) were removed before the LMM analysis. We conducted the same LMM analysis for two RT models. Whereas in the first average RT model, the switch condition comprised all possible orientation changes, the second weighted RT model included only switch trials in which the prior target orientation was associated with a high selection weight (i.e., changes away from the most frequent tilt). Descriptive mean RTs and SEs of both models for switch vs. repeat trials for each level of awareness are summarized in Table 2.

To begin, we conducted the LMM analysis for the *average* RT model in which the mean of the switch condition included all possible switch trials. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity nor normality. Estimated RTs appeared to be sensitive to changes in the level of visual awareness indicated by the significant fixed effect of *awareness*, $F(2,11.055) = 9.674$, $p = .004$. In line with our predictions, the post-hoc tests showed that RTs (averaged across the conditions *switch* and *repeat*) in AL1 trials were 157.6 ± 41.6 ms slower compared to AL2 trials, $t(12.00) = 3.783$, $p = .008$, 95% CI (41.8, 273.5), and 225.6 ± 54 ms slower compared to AL3 trials, $t(11.87) = 4.177$, $p = .004$, 95% CI (75.2,

Table 2: *RT data*

	Level of awareness (PAS)					
	AL1		AL2		AL3	
	switch	repeat	switch	repeat	switch	repeat
Average model						
M	1.110	1.068	0.945	0.917	0.889	0.869
SD	0.225	0.231	0.182	0.212	0.130	−0.128
Weighted model						
M	1.177	1.068	0.947	0.917	0.879	0.869
SD	0.242	0.231	0.194	0.212	0.129	0.128

376.1). RTs in AL2 and AL3 trials did not differ significantly, p = .368, 95% CI (–46.0, 182.0). Thus, RTs of the average RT model was indeed sensitive to changes in visual awareness and decreased with increasing stimulus visibility. There was, however, no significant main effect of switch, $F(1,35)$ = 3.1709, p = .084, nor a significant interaction $F(2,35)$ = 0.141, p = .869, showing that RTs appeared to be unaffected by changing stimulus orientations in this RT model. About 20% of the total variance was explained by the model's fixed effects, R^2m = 0.199, and 88% by the model's fixed and random effects, R^2c = 0.884.

Next, we used the same LMM to analyze the *weighted* RT model in which the switch condition comprised only switch trials where the prior orientation was the heavily weighted one. Again, visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity nor normality. The LMM analysis showed, also in this model, that estimated RTs increased with decreasing visual awareness, $F(2,10.93)$ = 10.989, p = .0024. The post-hoc tests with Bonferroni correction indicated that mean RTs across both switch and repeat trials were on average about 190.5 ± 45.6 ms significantly slower in AL1 trials compared to AL2, $t(12)$ = 4.177, p = .004, 95% CI (63.7, 317.3) and on average 263.1 ± 58.7 ms slower compared to AL3 trials $t(11.92)$ = 4.482, p = .002, 95% CI (99.7, 426.4). Mean RTs in AL2 and AL3 trials did not differ significantly, p = .300, 95% CI (–40.8, 185.9). Importantly, now also a switch of the target orientation affected RTs: the analysis revealed a significant fixed effect predictor switch, $F(1,35.00)$ = 6.030, p = .019. Here the post-hoc tests suggested that only in unaware trials (AL1) were RTs in response to a novel orientations on average 109.5 ± 34.5 ms significantly slower compared to trials in which the orientation was repeated, $t(35)$ = –3.171, p = .003, 95% CI (–179.7, –39.4). In trials with higher levels of visual awareness, switch costs were not significant, AL2, p = .403, 95% CI (–99.4, 40.9); AL3, p = .779, 95% CI (–83.1, 62.8). Yet, there was no significant interaction between the two fixed-effect predictors awareness and switch, $F(2,35.00)$ = 2.280, p = .117. Together, about 26% of the total variance was explained by the two fixed effects awareness and switch, R^2m = 0.2568, and about 85% was explained by all fixed and random effects, R^2c = 0.8532. RTs for

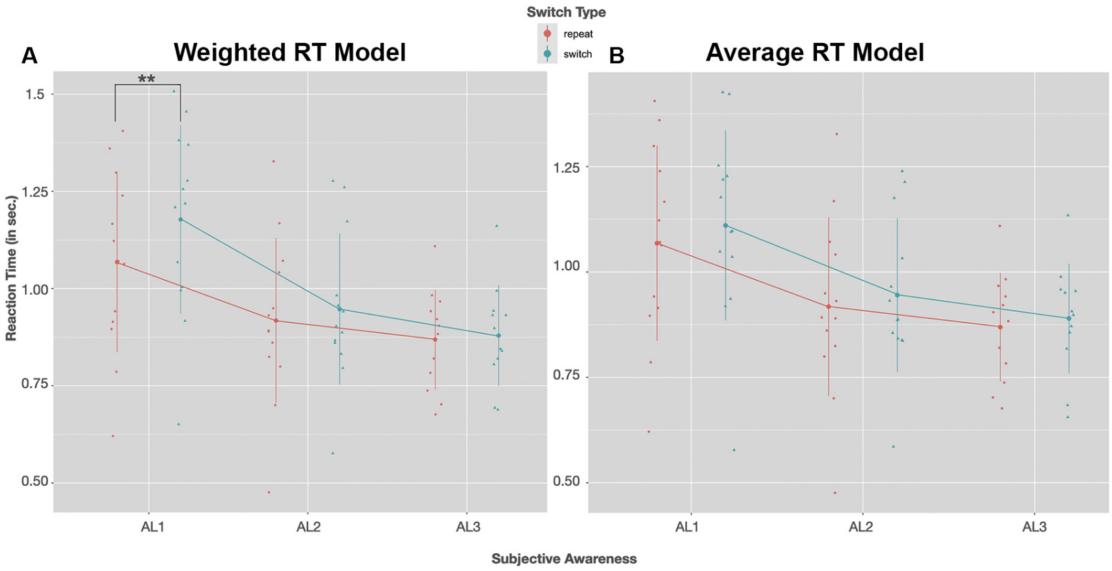


Figure 6: Group mean RTs (in seconds) for a switch (blue) and repeat trials (red) as a function of visual awareness (AL1–AL3); on the left (A) RTs of the weighted switch model, on the right (B) RTs of the average switch model is shown. Dots and triangles indicate individual participant data points. Vertical lines show the range of 1 SD ± the mean. In both reaction time (RT) models post-hoc tests with Bonferroni correction revealed that RTs speeded up with increasing visual awareness (AL2–AL1: $p <$ .01, AL3–AL1: $p <$ .01). A significant slowing of RTs in switch compared to repeat trials was observed only for the weighted model in unconscious trials (AL1); $^{**}p <$ .01 (Post-hoc tests with Bonferroni correction).

both switch and repeat trials as a function of visual awareness for the weighted and the exhaustive RT model are plotted in Figure 6 A & B. The LMM solutions for the fixed and random effects for the two RT models are given in Table 3 A & B.

In sum, the LMM analysis suggests that not only were RTs sensitive to decreasing visual awareness but also changes in the stimulus orientation. However, RT costs due to such changes were observed only in the weighted RT model which included only those switch trials in which the novel orientation changed away from the highly biased orientation (highly frequent tilt) fostering the conclusion that the prior visual selection bias had boosted behavioral switch costs in response to a change in the target orientation. As significant switch costs were observed in unaware trials only, the impact of the prior selection bias boosting behavioral switch costs during attentional re-selection appeared to be most prominent in the full absence of visual awareness.

Given that under unconscious conditions we had fewer trials included in the analysis, outliers could have a stronger effect on the results. Since 3 SD is not a rigid cutoff for outliers, we, therefore, repeated the analysis with a 2 SD, and 2.5 SD cutoff but the results did not change in terms of significant fixed effects. Most relevant to our research question, we found a significant switch effect for the weighted RT data

Table 3: *LMM parameter estimates of RT analysis*

Fixed effects	Estimate (in seconds)	SE	*t*-value	*p*
A) Average				
AL1 (intercept)	1.06823	0.05965	17.203	1.62e-10***
AL1-AL2	-0.15035	0.04457	-3.241	0.00452**
AL1-AL3	-0.21491	0.05536	-3.726	0.00254**
Switch (Intercept)	0.04209	0.02766	1.460	0.15309
AL2: Switch-Repeat)	-0.01465	0.03912	-0.359	0.72144
AL3: Switch-Repeat	-0.02151	0.03992	-0.539	0.60840
Random effects	Variance	SD		
Subject	0.044852	0.21148		
AL1-AL2	0.017186	0.13109		
AL1-AL3	0.031373	0.17712		
Residual	0.005399	0.07348		
B) Weighted				
AL1 (intercept)	1.06823	0.06362	16.792	1.19e-10**
AL1-AL2	-0.15035	0.04970	-2.905	0.00896**
AL1-AL3	-0.21338	0.06106	-3.354	0.00451**
Switch (Intercept)	0.10959	0.03316	3.171	0.00315**
AL2: Switch-Repeat)	-0.08035	0.04690	-1.644	0.10905
AL3: Switch-Repeat)	-0.09941	0.04786	-2.077	0.05408
Random effects	Variance	SD		
Subject	0.044852	0.21178		
AL1-AL2	0.017821	0.13889		
AL1-AL3	0.035990	0.18971		
Residual	0.007761	0.007761		

Note: Significance codes: *** $p < .001$; ** $p < .01$; * $p < .05$. (A) Average RT model: fixed effect predictor *switch* includes all types of orientation changes. (B) Weighted RT model: fixed effect predictor *switch* comprises only the changes away from the heavily weighted orientation. *P*-values indicate the difference between each factor level compared to baseline (intercept). For both models: intercept switch equals the estimated mean difference of switch trials compared to repeat trials across all three levels of awareness, intercept AL1 equals the mean of all switch and repeat trials rated as subjectively unaware. Random effects AL1-AL2, and AL1-AL3 indicate the amount of variation in the fixed effect switch between the two AL1 and AL2, and AL1 and AL3, respectivelv.

Table 4: *LMM parameter estimates of analysis of the weighted RT model after removing those trials preceding a weighted switch in which the stimulus was consciously perceived*

Fixed effects	Estimate (in seconds)	SE	*t*-value	*p*
Weighted				
AL1 (intercept)	1.09469	0.07955	13.761	4.01e-10
AL1-AL2	–0.17603	0.07667	–2.296	0.03137*
AL1-AL3	–0.24179	0.07637	–3.166	0.00447**
Switch (Intercept)	0.21353	0.07067	3.022	0.00414**
AL2: Switch-Repeat)	–0.114745	0.09333	–1.229	0.22556
AL3: Switch-Repeat	–0.29642	0.09740	–3.043	0.00398**
Random effects	Variance	SD		
Subject	0.05811	0.2411		
AL1-AL2	0.02811	0.1677		
AL1-AL3	0.02536	0.1592		
Residual	0.02416	0.1554		

Note: Significance codes: *** $p < .001$; ** $p < .01$; * $p < .05$. *P*-values indicate the difference between each factor level compared to baseline (intercept). The intercept of switch type equals the estimated mean difference of weighted switch trials compared to repeat trials across all three levels of awareness, intercept AL1 equals the mean of all weighted switch and repeat trials rated as subjectively unaware. Random effects AL1–AL2, and AL1–AL3 indicate the amount of variation in the fixed effect switch between the two AL1 and AL2, and AL1 and AL3, respectively.

model in AL1 but neither in AL2 nor in AL3 trials for all three cutoffs. We conducted further control analyses that are reported in the Appendix A.4, in which we matched the number of trials between AL1, AL2, and AL3 trials by random sampling to prove that the low amount of AL1 trials could not account for the observed switch effect, and used the numbers of trials obtained for the weighted switch trials rated as fully unaware (AL1) to do a Bayesian-based prediction to show that the switch costs in the weighted RT model were not associated with individual trial numbers.

Finally, to rule out the possibility that inter-trial response priming instead of attentional weighting could account for the observed switch effect, we repeated the LMM analysis for the weighted switch model after removing all weighted switch trials preceded by trials in which the orientation had been perceived consciously to some extent (i.e., AL2, and AL3 "pretarget" trials). This we did because inter-trial response priming is thought to necessitate awareness of the stimulus in the preceding trial (e.g., Peremen et al., 2013). Using the same LMM we found only a marginal switch effect, $F(1,42.301) = 4.011$, $p = .051$, a significant fixed effect of visual awareness, $F(2,14.399) = 17.561$, $p < .001$, and a significant interaction of the two fixed effects *switch* and *awareness*, $F(2,42.198) = 4.776$, $p = .0134$. The fixed and random effect solutions of this analysis are given in Table 4. Importantly, paired comparisons replicated our previous finding showing that unaware weighted switch trials were

significantly slower compared to unaware repeat trial, $t(33.24) = –2.954$, $p = .006$, 95% CI (–360.5, –66.4), while there were no differences between switch and repeat trials for AL2, $p = .1157$, nor for AL3, $p = .230$.

3.6 Discussion

3.6.1 Evidence for the unconscious reallocation of visual attention and methodological advantages

When volunteers engaged in our discrimination task of masked gratings, RTs were sensitive to orientation changes. However, significant switch costs were obtained only if the selection weight for the prior orientation was high (i.e., the highly frequent tilt) and if the novel orientation was unconsciously perceived. Importantly, our criteria for lack of awareness were based on the combination of subjective and objective measures, i.e., no experience reports and no ability to discriminate the relevant target features in a forced-choice test. To the best of our knowledge, this, therefore, is the first study investigating the effects of unaware targets on feature-based attention weighting by using a combination of objective sensitivity measures and subjective measures of visual (un-)awareness collected during the experimental task. This is a very important advantage for two reasons: first, to account for fluctuations of the perceptual threshold before, during, and after the actual experimental task it is extremely important to use an "online" measure of visual awareness during the task performance. This way, one ensures that the stimulus perception and the effect of the stimulus are measured in the same context (e.g., Avneon & Lamy, 2018). Second, studies that define unconscious processing only employing subjective awareness measures (e.g., Cheesman & Merikle, 1986) suffer from the criterion problem that arises when conscious knowledge is held with low confidence, hence objective measures that can ensure a clear absence of visual awareness (i.e., if d' = 0) are critical to studying unconscious information processing, which would then be pinpointed by information-based analyses of neural measures (Soto et al., 2019). Yet, to come up with an exhaustive means that measures visual awareness and unawareness equally well, the joint use of both the objective and subjective measures seems optimal (e.g., Wiens, 2007).

Taken together, our results indicate that prior feature likelihood differences modulated attentional weighting by introducing a competitive bias favoring (i.e., increasing the selection weight for) the most likely event (i.e., frequent tilt). Only when the orientation associated with a high selection weight was present in the prior trial did attentional re-selection towards a novel target orientation result in behavioral switch costs. This was indicated by significant RT differences between stay and switch trials in the weighted model. Switch costs due to changing features within a single feature dimension may be relatively small compared to cross-dimensional switch costs (see Müller et al., 1995), which could explain why there was no switch effect in the averaged model in any level of awareness. Thus, the presumably smaller effect of within-dimensional switches may require a strong prior feature weighting to emerge.

3.6.2 Integrating the absence of the effect in visible trials

Remarkably, the behavioral switch costs were observed only in trials reported as fully unaware, in which subjects had zero sensitivity for the stimulus orientation. According to Bundesen's (1990) TVA, the influence on sensory processing given by an attentional template that contains goal-relevant information (i.e., history-guided) becomes particularly strong if the sensory evidence of the to-be-processed stimulus is low. In such a case there is little stimulus information that could form the selection weight in a "bottom-up" fashion so that knowledge about the importance of attending to a certain stimulus category (e.g., because this category is more likely to occur) gains influence on stimulus processing. Hence, one could predict that the behavioral effect due to attentional weighting and re-weighting should be most pronounced in unaware trials in which decision-making may especially rely on implicit knowledge (i.e., prior beliefs about likelihoods), (Bohil & Wismer, 2015) maintained in the form of an attentional template because the weak sensory evidence given by the unconscious stimulus does not suffice to strongly bias its selection.

3.6.3 Information derived from visible stimuli may drive the response to invisible targets

Importantly, this conclusion does not imply that the information (i.e., prior beliefs about likelihoods) upon which the trial history-guided attentional selection is built is derived from invisible stimuli. Certainly, in our paradigm, a significant amount of targets were perceived partially and almost fully consciously. Thus, even if the subjects' reports suggest that the knowledge about the likelihood differences was rather implicit, it was likely to be derived from visible stimuli. Still, a shift of attentional selection weights against a prior bias was elicited by an invisible novel target. This clearly shows that invisible feature changes can indeed trigger a shift of visual attention.

3.6.4 The importance of feature weighting

Peremen et al. (2013) reported the opposite pattern of results: strong intertrial feature priming if primes and probes were consciously perceived but no such repetition effects under masking conditions. However, by using prior likelihood differences of the three orientations, we introduced a feature weighting that evidently boosted the switch effect, deliberately chose simple Gabor patches that are known to be readily processed, even if unconsciously (e.g., King et al., 2016; Rajimehr, 2004; see also Soto et al., 2011), and tested participants in a simple discrimination task in which the focus of spatial attention was always directed to the relevant location, instead of using a visual search paradigm. Importantly, the prior attention bias was essential to induce prior attentional weighting impeding consequent attentional re-weighting in response to a novel target. In conclusion, this suggests that a prior likelihood weighting indeed can induce an attentional selection bias for a feature even if the respective target is invisible and that a novel target can trigger the re-shifting of attentional resources even if the novel stimulus is invisible. Thus, our findings support the view that attentional selection and consciousness can be dissociated (e.g., Koch & Tsuchiya, 2007; Lamme, 2003; Van Gaal & Lamme, 2012). They also show that the

covert reallocation of feature-based attention can be studied by presenting a series of invisible targets at least if a prior selection bias had been introduced (i.e., likelihood weighting) to boost the inter-trial facilitation. Therefore, our study puts forward a parsimonious methodological approach using single-item displays with masked targets and a discrimination task to examine the effects of attentional feature weighting in the absence of visual awareness. Importantly, we used volunteers' discrimination ability to measure visual consciousness objectively but examined the effects of visual attention using discrimination response times, thereby guaranteeing a clear methodological separation between consciousness and attention.

3.6.5 Why intertrial response priming is unlikely explaining the pattern of results of the RT data

The observed switch effect for the weighted RT model could in theory be explained by intertrial response priming. That is, the orientation perceived in the recent past (trial n – 1) could have primed the response to the current target (in trial n) so that responses speed up following repeated target orientations and slow down once a novel target is presented. This prediction is in line with our observation and challenges the attentional weighting account that we proposed to explain the effect. However, if intertrial response priming was responsible for the effect, one would expect a significant slowing of responses following a novel orientation to occur independently of the feature weighting. In other words, significant switch costs should have been observed also in the average RT model which was not the case. Intertrial response priming is contingent on awareness of the "pretarget" stimulus (e.g., Peremen et al., 2013). Accordingly, the observed switch effect should rely on pretarget AL2 and AL3 trials but not on fully unconscious pretarget trials if response priming was the underlying mechanism. To test this account, we reanalyzed the RT data using the same mixed model approach after removing those switch trials that were preceded by AL2 and AL3 pretarget trials. Importantly, the switch effect was preserved for the weighted RT model even when this time only fully unconscious trials preceded an orientation change. This finding, together with the fact that the switch effect was missing in the average RT model, makes it rather unlikely that response priming could alternatively explain the effect we observed.

3.6.6 Conclusions

We demonstrated that unconscious feature changes of invisible targets can induce attentional re-weighting against a prior attentional selection bias, suggesting that the shifting of attentional selection weights during the behavioral performance does not necessitate visual awareness. This finding supports previous studies stressing the dissociation of attention and visual consciousness (e.g., McCormick, 1997), however, prior studies predominantly report how unconsciously perceived cues affect shifts in spatial attention (e.g., Mulckhuyse et al., 2007). To our knowledge, this is the first study to investigate the effect of unconsciously perceived feature changes on visual attention. Importantly, the methodological advantage of combining subjective and objective measures of visual awareness helps to ensure that the target stimuli were truly unconsciously processed. In the next step, it will be important to shed

light on the neural underpinnings supporting attentional feature-based re-weighting in the absence of visual awareness. Here, particularly the role of the frontopolar cortex (FPC) should be examined as previous findings consistently have linked it to exploratory attention shifts (for an extensive review see Mansouri et al., 2017), yet evidence showing that FPC supports attentional reallocation in the full absence of visual awareness is still missing.

4 Experiment 2: Frontopolar activity carries feature information of novel stimuli during unconscious re-weighting of selective attention

The results of this experiment were first published in: Güldener, L., Jüllig, A., Soto, D., & Pollmann, S. (2022). Frontopolar activity carries feature information of novel stimuli during unconscious re-weighting of selective attention. Cortex, 153, 146-165.

4.1 Introduction

The FPC is uniquely large in the human brain and possesses a distinctive cytoarchitecture (Petrides et al., 2012; Ramnani & Owen, 2004; Semendeferi et al., 2001). With its high number of spines and synapses it appears particularly suited for the integration of information (Jacobs et al., 2001; Ramnani & Owen, 2004). FPC plays a pivotal role in human cognition, where it ranks at the top of a high-level executive control system orchestrating our behavior by temporally organizing top-down strategic processing for goal-directed action (Cohen et al., 2000; Fuster, 2002; Ramnani & Miall, 2004). Only recently, frontopolar function has also been investigated in non-human primates. Bilateral FPC lesions increased conflict adaptation in a Wisconsin Card Sorting-like task (WCST), (Grant & Berg, 1948) in which the animals needed to adapt to frequently changing task rules (Mansouri et al., 2015). Importantly, FPC lesions did not affect the ability to follow the rule switches of the WCST, in contrast to frontal lesions posterior to FPC. In a human fMRI study utilizing a comparable WCST task, FPC activation signaled the presence of interfering task rules (Konishi et al., 2005). Again, FPC activation was not affected by rule changes per se, whereas this was observed more posteriorly, in the left inferior frontal cortex. These data exemplify a pattern that suggests a vital role of FPC in exploratory shifts of attentional selection, which is further supported by findings from the literature on decision-making (Beharelle et al., 2015; Boorman et al., 2009; Daw et al., 2006; Kovach et al., 2012).

In line with this notion, studies in the visual search domain showed that attention changes between feature dimensions (Pollmann et al., 2000) or, likewise, between locations (Lepsien & Pollmann, 2002) went along with increased BOLD signal in the FPC. Importantly, exploratory attention shifts were assumed to be implicit, namely, to occur without volitional orienting of attention to the new feature. In line with this assumption, FPC activation was also observed in response to changes in target-distractor contingencies that were learned implicitly: even though distractor configurations were not remembered explicitly, the violation of contingencies between learned target locations and specific distractor configurations activated FPC (Pollmann & Manginelli, 2009a, 2009b). Yet, the search tasks used in the above studies employed fully visible stimuli that were consciously seen and attended to on every trial. Hence it is unclear whether the role of the FPC in re-weighting of selection biases extends to changes of unconsciously processed stimuli.

To tackle this question, we developed a novel visual masking paradigm in which a Gabor patch was presented centrally followed by a backward mask to minimize the patch's visibility. The spatial orientation of the target stimulus randomly repeated or changed on a trial-by-trial basis and volunteers were obliged to distinguish between vertical and non-vertical orientations. At the end of each trial, we asked them to rate the target visibility using an adaptation of the perceptual awareness scale (PAS), (Ramsøy & Overgaard, 2004). Prior to this study, we provided behavioral evidence that attentional re-selection in response to a target change occurred in the full absence of visual awareness (Güldener et al., 2021). However, the role of the FPC in supporting this process remained untested. Here we used functional MRI to address this question using the experimental paradigm described above.

In line with our previous findings, we expected attentional adaptation to occur as soon as a given grating possessed the same orientation as the previous one (one-trial learning), resulting in a selection bias favoring the repeated orientation (Boschin, et al., 2015). Conversely, this attentional bias should be disrupted and adjusted as soon as the novel grating's orientation differed. Such reorienting of attentional resources in response to an orientation change in these switch trials was expected to result in RT switch costs on the behavioral level and to increase the BOLD response in FPC (Pollmann et al., 2000). Consequently, we used the RTs obtained in the orientation discrimination task as a proxy to measure attentional reorienting processes. Participants' perceptual decisions were analyzed by means of signal detection theoretic measures in combination with visibility ratings to measure visual (un)awareness and isolate unconscious information processing (Soto et al., 2019; Wiens, 2007). Visual unawareness was associated here with null perceptual sensitivity in those trials subjectively rated as unaware. This approach precluded confounds arising from individual response criterion shifts in reporting subjective awareness.

Importantly, we manipulated the proportions of the two non-vertical gratings (left vs. right) presented in a single block by presenting one tilt twice as often as the other tilt. We hypothesized that the increase of the frequency at which a certain tilt (i.e., a grating tilted to the left or right) was presented will boost the attentional selection weight for this tilt, consistent with feature-based statistical learning (Chetverikov et al., 2017; Turk-Browne & Scholl, 2009). Thus, particularly switch trials in which the prior orientation was the highly frequent tilt should show increased behavioral response latencies (e.g., Chetverikov et al., 2017; Leber et al., 2009).

Critically, we also tested whether the FPC activation pattern carried information about the grating's orientation in the non-conscious trials by using multivariate pattern analyses. Recent research has shown that consciously processed stimuli in working memory can be decoded from BOLD activity patterns from regions across the entire attention network (Corbetta et al., 2008) like the left superior precentral gyrus (SPG), bilateral SPL (Ester et al., 2015), and representations of task-relevant feature dimensions can be found in the frontal eye field and left prefrontal cortex

(lPFC) including FPC (Reeder et al., 2017). Furthermore, unconscious perceptual contents (i.e., living versus nonliving categories) can be decoded from brain activity patterns in prefrontal regions (Mei et al., 2022). Here, we tested the role of FPC in representing the relevant informational content during reorienting of attention across different states of visual (un)awareness. It has been shown that cortical representations of subjectively versus objectively invisible stimuli may differ (Stein et al., 2021; but see also Mei et al., 2022). Stein and colleagues asked participants to perform a visual discrimination task distinguishing between masked houses and faces. The key finding was that the processing of objectively invisible stimuli was restricted to visual (shape-related) object properties processed in early, lower-level visual areas, while the processing of subjectively invisible stimuli reached up to more categorical levels of representation in higher-level category-selective areas. However, this pattern of results may change once the stimulus processing is affected by attentional modulation (i.e., difference of goal relevance between stimulus types). Here we aim at testing whether feature representations of objectively unaware stimuli can reach a more global level of processing extending from occipital cortex up to FPC if the represented object feature is associated with a higher attentional weight.

4.2 Methods

4.2.1 Participants

Based on the previous behavioral study (Güldener et al., 2021) we recruited in total 25 native German students (11 female) from the University of Magdeburg, Germany. The volunteers were 20 to 39 years old (M. 24.08 years) and were either monetarily reimbursed (8 euros per hour) or received course credits for the 2 h of participation. A total of 8 participants were excluded prior to the main fMRI experiment: 3 participants interrupted the session during the calibration or the main experiment and were thus excluded and 2 other participants were excluded as they reported insufficient correction of their impaired vision using the MR compatible lenses. Three participants did not successfully pass the calibration, i.e., even after multiple repetitions we were not able to determine a stable threshold of the stimulus' luminance contrast. Out of the 17 participants that took part in the fMRI experiment, three reported a very low number of subjectively invisible trials despite the initial calibration (less than 5% of all trials); this was insufficient for statistical analysis and they were thus excluded. Hence, the following report is based on a final sample size of n = 14.

4.2.2 Apparatus & stimuli

The stimulus display and responses were controlled with PsychoPy (Peirce et al., 2019). The stimuli were back-projected onto an 18-inch screen placed in the bore of the magnet behind the participant's head. The projector's resolution was 1920 × 1080 pixels with a 60 Hz refresh rate. Participants viewed the screen via a mirror placed on top of the head coil. Stimuli were Gabor gratings with an individually calibrated Michelson contrast and a spatial frequency of 3.703 cycles per degree. They were centrally presented on a grey background and subtended 3.437° visual angle. The

gratings' orientation was either vertical (180°), 165°, 150° or 135° if it was a left-tilted, non-vertical Gabor patch, and 195°, 210° or 225° if it was a non-vertical patch tilted to the right. To further reduce the visibility of the Gabor patch we used a circular backward mask of black and white random dots (3.437° visual angle).

4.2.3 Experimental procedure

4.2.4 Threshold determination

All experimental sessions took place in the MR scanner (Siemens Prisma, Erlangen, Germany) of the Neurology Department of the University of Magdeburg, Germany. After placing the participant inside the scanner, the session started with a 1-up:1-down adaptive staircase procedure (adopted from Jachs et al., 2015) to determine the stimulus´ luminance contrast for the first trial of the main experiment. The details of this procedure are described in Chapter 2., in section 2.2.3.

4.2.5 Main experiment

In the main experiment, volunteers performed the orientation categorization task based using masked Gabor patches. The paradigm is described in detail in Chapter 2., section in 2.2.2.

4.2.6 Design

Although the categorization task demanded participants only to discriminate vertical from tilted orientations, irrespective of the specific direction of tilt, we expected attentional weighting of left versus right tilt, so that attentional resources would be allocated to discriminating the most recent tilt direction from vertical based on analogous attention weighting effects observed in visual singleton search tasks (Müller et al., 1995). This attentional weighting was expected to lead to reduced response times when the tilt direction repeated (e.g., left following left tilt) irrespective of the exact orientation of the grating (e.g., 165°, 150°,or 135°) compared to longer response times if the tilt direction changed.

To boost tilt-based attention weighting, we manipulated the likelihood of the two non-vertical gratings (left versus right): by increasing the frequency at which left or right-tilted gratings occurred, the attentional weighting of this orientation should be enhanced, while it should be reduced for the less frequent orientation (Desimone, 1996; Henson & Rugg, 2003). Thus, tilt-change costs were expected to be higher if the change occurred from the frequent to the infrequent orientation than vice versa. Additionally, if attentional weighting towards the frequent tilted orientation is used to facilitate the discrimination between a vertical and a tilted grating, this should result in higher tilt-change costs following the switch from the frequent tilt to vertical compared to the change from infrequent tilt to vertical as in the former case more attentional weight needs to be re-weighted, while in the latter case attentional reorienting should be relatively easy because the attentional weighting for the target preceding the orientation change should be only weak. Hence, within a single run consisting of 36 trials, the stimulus orientation was set to be vertical in 12 trials

(∼33%). The two non-vertical orientations, however, occurred in uneven proportions with 18 trials (50%) and six trials (∼16%), respectively. Thereby we obtained either left- (75% of all non-vertical trials with a left tilted grating) or right-weighted (75% of all non-vertical trials with a right-tilted grating) blocks, each containing frequent and infrequent non-vertical orientations in random order.

4.3 Statistical analysis

The exact same procedure as in experiment 1 was carried out to analysis the participants' behavioral performance. It is described in detail in Chapter 3, section 3.4. Participant's subjective awareness was again determined based on each individual's awareness reports (see chapter 3.4.1.) and individual response biases (criterion location C) and sensitivities (A') were calculated to determine participants' objective ability to categorize vertical and non-vertical gratings for each level of subjective awareness (Stanislaw & Todorov, 1999). All details are reported in section 3.4.2. We tested these measures on group level with Bayes factor analysis (e.g., C = 0 and A' = .5). All statistical analyses were carried out with R (Version 3.5, R Core Team, 2014): for the Bayes factor (BF) analysis (Rouder et al., 2009) we used the R-package *BayesFactor*.

4.3.1 Analysis of RT data

Consistent with the first experiment (section 3.4.3), linear mixed model analyses were conducted to analyze the RT data using packages *lme4* (Bates et al., 2014) as well as lmerTest.

RTs of trials in which incorrect responses had been given were discarded and each participant's individual outliers (mean RTs ± 2.5 SD) were removed prior to the analysis. The model was fitted using a restricted maximum likelihood estimation and the influence of the fixed effect predictors was tested with a type III ANOVA as implemented in the *lmer* and *anova* function of the *lme4* package (Version 1.1e23). The *P*-values were obtained using Satterthwaite approximations to degrees of freedom using the anova function of the package *lmerTest* (Version 3.1-2), (Kuznetsova et al., 2017). Post-hoc tests (least squared means of the contrasts with Bonferroni correction) were performed using the R package *emmeans* (Version 1.4.7). Prior to the statistical assessment of the factors of interest, we defined the full random effect structure of the mixed model with likelihood ratio tests (Baayen et al., 2008). The final LMM used for the analysis model was again defined as RT $\sim$ *switch* + *awareness* + *switch:awareness* + (1 + *awareness* + *switch* | *sub*) The details of the model selection were the same as in the first experiment described in detail 3.4.3 and in Appendix A.1).

Again, in the LMM analysis of RT data, the switch condition was defined in two distinct ways: the first LMM (*weighted* switch model) included only those trials in which the highest RT costs were expected to occur: due to the frequency differences between the three orientations, attentional weighting was expected to be boosted for the highly frequent non-vertical orientations (either left or right). Conse-

quently, re-weighting to the infrequent non-vertical orientation should cause higher switch costs than vice versa. Similarly, the change away from the heavily weighted to the vertical orientation should require more pronounced attentional re-weighting. Changes away from the low-frequent tilted orientation to vertical, on the other hand, should result in lower switch costs since attentional weighting for this tilted orientation is weaker, facilitating attentional reallocation, and were thus not included in the weighted dataset. For comparison, we repeated the LMM analysis using a fixed effect predictor (switch) that, this time, comprised all types of switch trials (*average* switch). Results are given separately for the LMM analyses using the *weighted* switch and the *average* switch model.

4.3.2 fMRI analysis

4.3.3 fMRI measurements and pre-processing

All parameters for image acquisition are reported in detail in Section 2.3.

A single scanning session was split into ten runs of 246 sec each. 123 volumes were sampled. The imaging data was pre-processed and analyzed by means of tools of the FSL package (Jenkinson et al., 2012). The anatomical scans underwent a non-brain removal with BET (Brain Extraction tool), (Smith, 2002) in preparation for the realignment. The functional images were motion-corrected to an image in the middle of each run with a normalized correlation ratio (MCFLIRT; FMRIB's Linear Image Registration Tool), (Jenkinson & Smith, 2001; Jenkinson et al., 2002) and slice time corrected (temporally aligned to the middle slice of the 3D volume). To ensure the validity of Gaussian random field theory, the functional data was spatially smoothed using a Gaussian kernel with a size matching the double of the voxel dimensions (FWHM = 6 mm). To remove low-frequency drifts (Smith et al., 1999), we temporally filtered the data using a highpass filter with a cutoff value of 90 sec.

4.3.4 GLM-analysis

For statistical analyses of the functional brain scans, we defined the onsets of the experimental events as explaining variables (EV) to model the BOLD response by means of a general linear model. The Gabor onsets of trials in which the orientation had changed compared to the previous trial (switch) and the Gabor onsets of trials on which the orientation had remained unchanged (repeat) were modeled for each awareness level separately (1e3), as well as the onsets of the start fixation, mask, and categorization response which were also defined as regressors. These were convolved with a hemodynamic response function (double gamma HRF) and regressed against the observed fMRI-data. Collinearity was checked for the modeled time series for each voxel ensuring a variance inflation factor (VIF) smaller than 5 (Mumford, et al., 2015). Each regressor was paired with a temporal derivative allowing for temporal flexibility, and motion parameter estimates were added as nuisance regressors. Serial voxel-wise autocorrelations were controlled with prewhitening by the FSL tool FILM (Monti, 2011; Woolrich et al., 2001).

In the first-level analysis, all contrasts of interest were tested for significance under mixed-effect assumptions and contrast images were processed for each participant (voxelwise Z threshold of 3.1 and a cluster significance threshold of p = .001, family-wise error (FWE) corrected). These images were consequently used in the second-level analysis estimating individual mean contrasts for the parameters across all runs using a fixed-effect model with the same voxelwise and cluster significance threshold (Z = 3.1, p_{FWE} = .001). In a two-step post-statistical normalization, prior to group analysis, the functional data was firstly co-registered to the individual, anatomical scan using boundary-based registration (BBR), and secondly normalized to the Montreal Neurological Institute standard space (MNI 152 2mm). Thus, statistical modeling on the subject level was carried out in native space. The statistical modeling at the group level was performed using FLAME 1 + 2 (FMRIB's Local Analysis of Mixed Effects) as implemented in FSL's FEAT (Version 6.00). Results are given by means of whole-brain maps of BOLD responses thresholded using clusters determined by a voxelwise Z threshold of 3.1 and a corrected cluster-forming significance threshold of p_{FWE} = .001, across the whole brain (Eklund et al., 2016; Worsley, 2001).

4.3.5 MVPA searchlight

Our goal was to test whether the involved regions convey reproducible spatial patterns of activity that differentiate between the specific orientations in the absence of awareness. Thus, we made use of MVPA in combination with a searchlight algorithm (Kriegeskorte et al., 2006) in order to further examine the nature of the brain signals that we observed in response to invisible orientation changes in the GLM analysis. Therefore, we carried out searchlight analyses within those brain regions that had been identified previously in the GLM analysis as to be particularly responsive to invisible orientation change. To do so, we created binary masks of these regions and used them as ROI in the consequent searchlight analysis. Note that GLM results were orthogonal to the decoding analysis as we chose those clusters as ROIs for the searchlight analysis that showed increased BOLD signal in response to unaware orientation changes but not to the different orientations per se. In addition, we conducted a whole-brain searchlight analysis to test whether feature information is represented in regions beyond those identified in the GLM analysis.

Prior to decoding, the individual fMRI data were motion corrected and smoothed (FWHM Gaussian kernel = 6 mm) to reduce noise and the impact of fine-scale signal patterns (Gardumi et al., 2016; Op de Beeck, 2010). Note that the MVPA analysis was conducted for both smoothed and unsmoothed data. Both analyses led to comparable results. A rather positive effect of smoothing was previously reported for MVPA in prefrontal cortex and sensory regions (Hendriks et al., 2017), thus we report the MVPA-results based on the smoothed fMRI data. After smoothing, we transformed subjects' data into MNI standard space and fitted a standard hemodynamic response function model to estimate the statistical parameters (scaling parameters, β-values) for each of the experimental conditions, resulting in one β-map for every run per experimental condition. The resulting datasets were detrended and z-scored per voxel within each run.

The searchlight analysis was implemented by extracting the z-scored β-values from spheres centered on each voxel in the ROI masks. For the accuracy maps, the classification accuracy (the mean of the proportion of correctly classified targets) for each sphere was assigned to the sphere's central voxel. To test the sensitivity as a function of sphere radius, we carried out the analysis with the radii of 6 and 9mm. If different searchlight radii reveal the same or similar overlapping clusters, it is more likely that these clusters are indeed not spurious (Etzel et al., 2013). For classification within a single sphere, we chose a linear support vector machine to classify the three stimulus orientations (LIBSVM; with fixed regularization hyperparameter C = -1). We selected this type of classifier as it tends to perform better or at least equivalent compared to other algorithms on fMRI data, and, due to its limited complexity, it reduces the probability of over-fitting (Lewis-Peacock & Norman, 2014; Pereira & Botvinick, 2011). Eventually, an n-fold cross-validation (leave-one-run-out) was carried out, using the *PyMVPA* software package (Hanke et al., 2009); with n as the given number of runs, the training dataset comprised all unaware trials of the first run (fold) to run n-1, while the unaware trials of run n constituted the test dataset. Note the total number of runs per subject varied between four and ten depending on the number of subjectively unaware trials that were obtained (i.e., in some subjects some runs did not include trials rated as subjectively unaware). This splitting was repeated until each of the n folds served once as the test dataset. To test if brain activity on unaware trials conveyed local information sufficient to discriminate the stimulus orientations, trials with higher levels of subjective awareness (AL2 & AL3) were omitted in this procedure. Furthermore, all unaware trials were included without differentiating between switch and no switch trials to maximize the total number of trials serving training and testing the classifier (26.88% of all trials). Likewise, we included correct as well as incorrect subjectively unaware trials (AL1) in the MVPA to maximize the chances of decoding.

In a two-step analysis with permutation tests on the subject level and bootstrapping on the group level, we aimed for finding final group-level clusters with decoding accuracies significantly exceeding the chance level (Stelzer et al., 2013). First, permutation tests for each subject (100 permutations) were carried out to assess chance distributions and to obtain individual chance accuracy maps (Chen et al., 2011; Golland & Fischl, 2003; Stelzer et al., 2013). To do so, we created a random permutation of the observation order of the orientations (labels) and applied this scheme to the data set. Next, the cross-validation was performed on the permuted data set, which was repeated 100 times. This resulted in a sampling distribution of the mean classification accuracy under the null hypothesis (i.e., no information of orientation representations present in the multivoxel activity patterns). The significance level (*P*-value) was estimated by the fraction of the permutation samples that were greater than or equal to the classification accuracy from the data without label shuffling. This "chance" map of decoding accuracies was saved each time for each participant. Importantly, balanced partitions containing the same number of items per orientation class were initially created within each subject and each cross-validation fold. At the group level, we recombined the individual null distribution maps into group accuracy maps (Stelzer et al., 2013). For this, we randomly drew (with replacement)

one of the 100 chance accuracy maps of each subject and averaged this selection of 14 chance maps (one for each participant) voxel-wise to one permuted group accuracy map. Repeating this 10^5 times with replacement we obtained a distribution of 10^5 permuted group accuracies. For statistical testing, we next calculated the probability of the unpermuted mean decoding accuracies across all 14 volunteers in the distribution of the permuted group accuracies (one-tailed) with a voxel-wise threshold of $p < .001$. Cluster *P*-values were calculated for the unpermuted accuracies that referred to the probability of observing a particular cluster size or a larger one given the Null hypothesis, controlling for multiple comparisons using false discovery rate correction (FDR, $p_{\text{Cluster}} < .05$). Group cluster brain maps containing clusters with above chance decoding accuracies were saved as well as classification accuracy maps.

4.4 Results

4.4.1 Visual (un-)awareness & RT data

To assess subjective awareness, we calculated the number of trials for each level of awareness for each participant using the trial-by-trial PAS-rating. In the majority of trials, participants' subjective awareness of the to-be-categorized orientation was low (AL2; 34.71%) or even fully absent (AL1; 24.97%). In 31.1% of all trials, subjects reported an almost clear perception of the grating and its orientation (AL3) and in only 9.23% they clearly saw the grating and its orientation (AL4). The mean numbers of trials for each level of awareness and trial type are summarized in Table 5. As the number of fully aware trials (AL4) was overall very low with less than 10 trials in 64% of all subjects, we excluded these trials from further analyses.

Table 5: *Number of switch and repeat trials*

	Level of awarness (PAS)							
	AL1		AL2		AL3		AL4	
	switch	repeat	switch	repeat	switch	repeat	switch	repeat
M	55.9	32.8	75.1	48.14	68.6	41.8	21.6	11.1
SD	21.9	14.2	28.0	16.8	19.2	11.6	17.4	9.3

4.4.2 Discrimination ability depends on subjective awareness

According to the individual reports after the experiment, some participants had noticed that non-vertical Gabors had occurred more often than the vertical Gabor. None of the participants, however, noticed a difference in the frequency between left- and right-tilted orientations. The averaged sensitivity and bias measures for each level of subjective awareness are reported in Table 6. Participants' sensitivity to discriminate between a non-vertical and a vertical grating decreased with vanishing subjective awareness. On trials with almost full (AL3) and partial awareness (AL2) participants maintained considerable perceptual sensitivity regarding the Gabor's orientation: Bayes-factors provided strong evidence for the mean A' of .675 ± .056 to be greater than .5 in AL3 trials, $BF_{10} = 12.89$, 95% CI (.553, .797) and anecdotal

evidence for the mean A0 of .585 ± .049 (SE) to be truly greater than .5 in AL2 trials, BF_{10} = 1.671, 95% CI (.479, .690). In contrast, in unaware trials (AL1), the mean A' was .48 ± .030 with the Bayes factor providing moderate evidence for the H0 suggesting that volunteers' perceptual discrimination ability was at chance, BF_{10} = .178, 95% CI (.414, .545), (Quintana & Williams, 2018). Comparing the sensitivity between the three awareness levels Bayes factors provided strong evidence that A' in AL1 trials was truly smaller than the mean A0 in AL2 trials, BF_{10} (AL1 $<$ AL2) = 19.957, and extreme evidence that it was smaller than the mean A' in AL3 trials, BF_{10} (AL1 $<$ AL3) > 100. Comparably, there was strong evidence for A' of AL2 trials to be smaller than A' of AL3 trials, BF_{10} (AL2 $<$ AL3) = 34.336. Violin plots of the sensitivity distribution of each level of awareness are depicted in Figure 7 a).

We assumed a yes/no discrimination task set up in our paradigm and calculated A' as the sensitivity measure. However, given the fact that volunteers were required to map left and right tilted Gabors with differing angles to the same response (i.e., non-vertical), a classification scenario may be more appropriate (Snodgrass et al., 2004). On the cognitive level, such a scenario demands volunteers to establish two rather than one decision criteria, thereby increasing the decision uncertainty. The proportion of correct responses (i.e., proportion correct, p(c)) then serves as the means to measure perceptual sensitivity (Macmillan & Creelman, 2004, pp. 190-191). Hence we next calculated p(c) for each level of subjective awareness, where p(c) was defined by using the presentation probabilities of the two non-vertical targets as weights for the hit rate and adding this to the product of the 1-false alarm rate (i.e., correct rejection rate) and the presentation probability of the vertical target (i.e., p(c) = (8/36)*H + (16/36)*H + (12/36)*(1-F); Macmillan & Creelman, 2004, p. 89).

In agreement with the results of the sensitivity analysis using A', we observed a mean p(c) of .514 ± .042 in AL1 trials with a Bayes factor analysis providing moderate evidence for p(c) to be equal to chance (50%), BF_{10} = .310, 95% CI (.423, .605), (Quintana & Williams, 2018). In AL2 trials the p(c) was .609 ± .060 and the Bayes factor gave anecdotal evidence that it was truly above the chance level, BF_{10} = 1.885, 95% CI (.480, .739). In AL3 trials the mean p(c) was .695 ± .272 and the Bayes factor showed moderate evidence for a p(c) above chance, the BF_{10} = 6.738, 95% CI (.538, .852). Group distributions of the p(c) for each level of subjective awareness are depicted in Figure 7 b).

Finally, we analyzed individual response biases. On unaware trials (AL1) we observed a negative mean C of .471 ± .187 and the Bayes factor provided only anecdotal evidence in favor of a C smaller than zero, BF_{10} = 1.720, 95% CI(0.821, 0.012), (Quintana & Williams, 2018), tentatively suggesting that volunteers were biased to report a non-vertical orientation more often. In trials with residual awareness (AL2) the mean C of 0.098 ± 0.147 was associated with a Bayes factor providing moderate evidence for a C truly at zero indicating unbiased responses, BF_{10} = .328, 95% CI (0.415, 0.219). Similarly, in almost fully aware trials (AL3) the Bayes Factor for the mean C of .242 ± .159 provided anecdotal evidence for a C equal to zero, BF_{10} = .692, 95% CI (0.102, 0.586). Next, we computed a Bayesian mixed model using the subjective measures of awareness (AL1-AL3) as fixed effect predictor and a by-subject

random intercept to test for variations in C across the three levels of awareness. The analysis resulted in BF_{10} = 52.89, providing strong evidence for variations in C across the levels of subjective awareness. Violin plots showing the distributions for the criterion location for the three awareness levels are shown in Figure 7 c).

Table 6: *Perceptual sensitivity and response bias*

	Level of awarness (PAS)					
	AL1		AL2		AL3	
	A′	C	A′	C	A′	C
M	0.513	−0.254	0.616	−0.060	0.685	0.0
SD	0.048	0.538	0.128	0.399	0.139	0.422

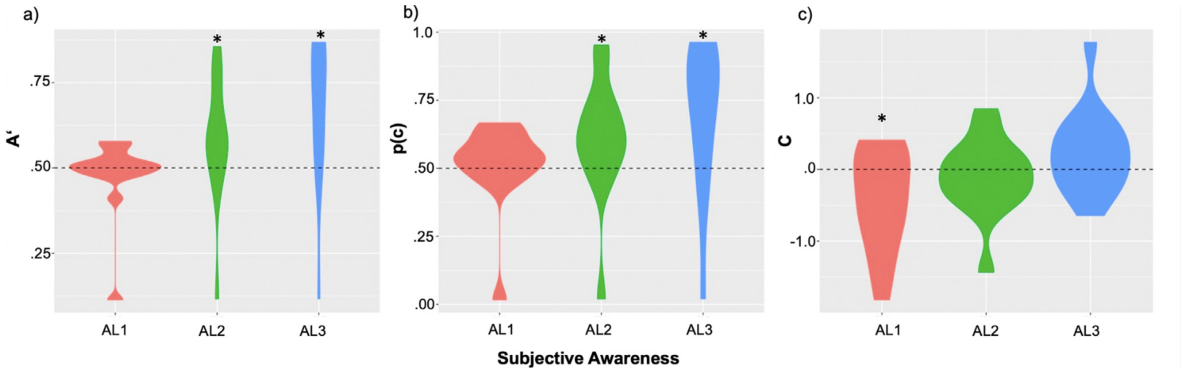


Figure 7: a) Violin plots shows the sensitivity parameter A′ as a function of subjective awareness. Black dashed line shows the level of zero sensitivity. Black asterisks indicate BF_{10} providing evidence for a mean A′ truly greater than .5. b) Sensitivity parameter p(c) as a function of subjective awareness. Black dashed line shows the level of zero sensitivity. Black asterisks indicate Bayes factors providing evidence for a mean p(c) truly greater than .5. c) Response bias C as a function of subjective awareness. Black dashed line appears at the level of no response bias. Black asterisk indicates a Bayes factor providing evidence for a mean C truly smaller than 0. To illustrate distributions of numeric data, violin plots make use of density curves where the width matches the approximate frequency of data points in each region. The lower and upper limits of each plot is determined by the distribution's minimum and maximum value.

Together, the data show that the ability to distinguish the two types of orientations (non-vertical versus vertical) strongly depended on subjective visibility. Importantly, we found a concordance of subjective visibility and the objective measure of awareness: the lower volunteers rated their subjective awareness, the worse their ability to correctly identify the stimulus orientation, being at random when subjective unawareness was reported. Trials with higher subjective levels of awareness (AL2-AL3) showed substantial sensitivity above chance level and were thus counted as aware. However, the variation in the response bias across the three levels of visual

awareness may suggest that volunteers' perceptual decision criteria were affected by their subjective awareness reports and variations in the sensitivity measure (A') could potentially be influenced by variations in the response bias. In fact, the low sensitivity (i.e., A' or p(c), respectively) in AL1 trials could have resulted from a response bias in this condition (Macmillan & Creelman, 2004). Thus, true absence of perceptual sensitivity in the subjectively unaware trials cannot be fully ascertained. An analysis of the luminance contrast values (i.e., the physical signal strength) of the presented Gabor patches can be found in the Appendix B.1.

4.4.3 Reaction times switch costs in response to unaware orientation changes

Critical for the purpose of this study was to examine if there was an effect on decision RTs in switch trials, i.e., due to changes in the orientation between a given trial and the previous trial. Descriptive mean RTs and SEs for switch versus repeat trials for each level of awareness are summarized in Table 7. First, we analyzed the *weighted* switch model containing only those trials with the highest expected switch costs. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Here the LMM analysis showed that estimated RTs increased with decreasing visual awareness, $F(2, 13.671) = 24.118$, $p < .001$; Post-hoc tests after Bonferroni correction showed that mean RTs across both switch and repeat trials significantly slowed down about 119.7 ± 35.9 ms in AL1 trials compared to AL2, $t(12.5) = 3.335$, $p = .017$, 95% CI (20.57, 219.0), about 186.5 ± 37.2 ms compared to AL3 trials, $t(12.0) = 5.019$, $p < .0001$, 95% CI (83.24, 290.0). Finally, mean RTs in AL2 trials were on average 66.8 ± 22.6 ms slower compared to AL3 trials, $t(15.5) = 2.961$, $p = .028$, 95% CI (6.26, 127.0).

More importantly, the orientation change (fixed effect switch) had also impacted RTs as indicated by the significant interaction between the factors switch and subjective awareness, $F(2, 35.999) = 5.406$, $p = .009$. Here the post-hoc tests showed that in the unaware condition (AL1) RTs in repeat trials were on average 115.35 ± 32.5 ms faster compared to switch trials $t(29.4) = -3.570$, $p = .001$, 95% CI (-182.4, -49.6). In trials with higher levels of visual awareness switch costs were not significant, AL2, $p = .343$, 95% CI (-97.7, 35.1); AL3, $p = .907$, 95% CI (-62.6, 70.2). There was also a statistical trend for the main effect of the fixed effect predictor switch, which was, however, not significant, $F(1, 13.278) = 3.889$, $p = .070$. In Figure 8 a) RTs for both switch and repeat trials are plotted as a function of visual awareness for the weighted switch model. The LMM solutions for the fixed and random effects are given in Table 8 a).

Next, we calculated the same LMM analysis for the *average* RT model in which the mean of the switch condition included all possible switch trials. Residual plots did not suggest deviations from homoscedasticity or normality. Again, estimated RTs appeared sensitive to changes in the level of visual awareness indicated by the significant fixed effect of *awareness*, $F(2, 12.575) = 6.073$, $p = .014$. The post-hoc tests with Bonferroni correction indicated that RTs in AL1 trials were on average 76.1 ± 31.0 ms slower compared to AL2 trials, but this difference was statistically not significant, $(12.1) = 2.454$, $p = .091$, 95% CI (-10.05, 162.0). However, RTs in AL1 trials

were on average 133.2 ± 40.4 ms slower compared to AL3 trials, $t(12.0) = 3.301$, $p = .019$, 95% CI (21.06.49, 254.0). Finally, mean RTs in AL2 were about 57.1 ± 18.7 ms slower compared to AL3 trials, $t(12.4) = 3.046$, $p = .029$, 95% CI (5.29, 109).

Orientation changes, however, did not significantly impact RTs: There was no significant main effect of *switch*, $F(1,14.515) = 2.633$, $p = .126$, nor a significant interaction $F(2, 36.00) = .267$, $p = .767$. RTs for both switch and repeat trials as a function of visual awareness for the *average* switch model are depicted in Figure 8 b). The LMM solutions for the fixed and random effects for the exhaustive switch model are given in Table 8 b).

Taken together the LMM analysis showed that RTs were sensitive to decreasing visual awareness as well as to changes in the stimulus orientation. Yet, orientation changes impacted RTs only if those switch trials were considered in which the novel orientation changed away from the highly biased orientation (highly frequent tilt) suggesting that prior visual selection had boosted behavioral switch costs. RT data were best described by an interaction of visual awareness and changes in the stimulus orientation with significant slowing of RTs only in unaware switch trials.

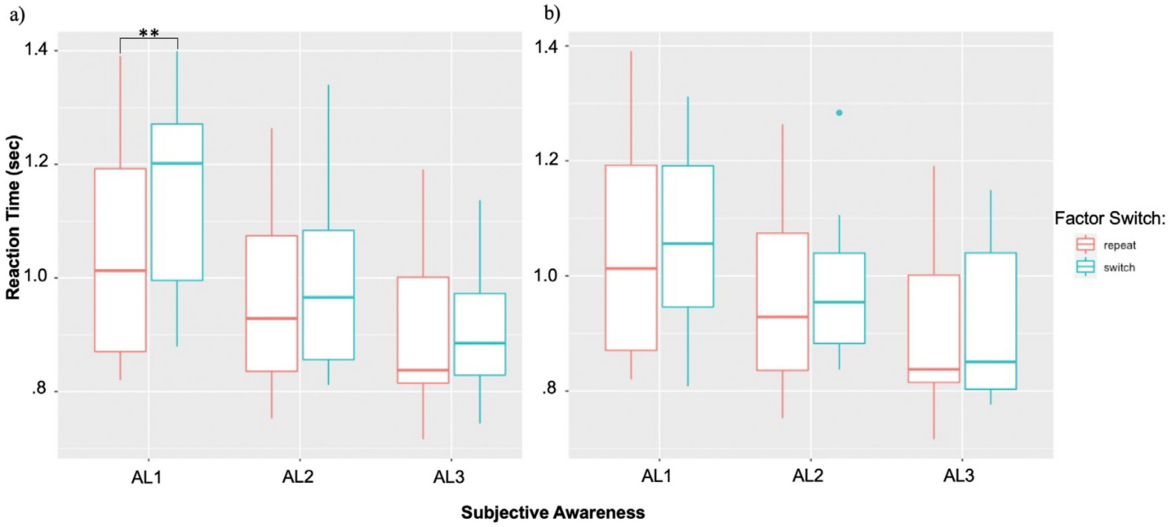


Figure 8: Boxplots depicting RTs in seconds as a function of visual awareness plotted for switch (blue) and repeat trials (red) in the weighted model (a) in which the switch factor comprised only those switch trials away from the strongly weighted (frequent tilt); b) average model in which all switch trials were included; black asterisks = significant difference between switch and repeat trials, $p < .01$.

Table 7: *RT data*

	Level of awareness (PAS)					
	AL1		AL2		AL3	
	switch	repeat	switch	repeat	switch	repeat
Weighted model						
M	1.152	1.036	0.990	0.959	0.906	0.910
SD	0.185	0.178	0.151	0.162	0.124	0.145
Average model						
M	1.061	1.036	0.986	0.959	0.921	0.910
SD	0.158	0.178	0.125	0.162	0.140	0.145

Note: Values are reported in seconds.

Table 8: LMM parameter estimates of RT analysis

Fixed effects	Estimate (in seconds)	SE	*t-value*	*p*
A) Weighted				
AL1 (intercept)	1.03669	.04584	22.617	3.91e-12***
AL1-AL2	-.07737	.03883	-1.916	.070276
AL1-AL3	-.12661	.04202	-2.980	.008178**
Switch (Intercept)	.11599	.03249	3.570	.00116**
AL2: Switch-Repeat)	-.08469	.03746	-2.261	.016213*
AL3: Switch-Repeat	-.11981	.03746	-3.198	.00288**
Random effects	Variance	SD		
Subject	.022751	.15084		
AL1-AL2	.012188	.11040		
AL1-AL3	.013392	.11572		
Residual	.004604	.06754		
B) Average				
AL1 (intercept)	1.036688	.047479	21.835	2.51e-11***
AL1-AL2	-.077365	.033110	-2.337	.0.03326 *
AL1-AL3	-.126607	.041977	-3.016	.0.00924**
Switch (Intercept)	.024623	.018619	1.322	.19467
AL2: Switch-Repeat)	.002497	.023136	.108	.91464
AL3: Switch-Repeat)	-.013223	.023136	-.572	.57120
Random effects	Variance	SD		
Subject	.027566	.16603		
AL1-AL2	.010772	.10379		
AL1-AL3	.019427	.13938		
Residual	.001027	.03205		

Note: Significance codes: *** $p < .001$; ** $p < .01$; * $p < .05$. Estimates, (beta values, in seconds), SEs, *t*-ratios, and *P*-values for the fixed effect predictors in the final LMM. a) *T*- and *p*-values indicate the difference of each factor level (and factor level combination) compared to baseline (intercept AL1 for the fixed effect awareness and switch for the fixed effect switch). The intercept is tested against zero. The model results based on the fixed effect predictor switch comprising only the changes away from the heavily weighted orientation are given in a); b) shows the results of the model using the fixed effect predictor switch including all types of orientation changes.

4.5 fMRI Results

4.5.1 GLM results - switch-related fronto-parietal activity linked to unaware information processing

To show brain regions that were particularly sensitive to unawareness, first, both unaware switch and repeat trials were combined and contrasted against both types of trials on higher levels of awareness *(2 x switch AL1 + 2 x repeat AL1) > (switch AL2 + repeat AL2 + switch AL3 + repeat AL3)*. Here we observed three clusters showing significant signal change in unaware (AL1) compared to almost fully aware trials (AL3) and in trials with residual awareness (AL2) located in the right lateral FPC, the right SMG and the right angular gyrus (AG), including the right TPJ (see Figure 9 a) and Table 9 a)). The respective parameter estimates (β-values) converted to percentage change (see Figure 10) indicated an elevated BOLD response in these regions. The behavioral analysis already showed switch costs to be more pronounced for unconscious orientation changes compared to consciously perceived changes. Thus, we next looked for analogous activation increases in the fMRI data and contrasted unaware switch trials against unaware repeat trials using the interaction contrast *(switch > repeat) AL1 > (switch > repeat) AL3* to test if the difference between switch and repeat trials differed between AL1 and AL3. It revealed an anterior cluster of 409 voxels located in the frontal lobe spanning parts of the right MFG as well as the right IFG and reaching posteriorly beyond the right central sulcus to the precentral gyrus (PCG). We found a second cluster located more posteriorly comprising 398 voxels that extended from the superior division of the right lateral occipital cortex (LOC) extending anteriorly to the right SPL and, more inferior, to the right AG, reaching down to the right precuneus (PCC) and cuneal cortex (CC) (see Figure 9 b) and Table 4 b). The positive percentage change in the signal observed in all but two participants (see Figure 10) suggested an elevated activation in the unaware switch condition in these regions.

In sum, right MFG, IFG as well as the right AG were modulated by invisible target changes. Together these findings suggest that right frontal and parietal regions together serve the detection of unconsciously perceived alterations of a critical stimulus supporting attentional shifting in a situation that requires the disruption of rule-based resource allocation and the exploration of new environmental aspects. Yet, while clusters reaching in the dorsal parts of the right lateral FPC showed an increased BOLD response in trials in which the target was not perceived consciously, this region did not respond specifically to invisible target changes.

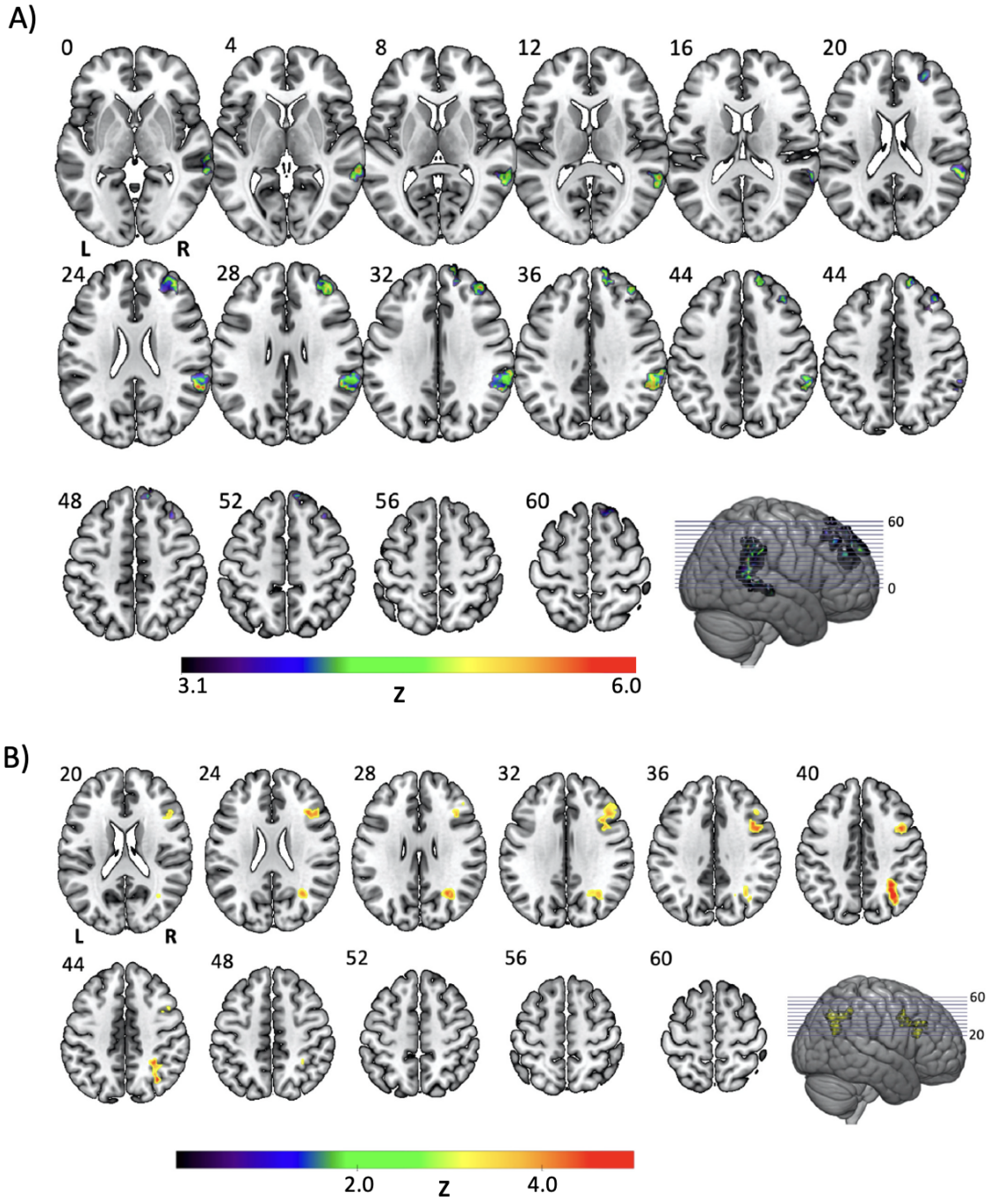


Figure 9: Z-score group-level activation maps overlaid on an MNI152 template ($Z > 3.1$; $p_{FWE} < .001$); A) Areas with stronger BOLD signal on unaware compared to aware trials combined for both, switch and repeat trials; B) Regions with a higher BOLD-response on unaware switch trials compared to unaware repeat trials tested via the interaction (*switch* > *repeat*) *AL1* > (*switch* > *repeat*) *AL3*.

Table 9: *Table of activations*

a) 2AL1 > (AL2 + AL3)

#	K	Z Max Loc	Z Max	*p*	Structure
1	1090	56,-46, 26	7.19	8.32e-11	rAG, rSMG (rTPJ)
2	496	38, 42, 26	5.95	3.7e-06	rFPC
3	319	14, 50, 38	5.75	.000191	rFPC

b) (switch AL1 > switch AL3) > (repeat AL1 > repeat AL3)

1	409	44, 6, 38	4.61	.000601	rMFG
2	398	34,-62, 40	6.16	.000719	rLOC, rAG, rSPL (rTPJ)

Note: K = cluster size in voxels, Z Max Loc = MNI coordinates of the location with the maximal Z-value, Z Max = maximal Z-value, r = right, structure determined using the Harvard–Oxford Cortical Structural atlas).

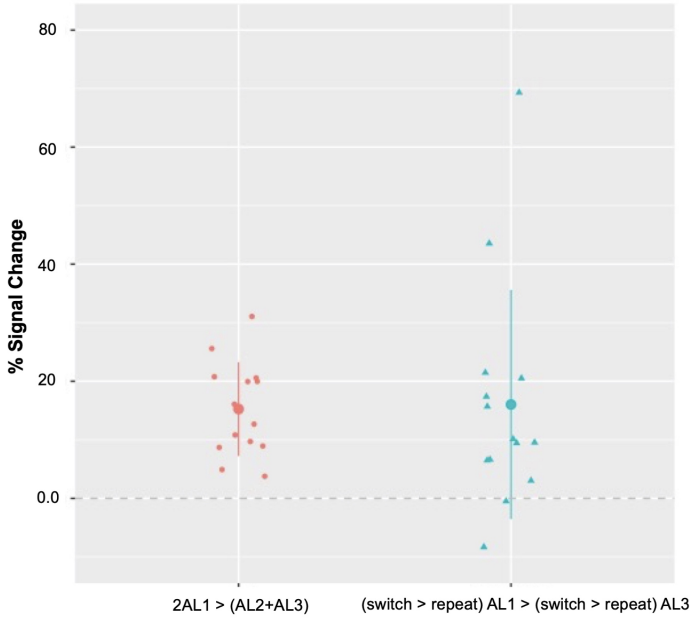


Figure 10: Average parameter estimates (percentage (%) signal change) observed across all clusters with a significant BOLD response on the left for the contrast *2AL1 > AL2+AL3*, and, on the right, for the interaction contrast (*switch > repeat*) *AL1 >* (*switch > repeat*) *AL3*.

4.5.2 Representations of unconsciously perceived target orientations outside the visual cortex in the right parietal and frontal cortex

We next carried out multivariate searchlight analysis to assess where in the brain the multivoxel activation patterns carried information about the critical stimulus even though that stimulus was not perceived consciously. On the one hand we were interested to assess if areas that were particularly responsive to invisible orientation changes (i.e., those clusters with an increased BOLD signal for the interaction contrast *(switch > repeat) AL1 > (switch > repeat) AL3* shown in Figure. 9 b)) also maintained the orientation information. Thus, we performed a searchlight analysis on these regions of interest (number of non-zero voxels = 807). One the other hand, we wanted to test where else the feature information of the invisible stimulus was represented, extending those areas revealed in the GLM analysis, wherefore we also conducted a whole-brain searchlight analysis (number of non-zero voxels = 221432). The behavioral task required participants to make two types of responses, i.e., one for vertical and another one for tilted gratings. Thus, if the switch was between vertical and tilted, not only was there a change in the orientation but also a change of the behavioral response. Therefore, we included only left- and right-tilted Gabors rated as unaware in the multivariate analysis. For the searchlight we chose a 9mm radius. The same analysis we repeated using a 6 mm searchlight radius which is reported in the Appendix B.2. The first ROI-based searchlight analysis showed that parts of those cluster showing an increased BOLD response in the interaction contrast of the GLM analysis (i.e. *(switch > repeat) AL1 > (switch > repeat) AL3*, Figure. 9 b)) also carried informative clusters on the group level with significant searchlight centers ($p_{\text{Cluster}} < .05$) located in the right MFG, right IFJ as well as in right SPL, AG and right LOC with a mean decoding accuracy of 66.7 ± 1.9%. Figure 11 a) contains dot plots showing the individual decoding accuracies for the 14 subjects. Dot plots showing the decoding accuracies of all clusters that survived the significance testing by bootstrapping on group level are depicted in Figure 11 b). Figure 12 a), shows the accuracies of clusters with significant searchlight centers on group-level mapped on an MNI152 standard brain (red clusters))

Remarkably, the second whole-brain searchlight analysis revealed local information that discriminated between the two tilted orientations on unaware switch trials located even more anteriorly in the right lateral FPC (see Figure 12 b), transverse slices 12 and 16). Next, confirming the ROI-based searchlight, we also found informative clusters in the right MFG and IFJ. More posteriorly, the analysis revealed clusters with significant searchlight centers located bilaterally in the TPJ, precuneus, and intraparietal sulcus (IPS), as well as in visual cortex including the lingual gyrus, cuneus, LOC, and occipital pole ($p_{\text{Cluster}} < .05$). There were also group clusters with significant searchlight centers ($p_{\text{Cluster}} < .05$) in anterior cingulate gyrus, bilaterally in the temporal pole and orbitofrontal cortex (OFC). The mean decoding accuracy among all clusters carrying local information informative about the orientation of the invisible stimulus was 71.5 ± 4%. Figure 11 a) contains dot plots showing the individual decoding accuracies for the 14 subjects for the whole-brain searchlight. Dot plots showing the decoding accuracies of all clusters that survived the significance testing by bootstrapping on group level are depicted

in Figure 11 b). The brain clusters revealed in the whole-brain searchlight analysis are depicted in Figure 12 b).

We then assessed whether decoding in frontal and parietal regions depended on whether the target grating was visible or invisible. Therefore, we repeated the ROI-based searchlight analysis following the same methodological procedure as outlined in the method's section 4.3.5 in the ROI derived from the GLM contrast (Figure 4 b)) comprising the right MFG, right AG and right LOC. This time left- and right-tilted Gabors of only AL2 and AL3 trials were included in the analysis. Group clusters with significant searchlight centers ($p_{\text{Cluster}} < .05$) were located in the posterior right middle frontal gyrus (rMFG). For these clusters a mean decoding accuracy of 63.0 ± 2% was observed. However, this analysis did not reveal the maintenance of feature information in clusters located in the parietal areas that we found when decoding the orientation of invisible targets (i.e., rAG, SPL, and LOC). Thus, albeit the decoding accuracies, obtained in the searchlight analysis using aware trials only, did not differ statistically from those obtained in the analysis of AL1 trials, $t(49) = -1.38$, $p = .1730$, the spatial distribution of clusters with local information discriminating the target orientations was distinct for invisible versus visible targets. Decoding accuracies mapped on a MNI 152 standard brain are depicted in Figure. 12 a, blue clusters).

Taken together, these results show that the orientation could be decoded even for unconsciously processed gratings. The informative voxel clusters on the group level were distributed in a network reaching from the frontal to the occipital cortex including the right lateral, MFG, IFJ, SPL, IPS and TPJ. Most importantly, we showed that clusters in the prefrontal and parietal cortex were not only modulated by changes of the unconsciously perceived target's orientation compared to visible orientation changes, but also that the activity patterns within these regions carried information about the stimulus orientation.

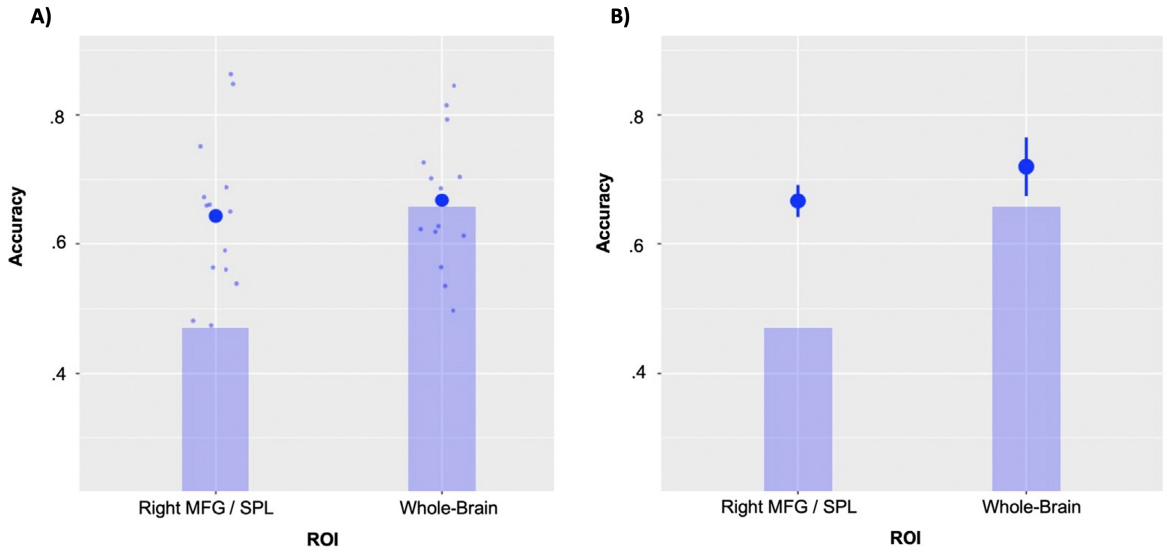
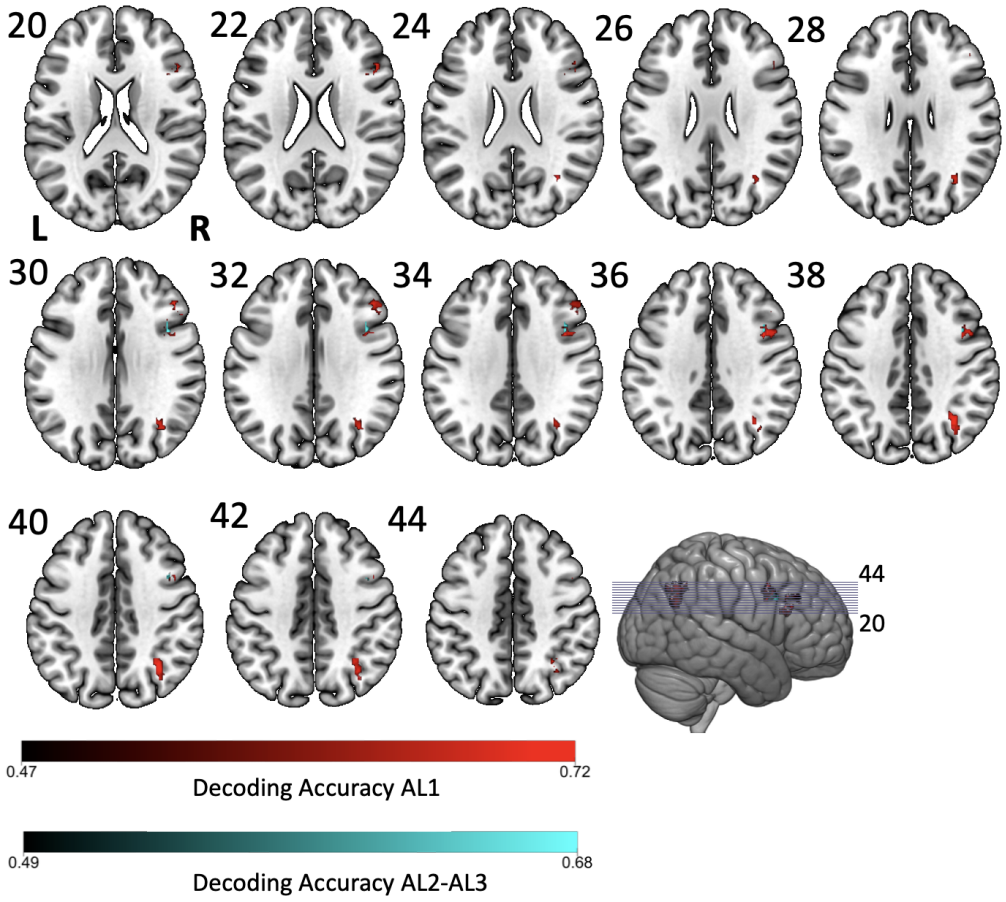


Figure 11: Dot plots of individual decoding accuracies in the searchlight analysis with a 9 mm SLR (3 voxels). On the left it shows subjects' decoding accuracies obtained for ROI-base searchlight analysis restricted to those clusters that showed a significant BOLD signal change in the *(switch > repeat) AL1 > (switch > repeat) AL3* contrast (left side). One the right the individual decoding accuracies of the whole-brain searchlight are depicted. The circles bigger in size show the mean accuracy across all subjects. The shaded areas show individual accuracies that fall below the 99th percentile of the null distribution computed by the permutation test. B) reports the mean decoding accuracies calculated of all informative clusters with significant searchlight centers ($p_{\text{Cluster}} < .05$) observed after bootstrapping on the group level for the ROI-based as well as for the whole-brain searchlight. The algorithm that we used implements a two-stage procedure using the results of within-subject permutation analyses, estimates a per feature cluster forming threshold (via bootstrap), and uses the thresholded bootstrap samples to approximate the distribution of cluster sizes in group-average accuracy maps under the Null hypothesis (i.e., no information present in the multi-voxel patterns). The big circles indicate the mean accuracy across these clusters. The shaded areas, again, reach the upper bound of the 99th percentile of the null distribution obtained from the permutations after bootstrapping on the group level.

a)


20
22
24
26
28
L
R
30
32
34
36
38
40
42
44
44
20
0.47
0.72
Decoding Accuracy AL1
0.49
0.68
Decoding Accuracy AL2-AL3

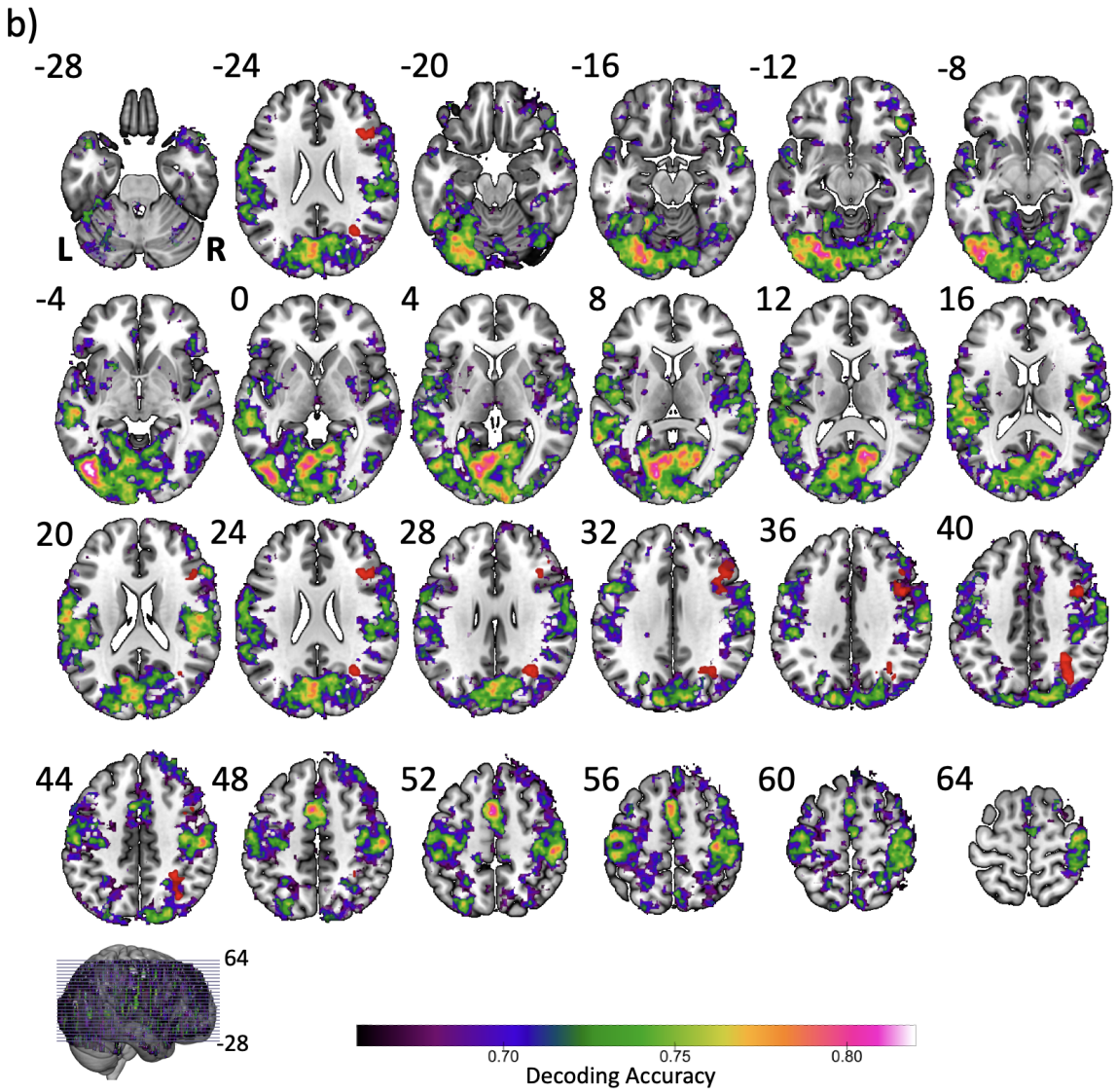


Figure 12: Decoding accuracies obtained with a 9 mm searchlight radius of clusters with significant searchlight centers (green) surviving the thresholded two-steps permutation procedure ($p_{\text{Cluster}} < .05$) mapped on a MNI 152 standard brain. A) shows the resulting clusters of the searchlight analysis restricted to those regions revealed in the interaction contrast *(switch > repeat) AL1 > (switch > repeat) AL3* (see Figure 9 b)). Red clusters are those we obtained in the analysis that included only AL1 trials, while the blue clusters are those, we observed in the analysis using AL2 and AL3 trials. For only those accuracies above the 99th percentile are shown. B) shows the results of the whole brain searchlight analysis: decoding accuracies that exceed the 99th percentile (i.e., 65.7%) of the null distribution are mapped on a MNI 152 standard brain. Red clusters show the areas with a significant signal change in the univariate interaction contrast of the GLM analysis (see Figure 9 b)).

4.6 Discussion

In this study we investigated exploratory shifts of attention in the absence of visual awareness and hypothesized, on the brain level, the right FPC to be crucially involved. At the behavioral level, we indeed observed that participants' RTs in the categorization task were prolonged when the orientation of a central bar stimulus changed away from the most frequent orientation, while both subjective and objective measures indicated the absence of conscious perception.

Thus, our participants appeared to have optimized perception by attentional weighting of the most frequent orientation although the stimuli were not consciously perceived. This, in turn, caused RT costs when less frequent orientations were presented. Previously, we had observed left lateral FPC activation when attention was reweighted following visible target changes. Here, we investigated if the same pattern would be observed for unconsciously processed stimuli. However, FPC was not more strongly activated in unconscious change trials. Thus, a specific role of the FPC for attentional modulation of unconsciously perceived stimuli could not be confirmed. Instead, however, FPC activation represented the stimulus orientation of unconsciously perceived stimuli. Thus, the FPC would be able to send feedback to posterior regions such as the TPJ and IPS when target changes occur. Due to the lack of change-related FPC activation, we have no evidence that such change signaling occurs for unconsciously perceived target changes, but it cannot be ruled out that such change signals might have been too weak to be observed in the present experiment, or that they might occur for target dimension changes instead of feature changes (Pollmann et al., 2000; Weidner et al., 2002) or spatial attentional changes (Lepsien & Pollmann, 2002) even if these occur unconsciously.

4.6.1 Nodes of the ventral attention network detect invisible stimulus changes

Increased activation during unconsciously perceived switch trials was observed in the right posterior parietal cortex. This aligns well with many studies showing posterior parietal involvement in conscious attention changes (reviewed by Corbetta et al., 2008; Wager et al., 2004). Activation along the posterior, descending segment of the IPS was recently found to be increased for salient distractor stimuli rather than for equally salient targets (Jamoulle et al., 2021). Thus, posterior IPS activation was not driven by salience per se, but discriminated between the task-relevant, attended targets and the task-irrelevant, non-attended distractors. In the present experiment, we find a similar pattern in that this area was more strongly activated by the non-attended, new orientation than the attended orientation. In both cases, the underlying function may be a response to stimulus changes that may potentially require a reallocation of attention. Certainly, the visual search task used by Jamoulle et al. (2021) and our task differ in many respects. Thus, the hypothesis that both studies induced the same functional process in posterior IPS needs to be confirmed by further studies where consciousness is varied within the same paradigm. In addition, areas that partially define the anatomy of the right TPJ, such as the right AG as well as the right LOC (Schurz et al., 2017) showed increased activation in response to invisible target changes. This observation is consistent with the idea that

TPJ serves to facilitate attentional weighting of the novel target orientation. We also found decodable feature information of the invisible stimulus in TPJ suggesting that, by representing the novel invisible stimulus in the multivoxel activity patterns, TPJ supports participants' disengagement of attention (Corbetta et al., 2008) from the previous orientation to facilitate subsequent attentional weighting of the new orientation. The cortex along the posterior inferior frontal sulcus (including the inferior frontal junction (IFJ))- consistently activated following attentional switch processes in the literature (e.g., Derrfuß et al., 2009; Dove et al., 2000) - showed increased activity following an invisible target change and represented the orientation of the novel target which supports the idea that the ventral attention network sends a reorienting signal to the dorsal attention network through MFG (Corbetta et al., 2008; Japee et al., 2015).

4.6.2 Dissociation between attention and visual awareness

Our findings support the distinction between attention and consciousness: the reweighting of attention in response to a change of an unconscious target is in keeping with the view that attention and consciousness can be dissociated (Koch & Tsuchiya, 2007; Lamme, 2003). Our findings tie in well with other reports, showing visual selection biases can occur in the absence of visual awareness (Kanai et al., 2006; Pan et al., 2014; Zhang & Fang, 2012; for a review see Mulckhuyse & Theeuwes, 2010) and are thus in keeping with emerging evidence that higher-order cognitive control mechanisms can be deployed without conscious awareness (Soto & Silvanto, 2014; Van Gaal & Lamme, 2012; Van Gaal et al., 2010). Previous studies provided evidence for mid-dlPFC (BA 46) to be responsive during unconscious priming of task response settings (Lau & Passingham, 2007) and for left lateral FPC to serve implicit attention guidance for visible items (Pollmann & Manginelli, 2009b), but the activity was not tested for task-relevant representational content. Frontal involvement in dorsolateral and anterior PFC has been shown during a visual short-term memory task for masked stimuli (Dutta et al., 2014), however in this study visual awareness was defined by means of a subjective measure (items reported as 'unaware') while here we used both objective and subjective measures to establish the lack of visual awareness. Critically, our multivariate pattern analyses showed that the FPC was involved in the representation of the task-relevant feature (orientation).

4.6.3 Deconding of masked stimuli

Previous work demonstrated that masked stimulus information (i.e., orientation) is processed and maintained in the visual cortex (e.g., Haynes & Rees, 2005). Yet, our classification analysis provides novel evidence that the representation of an unconscious stimulus orientation was maintained across a distributed set of brain areas in the right lateral FPC, right IFG, right MTL, right parietal and visual cortex. This pattern of results aligns with recent evidence showing that unconscious perceptual content can be decoded from activity patterns in a distributed set of brain regions, including parieto-frontal areas (Mei et al., 2022) and further elaborates on recent evidence showing that the frontoparietal cortex is implicated in the maintenance of conscious feature information during visual working memory and search tasks (Ester et

al., 2015; Lee & Baker, 2016; Reeder et al., 2017). Importantly, our results foster the notion that unconscious information processing implicates supra-modal areas typically linked to conscious processing such as the pre-frontal cortex (for reviews see Soto & Silvanto, 2014; Van Gaal & Lamme, 2012). The present results align with the view that engagement of prefrontal areas is not restricted to conscious processing.

4.6.4 Switch costs in RTs restricted to unaware trials

A somewhat unexpected result was the absence of significant switch costs in trials in which the target was consciously perceived (AL2-AL3), a pattern that we already observed in a previous behavioral study (Güldener et al., 2021). The switch costs observed in unaware trials appear to underline the necessity for attentional re-weightingeat least under difficult viewing conditions. We assume that attention modulates contrast gain in our paradigm (e.g., Reynolds & Heeger, 2009). Modulating contrast gain can in turn increase sensitivity (Carrasco, 2006), which would be particularly helpful for AL1 stimuli. Furthermore, attentional weighting processes at high stimulus contrast appear to play a minor role for within-dimension feature changes (such as the orientation changes in the current experiment) compared to changes between feature dimensions (Müller et al., 1995; Pollmann et al., 2000). Therefore, contrast gain modulation by attention may have been particularly present at AL1, explaining the increased response times when a changed orientation did not fit the attention template, requiring re-weighting of feature attention.

4.6.5 Expectation about the upcoming target driving the decoding?

It is possible that the information about the target orientation decoded from the MVPs of brain activity may be related to participants' expectations regarding the target's orientation in addition to the actual stimulus orientation. Further studies are needed to test this hypothesis by using MVPA for decoding participants' predictions regarding the incoming perceptual input (e.g., based on the categorization response given to each stimulus orientation). As we did not specify different responses for the left versus right tilt, our study was not designed to pinpoint this issue. Note, however, that, since correct and incorrect responses were used for the MVPA in the unconscious trials, and volunteers performed at chance (A' = .5) in these trials, it is unlikely that the results presented here were driven by categorization responses or participants' expectations driving their responses rather than the actual stimulus orientation.

4.6.6 Caveats of combining objective and subjective measures of visual awareness

Our methodological approach in which we combined subjects' perceptual sensitivity as the objective measure of the stimulus visibility with subjects' awareness ratings may potentially be criticized because there is only a single distribution of vertical and tilt responses that we arbitrarily divided into the four levels of subjective awareness. Thus, the sensitivity measure (A') is biased if the response bias differs across the levels of subjective visual awareness (AL1-4). The analysis of subjects' response bias (C) indeed revealed variations and we therefore cannot rule out that the A' was

biased which may have led to an under-estimation of A' in subjectively unaware trials (AL1), (Stein et al., 2016). Hence, even if combining subjective and objective measures to determine the stimulus visibility appears conceptually appealing, we cannot fully rule out that there is some variance in sensitivity across conditions.

4.6.7 Why neural repetition suppression as the driving mechanism is unlikely

Could other mechanisms than attentional re-weighting explain our results? It might be argued that a simple repetition effect (Bertelson, 1961, 1963), leading to neural repetition suppression (Henson & Rugg, 2003), could alternatively explain the observed response facilitation and lower activation in unaware repeat trials compared to unaware switch trials. In previous work, we have demonstrated modulation of activity in visual areas processing color and motion in support of attentional weighting of the target dimension (Pollmann, 2016). In the present experiment, such a proof is difficult because stimulus changes occurred between features of the same stimulus dimension, which are neurophysiologically represented in neural columns within the same brain areas (Hubel & Wiesel, 1962). However, our central claims, that orientation changes can be processed in invisible stimuli and that orientation was represented in several brain areas up to FPC are valid both for attentional weighting and repetition suppression accounts.

4.6.8 Conclusions

Orientation changes in unconsciously perceived stimuli induced both behavioral and neural effects. Behaviorally, we observed switch-costs that went along with increased activation in the posterior parietal cortex. In addition, the orientation of invisible targets was represented in a number of brain areas, reaching anteriorly up to the frontopolar cortex. We conclude that while change-related activation was restricted to the posterior cortex, the information about target feature changes was available in a network of brain areas including prefrontal cortex.

5 Experiment 3: Behavioral bias for exploration is associated with enhanced signaling in the lateral and medial frontopolar cortex

The results of this experiment were first published as a preprint at psyarxiv: Güldener, L., & Pollmann, S. (2023). Behavioral bias for exploration is associated with enhanced signaling in the lateral and medial frontopolar cortex.

5.1 Introduction

In everyday life, we are constantly confronted with the need to observe our own action strategies and decisions and to readjust them to changing contexts. One example is to flexibly switch between a state of exploitation, in which we hold on to the same, known, strategy and exploration, that allows us to turn away from our previous strategy and examine alternative choices. A central question in cognitive neuroscience is how the brain balances between these two states for an optimal behavioral outcome.

Here, we report results from an fMRI study in which we tested 20 human young adults in a probabilistic foraging task based on the visual search paradigm and recorded the task-related BOLD-signal. The aim was to identify the neural correlates supporting exploratory shifts of attention during a visual foraging paradigm and to link brain activity related to exploration to behavioral parameters indexing rules of thumbs the participants used to make exploratory foraging decisions (specifically patch-leaving decisions). During a 60 minute search participants 'foraged' in search displays (i.e., patches) for target items among distractors. They earned a monetary reward each time they located a target item using an MR-compatible PC-mouse inside the scanner. Importantly, with each new reward capture, the remaining reward probability in the current display decreased exponentially. Thus, foraging became less rewarding the longer participants kept searching in the same display. To compensate for this, participants could switch by mouse-click to a new display at any time. Previous neuroimaging studies examining the exploration-exploitation dilemma typically employed variants of the n-armed bandit gambling task stemming from the principles of a casino's slot-machine (e.g., Jones, 1975). In the non-stationary version, the reward probabilities of the arms to be pulled change unpredictably (i.e., restless bandit; i.e., Daw et al., 2006). Due to this unpredictability, this version requires constant balancing between exploitation and exploration. Similar to this, we introduced random and unpredictable variations of the initial reward probability. This should prevent participants from using a simple stopping rule where they always capture a fixed number of rewards, e.g., 5 rewards, and leave the patch once that number is reached (i.e., the 'fixed-N' rule), (e.g., Wilke et al., 2009). Instead, due to the quick reward depletion, participants in our task were similarly required to constantly balance their search behavior between exploiting the current or exploring a new display.

Animals as well as our ancestors who lived as hunter-gatherers would typically

face foraging environments structured in patches that are allocated in space and time (e.g., meadows or forest districts differing in their richness of prey at different locations that vary in their distance within one habitat). Foraging thus becomes a rather serial stopping-or-switching task entailing temporal traveling costs that result from roaming in the environment from patch to patch. This characteristic is not very well captured in bandit-like gambling tasks in which the decision problem equals a decision between simultaneous choices. Importantly, evidence suggests that the brain mechanisms that are involved in solving these two types of problems (serial versus simultaneous) may differ to some extent (Constantino & Daw, 2015; Kolling et al., 2012; Rushworth et al., 2012). While there is a large body of research reporting on the neural structures involved in the exploration-exploitation dilemma in bandit-like tasks, less evidence exists reporting on the functional brain anatomy supporting exploratory decisions making in patch-foraging tasks. At the same time, choosing the patch-based serial stopping-or-switching task may increase the ecological validity facilitating the translation to ethological studies in other species (Garrett & Daw, 2020; Kolling et al., 2012; Mobbs et al., 2018).

The FPC capitalizes a key role in supporting exploratory shifts of visual attention (Pollmann, 2016; Pollmann & Manginelli, 2009a; 2009b), presumably by maintaining relevant information about the novel goal attentional resources are redirected to (Güldener et al., 2022). Importantly, FPC activity is also consistently linked to exploratory decisions in probabilistic gambling (e.g., Boorman et al., 2009; Daw et al., 2006; Laureiro-Martínez et al., 2015; Zajkowski et al., 2017). Also subregions within the parietal cortex such as the TPJ and IPS as well as the dACC, and the insula were reportedly related to attentional exploration during gambling (Addicott et al., 2014; Blanchard & Gershman, 2018; Chakroun et al., 2020; Daw et al., 2006; Laureiro-Martínez et al., 2015) Thus, the FPC appears to be a crucially involved in a neurocognitive circuit containing regions of the ventral and dorsal attention network that allows subjects to explore alternative sources of reward in the environment by disengaging attentional resources from the current focus (Hogeveen et al., 2022; Mansouri et al., 2015). Therefore, we expected to observe a spatially similar pattern of positive BOLD signal changes when our subjects decided to explore a new display. These decisions obey so-called 'patch-leaving rules', behavioral heuristics or rules of thumb subjects may use to determine the optimal time point for exploring a new patch (Charnov, 1976; Krebs et al., 1974; Mcnair, 1982; Wilke et al., 2009). We were particularly interested in the participants' giving-up times (GUT). They are defined as the time since the last reward capture and leaving the current patch. The 'GUT rule' states that a patch is left once the time, since the last capture, exceeds a certain subjective threshold (Mcnair, 1982). Intriguingly, elderlies' declining tendency to explore is associated with increased GUTs during foraging (Mata et al., 2013). Thus, GUTs may serve as an index of behavioral exploitation. Due to the quick depletion of reward in a given display, subjects with a bias towards longer exploitation indexed by relatively longer GUTs should yield less monetary earnings. At the same time we expected a reduced signaling on the brain level during exploration in exploitation-biased participants. Conversely, stronger positive BOLD-signal changes during exploration, particularly in FPC, were expected to be found in subjects with higher behavioral propensity to explore, indexed by short GUTs.

5.2 Methods

5.2.1 Participants

20 native Germans (8 male) participated in the experiment. All volunteers were between 19 and 37 years old (M = 26,5 years), right-handed by self-report, had normal or corrected to normal vision, and no history of mental illness. They all gave written consent in accordance with the local ethics committee of the Otto-von-Guericke University and received the earnings they made in the foraging task as reimbursement. The sample size was calculated using the R library *WebPower* (Zhang & Yuan, 2018). Assuming an effect size of Cohen's f = 1.31 (Wolfe, 2013, experiment 5), and one independent variable with three levels (patch-quality, i.e., initial reward probability), a minimum of 9 participants was required to be able to detect a true effect of patch-quality on the participants' foraging behavior tested in a single-factor repeated measure ANOVA with a power of 90%.

5.2.2 Apparatus and Stimuli

We used the Python toolbox *PsychoPy 3* (Python 3.6) to control the stimulus presentation and response collection (Peirce et al., 2019). In the fMRI scanning sessions the stimuli were back-projected onto an 18-inch screen placed in the bore of the magnet just behind the participant's head. The projector had a resolution of 1920:1080 pixels with a 60 Hz refresh rate. Viewing of the display was enabled via a mirror placed on top of the head coil. For navigation participants used a MR-compatible Fiber Optic Mouse with two standard buttons (left and right; FOM-2B-10B fMRI Mouse System, Nata Technologies Inc.). The mouse was placed on a 1800dpi Mouse pad (8.5 x 8.5 inch) that rested on the participant's upper body roughly above waist level. The stimuli consisted of geometrical squares or circles of the color blue and green. Training stimuli were red and yellow. A single stimulus subtended 0.59° visual angle. One stimulus type (e.g., all blue circles) was randomly assigned as the target type while the three remaining stimulus types served as distractors. A search display consistent of 40 target stimuli and 120 distractors. The spatial locations were randomly assigned on a spatial grid spanning a rectangle field of 12.9° * 14° visual angle. As the reward indication, we used an image of a 5 Euro Cent subtending 0.70° visual angle.

5.2.3 Procedure

The experimental task was based on a visual search paradigm. Figure 13 shows a trial sequence in detail. Participants were required to search for multiple targets randomly located in the search display among distractor stimuli. Since the target object was defined by color and shape, subjects performed a feature conjunct search. At the beginning of the experimental session, after the training, they were shown the target item they had to look out for in the subsequent search (i.e., all blue circles occurring next to blue squares, green circles, and green squares). Participants used a MR-compatible PC-mouse to search through the display. To obtain a reward, they merely navigated the mouse point to a target. Once a target had been fixated for 300 ms, the target turned into a reward indicator (i.e., 1 Euro Cent) and then returned to

its previous appearance. This served as the feedback that the target had been "foraged", and a reward had been received. The participants could then continue the search for the next target in the display. Each target item could be collected only once and it was deactivated following its detection so that a second fixation of the very same target would remain unrewarded. With each collected target item, an additional number of targets was deactivated too. These targets remained in the display but would not result in a reward following a fixation. The number of these additional deactivations was determined by an exponential decay function mimicking a quickly depleting food source (see upper box in Figure 13). In addition, the entire spatial configuration of target and distractor locations changed randomly following a reward capture. This hindered participants from building any kind of search memory for old target locations and the search became increasingly difficult and quickly inefficient (Horowitz & Thornton, 2008). To compensate for this, participants could choose to switch to a new display by pressing the right mouse button at any time (i.e., patch leaving). The countdown did not pause when volunteers switched to the next display. The time needed to move to the next patch (i.e., display), known as the 'travel time', was set to a minimum of 2 s (central fixation cross for 1 s followed by a 1 s blank). Because of the time required to save data and calculate the next display, measured travel time between the last collection in a patch and the appearance of the next patch averaged approximately 5 s. We did not manipulate travel time as a dependent variable and kept it short to support exploratory foraging choices (see Wolfe, 2013). In the control condition the search in the current display ended automatically after two to four reward earnings. Each new trial had a 25% probability to end automatically. At the display's left bottom corner, the participants were able to constantly track the total number of rewards they had already earned. At the bottom's right corner the countdown timer was displayed allowing participants to track their remaining search time. All participants performed six runs of a 10 minute search. Between runs they were able to take short breaks.

5.2.4 Design

As we aimed to yield a high number of trials in which volunteers performed self-initiated patch leaving behavior we introduced the exponential decay of the reward probability adopted from Lottem et al. (2018). It ensured that the number of remaining rewards rapidly decreased within a short amount of time. Due to the time constraint, participants should decide to switch to a new display instead of wasting time in the current display in which a further reward capture quickly becomes very unlikely. We also varied the initial reward probabilities by applying three conditions from high (100% reward probability, i.e., all targets were active), middle (75%) to low probabilities (50%). The initial reward probability was not cued and changed randomly from trial to trial creating a foraging environment with an unpredictable underlying reward structure consistent with the condition of restless 'n'-armed bandit task (e.g., Daw et al., 2006). We used following decay function that was used to determine the number of remaining targets after each new reward capture (Lottem et al., 2018):

$$P(O_{\mathrm{n}} = 1|t_{\mathrm{i}}) = A_{\mathrm{i}} e^{(-(\mathrm{n}-1))/5}$$

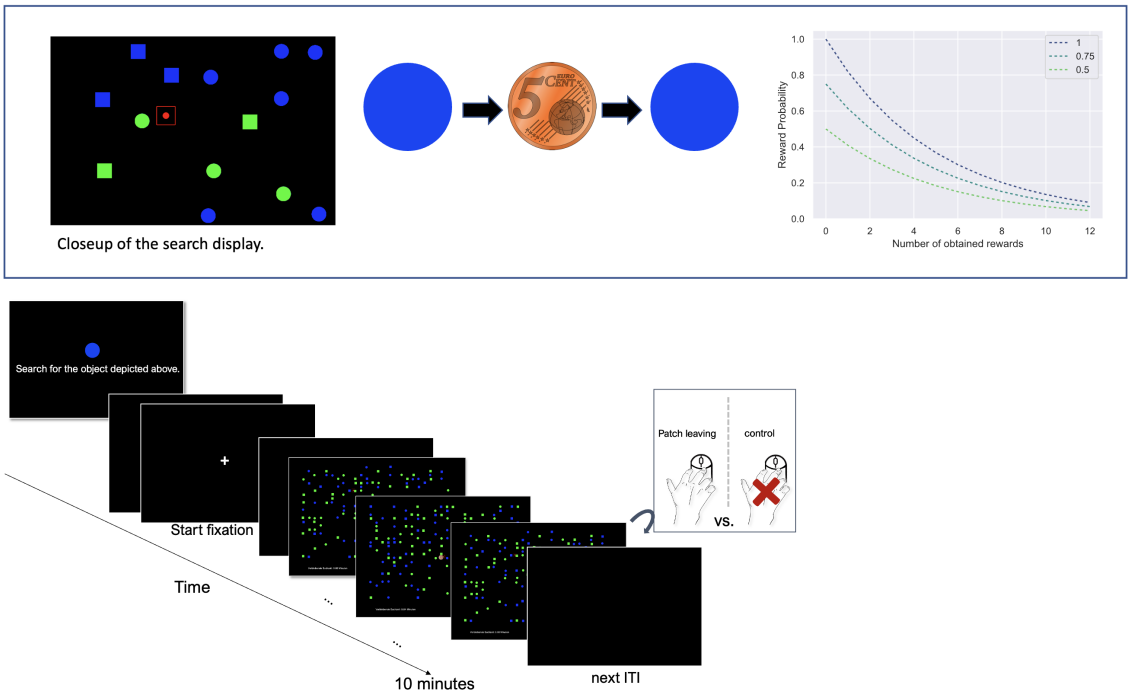


Figure 13: Diagonal sequence represents a trial. At the beginning of a session, the target object (here a blue circle) was introduced. The beginning of a trial was cued by a central fixation (1 s) followed by a blank (1 s). Next, the search display (i.e., a patch) appeared. By navigating the mouse cursor to a target (red square with red dot at center in lower left), participants yielded a reward capture. Upon such a capture, the collected target turned into a reward for 500 ms (5 euro-cent image) and then changed back into its previous appearance. An already collected target would not turn into a reward again if fixated again. With each reward capture, all items changed positions randomly. At any time, participants were able to switch to a new display by button-press. In the control condition, these cancellations occurred automatically after a random number of targets had been captured. The graph in the box in the figure's top shows the exponential decay function of the reward probability for the three patch qualities.

here t_i is the ith trial type, i.e., low-, medium-, and high-quality trials. These trial types had different exponential scaling factors A1 = 0.5, A2 = 0.75, A3 = 1. N indicates the number of already yielded reward captures (previous target fixations that resulted in an earning) within a trial. On is the positive outcome of the nth target fixation (1 for reward). The decaying reward probabilities as a function of target captures are shown in Figure 13 for all three patch qualities.

We also varied the initial reward probabilities by applying three conditions from high (100% reward probability, i.e., all targets were active), middle (75%) to low probabilities (50%). The initial reward probability was not cued and changed randomly from trial to trial creating a foraging environment with an unpredictable underlying reward structure. Otherwise, in the case of constant and predictable reward probabilities, the most optimal patch-leaving rule would be to capture a fixed number of

rewards, e.g., 3 rewards, and leave the patch once that number is reached (Wilke et al., 2009). Now, if this number happens to be rather small due to the quick reward depletion, it is possible that volunteers using this patch-leave rule would leave a given display even before the automated switches would occur resulting in a decreased number of control trials.

5.2.5 Data Analysis

5.2.6 Behavioral Data

Custom-written code in Python (version 3.6) was used to analyze participants' behavioral data. The behavioral data obtained in the control condition (automated search cancellations) were excluded from the behavioral analysis as we were interested in analyzing the self-initiated patch-leaving behavior. Single factor repeated-measures ANOVAs with patch quality (i.e., the start reward probability) as the single repeated-measures factor and two-sided post-hoc t-tests with Tukey HSD correction for multiple comparisons using the Python package *pingouin* (Vallat, 2018) were used to test for differences between the patch qualities. QQ-Plots served for screening for violations of normality. In case of violations non-parametric alternatives were used for significance testing. We applied Greenhouse-Geiser corrections in case of non-sphericity. Linear regressions were used for within-subject analyses using the *linregress* function of python's *stats* library (Version 1.10.1) In all analyses, the averaged individual medians were used with no outlier correction to report descriptive statistics as well as for statistical testing.

5.2.7 FMRI Data

Image acquisition All parameters for image acquisition are reported in detail in Section 2.3.

Pre-processing A single scanning session was split into ten runs of 600 s each and 300 volumes were sampled per run. The imaging data was pre-processed and analyzed by means of tools of the FSL package (Jenkinson et al., 2012). The anatomical scans underwent a non-brain removal with BET (Brain Extraction tool), (Smith, 2002) in preparation for the realignment. The functional images were motion-corrected to an image in the middle of each run with a normalized correlation ratio (MCFLIRT; FMRIB's Linear Image Registration Tool), (Jenkinson & Smith, 2001; Jenkinson et al., 2002) and slice time corrected (temporally aligned to the middle slice of the 3D volume). To ensure the validity of Gaussian random field theory, the functional data was spatially smoothed using a Gaussian kernel with a size matching the double of the voxel dimensions (FWHM = 6 mm). To remove low-frequency drifts (Smith et al., 1999), we temporally filtered the data using a highpass filter with a cutoff value of 90 sec. Prior to the mass-univariate analysis, run 1, 2, and 3 as well as run 4, 5, and 6 were merged into two epochs 1 and 2 using *fslmerge*. This was done to ensure sufficient numbers of experimental and control trials for statistical analysis and was carried out after motion and slice time correction.

GLM analysis For statistical analyses of the functional brain scans, we defined the onsets of the experimental events as explaining variables (EV) to model the BOLD response by means of a general linear model. To be able to test for BOLD-signal changes specific to patch-leaving decisions, we added the onsets of the end of search for patch-leaving and control trials (i.e., the onsets when the current display was aborted) as the first two regressors to the linear model. To contrast BOLD-signal changes associated with patch-leaving with search-related activity, the onset of all target detections that occurred during the search phase of patch-leaving and control trials were also added to the model. For convolution we used a hemodynamic response function (double gamma HRF). To all six regressors a temporal derivative allowing for temporal flexibility was added. Serial voxel-wise autocorrelations were controlled with prewhitening by the FSL tool FILM (Monti, 2011; Woolrich, et al., 2001).

In the first-level analysis, all contrasts of interest were tested for significance under mixed-effect assumptions and contrast images were processed for each participant (voxelwise Z threshold of 2.3 and a cluster significance threshold of p = .05, family-wise error (FWE) corrected). Next, the resulting images containing the lower-level contrast of parameter estimates (COPEs) were used in the second-level analysis estimating individual mean contrasts for the parameters across all runs using a fixed-effect model with the same voxelwise and cluster significance threshold (Z = 2.3, p_{FWE} = .05). In a two-step post-statistical normalization, prior to group analysis, the functional data was firstly co-registered to the individual, anatomical scan using boundary-based registration (BBR), and secondly normalized to the Montreal Neurological Institute standard space (MNI 152 2mm). Thus, statistical modeling on the subject level was carried out in native space. The statistical modeling at the group level was performed using FLAME 1+2 (FMRIB's Local Analysis of Mixed Effects) as implemented in FSL's FEAT (Version 6.00). Results are given by means of whole-brain statistical Z-maps on group-level thresholded at $Z > 2.3$ and corrected for multiple comparisons on a cluster level at $p_{\text{FWE}} < .01$ (Worsley, 2001).

5.3 Behavioral results

5.3.1 Participants adapted to changing patch qualities

Figure 14 a) shows dotplots of the number of obtained rewards as a function of patch quality. The single factor repeated measure ANOVA revealed a main effect of *patch quality*, $F(2,38) = 220.90$, $p < .001$. As expected, the most rewards were obtained in the high quality patches with an average of 52 ± 13 cents. Compared to this participants earned on average 28 ± 4 cents less in low quality patches, $t(18) = -8.185$, $p_{\text{Tuckey}} = .001$, and 13 ± 4 cents less in medium quality patches, $t(18) = -3.538$, $p_{\text{Tuckey}} = .003$. Earnings in low quality patches were also on average 15 ± 3 cents less compared to medium quality patches, $t(18) = -4.648$, $p_{\text{Tuckey}} = .001$. The same pattern of results we observed for the residence times (i.e, search times spent per display). The higher the patch quality, the longer the participants spent foraging in it, Friedman $F(1.9,36.1) = 18.811$, $p < .001$, with longest average residence times of 48.364 ± 18.487 s in the high quality patches, followed by 42.623 ± 16.592 s in medium-quality, and

34.990 ± 18.768 in low-quality patches. Performing post-hoc contrasts with p-value correction (Tukey) showed that residence times in low-quality patches were significantly shorter compared to medium-, and high-quality patches, Nemenyi post-hoc, p = .019, and p = 001. Residence times as a function of patch quality are shown in Figure 14 b). Analyzing participants' residence times showed considerable variations across subjects. Given the underlying reward structure of the task, spending too much time searching in the current display instead of switching to a new one should be detrimental to overall earnings. Residence times averaged across patch types indeed correlated negatively with the averaged total earnings in Euros (€) demonstrating that the more participants tended to overexploit the current display, the less earnings they made in the total search time of 60 minutes, r_{pearson} = -0.653, p = .002 (see Figure 14 c). This confirmed that optimal performance under the given task depended on a readiness to switch displays quickly.

5.3.2 Reward captures increased residence times incrementally

To test what could have driven participants to extend their residence times, we next regressed individuals' residence times on target captures to model participants' foraging strategies (Hutchinson, et al., 2008; Mata et al., 2009; Wilke, 2006). This way we obtained a slope and intercept for each participant, where the intercept represented the initial time spent in the current display without a reward detection while the slope indexed the increase in the residence time with each new reward capture. All slopes were positive and significantly above zero on group level, mean slope = 3.533 ± 0.760, $t(19)$ = 20.252, $p < .001$, indicating that participants incrementally extended their stay with each new target detection (see Figure 14 d)).

5.3.3 Prolonged giving-up times negatively impacted search performance

Considerable variation in participants' GUTs suggested that the participants differed in their behavioral tendency to either explore or exploit. Similar to the residence times, GUTs were negatively correlated with the task performance, measured by monetary earnings, r_{Pearson} = -0.676, p = .001, (Figure 15 a)). The GUT can serve as a rather simple heuristic for participants to time their patch-leaving decisions. According to an optimal GUT rule (Mcnair, 1982), GUTs should be adjusted depending on patch quality with longer GUTs in patches of higher quality. We, however, observed the opposite: the average GUTs were 7.927 ± 3.250 s in high-quality patches, 8.531 ± 3.660 s in medium-quality patches, and 9.062 ± 4.025 s in low-quality patches, Friedman $F(1.9 , 36.1)$ = 6.083, p = .006 (see Figure 15 b)). Post-hoc contrasts confirmed that GUTs increased with decreasing patch quality at least between high- and medium-quality patches, p = .007, [high - low, p = .067, medium-low, p = .691]. Participants thus appeared to adjust their GUTs but in a non-optimal fashion. Plotting GUTs against the residence times (Figure 15 c)) also showed that GUTs were positively correlated with the residence times, r_{Pearson} = 0.63, p = .002. This finding, we followed up by regressing the GUTs averaged across patch qualities on the averaged residence times on a trial-by-trial basis within each subject. The resulting intercepts indexed the minimum GUT individuals accepted and the slopes represented the increment added to the GUT with each second more spent in the given display. We

observed positive slopes for all participants [$t(19) = 5.656$, $p < .001$] suggesting that the longer participants had already searched in the current display, the more likely they to invest even more time consistent with a sunk-cost effect (see Figure 15 c)).

5.3.4 Still optimal timing for patch-leaving on group level

Lastly, we analyzed participants' collection rate, that is, at what speed target detections occurred. The marginal value theorem (MVT) states that optimal foragers time their patch-leaving to the moment at which the instantaneous collection rate (ICR) approximates the average collection rate of the entire environment. The ICRs as a function of number of reward earnings for the last seven rewards for each patch quality are depicted in Figure 15 d). They are given by the inverse of the time that passed between two consecutive target captures. To test the MVT prediction, we estimated the ICR at the moment of patch leaving by dividing 1 by the GUT instead (Mcnair, 1982). To obtain individuals' MCRs, we divided the total number of rewards received across the entire experimental session by the 60 minutes of total search time (see Wolfe, 2013). Note that this way the MCR also included the travel time during which the collection rate equals zero. Consistent with the MVT, Wilcoxon signed-ranks tests showed that ICRs, i.e., reward earnings per second, of $0.158 \pm 0.1/s$ (SD) in high-quality patches did not differ statistically from the average rate of $0.164 \pm 0.02/s$, $Z = 63.0$, $p = .126$. The same was true for medium quality patches, ICR = $0.152 \pm 0.09/s$, $Z = 70.0$, $p = .202$, and in low-quality patches, ICR = $0.145 \pm 0.09/s$, $Z = 62.0$, $p = .114$.

In summary, the incremental relationship between individual reward captures and residence times suggested that single reward events were read as cues indicating that it was worthwhile to continue searching the current display. Thus, subjects adapted their foraging to the changing reward probabilities and spent more time foraging in high-quality patches. Residence times as well as GUTs varied greatly across subjects implying that they differed in their behavioral tendency to explore. In contradiction with an optimal GUT rule, the participants tended to prolong their GUTs with decreasing patch quality and depending on sunk costs. Despite this, patch departures were still optimally timed according to the MVT.

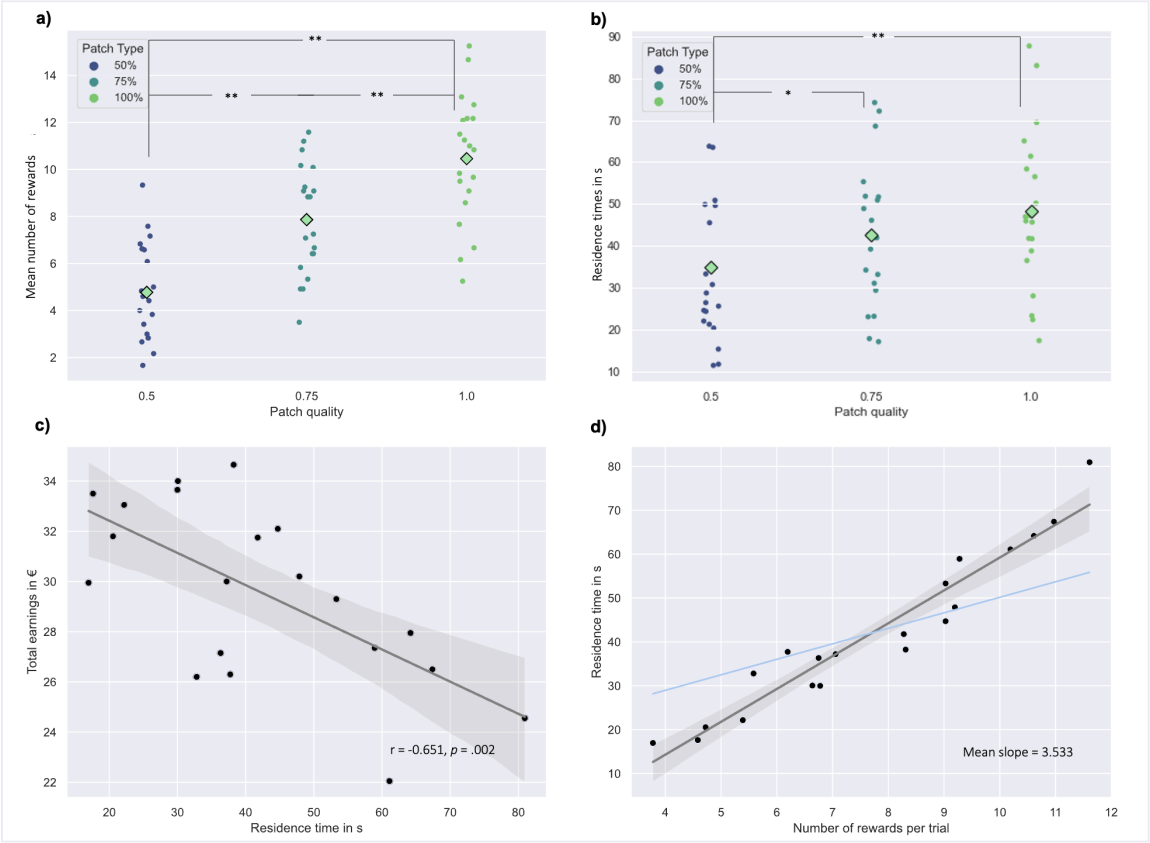


Figure 14: The Average (diamonds) number of rewards (a)), and average trial durations (b)) are shown together with subjects' individual means (small dots) as a function of patch quality. The number of rewards increased significantly with increasing patch quality. Trial durations indicated the same trend but only low- and high-quality patches differed significantly. In c) the regression plot shows the relation between the total earnings in € participants achieved in the entire experiment and residence times, both averaged across all patch-qualities. The negative correlation showed that participants who decided to explore a new display more quickly made also more earnings. d) Shows the relationship between the number of obtained rewards per trial and residence times. Within-subject regressions showed that the residence time was extended incrementally with each novel reward capture. The blue line shows the averaged regression line of the within-subject regressions, based on the mean intercept (residence time) and the mean slope, that is, the amount of time by which the residence time increased with each new reward capture.

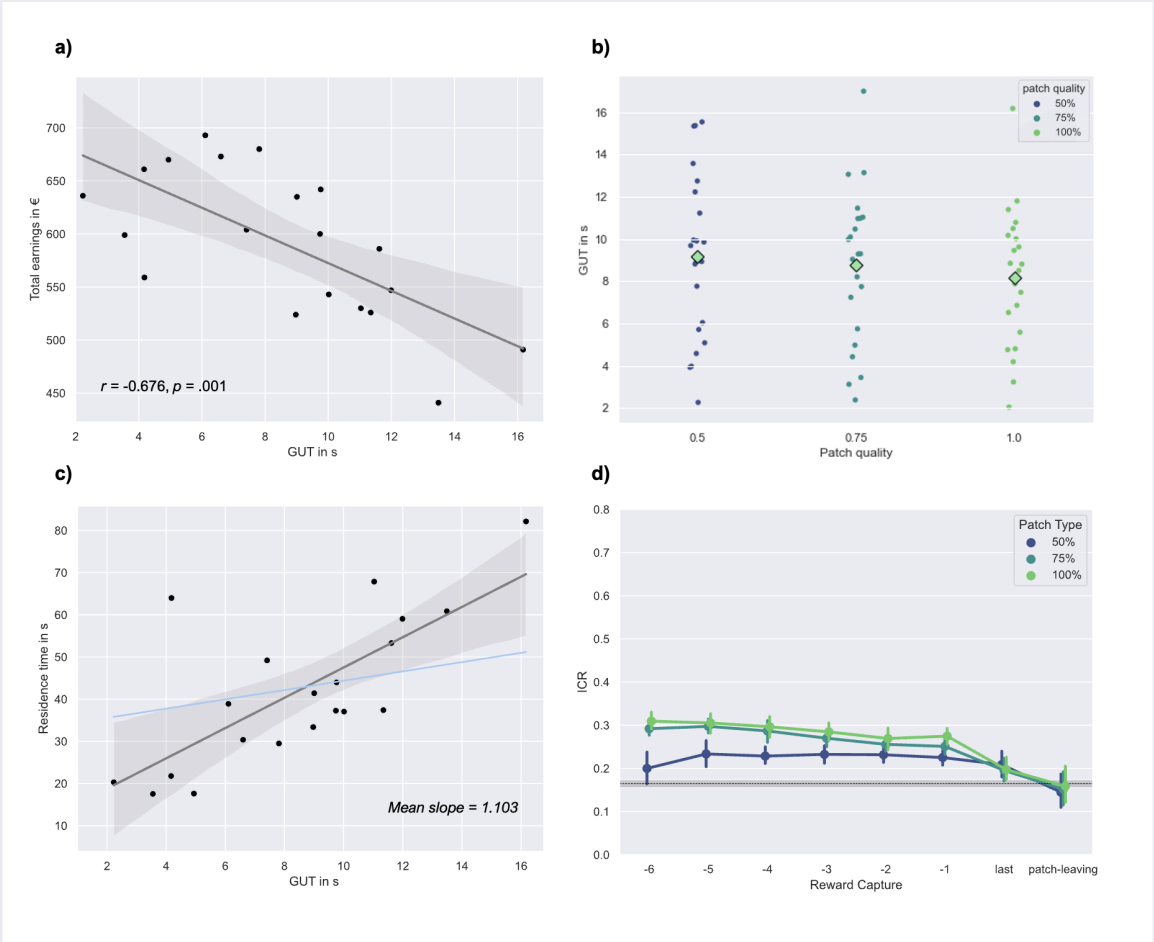


Figure 15: a) shows the negative correlation between GUTs on the total earnings. Participants with the longest GUTs made on average the least earnings. b) shows the average (diamonds) GUTs together with subjects' individual mean GUTs (small dots) as a function of patch quality. c) shows the positive relation between residence times and GUTs indicating a sunk-cost effect. The fitted blue line is the regression line derived from the mean intercept (minimum GUT) and the mean slope (residence time) that was obtained from the within-subject regressions. The slope represents the increment in time added to the GUT with each second more spent in the current display. d) The instantaneous collection rates (ICR, i.e., reward earnings / s) are plotted for the seventh last ('-6' on the x-axis) up to the time point of patch-leaving for the three patch types, where the estimated ICR was given by the inverse of participants' GUTs. Consistent with optimal foraging according to the MVT, the ICRs decreased as a function of target captures and approximated the mean collection rate (MCR, dashed black line, shaded area in gray indexes the 95% confidence interval) in all three patch types.

5.4 FMRI results

5.4.1 Positive BOLD-signal changes associated with patch leaving decisions

To test for brain regions particularly responsive to self-initiated switches between displays (i.e., patch-leaving), we contrasted these trials against those in which the search in the current display was automatically canceled (*patch-leaving - control*). This resulted in a single large cluster with 4853 voxels located in the medial frontal cortex, spanning bilaterally from the supplementary motor cortex (SMC) over the precentral to the postcentral gyrus. Left-laterally in the rostral direction the cluster extended ventrally to the anterior division of the cingulate cortex (dACC). Dorsally it reached up to the posterior parts of the SFG. Posteriorly, also in the left hemisphere, it included parts of the anterior division of the SMG. A second small cluster containing 760 voxel was located in the right temporo-occipital fusiform cortex (TOFC) reaching ventrally into the right cerebellum. The reverse contrast (*control - patch-leaving*) did not yield any significant results. The statistical z-map mapped on an MNI152 template are shown in Figure 16 a). Voxel locations of maximum z-values are listed in Table 10 a).

Next, we contrasted the onsets at the moment of patch-leaving against the onsets of the target detections of both patch-leaving and control trials to identify brain areas being more responsive during patch-leaving compared to visual search (*patch-leaving - target detections*). The average time between the last target detection and the moment of patch-leaving was 7.5 s. This relatively large time window between the onsets of these regressors of interest, and the random temporal variations in their onsets precluded concerns about collinearity between patch leaving decisions and those target detections that had preceded the patch-leaving decision in the same trial. Also the stimulus onset asynchrony (SOA) between two to target detections was on average about 3 s so that the regressors were sufficiently spaced in time. This contrast revealed three clusters in the frontal, parietal, and occipital cortex. Located in the prefrontal cortex, the smallest cluster of the three with 1474 voxels bordered dorsally into the right lateral FPC. Ventrally, the cluster extended bilaterally along the paracingulate gyrus (PCG) down to the dACC, i.e., the rostral cingulate zone (RCZ). More posteriorly another cluster with a size of 1558 voxels was located predominantly in the left parietal cortex. Its anterior borders resided in the left precentral gyrus (PreCG). From there it spanned posteriorly over AG, to the SMG where it reached ventrally to the temporo-occipital part of the left middle temporal gyrus (MTG). Posteriorly from the left SMG, it also extended to the superior division of the left LOC. In addition, it also included voxels located in the left anterior insular and the operculum. The largest cluster comprising 3603 voxels spanned parts of the posterior parietal cortex up to the anterior parts of the occipital cortex. It reached bilaterally from the precuneus (Pcu) to the cuneus where it extended ventrally to the lingual gyrus (LG), also in both hemispheres. Neither the inverse contrast *(control - detections)* nor the interactions *(patch-leaving - detection) - (control - detection)* or *(control - detection) - (patch-leaving - detections)* yielded significant activation clusters. The statistical Z-map for *patch-leaving - detections* mapped on an MNI152 template is shown in Figure 16 b). Individuals' parameter estimates retrieved from significant

activation clusters are reported in Figure 17 a) for both contrasts. Again, the beta weights indicated on average a positive signal change similar to the beta weights that resulted from the patch-leaving - control contrast.

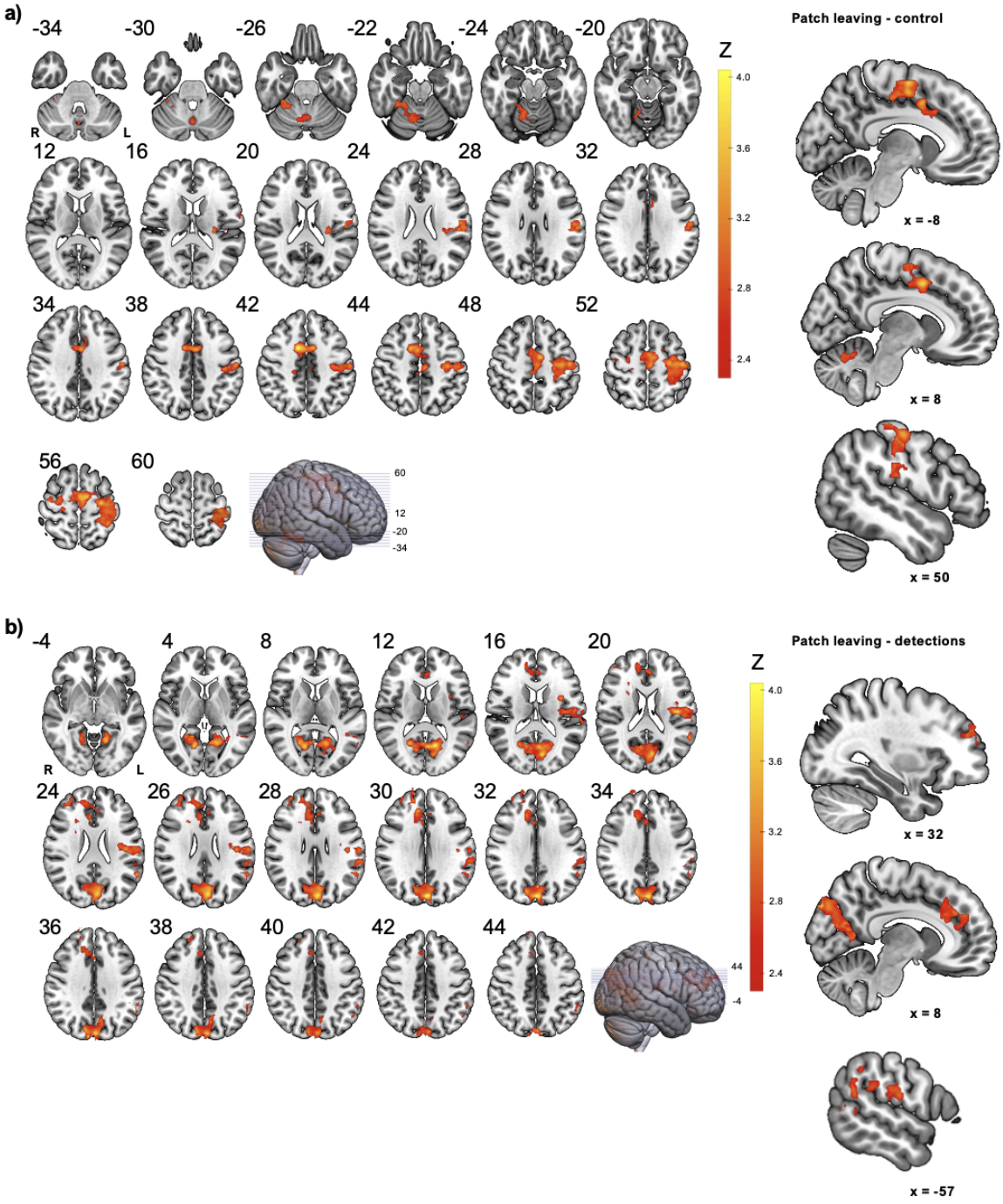


Figure 16: FMRI activation maps show clusters of activation with significance on cluster level $p < 0.01$, cluster forming threshold $z > 2.3$, left). a) shows clusters of activation revealed by the patch-leaving - control contrast. b) Shows activations clusters that resulted from *patch-leaving - detections*.

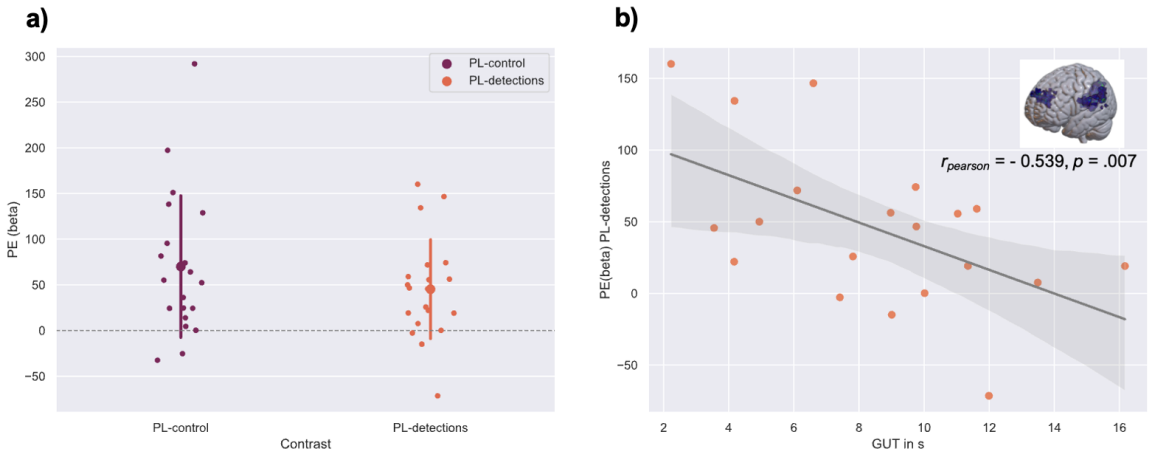


Figure 17: a) shows the average standardized parameter estimates (PE, big dots, vertical bars indicating a range of ± SD) and individual parameters (small dots) that were derived from voxels showing significant signal changes in the contrasts *patch-leaving - control* (left), and *patch-leaving - detection* (right). b) shows the correlation plot between the average giving-up times (GUT) and beta estimates of patch leaving - detection. Participants with longer GUTs showed weaker activations in attention-related areas in the moment of patch-leaving.

Table 10: *Table of activations observed during patch-leaving*

#	K	Z Max Loc	Z Max	*p*	Structure
a) patch-leaving - control					
1	4853	12, 6, 44	4.5	1.28e-13	bil PreCG to PCG, dACC, lSFG, lSMG
2	760	30,-40,-26	3.32	.0027	TOFC, rCerebellum
b) patch-leaving - detections					
1	4075	-2, -70, 32	4.57	1.06e-09	bil Pcu, cuneus, LG
2	1598	-44,-16, 20	3.65	< .001	lPreCG, SMG, AG, AIC, OPC
3	1474	10, 34, 30	3.76	< .001	rlatFPC, dACC/RCZ, PCG

Note: K = cluster size in voxels, Z Max Loc = MNI coordinates of the location with the maximal Z-value, Z Max = maximal Z-value, bil = bilateral, r = right, l = left, lat= lateral, structure determined using the Harvard–Oxford Cortical Structural atlas.

5.4.2 Brain signals of exploration correlate negatively with behavioral giving-up times

In keeping with our expectations we found activations, time-locked to self-initiated patch-leaving decisions, in the dACC, in parts of the dorsal (AIC and OPC) as well as in parts of the ventral attention network (left TPJ). While one cluster extended with a few voxels to the right lateral FPC (see Figure 16 b), transverse slice z = 16 and sagittal slice x = 32), the activation in this area during exploratory decision-making was less pronounced than anticipated. Yet, intriguingly, beta weights derived from the *patch-leaving - detections* contrast correlated, as expected, negatively with the participants' GUTs, r_{Pearson} = -0.539, $t(18)$ = 1.405, one-sided p = .007 (Figure 17 b)). This suggested that those participants with a higher behavioral tendency to exploit the current display showed weaker signaling in the fronto-parietal attention network during exploratory decision-making.

We followed this correlation analysis up by calculating a two-sample t-test on the *patch-leaving - detections* contrast between those subjects with GUTs below the median GUT (8.9 s) (short-GUT) and those subjects with GUTs above the median (long-GUT). Testing short-GUT - long-GUT yielded cluster activations in brain regions responding more strongly during exploration in the short-GUT group compared to the long-GUT group. Z-maps of cluster activation of this analysis are shown in Figure 18. The analysis resulted in three clusters of which the largest with 2985 voxels was located in the anterior prefrontal cortex with its maximum z-scores located in the medial FPC. It spanned bilaterally over both the medial and lateral FPC. Its ventral borders tapped into parts of ACC, while its center mostly spanned from FPC over the PCG dorsal-posteriorly to the SFG. The second cluster comprising 1646 voxels was located in the parietal cortex with its center spanning bilaterally along the IPS. Ventrally it reached into the posterior division of the cingulate cortex. Dorsally with significant contrast activation were also located in the right SPL. The third cluster with a size of 902 voxels spanned parts of the left SMG and AG containing the left TPJ, matching locations previously detected in the within-subject analysis (see Figure 17 b). The findings are summarized in Table 11. Calculating the same two-sample t-test on the patch-leaving - control contrast did not yield any results.

If the SOAs between targets varied significantly between the short- and the long-GUT group, it is possible that successive BOLD responses for single target detections could be more or less overlapping. A higher degree of overlap due to shorter SOAs in the short-GUT group could increase the activity strength in the moment of patch-leaving, thus confounding the between-group effect. In the low-quality patches SOAs of long-GUT subjects were on average 6.196 ± 1.002 s and 5.442 ± 1.036 s in short-GUT subjects, with anecdotal evidence comparable SOA lengths between the two groups, Bayes Factor 10 (BF_{10}) = 0.733. In medium-quality patches SOAs were on average 4.881 ± 0.509 s in long- and 4.297 ± 0.656 s in short-GUT subjects with anecdotal evidence for equal SOAs, BF_{10} = 0.919. Only in high-quality patches the analysis provided anecdotal evidence for the longer SOAs in long- [4.383 ± 0.492 s] compared to short-GUT subjects [3.770 ± 0.629 s], BF_{10} =

1.239. Given these results, it is rather unlikely that the observed group-difference in the BOLD signals during patch-leaving was driven by a difference in the degree of overlap of the BOLD responses building up in the search period.

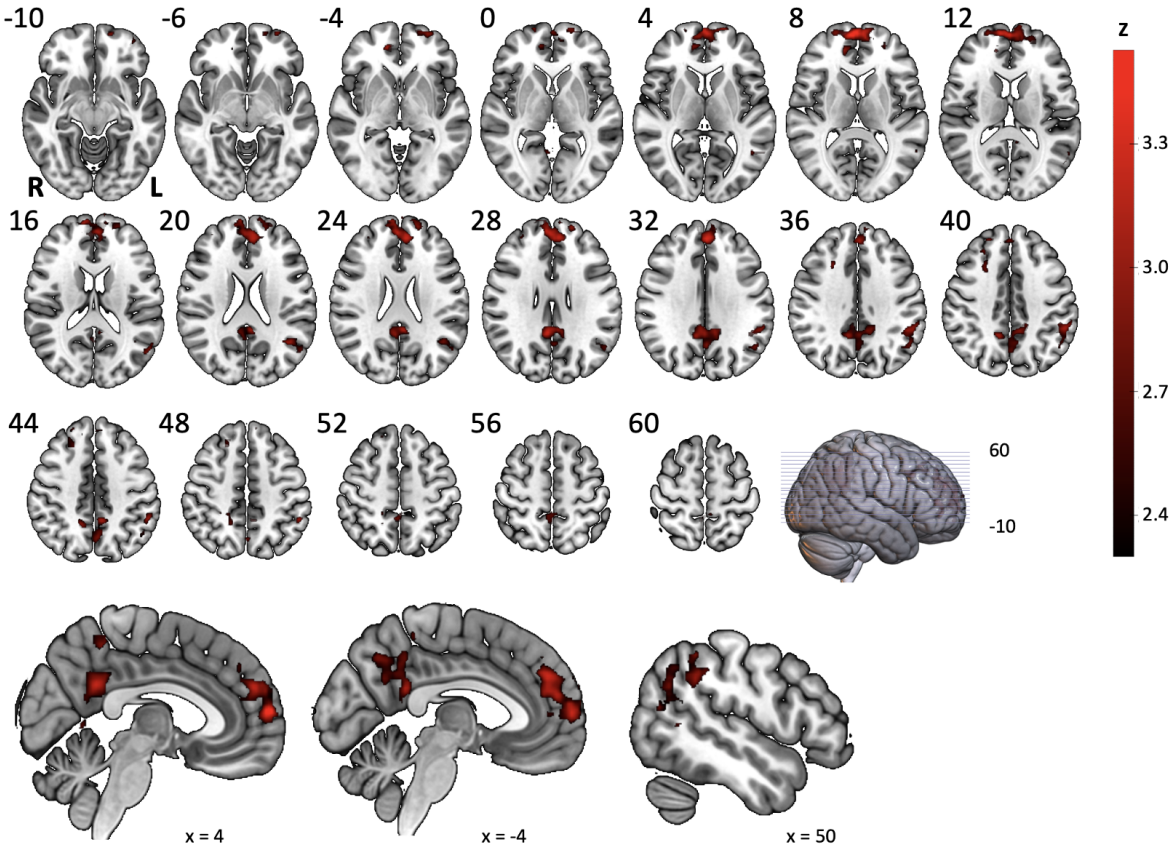


Figure 18: Z-maps resulting from the follow-up two-sample t-test. The clusters indicated in which regions participants in the short GUT group responded showed stronger activation during patch leaving than during visual search compared to the long GUT group of participants. Sagittal slices at the bottom show MNI152 coordinates.

Table 11: *Table of activations resulted from the two-sample between group t-test on patch-leaving - detections*

short-GUT - long-GUT x patch-leaving - detection

#	K	Z Max Loc	Z Max	*p*	Structure
1	2985	4, 62, 12	3.55	5.96e-08	bil medFPC, latFPC, PCG, rSFG
2	1646	4,-48, 28	3.26	3.65e-05	bil IPS, PCC,rSPL
3	902	-54,-44, 42	3.1	0.004	lAG, lSMG (lTPJ)

Note: K = cluster size in voxels, Z Max Loc = MNI coordinates of the location with the maximal Z-value, Z Max = maximal Z-value, bil = bilateral, r = right, l = left, lat = lateral, med = medial, structure determined using the Harvard–Oxford Cortical Structural atlas.

5.5 Discussion

We investigated the neural correlates of exploratory patch-leaving decisions during foraging in a human. To this end we designed a novel visual search paradigm that allowed us to study visual foraging behavior of human participants inside the MR scanner. Of particular interest were the time points in which the participants performed self-initiated switches to a new search display. In line with our predictions, the analysis of participants' BOLD signal time-locked to exploratory foraging choices (patch-leaving) revealed brain clusters with positive signal changes across both the dorsal and ventral attention networks and reached up to the right lateral FPC, resonating with findings from previous bandit-like gambling tasks (e.g., Daw et al., 2006). Remarkably, the strength of the positive signal change, time-locked to patch-leaving, within these areas correlated negatively with the participants' behavioral exploration bias. A follow-up analysis provided further insights showing that the BOLD-signal in medial and lateral FPC, SFG, TPJ, and Pcu was more strongly modulated during patch-leaving in participants with below median GUTs compared to participants with above median GUTs.

5.5.1 Relating behavioral signs of exploitation to signals of exploration

Our key aim was to identify possible links between participants' internal rules guiding patch-leaving decisions and brain activity. In keeping with previous gambling studies (e.g., Chakroun et al., 2020; Daw et al., 2006; Laureiro-Martínez et al., 2015), we could show that the right lateral FPC together with dACC and left TPJ was active when participants made patch-leaving decisions during foraging in a visual search task. The signal strength in the right lateral FPC reportedly correlates

positively with attentional control and decision-making during gambling (Daw et al., 2006; Laureiro-Martínez et al., 2015; Zajkowski et al., 2017) and is known to serve shifts of visual attention in non-foraging tasks (e.g., Pollmann & Manginelli, 2009a).

Extending this, we provide new evidence that participants who differ in their readiness to explore, indexed by different giving-up time durations, can also be distinguished by the degree to which lateral FPC, among other regions of the fronto-parietal attention network, is active at the moment participants decide to leave the patch. It was expected that participants with a higher readiness to make exploratory decisions also showed stronger FPC activations during patch leaving, because this region is known to be central for exploration, specifically the assessment of the potential value of alternative strategies and for the consequent re-distribution of cognitive resources (Badre et al., 2012; Boorman et al., 2009; Boorman et al., 2011; Cavanagh et al., 2012; Mansouri et al., 2017). Thus, heightened FPC activation may play a central role in facilitating the decision to leave a current patch. Thus, FPC may serve as a central hub in redistributing attentional resources by evaluating the potential benefits of staying versus leaving a patch (Mansouri et al., 2017). It may be that individuals with stronger FPC activation at the moment of patch-leaving have enhanced cognitive flexibility, allowing them to quickly weigh competing values of the current behavioral choice versus foregone alternatives and make more efficient decisions about exploration. This interpretation finds support from studies in elderly who show a decline in cognitive flexibility (e.g., Giller & Beste, 2019) as well as prolonged GUTs during foraging (Mata et al., 2013). Other functional neuroimaging studies have demonstrated age-related changes in activation patterns within the prefrontal cortex during working memory and attention tasks, often with less activity in elderly compared with young adults (Milham et al., 2002; Reuter-Lorenz, 2002). These age-related changes in brain activity are associated with changes in the dopaminergic system (Berry et al., 2016) and anatomical changes of the prefrontal surface area and thickness (Dotson et al., 2016). Similarly, it may be that already young adults show subtle neurofunctional or even anatomical differences resulting in a stronger or weakened activity in the brain network supporting attention and cognitive control. Clearly, our finding strongly suggests that FPC is involved in regulating the trade-off between exploitation and exploration. Individuals with higher FPC activation may have a greater inclination towards exploration, leading to shorter GUTs times. On the other hand, it may be that participants with prolonged GUTs merely adopted a suboptimal exploitative foraging strategy which could have led to a suppression of brain regions related to attentional exploration. By instructing a specific patch-leaving rule (Wilke et al., 2009), it is possible that their behavior as well as the signaling in the FPC may change accordingly. More research will be needed to test these hypotheses. Yet another explanation could be that participants differ in their sensitivity to prediction errors (Chowdhury et al., 2013): subjects may differ in how strongly they judge accumulating reward omissions so that its decreasing effect on the value of the current behavioral choice may be more or less strong. If reward omissions are judged less negatively, the value attributed to the current behavioral choice (i.e., exploitation) would then remain relatively high, while the assessment of the potential value of alternative choices (in our case exploring a new display), on the other hand, would be relatively suppressed. This could explain

later patch-leaving decisions and decrease brain signals of attentional control in the moment of patch-leaving. It will be highly interesting to focus more on reward omission during the giving-up time phase and further examine the dopaminergic system using PET imaging.

5.5.2 Exploratory foraging choices rely on an interplay of different neural correlates

The decision to switch to a new display in our experiment certainly entailed a computational process in which the value of forgone options (i.e., exploring a new display), likely represented by the FPC (Boorman et al., 2009; see Mansouri et al., 2017), must be compared to and contrasted against the value of the current choice, supposedly represented and updated during performance monitoring executed by the dACC (e.g., Daw et al., 2006; Ridderinkhof et al., 2004). Consistent with this, dACC was shown to encode both the average value of the foraging environment and the costs of foraging (Kolling et al., 2012). The activity we observed in dACC during patch-leaving decision may also be related to the evaluation of the unrewarded target fixations preceding the actual patch-leaving that may have led to a prediction error due to increasing number of omissions of expected rewards in the giving-up time phase (Rushworth & Behrens, 2008; Silvetti et al., 2011). This notion is bolstered by evidence showing that neural firing rates in dACC of monkeys ramp up with the gradually depleting reward up to the point of patch-leaving (Hayden et al., 2011). Co-activation together with the dACC we also observed in the left AIC which corresponds with other studies reporting co-activity of these areas arose such as during conflict or pain processing (Craig, 2009; Medford & Critchley, 2010).

Patch-leaving not only relies on reallocating attentional resources, but also requires disengaging from the existing attentional focus. Thus, in line with our predictions, we also found positive signal changes in the left TPJ: suppressed during visual search (e.g., Shulman et al., 2003; Shulman et al., 2007), the region is thought to be as a "circuit breaker" supporting the disengagement of attention from the current focus of attention, e.g., in response to an unforeseen and surprising event (Corbetta et al., 2008). Thus, the activity in the left TPJ during patch-leaving decisions may have served the disengagement from the current display which is in keeping with the idea that that the ventral attention network sends such reorienting signals to the dorsal attention network (Corbetta et al., 2008; Japee et al., 2015). Yet the largest cluster that we observed to be active during patch-leaving decisions was bilaterally in the Pcu extending into the occipital cortex. We did not have particular hypotheses about Pcu activity during exploratory foraging choices but it may be that the activity we observed was related to attention shifting. A transient increase of neural activity in the Pcu time locked with attentional shifting was reported in a study in which subjects performed a card sorting task (Nagahama et al., 1999). More recently, Li et al. (2020) showed a decrease in the functional connectivity between the right Pcu and MFG related to attentional decline in response to acute sleep deprivation.

5.5.3 Behavioral mechanisms driving patch-leaving decisions

The large variation in GUTs showed that subjects differed in their GUT threshold with longer GUTs indexing a behavioral bias towards exploitation (e.g., Mata et al., 2013) which had a negative impact on task performance. Although not optimal, the participants prolonged their GUTs with decreasing patch quality (see Mcnair, 1982), and extended them depending on the time they had already invested in the current display, consistent with a sunk cost effect (Arkes & Ayton, 1999). Such fallacy during patch-based foraging was reported before (Hutchinson et al., 2008; Redish et al., 2022; Sweis et al., 2018) and is reflected in the increased propensity of individuals to keep investing resources into a current endeavor that has already incurred significant costs even if that endeavor is unlikely to succeed or was already doomed to fail in the beginning. Marginal costs and benefits, but not past costs, should drive the decision-making, which is why the fallacy is maladaptive (Navarro & Fantino, 2005).

At the same time, participants' behavioral data showed that subjects increased their residence time in a current display incrementally in response to a new reward capture. This is consistent with the incremental rule of patch leaving. It assumes that once a patch is entered, the tendency to stay steadily declines with time. Each reward encounter, however, increments the current staying tendency, but if the decreasing tendency to stay crosses a lower threshold, the current patch is left (e.g., Wajnberg et al., 2000). Thus, residence times in poor-quality patches remain short compared to the time spent in patches of higher quality, consistent with our data. Moreover, the incremental rule is well suited in environments with an aggregate reward structure in which the reward structure is difficult to assess before entering the patch (Hutchinson et al., 2008). Such an environment would offer a few really good patches with very high but rare reward probabilities (i.e., most of the reward accumulates in a few patches). The argument would then be that a random encounter of a reward in the current patch could be used as a cue indicating that the forager may have hit one of the few rich patches, and, thus, should stay in it. In contrast, in a poor patch, the tendency to stay would decrease rapidly until the lower threshold for patch-leaving is met. This would result in longer residence times in high- compared to low-quality patches as well as to an incremental relationship between single rewards encounters and residence times consistent with our data. For our participants the rule worked well given the near optimal time points of patch-leaving according to the MVT (see Figure 15 d). Lottem et al. (2018) designed a probabilistic foraging task from which we copied the underlying reward structure. Intriguingly, their study also showed that foraging mice behaved in keeping with the incremental rule. The mice's tendency to stay also decreased as a function of time and was incremented again with each reward capture, creating a step-wise trajectory of the probability to reside in the current patch as a function of residence time similar to the behavior of our human participants.

5.5.4 Concluding remarks

Tying in with previous studies examining the neural substrates of exploratory choices in the exploration-exploitation dilemma, FPC along with regions of the

fronto-parietal attention network were implicated in behavioral exploration tested in a foraging paradigm. The association between behavioral giving-up times and frontopolar BOLD signal suggests that the FPC plays a central role in decision-making during attentional exploration. Individuals with higher FPC activation at the moment of patch-leaving exhibited shorter giving-up times, implying a stronger bias towards exploration and more efficient attentional control. The use of a foraging paradigm may enable comparisons with foraging tasks in other species.

5.5.5 Author contributions

LG, and SP designed the study. LG managed data collection. LG wrote all formal analysis scripts and analyzed the behavioral and fMRI data. LG wrote the manuscript. SP commented on the final version.

5.5.6 Code and data availability

All relevant code used for the stimulus presentation as well as for data analysis is available at https://github.com/LGparrot/exploratory-attention-in-visual-foraging.

6 Summary and general conclusions

In summary, this dissertation investigated the role of the frontopolar cortex (FPC) in attentional exploration. To this end, two different behavioral paradigms were conceived and tested. The main chapters cover one behavioral study (Chapter 3) and and two neuroimaging experiments, utilizing the BOLD response to elucidate the role of the FPC during shifts of visual attention in the absence of visual awareness (Chapter 4) as well as shifts towards an exploratory mode of cognitive control in a naturalistic foraging task (Chapter 5).

6.1 Summary of the experimental procedure and results

The initial behavioral experiment (experiment 1) introduced a novel feature categorization task using masked gratings as a means to investigate whether shifts of attentional resources occur in response to invisible feature changes, independent of participants' awareness of the stimulus. The results showed that orientation changes (switches) slowed down participants' response in the categorization task if the switch occurred from the heavily weighted feature (i.e., the spatial orientation that was most likely to occur) and if the presented grating was not consciously perceived, consistent with attentional selection weights being re-weighted towards the novel invisible stimulus. Following up the first behavioral experiment, using fMRI the second experiment replicated the behavioral switch effect following invisible feature changes away from weighted spatial orientations in a new group of healthy young adults. In addition, analysing participants' BOLD signal revealed that the switch-costs in response to invisible feature changes were registered in the right MFG, right AG and SPC (right TPJ) but not in the FPC. Thus, FPC was not specifically involved in detecting stimulus changes outside of visual awareness. However, the subsequent searchlight MVPA showed that FPC, together with central nodes of the frontoparietal attention network, maintained feature information of the novel stimuli, promoting the notion that FPC takes a role in updating and retaining relevant information during exploratory shifts of attention independent of visual awareness.

Whereas the first experiment was conducted to methodologically paved the way for the second experiment that served to investigate the FPC's involvement in the redistribution of attentional resources necessitated by invisible stimulus changes, the third experiment aimed to examine the role of the FPC to execute shifts from an exploitative towards an exploratory state of cognitive control. To this end, we designed a more ecologically derived visual-search-based foraging task allowing to study how participants constantly weigh between exploring and exploiting (Cohen et al., 2007). The key findings were that FPC activity was indeed linked to shifts away from an exploitative towards an exploratory mode of attention, which has previously only been found in bandit-like tasks. Moreover, the study provides important novel insights into how a behavioral idiosyncrasy in a tendency for behavioral exploration modulates the degree to which FPC gets involved during attentional exploration. Specifically, experiment 3 revealed that inter-individual differences in GUTs, indexing an individual's propensity to exploit, modulated the signal strength of the FPC time-

locked to the moment of exploration (i.e, patch-leaving). Individuals with a stronger behavioral inclination for exploration exhibited greater FPC activation when deciding to move on to a new patch.

6.2 Contributions and implications

Earlier work has consistently linked the FPC to exploratory attentional shifting, where it was shown to serve attention changes between spatial locations as well as between feature dimensions (Lepsien & Pollmann, 2002; Pollmann et al., 2000; Weidner et al., 2002). Tying in with these findings in healthy adults, patients with anterior prefrontal lesions show a deficit in visual dimension weighting (Pollmann et al., 2007). During contextual cueing, FPC responds to violations of implicitly learned contingencies between visible stimuli (Pollmann & Manginelli, 2009a, 2009b) and enables the behavioral adaptation to newly introduced contingencies (Zinchenko et al., 2018). Based on these findings, experiment 2 tested for the first time if FPC also supports the re-shifting of attentional selection weights in response to feature changes of fully invisible stimuli.

6.2.1 FPC activity not specifically linked to the detection of invisible feature changes

Against our prediction, the results of the second experiment indicated not specific role of FPC in the detection unconsciously perceived feature changes, but rather suggest that these changes are primarily detected by the ventral attentional network (TPJ and IFJ). This latter finding integrates well with the idea that the ventral attention network (see Figure 1, Section 1.2.2) supports the detection of unexpected but relevant or previously unattended but salient stimuli (Corbetta et al., 2008), where particularly the TPJ is considered to 'break' the current attentionl focus (e.g., Shulman et al., 2007). Importantly, we showed that this detection can occur outside of visual awareness, stressing the dissociation of visual attention and consciousness (Koch & Tsuchiya, 2007; Lamme, 2003). However, we could not confirm that FPC capitalizes a specific role in the re-shifting of attentional selection weights in response to unconsciously perceived stimuli, which was shown before to be the case for visible changes between feature dimensions or spatial locations (Pollmann et al., 2000; Weidner et al., 2002). At odds with these early findings, but similar to our finding of no specific FPC involvement, is the observation in lesion monkeys, that the ability to shift attention between feature dimensions remains intact following focal lesions of the FPC, whereas lesions within the ACC, dlPFC, or OFC impeded this ability (Mansouri et al., 2020). Nevertheless, it may be still worthwhile to test if change-related FPC activation occurs for invisible changes of target dimension (Pollmann et al., 2000; Weidner et al., 2002) or for changes of spatial locations (Lepsien & Pollmann, 2002).

6.2.2 Unconscious feature representation maintained in the FPC

Yet, what do the results of the second experiment tell us about the role of FPC in the adaptation to a novel yet invisible stimulus? Although clusters conveying the

information about the unconscious feature information of the novel target were not very prominent within FPC, transverse slices 12 and 16 in Figure 12 do show that informative clusters were at least bordering into the dorsal part of the right Fp1 and yet smaller clusters were also located in the right Fp2 (slices -4 and 0 in Figure 12). This activity carrying information of the novel invisible stimulus attention needed to be redirected to, could be the basis of a feedback signal presumably to more posterior regions such as the TPJ and IPS, when target changes occurred, directing the reallocation of attentional selection weights towards the new orientation. The usage of methods with a higher temporal resolution, such as EEG, could provide further insights into the temporal dynamics of the interplay between FPC and the ventral attention network during unconscious shifts of attention. Such interpretation is consistent with the hypothesis that the FPC supports the shifting toward a new behavioral strategy or an alternative action goal (Hogeveen et al., 2022; Mansouri et al., 2017). FPC maintaining a representation of action-relevant information about the new stimulus feature is consistent with previous decoding studies showing that FPC conveys unconscious patterns of brain activity predictive for the upcoming choice participants were about to make in the absence of significant changes in the overall BOLD response (Haynes, 2012). This suggests that the configuration of activity patterns in FPC change in a choice-specific manner, not necessarily related to a net change in the BOLD signal strength. Likewise, lesions in the FPC disrupt humans' prospective memory, effectively eliminating their capacity to retain action plans in memory for subsequent execution (Burgess et al., 2000). Given this evidence, it may be that the above chance decoding in FPC that we observed, reflects the maintenance of the novel action goal, in this case the novel stimulus orientation, requiring an adaptation of the behavioral response. Such conclusion would indeed be in support of the idea that FPC is essential for a switch away from the currently pursued action towards an alternative action by maintaining a representation of the alternative choice (Hogeveen et al., 2022; Mansouri et al., 2017).

6.2.3 FPC facilitates serial exploratory choices during foraging

Given the lack of evidence for FPC's specific involvement in the redistribution of attentional selection weights, but rather for its support in maintaining a representation of the novel target attention needs to be redirected to, the third experiment aimed to test if the FPC capitalizes a more general role in cognitive control by switching from an exploitative towards an exploratory state of attentional control (Boorman et al., 2009; Cohen et al., 2007; Mansouri et al., 2015; Mansouri et al., 2017).

Previous work studying the exploration-exploitation dilemma could identify a network of brain regions including the dACC, FPC, TPJ and IPS linked to exploratory choices during virtual gambling (e.g., Boorman et al., 2009; Daw et al., 2006; Laureiro-Martínez et al., 2015; Zajkowski et al., 2017). Tying in with these reports, in experiment 3, a similar set of brain regions including the FPC was found to show enhanced signaling timed-locked to exploratory choices. Yet, all studies mentioned above employed bandit tasks to study the exploration-exploitation dilemma, while approaches using patch-based foraging tasks in which choice options are encountered serially but not simultaneously have received only limited

attention (e.g., Garrett & Daw, 2020; Wolfe, 2013). Thus, experiment 3 provides an important extension to the existing literature by replicating its findings outside of the predominant domain of virtual gambling tasks that may lack a high ecological validity (Mobbs et al., 2018). In addition to this, novel insights in FPC function for exploratory decision-making are provided by linking participants' giving-up times (GUTs) with BOLD signal changes related to exploratory choices during foraging. Existing reports, again from bandit tasks, show that the FPC signal, related to exploratory choices, predicts how effectively subjects adapt their behavior to changing task conditions (e.g., Boorman et al., 2009), and that signal strength in FPC during exploration positively correlates with subjects' decision-making performance measured by their yield of total earnings (Laureiro-Martínez et al., 2015). Consistent with this is the significant positive correlation between participants' behavioral bias towards exploration (indicated by their GUTs) and the strength of the BOLD signal change in the right lateral FPC time-locked to exploratory foraging choices found in the third experiment. Moreover, a significant main effect of group (short- GUT versus long-GUT subjects) indicated that participants with a higher behavioral propensity to make exploratory choices showed significantly stronger activity bilaterally in both Fp1 and Fp2, as well as the IPS and TPJ in the moment of behavioral exploration. In other words, subjects who were more inclined to keep exploiting a current patch had weaker positive signal changes in the lateral and medial FPC, as well as in the ventral attention network time-locked to exploratory foraging choices. This finding connects well with a very similar observation in apes showing that FPC lesion monkeys have a decreased tendency to explore the potential value of an unforeseen novel task or reward (Mansouri et al., 2015). Hence, the third experiment provides convergent evidence promoting the hypothesis that FPC capitalizes a central role in the redistribution of cognitive resources away from the currently pursued strategy or goal towards an alternative opportunity (Mansouri et al., 2017).

6.2.4 A functional distinction between Fp1 and Fp2?

Evidence exists that promotes a functional distinction between Fp1 (lateral FPC) and Fp2 (medial FPC) in that Fp1 presumably drives directed whereas Fp2 is thought to support random exploration (Mansouri et al., 2017; Tomov et al., 2020; Zajkowski et al., 2017), although Fp1 also was implicated in random exploration before (Daw et al., 2006). Experiment 3 was not designed to distinguish between the two types of exploration and it may be rather difficult to tell if participants used one ore the other strategy. Given that participants timed their patch leaving rather optimally than randomly as indicated by patch-leaving onset times consistent with the predictions of the MVT, the notion that participants engaged in a more directed than random exploration is favored (Wilson et al., 2014). Consequently, assuming that participants engaged in directed exploration would explain why the FPC signal, time-locked to patch-leaving, was observed in rFp1 but not Fp2 (Zajkowski et al., 2017). However, signals of exploration in *both* subregions were modulated by inter-individual differences in the behavioral propensity to explore. Directed exploration means that exploratory choices are made to gain information but not to obtain an immediate reward. In other words, information is favored over immediate reward (Wilson et

al., 2014). In our experiment, however, foragers likely made a decision to switch to the next display to increase the chance of a new reward, driven by the experience that the search in the current display had become increasingly difficult. Thus, the exploration may have been driven mostly by the opportunity of an immediate new reward rather than by information-seeking which would be inconsistent with directed exploration. Yet, only by switching to a new display participants could potentially learn more about the foraging environment, thus suggesting that reward and information may have been confounded which makes it more difficult to identify directed exploration in experiment 3. Therefore it will be important to directly decouple information and reward to be able to test directed exploration more precisely (e.g., Wilson et al., 2014).

6.2.5 Further directions

Future research should target a deeper comprehension of the relationship between participants' behavioral strategies, particularly the balance between exploration and exploitation during naturalistic activities like foraging, and brain function. A crucial emphasis should be placed on investigating the *direction* of this relationship. As previously discussed, a crucial question revolves around whether adopting a specific foraging strategy influences the extent of FPC engagement, or if inter-individual differences in FPC function, potentially stemming from genetic variations impacting the catecholamine system (Gershman & Tzovaras, 2018), guide the preference for exploratory or exploitative behavioral tendencies. Exploring this dynamic relationship could also reveal a continuum along which individuals' inclinations toward exploration or exploitation exist. Furthermore, insights from studies focusing on elderly individuals suggest an age-related decline in exploratory behavior, accompanied by increased GUTs during foraging (Mata et al., 2009). This decline coincides with observed decreases in working memory and attentional control (Milham et al., 2002; Reuter-Lorenz, 2002), which are linked to compromised PFC function, as well as changes in PFC surface area and thickness (Dotson et al., 2016). Hence, it may be worthwhile for forthcoming investigations to explore whether individual factors such as personality traits and cognitive abilities, which may be associated with both functional and anatomical differences in the brain, contribute significantly to an individual's propensity for exploration or exploitation. Understanding these interplays could shed light on the mechanisms shaping decision-making strategies and guide potential interventions for optimizing such strategies.

Understanding the nuanced role of the FPC, alongside its susceptibility to inter-individual differences, also holds significant potential for shaping clinical research and practice, particularly in attention-related psychiatric disorders such as the attention deficit (hyperactivity) disorder (ADHD). Notably, recent research revealed that individuals with ADHD, engaged in a conventional task-switching paradigm, exhibited both reduced task switching costs and amplified activity changes within the FPC compared to healthy controls (Li et al., 2023). This resonates with previous studies in healthy individuals, where FPC signals aligned with exploratory choices were positively linked to behavioral adaptation to changing task conditions (Boorman et al., 2009; Laureiro-Martínez et al., 2015). It is also consistent with the outcomes of Ex-

periment 3, which unveiled that variations in individuals' propensity to explore correlated with differences in FPC signal strength during exploratory decision-making. Critically, the findings by Li et al. (2023) present a novel perspective on ADHD, spotlighting the mobilization of cognitive resources in contrast to the traditional emphasis on limitations imposed by the condition. This paradigm shift underscores the potential for further investigations into FPC function to yield innovative strategies to empower individuals with ADHD to maximize their cognitive abilities.

6.3 Final conclusions

Altogether, the experimental work of this dissertation confirms the current view in FPC function suggesting that the FPC is involved in a more general role of switching between an exploitative towards an exploratory mode of cognitive control. While the FPC does not appear to be specifically tasked with detecting fully invisible changes in our environment, it does appear to contribute to attentional adaptation by encoding pertinent information about what has unconsciously changed. Beyond mere external changes in the environment necessitating attention shifts, the realization that a pursued action strategy has become ineffective also calls for adaptive responses. In this context, the research presented in this thesis underscores the pivotal role of the FPC in facilitating the transition from exploitation to exploration. Intriguingly, I could show that a heightened inclination towards exploitative behavioral strategies correlates with reduced FPC activity during these transitions. This finding highlights the significance of accounting for inter-individual differences in behavioral inclinations, which can impact the extent of FPC involvement during the shift from exploitative to exploratory cognitive control. This underscores consequently the need for hypotheses to be carefully aligned. Behavioral tendencies, as potential modulators of FPC involvement in exploratory attentional control, warrant consideration. Moreover, it emphasizes the urgency of comprehending the directionality of the relationship between the propensity to explore behaviorally and the signaling of the FPC during attentional exploration. Such insights not only contribute to the evolving understanding of the FPC's multifaceted role but also hold the promise of refining the treatment strategies for conditions governed by cognitive control dilemmas.

References

Addicott, M. A., Pearson, J. M., Froeliger, B., Platt, M. L., & McClernon, F. J. (2014). Smoking automaticity and tolerance moderate brain activation during explore–exploit behavior. *Psychiatry Research: Neuroimaging*, *224*(3), 254–261. https://doi.org/10.1016/j.pscychresns.2014.10.014

Arkes, H. R., & Ayton, P. (1999). The sunk cost and concorde effects: Are humans less rational than lower animals? *Psychological bulletin*, *125*(5), 591. https://doi.org/10.1037/0033-2909.125.5.591

Avneon, M., & Lamy, D. (2018). Reexamining unconscious response priming: A liminal-prime paradigm. *Consciousness and cognition*, *59*, 87–103. https://doi.org/10.1016/j.concog.2017.12.006

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*, 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412. https://doi.org/https://doi.org/10.1016/j.jml.2007.12.005

Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, *73*, 595–607. https://doi.org/10.1016/j.neuron.2011.12.025

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course epi of human brain function during task activation. *Magnetic resonance in medicine*, *25*(2), 390–397. https://doi.org/10.1002/mrm.1910250220

Bannister, A. P. (2005). Inter-and intra-laminar connections of pyramidal cells in the neocortex. *Neuroscience research*, *53*(2), 95–103. https://doi.org/10.1016/j.neures.2005.06.019

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00328

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. https://doi.org/10.48550/arXiv.1406.5823

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*. https://doi.org/10.18637/jss.v067.i01

Beharelle, A. R., Polanía, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial stimulation over frontopolar cortex elucidates the choice attributes and neural mechanisms used to resolve exploration–exploitation trade-offs. *Journal of Neuroscience*, *35*, 14544–14556. https://doi.org/10.1523/JNEUROSCI.2322-15.2015

Berens, P., Logothetis, N., & Tolias, A. (2010). Local field potentials, bold and spiking activity–relationships and physiological mechanisms. *Nature Precedings*, 1–1. https://doi.org/10.1038/npre.2010.5216.1

Berry, A. S., Shah, V. D., Baker, S. L., Vogel, J. W., O'Neil, J. P., Janabi, M., Schwimmer, H. D., Marks, S. M., & Jagust, W. J. (2016). Aging affects dopaminergic neu-

ral mechanisms of cognitive flexibility. *Journal of Neuroscience*, *36*(50), 12559–12569. https://doi.org/10.1523/JNEUROSCI.0626-16.2016

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, *13*(2), 90–102. https://doi.org/10.1080/17470216108416478

Bertelson, P. (1963). Sr relationships and reaction times to new versus repeated signals in a serial task. *Journal of experimental psychology*, *65*(5), 478. https://doi.org/10.1037/h0047742

Betz, T., Kietzmann, T. C., Wilming, N., & Koenig, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *Journal of vision*, *10*(3), 15–15. https://doi.org/10.1167/10.3.15

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual review of neuroscience*, *33*, 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823

Bisley, J., Ipata, A., Krishna, B., Gee, A., Goldberg, M., Jenkin, M., & Harris, L. (2009). The lateral intraparietal area: A priority map in posterior parietal cortex. *Cortical mechanisms of vision*, 9–34.

Blais, C., Hubbard, E., & Mangun, G. R. (2016). Erp evidence for implicit priming of top–down control of attention. *Journal of Cognitive Neuroscience*, *28*(5), 763–772. https://doi.org/10.1162/jocn_a_00925

Blanchard, T. C., & Gershman, S. J. (2018). Pure correlates of exploration and exploitation in the human brain. *Cognitive, Affective and Behavioral Neuroscience*, *18*, 117–126. https://doi.org/10.3758/s13415-017-0556-2

Bludau, S., Eickhoff, S. B., Mohlberg, H., Caspers, S., Laird, A. R., Fox, P. T., Schleicher, A., Zilles, K., & Amunts, K. (2014). Cytoarchitecture, probability maps and functions of the human frontal pole. *NeuroImage*, *93*, 260–275. https://doi.org/10.1016/j.neuroimage.2013.05.052

Bodner, G. E., & Lee, L. (2014). Masked response priming across three prime proportions: A comparison of three accounts. *Perceptual and Motor Skills*, *119*(1), 59–68. https://doi.org/10.2466/22.23.PMS.119c18z0

Bohil, C. J., & Wismer, A. J. (2015). Implicit learning mediates base rate acquisition in perceptual categorization. *Psychonomic Bulletin and Review*, *22*, 586–593. https://doi.org/10.3758/s13423-014-0694-2

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS biology*, *9*(6), e1001093. https://doi.org/10.1371/journal.pbio.1001093

Boorman, E. D., Behrens, T. E., Woolrich, M. W., & Rushworth, M. F. (2009). How green is the grass on the other side? frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*, 733–743. https://doi.org/10.1016/j.neuron.2009.05.014

Boschin, E. A., Piekema, C., & Buckley, M. J. (2015). Essential functions of primate frontopolar cortex in cognition. *Proceedings of the National Academy of Sciences*

of the United States of America, *112*, E1020–E1027. https://doi.org/10.1073/pnas.1419649112

Bourgeois, A., Chelazzi, L., & Vuilleumier, P. (2016). How motivation and reward learning modulate selective attention. *Progress in brain research*, *229*, 325–342. https://doi.org/10.1016/bs.pbr.2016.06.004

Brodmann, K. (1909). *Vergleichende lokalisationslehre der grosshirnrinde in ihren prinzipien dargestellt auf grund des zellenbaues*. Barth.

Bundesen, C. (1990). A theory of visual attention. *Psychological review*, *97*(4), 523. https://doi.org/10.1037/0033-295X.97.4.523

Burgess, P. W., Veitch, E., de Lacy Costello, A., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, *38*(6), 848–863. https://doi.org/10.1016/S0028-3932(99)00134-7

Buxton, R. (2010). Interpreting oxygenation-based neuroimaging signals: The importance and the challenge of understanding brain oxygen metabolism. *Frontiers in neuroenergetics*, *2*, 1648. https://doi.org/10.3389/fnene.2010.00008

Carlson, K. S., Gadziola, M. A., Dauster, E. S., & Wesson, D. W. (2018). Selective attention controls olfactory decisions and the neural encoding of odors. *Current Biology*, *28*(14), 2195–2205. https://doi.org/10.1016/j.cub.2018.05.011

Carmichael, S. T., & Price, J. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology*, *371*(2), 179–207. https://doi.org/10.1002/(SICI)1096-9861(19960722)371:2<179::AID-CNE1>3.0.CO;2-%23

Carrasco, M. (2006). Covert attention increases contrast sensitivity: Psychophysical, neurophysiological and neuroimaging studies. *Progress in brain research*, *154*, 33–70. https://doi.org/10.1016/S0079-6123(06)54003-8

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, *49*(2), 220–238. https://doi.org/10.1111/j.1469-8986.2011.01293.x

Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F., & Peters, J. (2020). Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. *eLife*, *9*, 1–44. https://doi.org/10.7554/eLife.51260

Chang, C.-F., Hsu, T.-Y., Tseng, P., Liang, W.-K., Tzeng, O. J., Hung, D. L., & Juan, C.-H. (2013). Right temporoparietal junction and attentional reorienting. *Human brain mapping*, *34*(4), 869–877. https://doi.org/doi.org/10.1002/hbm.21476

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical population biology*, *9*(2), 129–136. https://doi.org/10.1016/0040-5809(76)90040-X

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *40*(4), 343. https://doi.org/10.1037/h0080103

Chen, J. E., Glover, G. H., Fultz, N. E., Rosen, B. R., Polimeni, J. R., & Lewis, L. D. (2021). Investigating mechanisms of fast bold responses: The effects of stimulus intensity and of spatial heterogeneity of hemodynamics. *NeuroImage*, *245*, 118658. https://doi.org/10.1016/j.neuroimage.2021.118658

Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W., & Haynes, J.-D. (2011). Cortical surface-based searchlight decoding. *Neuroimage*, *56*(2), 582–592. https://doi.org/10.1016/j.neuroimage.2010.07.035

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Representing color ensembles. *Psychological science*, *28*(10), 1510–1517. https://doi.org/10.1177/0956797617713787

Chica, A. B., Bartolomeo, P., & Valero-Cabré, A. (2011). Dorsal and ventral parietal contributions to spatial orienting in the human brain. *Journal of Neuroscience*, *31*(22), 8143–8149. https://doi.org/10.1523/JNEUROSCI.5463-10.2010

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Düzel, E., & Dolan, R. J. (2013). Dopamine restores reward prediction errors in old age. *Nature neuroscience*, *16*(5), 648–653. https://doi.org/10.1038/nn.3364

Christoff, K., & Gabrieli, J. D. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*(2), 168–186. https://doi.org/10.3758/BF03331976

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. https://doi.org/doi.org/10.1098/rstb.2007.2098

Cohen, J. D., Botvinick, M., & Carter, C. S. (2000). Anterior cingulate and prefrontal cortex: Who's in control? *Nature neuroscience*, *3*(5), 421–423. https://doi.org/10.1038/74783

Constantino, S. M., & Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective and Behavioral Neuroscience*, *15*, 837–853. https://doi.org/10.3758/s13415-015-0350-y

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature neuroscience*, *3*(3), 292–297. https://doi.org/10.1038/73009

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*, 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215. https://doi.org/10.1038/nrn755

Costa, V. D., Mitz, A. R., & Averbeck, B. B. (2019). Subcortical substrates of explore-exploit decisions in primates. *Neuron*, *103*(3), 533–545. https://doi.org/10.1016/j.neuron.2019.05.017

Craig, A. D. (2009). How do you feel—now? the anterior insula and human awareness. *Nature reviews neuroscience*, *10*(1), 59–70. https://doi.org/10.1038/nrn2555

Craighero, L., Bello, A., Fadiga, L., & Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, *40*(5), 492–502. https://doi.org/10.1016/S0028-3932(01)00134-8

Crainiceanu, C. M., & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *66*(1), 165–185. https://doi.org/10.1111/j.1467-9868.2004.00438.x

Dagher, A., Owen, A. M., Boecker, H., & Brooks, D. J. (1999). Mapping the network for planning: A correlational pet activation study with the tower of london task. *Brain*, *122*(10), 1973–1987. https://doi.org/10.1093/brain/122.10.1973

Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fmri: Practices and pitfalls. *Annals of the New York Academy of Sciences*, *1296*(1), 108–134. https://doi.org/10.1111/nyas.12156

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879. https://doi.org/10.1038/nature04766

De Schotten, M. T., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D., & Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nature neuroscience*, *14*(10), 1245–1246.

Derrfuß, J., Brass, M., Von Cramon, D., Lohmann, G., & Amunts, K. (2009). Interindividuelle variabilität von aktivierungen im posterioren frontalcortex. https://doi.org/10.23668/psycharchives.896.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, *93*(24), 13494–13499.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193–222.

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer bayesian analyses over significance testing. *Psychonomic bulletin & review*, *25*, 207–218. https://doi.org/10.3758/s13423-017-1266-z

DiQuattro, N. E., & Geng, J. J. (2011). Contextual knowledge configures attentional control networks. *Journal of Neuroscience*, *31*(49), 18026–18035. https://doi.org/10.1523/JNEUROSCI.4040-11.2011

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron*, *35*(5), 989–996.

Donoso, M., Collins, A. G., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science*, *344*(6191), 1481–1486. https://doi.org/10.1126/science.1252254

Dotson, V. M., Szymkowicz, S. M., Sozda, C. N., Kirton, J. W., Green, M. L., O'Shea, A., McLaren, M. E., Anton, S. D., Manini, T. M., & Woods, A. J. (2016). Age differences in prefrontal surface area and thickness in middle aged to older adults. *Frontiers in aging neuroscience*, *7*, 250. https://doi.org/10.3389/fnagi.2015.00250

Dove, A., Pollmann, S., Schubert, T., Wiggins, C. J., & Von Cramon, D. Y. (2000). Prefrontal cortex activation in task switching: An event-related fmri study. *Cognitive brain research*, *9*(1), 103–109. https://doi.org/10.1016/S0926-6410(99)00029-4

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological review*, *96*(3), 433. https://doi.org/10.1037/0033-295X.96.3.433

Dutta, A., Shah, K., Silvanto, J., & Soto, D. (2014). Neural basis of non-conscious visual working memory. *Neuroimage*, *91*, 336–343. https://doi.org/10.1016/j.neuroimage.2014.01.016

Ebitz, R. B., Albarran, E., & Moore, T. (2018). Exploration disrupts choice-predictive signals and alters dynamics in prefrontal cortex. *Neuron*, *97*(2), 450–461. https://doi.org/10.1016/j.neuron.2017.12.007

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *Neuroimage*, *78*, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041

Failing, M., & Theeuwes, J. (2018). Selection history: How reward modulates selectivity of visual attention. *Psychonomic bulletin & review*, *25*(2), 514–538. https://doi.org/10.3758/s13423-017-1380-y

Farooqui, A. A., & Manly, T. (2015). Anticipatory control through associative learning of subliminal relations: Invisible may be better than visible. *Psychological Science*, *26*(3), 325–334. https://doi.org/10.1177/0956797614564191

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in cognitive sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human perception and performance*, *18*(4), 1030. https://doi.org/10.1037/0096-1523.18.4.1030

Formisano, E., De Martino, F., & Valente, G. (2008). Multivariate analysis of fmri time series: Classification and regression of brain responses using machine learning. *Magnetic resonance imaging*, *26*(7), 921–934. https://doi.org/10.1016/j.mri.2008.01.052

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051. https://doi.org/10.1073/pnas.0604187103

Fox, P. T., & Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences*, *83*(4), 1140–1144.

Fox, P. T., Raichle, M. E., Mintun, M. A., & Dence, C. (1988). Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, *241*(4864), 462–464. https://doi.org/10.1126/science.3260686

Frahm, J., Bruhn, H., Merboldt, K.-D., & Hänicke, W. (1992). Dynamic mr imaging of human brain oxygenation during rest and photic stimulation. *Journal of Magnetic Resonance Imaging*, *2*(5), 501–505. https://doi.org/10.1002/jmri.1880020505

Fuster, J. M. (2002). Prefrontal cortex in temporal organization of action. *The Handbook of Brain Theory and Neural Networks, 2nd edn. MIT Press, Cambridge*, 905–910.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological review*, *116*(2), 439. https://doi.org/10.1037/a0015251

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., & Uludağ, K. (2016). The effect of spatial resolution on decoding accuracy in fmri multivariate pattern analysis. *NeuroImage*, *132*, 32–42. https://doi.org/10.1016/j.neuroimage.2016.02.033

Garrett, N., & Daw, N. D. (2020). Biased belief updating and suboptimal choice in foraging decisions. *Nature communications*, *11*(1), 3417. https://doi.org/10.1038/s41467-020-16964-5

Gershman, S. J., & Tzovaras, B. G. (2018). Dopaminergic genes are associated with both directed and random exploration. neuropsychologia. *Neuropsychologia*, *120*, 97–104. https://doi.org/10.1016/j.neuropsychologia.2018.10.009

Giller, F., & Beste, C. (2019). Effects of aging on sequential cognitive flexibility are associated with fronto-parietal processing deficits. *Brain Structure and Function*, *224*, 2343–2355. https://doi.org/10.1007/s00429-019-01910-z

Goense, J. B., & Logothetis, N. K. (2008). Neurophysiology of the bold fmri signal in awake monkeys. *Current Biology*, *18*(9), 631–640. https://doi.org/10.1016/j.cub.2008.03.054

Golland, P., & Fischl, B. (2003). Permutation tests for classification: Towards statistical significance in image-based studies. *Biennial international conference on information processing in medical imaging*, 330–341. https://doi.org/10.1007/978-3-540-45087-0_28

Goodyear, B. G., & Menon, R. S. (1998). Effect of luminance contrast on bold fmri response in human primary visual areas. *Journal of Neurophysiology*, *79*(4), 2204–2207. https://doi.org/10.1152/jn.1998.79.4.2204

Gramann, K., Töllner, T., & Müller, H. J. (2010). Dimension-based attention modulates early visual processing. *Psychophysiology*, *47*, 968–978. https://doi.org/10.1111/j.1469-8986.2010.00998.x

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. *Journal of experimental psychology*, *38*(4), 404. https://doi.org/10.1037/h0059831

Green, A. E., Fugelsang, J. A., Kraemer, D. J., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain research*, *1096*(1), 125–137. https://doi.org/10.1016/j.brainres.2006.04.024

Güldener, L., Jüllig, A., Soto, D., & Pollmann, S. (2021). Feature-based attentional weighting and re-weighting in the absence of visual awareness. *Frontiers in Human Neuroscience*, *15*, 610347. https://doi.org/10.3389/fnhum.2021.610347

Güldener, L., Jüllig, A., Soto, D., & Pollmann, S. (2022). Frontopolar activity carries feature information of novel stimuli during unconscious reweighting of selective attention. *Cortex*, *153*, 146–165. https://doi.org/10.1016/j.cortex.2022.03.024

Güldener, L., & Pollmann, S. (2023). Behavioral bias for exploration is associated with enhanced signaling in the lateral and medial frontopolar cortex. https://doi.org/10.31234/osf.io/tpybj

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). Pymvpa: A python toolbox for multivariate pattern analysis of fmri data. *Neuroinformatics*, *7*, 37–53. https://doi.org/10.1007/s12021-008-9041-y

Haxby, J. V. (2012). Multivariate pattern analysis of fmri: The early beginnings. *NeuroImage*, *62*, 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature neuroscience*, *14*(7), 933–939. https://doi.org/10.1038/nn.2856

Haynes, J. D. (2012). Beyond libet: Long-term prediction of free choices from neuroimaging signals. in characterizing consciousness: From cognition to the clinic? Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-18015-6_10

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature neuroscience*, *8*(5), 686–691. https://doi.org/10.1038/nn1445

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current biology*, *17*(4), 323–328. https://doi.org/10.1016/j.cub.2006.11.072

He, B. J., Snyder, A. Z., Vincent, J. L., Epstein, A., Shulman, G. L., & Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron*, *53*(6), 905–918. https://doi.org/10.1016/j.neuron.2007.02.013

Hendriks, M. H. A., Daniels, N., Pegado, F., & Op de Beeck, H. P. (2017). The effect of spatial smoothing on representational similarity in a simple motor paradigm. *Frontiers in Neurology*, *8*. https://doi.org/10.3389/fneur.2017.00222

Henson, R., & Rugg, M. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*(3), 263–270. https://doi.org/10.1016/S0028-3932(02)00159-8

Herrojo Ruiz, M., Maudrich, T., Kalloch, B., Sammler, D., Kenville, R., Villringer, A., Sehm, B., & Nikulin, V. V. (2021). Modulation of neural activity in frontopolar cortex drives reward-based motor learning. *Scientific reports*, *11*(1), 20303. https://doi.org/10.1038/s41598-021-98571-y

Hogeveen, J., Medalla, M., Ainsworth, M., Galeazzi, J. M., Hanlon, C. A., Mansouri, F. A., & Costa, V. D. (2022). What does the frontopolar cortex contribute to goal-directed cognition and action? *Journal of Neuroscience*, *42*(45), 8508–8513. https://doi.org/10.1523/JNEUROSCI.1143-22.2022

Horowitz, T. S., & Thornton, I. M. (2008). Objects or locations in vision for action? evidence from the milo task. *Visual cognition*, *16*(4), 486–513. https://doi.org/10.1080/13506280601087356

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106. https://doi.org/10.1113/jphysiol.1962.sp006837

Hutchinson, J. M., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: Can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour*, *75*, 1331–1349. https://doi.org/10.1016/j.anbehav.2007.09.006

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10-12), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

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

Jachs, B., Blanco, M. J., Grantham-Hill, S., & Soto, D. (2015). On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 269–276. https://doi.org/10.1037/xhp0000026

Jacobs, B., Schall, M., Prather, M., Kapler, E., Driscoll, L., Baca, S., Jacobs, J., Ford, K., Wainwright, M., & Treml, M. (2001). Regional dendritic and spine variation in human cerebral cortex: A quantitative golgi study. *Cerebral cortex*, *11*(6), 558–571. https://doi.org/10.1093/cercor/11.6.558

Jamoulle, T., Ran, Q., Meersmans, K., Schaeverbeke, J., Dupont, P., & Vandenberghe, R. (2021). Posterior Intraparietal Sulcus Mediates Detection of Salient Stimuli Outside the Endogenous Focus of Attention. *Cerebral Cortex*, *32*(7), 1455–1469. https://doi.org/10.1093/cercor/bhab299

Japee, S., Holiday, K., Satyshur, M. D., Mukai, I., & Ungerleider, L. G. (2015). A role of right middle frontal gyrus in reorienting of attention: A case study. *Frontiers in systems neuroscience*, *9*, 23. https://doi.org/fnsys.2015.00023

Jeffreys, H. (1998). *The theory of probability*. OuP Oxford.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H., & Curtis, C. E. (2012). Prioritized maps of space in human frontoparietal cortex. *Journal of Neuroscience*, *32*(48), 17382–17390. https://doi.org/10.1523/JNEUROSCI.3810-12.2012

Johnson, P. C. (2014). Extension of nakagawa & schielzeth's r2glmm to random slopes models. *Methods in ecology and evolution*, *5*(9), 944–946. https://doi.org/10.1111/2041-210X.12225

Jones, P. (1975). The two-armed bandit. *Biometrika*, *62*(2), 523–524. https://doi.org/10.1093/biomet/62.2.523

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kanai, R., Tsuchiya, N., & Verstraten, F. A. (2006). The scope and limits of top-down attention in unconscious visual processing. *Current Biology*, *16*(23), 2332–2336. https://doi.org/10.1016/j.cub.2006.10.001

Katsuki, F., & Constantinidis, C. (2012). Early involvement of prefrontal cortex in visual bottom-up attention. *Nature neuroscience*, *15*(8), 1160–1166. https://doi.org/10.1038/nn.3164

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. *The Neuroscientist*, *20*(5), 509–521. https://doi.org/10.1177/1073858413514136

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (1999). Attention without awareness in blindsight. *Proceedings of the Royal Society B: Biological Sciences*, *266*, 1805–1811. https://doi.org/10.1098/rspb.1999.0850

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of neurophysiology*, *97*(6), 4296–4309. https://doi.org/10.1152/jn.00024.2007

King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. *Neuron*, *92*(5), 1122–1134. https://doi.org/10.1016/j.neuron.2016.10.051

Klink, P. C., Jentgens, P., & Lorteije, J. A. (2014). Priority maps explain the roles of value, attention, and salience in goal-oriented behavior. *Journal of Neuroscience*, *34*, 13867–13869. https://doi.org/10.1523/JNEUROSCI.3249-14.2014

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in cognitive sciences*, *11*(1), 16–22. https://doi.org/10.1016/j.tics.2006.10.012

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*(6732), 148–151. https://doi.org/10.1038/20178

Kolling, N., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2012). Neural mechanisms of foraging. *Science*, *336*(6077), 95–98. https://doi.org/10.1126/science.121693

Konishi, S., Chikazoe, J., Jimura, K., Asari, T., & Miyashita, Y. (2005). Neural mechanism in anterior prefrontal cortex for inhibition of prolonged set interference. *Proceedings of the National Academy of Sciences*, *102*(35), 12584–12588. https://doi.org/10.1073/pnas.0500585102

Kovach, C. K., Daw, N. D., Rudrauf, D., Tranel, D., O'Doherty, J. P., & Adolphs, R. (2012). Anterior prefrontal cortex contributes to action selection through tracking of recent reward trends. *Journal of Neuroscience*, *32*(25), 8434–8442. https://doi.org/10.1523/JNEUROSCI.5468-11.2012

Krebs, J. R., Ryan, J. C., & Charnov, E. L. (1974). Hunting by expectation or optimal foraging? a study of patch use by chickadees. *Animal behaviour*, *22*, 953–IN3. https://doi.org/10.1016/0003-3472(74)90018-9

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of statistical software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., & Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, *89*(12), 5675–5679. https://doi.org/10.1073/pnas.89.12.5675

Lambert, A., Naikar, N., McLachlan, K., & Aitken, V. (1999). A new component of visual orienting: Implicit effects of peripheral information and subthreshold cues on covert attention. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 321. https://doi.org/10.1037/0096-1523.25.2.321

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in cognitive sciences*, *7*(1), 12–18. https://doi.org/10.1016/S1364-6613(02)00013-X

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *Journal of Neuroscience*, *27*, 5805–5811. https://doi.org/10.1523/JNEUROSCI.4335-06.2007

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration-exploitation dilemma: An fmri study of attention control and decision-making performance. *Strategic Management Journal*, *36*, 319–338. https://doi.org/10.1002/smj.2221

Leber, A. B., Kawahara, J.-I., & Gabari, Y. (2009). Long-term abstract learning of attentional set. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1385. https://doi.org/10.1037/a0016470

Lee, S.-H., & Baker, C. I. (2016). Multi-voxel decoding and the topography of maintained information during visual working memory. *Frontiers in systems neuroscience*, *10*, 2. https://doi.org/10.3389/fnsys.2016.00002

Lepsien, J., & Pollmann, S. (2002). Covert reorienting and inhibition of return: An event-related fmri study. *Journal of cognitive neuroscience*, *14*(2), 127–144. https://doi.org/10.1162/089892902317236795

Lewis-Peacock, J. A., & Norman, K. A. (2014). Multi-voxel pattern analysis of fmri data. *The cognitive neurosciences*, *512*, 911–920.

Li, B., Zhang, L., Zhang, Y., Chen, Y., Peng, J., Shao, Y., & Zhang, X. (2020). Decreased functional connectivity between the right precuneus and middle frontal gyrus is related to attentional decline following acute sleep deprivation. *Frontiers in Neuroscience*, *14*, 530257. https://doi.org/10.3389/fnins.2020.530257

Li, Y., Chen, J., Zheng, X., Liu, J., Peng, C., & Liao, Y. (2023). Functional near-infrared spectroscopy evidence of prefrontal regulation of cognitive flexibility in adults with adhd. *Journal of Attention Disorders*, 10870547231154902. https://doi.org/10.1177/10870547231154902

Liesefeld, H. R., Liesefeld, A. M., Pollmann, S., & Müller, H. J. (2019). Biasing allocations of attention via selective weighting of saliency signals: Behavioral and neuroimaging evidence for the dimension-weighting account. *Processes of visuospatial attention and working memory*, 87–113. https://doi.org/10.1007/7854_2018_75

Logothetis, N. K. (2003). The underpinnings of the bold functional magnetic resonance imaging signal. *Journal of Neuroscience*, *23*(10), 3963–3971. https://doi.org/10.1523/JNEUROSCI.23-10-03963.2003

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fmri signal. *nature*, *412*(6843), 150–157. https://doi.org/10.1038/35084005

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the bold signal. *Annual Review of Physiology*, *66*, 735–769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M. O., & Mainen, Z. F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature Communications*, *9*. https://doi.org/10.1038/s41467-018-03438-y

Lukas, S., Philipp, A. M., & Koch, I. (2010). Switching attention between modalities: Further evidence for visual dominance. *Psychological Research PRPF*, *74*, 255–267. https://doi.org/10.1007/s00426-009-0246-y

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in r. *Behavior research methods*, *49*, 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Lynn, S. K., & Barrett, L. F. (2014). "utilizing" signal detection theory. *Psychological science*, *25*(9), 1663–1673. https://doi.org/10.1177/0956797614541991

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (lmmgui). *Frontiers in psychology*, *6*, 2. https://doi.org/10.3389/fpsyg.2015.00002

Manginelli, A. A., & Pollmann, S. (2009). Misleading contextual cues: How do they affect visual search? *Psychological Research*, *73*, 212–221. https://doi.org/10.1007/s00426-008-0211-1

Mansouri, F. A., Buckley, M. J., Fehring, D. J., & Tanaka, K. (2020). The role of primate prefrontal cortex in bias and shift between visual dimensions. *Cerebral Cortex*, *30*(1), 85–99. https://doi.org/10.1093/cercor/bhz072

Mansouri, F. A., Buckley, M. J., Mahboubi, M., & Tanaka, K. (2015). Behavioral consequences of selective damage to frontal pole and posterior cingulate cortices. *Proceedings of the National Academy of Sciences*, *112*(29), E3940–E3949. https://doi.org/10.1073/pnas.1422629112

Mansouri, F. A., Koechlin, E., Rosa, M. G., & Buckley, M. J. (2017). Managing competing goals - a key role for the frontopolar cortex. *Nature Reviews Neuroscience*, *18*, 645–657. https://doi.org/10.1038/nrn.2017.111

Mata, R., Wilke, A., & Czienskowski, U. (2013). Foraging across the life span: Is there a reduction in exploration with aging? *Frontiers in neuroscience*, *7*(53). https://doi.org/10.3389/fnins.2013.00053

Mata, R., Wilke, A., & Czienskowski, U. (2009). Cognitive aging and adaptive foraging behavior. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *64*, 474–481. https://doi.org/10.1093/geronb/gbp035

Mathiesen, C., Caesar, K., Akgören, N., & Lauritzen, M. (1998). Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *The Journal of physiology*, *512*(2), 555–566. https://doi.org/10.1111/j.1469-7793.1998.555be.x

Mathiesen, C., Caesar, K., & Lauritzen, M. (2000). Temporal coupling between neuronal activity and blood flow in rat cerebellar cortex as indicated by field potential analysis. *The Journal of physiology*, *523*(1), 235–246. https://doi.org/10.1111/j.1469-7793.2000.t01-1-00235.x

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of memory and language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 168–180. https://doi.org/10.1037//0096-1523.23.1.168

Mcnair, J. N. (1982). Optimal giving-up times and the marginal value theorem. *Am. Nat*, *119*, 511–529. https://doi.org/10.1086/283929

McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of reml and the kenward-roger correction. *Multivariate behavioral research*, *52*(5), 661–670. https://doi.org/10.1080/00273171.2017.1344538

Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: Awareness and response. *Brain structure and function*, *214*, 535–549. https://doi.org/10.1007/s00429-010-0265-x

Mei, N., Santana, R., & Soto, D. (2022). Informative neural representations of unseen contents during higher-order processing in human brains and deep artificial networks. *Nature Human Behaviour*, *6*(5), 720–731. https://doi.org/10.1038/s41562-021-01274-7

Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., & Cohen, N. J. (2002). Attentional control in the aging brain: Insights from an fmri study of the stroop task. *Brain and cognition*, *49*(3), 277–296. https://doi.org/10.1006/brcg.2001.1501

Mobbs, D., Trimmer, P. C., Blumstein, D. T., & Dayan, P. (2018). Foraging for foundations in decision neuroscience: Insights from ethology. *Nature Reviews Neuroscience*, *19*, 419–427. https://doi.org/10.1038/s41583-018-0010-7

Monti, M. M. (2011). Statistical analysis of fmri time-series: A critical review of the glm approach. *Frontiers in human neuroscience*, *5*, 28. https://doi.org/10.3389/fnhum.2011.00028

Morey, R. D., Rouder, J. N., & Jamil, T. (2018). Bayesfactor: Computation of bayes factors for common designs. r package version 0.9. 12-4.2.

Mulckhuyse, M., Talsma, D., & Theeuwes, J. (2007). Grabbing attention without knowing: Automatic capture of attention by subliminal spatial cues. *Visual Cognition*, *15*, 779–788. https://doi.org/10.1080/13506280701307001

Mulckhuyse, M., & Theeuwes, J. (2010). Unconscious attentional orienting to exogenous cues: A review of the literature. *Acta Psychologica*, *134*, 299–309. https://doi.org/10.1016/j.actpsy.2010.03.002

Müller, H., Krummenacher, J., & Heller, D. (2004). Dimension-specific intertrial facilitation in visual search for pop-out targets: Evidence for a top-down modulable visual short-term memory effect. *Visual Cognition*, *11*(5), 577–602. https://doi.org/10.1080/13506280344000419

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception Psychophysics*, *57*, 1–17. https://doi.org/10.3758/BF03211845

Mumford, J. A., Poline, J.-B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fmri models. *PloS one*, *10*(4), e0126255. https://doi.org/10.1371/journal.pone.0126255

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Sawamoto, N., Toma, K., Nakamura, K., Hanakawa, T., Konishi, J., et al. (1999). Transient neural activity in the medial superior frontal gyrus and precuneus time locked with attention shift between object features. *Neuroimage*, *10*(2), 193–199. https://doi.org/10.1006/nimg.1999.0451

Naselaris, T., & Kay, K. N. (2015). Resolving ambiguities of mvpa using explicit models of representation. *Trends in cognitive sciences*, *19*(10), 551–554. https://doi.org/10.1016/j.tics.2015.07.005

Navarro, A. D., & Fantino, E. (2005). The sunk cost effect in pigeons and humans. *JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR*, *83*, 1–13. https://doi.org/10.1901/jeab.2005.21-04

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fmri data. *Trends in Cognitive Sciences*, *10*, 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *89*(13), 5951–5955. https://doi.org/10.1073/pnas.89.13.5951

Öngür, D., Ferry, A. T., & Price, J. L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *Journal of Comparative Neurology*, *460*(3), 425–449. https://doi.org/10.1002/cne.10609

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fmri analyses? *NeuroImage*, *49*(3), 1943–1948. https://doi.org/https://doi.org/10.1016/j.neuroimage.2009.02.047

Ort, E., Fahrenfort, J. J., Reeder, R., Pollmann, S., & Olivers, C. N. (2019). Frontal cortex differentiates between free and imposed target selection in multiple-target search. *NeuroImage*, *202*, 116133. https://doi.org/10.1016/j.neuroimage.2019.116133

Pan, Y., Lin, B., Zhao, Y., & Soto, D. (2014). Working memory biasing of visual perception without awareness. *Attention, Perception, & Psychophysics*, *76*, 2051–2062. https://doi.org/10.3758/s13414-013-0566-2

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior research methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, *162*(1-2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pereira, F., & Botvinick, M. (2011). Information mapping with pattern classifiers: A comparative study. *Neuroimage*, *56*(2), 476–496. https://doi.org/10.1016/j.neuroimage.2010.05.026

Peremen, Z., Hilo, R., & Lamy, D. (2013). Visual consciousness and intertrial feature priming. *Journal of Vision*, *13*(5), 1–1. https://doi.org/10.1167/13.5.1

Petrides, M., Tomaiuolo, F., Yeterian, E. H., & Pandya, D. N. (2012). The prefrontal cortex: Comparative architectonic organization in the human and the macaque monkey brains. *cortex*, *48*(1), 46–57. https://doi.org/10.1016/j.cortex.2011.07.002

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: Basic concepts and examples. *Mixed-effects models in S and S-Plus*, 3–56.

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional mri data analysis*. Cambridge University Press.

Pollmann, S. (2001). Switching between dimensions, locations, and responses: The role of the left frontopolar cortex. *NeuroImage*, *14*(1), S118–S124. https://doi.org/10.1006/nimg.2001.0837

Pollmann, S. (2004). Anterior prefrontal cortex contributions to attention control. *Experimental psychology*, *51*(4), 270–278. https://doi.org/10.1027/1618-3169.51.4.270

Pollmann, S. (2016). Frontopolar resource allocation in human and nonhuman primates. *Trends in Cognitive Sciences*, *20*(2), 84–86. https://doi.org/10.1016/j.tics.2015.11.006

Pollmann, S., Mahn, K., Reimann, B., Weidner, R., Tittgemeyer, M., Preul, C., Müller, H. J., & von Cramon, D. Y. (2007). Selective visual dimension weighting deficit after left lateral frontopolar lesions. *Journal of Cognitive Neuroscience*, *19*(3), 365–375. https://doi.org/10.1162/jocn.2007.19.3.365

Pollmann, S., & Manginelli, A. A. (2009b). Anterior prefrontal involvement in implicit contextual change detection. *Frontiers in Human Neuroscience*, *3*, 703. https://doi.org/10.3389/neuro.09.028.2009

Pollmann, S., & Manginelli, A. A. (2009a). Early implicit contextual change detection in anterior prefrontal cortex. *Brain research*, *1263*, 87–92. https://doi.org/10.1016/j.brainres.2009.01.039

Pollmann, S., Weidner, R., Müller, H. J., & von Cramon, D. Y. (2000). A fronto-posterior network involved in visual dimension changes. *Journal of cognitive neuroscience*, *12*(3), 480–494. https://doi.org/10.1162/089892900562156

Pollmann, S., Weidner, R., Müller, H. J., Maertens, M., & von Cramon, D. Y. (2006). Selective and interactive neural correlates of visual dimension changes and response changes. *NeuroImage*, *30*(1), 254–265. https://doi.org/10.1016/j.neuroimage.2005.09.013

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3–25. https://doi.org/https://doi.org/10.1080/00335558008248231

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using jasp. *BMC psychiatry*, *18*, 1–8. https://doi.org/10.1186/s12888-018-1761-4

Rajimehr, R. (2004). Unconscious orientation processing. *Neuron*, *41*, 663–673. https://doi.org/10.1016/S0896-6273(04)00041-8

Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature neuroscience*, *7*(1), 85–90. https://doi.org/10.1038/nn1168

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature reviews neuroscience*, *5*(3), 184–194. https://doi.org/10.1038/nrn1343

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the cognitive sciences*, *3*, 1–23. https://doi.org/10.1023/B:PHEN.0000041900.30172.e8

Raymond, J. E., & O'Brien, J. L. (2009). Selective visual attention and motivation: The consequences of value learning in an attentional blink task. *Psychological Science*, *20*(8), 981–988. https://doi.org/10.1111/j.1467-9280.2009.02391.x

Redish, A. D., Abram, S. V., Cunningham, P. J., Duin, A. A., Cuttoli, R. D.-d., Kazinka, R., Kocharian, A., MacDonald, A. W., Schmidt, B., Schmitzer-Torbert, N., Thomas, M. J., & Sweis, B. M. (2022). Sunk cost sensitivity during change-of-mind decisions is informed by both the spent and remaining costs. *Communications Biology*, *5*. https://doi.org/10.1038/s42003-022-04235-6

Reeder, R. R., Hanke, M., & Pollmann, S. (2017). Task relevance modulates the cortical representation of feature conjunctions in the target template. *Scientific Reports*, *7*(1), 4514. https://doi.org/10.1038/s41598-017-04123-8

Reuter-Lorenz, P. A. (2002). New visions of the aging mind and brain. *TRENDS in Cognitive Sciences*, *6*. https://doi.org/10.1016/S1364-6613(02)01957-5

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185.

Richardson, A. M., & Welsh, A. H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics*, 1429–1439. https://doi.org/10.2307/2533273

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447. https://doi.org/10.1126/science.1100301

Roelfsema, P. R., Lamme, V. A., Spekreijse, H., & Bosch, H. (2002). Figure—ground segregation in a recurrent network architecture. *Journal of cognitive neuroscience*, *14*(4), 525–537. https://doi.org/10.1162/08989290260045756

Rosenbaum, D. A., Slotta, J. D., Vaughan, J., & Plamondon, R. (1991). Optimal movement selection. *Psychological Science*, *2*(2), 86–91. https://doi.org/10.1111/j.1467-9280.1991.tb00106.x

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*, 389–397. https://doi.org/10.1038/nn2066

Rushworth, M. F., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: One or many serial or parallel systems? *Current Opinion in Neurobiology*, *22*, 946–955. https://doi.org/10.1016/j.conb.2012.04.011

Sarkisov, S., Filimonoff, I., & Preobrashenskaya, N. (1949). Cytoarchitecture of the human cortex cerebri. *Moscow: Medgiz*.

Schurz, M., Tholen, M. G., Perner, J., Mars, R. B., & Sallet, J. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: A review using probabilistic atlases from different imaging modalities. *Human Brain Mapping*, *38*(9), 4788–4805. https://doi.org/https://doi.org/10.1002/hbm.23675

Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: A comparative study of area 10. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *114*(3), 224–241. https://doi.org/10.1002/1096-8644(200103)114:3<224::AID-AJPA1022>3.0.CO;2-I

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in cognitive sciences*, *10*(1), 38–45. https://doi.org/10.1016/j.tics.2005.11.008

Shomstein, S., & Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *Journal of neuroscience*, *24*(47), 10702–10706. https://doi.org/10.1523/JNEUROSCI.2939-04.2004

Shulman, G. L., Astafiev, S. V., McAvoy, M. P., d'Avossa, G., & Corbetta, M. (2007). Right tpj deactivation during visual search: Functional significance and support for a filter hypothesis. *Cerebral cortex*, *17*(11), 2625–2633. https://doi.org/10.1093/cercor/bhl170

Shulman, G. L., McAvoy, M. P., Cowan, M. C., Astafiev, S. V., Tansy, A. P., d'Avossa, G., & Corbetta, M. (2003). Quantitative analysis of attention and detection signals during visual search. *Journal of neurophysiology*, *90*(5), 3384–3397. https://doi.org/10.1152/jn.00343.2003

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in cognitive sciences*, *13*(11), 488–495. https://doi.org/https://doi.org/10.1016/j.tics.2009.08.005

Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and prediction error in medial frontal cortex: Integrating the single-unit and systems levels of analysis. *Frontiers in human neuroscience*, *5*, 75. https://doi.org/10.3389/fnhum.2011.00075

Smith, A. M., Lewis, B. K., Ruttimann, U. E., Frank, Q. Y., Sinnwell, T. M., Yang, Y., Duyn, J. H., & Frank, J. A. (1999). Investigation of low frequency drift in fmri signal. *Neuroimage*, *9*(5), 526–533. https://doi.org/10.1006/nimg.1999.0435

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Snodgrass, M., Bernat, E., & Shevrin, H. (2004). Unconscious perception: A model-based approach to method and evidence. *Perception and Psychophysics*, *66*, 846–867. https://doi.org/10.3758/BF03194978

Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology*, *21*, R912–R913. https://doi.org/10.1016/j.cub.2011.09.049

Soto, D., Sheikh, U. A., & Rosenthal, C. R. (2019). A novel framework for unconscious processing. *Trends in Cognitive Sciences*, *23*(5), 372–376. https://doi.org/10.1016/j.tics.2019.03.002

Soto, D., & Silvanto, J. (2014). Reappraising the relationship between working memory and conscious awareness. *Trends in cognitive sciences*, *18*(10), 520–525. https://doi.org/10.1016/j.tics.2014.06.005

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stein, T., Kaiser, D., Fahrenfort, J. J., & Van Gaal, S. (2021). The human visual system differentially represents subjectively and objectively invisible stimuli. *PLoS biology*, *19*(5), e3001241. https://doi.org/10.1371/journal.pbio.3001241

Stein, T., Kaiser, D., & Hesselmann, G. (2016). Can working memory be non-conscious? *Neuroscience of Consciousness*, *2016*, niv011. https://doi.org/10.1093/nc/niv011

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (mvpa): Random permutations and cluster size control. *Neuroimage*, *65*, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063

Strotzer, M. (2009). One century of brain mapping using brodmann areas. *Clinical Neuroradiology*, *19*(3), 179. https://doi.org/10.1007/s00062-009-9002-3

Sweis, B. M., Abram, S. V., Schmidt, B. J., Seeland, K. D., MacDonald III, A. W., Thomas, M. J., & Redish, A. D. (2018). Sensitivity to "sunk costs" in mice, rats, and humans. *Science*, *361*(6398), 178–181. https://doi.org/10.1126/science.aar8644

Swets, J. A. (2014). *Signal detection theory and roc analysis in psychology and diagnostics: Collected papers*. Psychology Press.

Team, J. (2019). Jasp (version 0.10. 1)[computer software]. *(No Title)*.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*, 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2018). Visual selection: Usually fast and automatic; seldom slow and volitional. *Journal of Cognition*, *1*, 1–15. https://doi.org/10.5334/joc.13

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97–101. https://doi.org/10.1016/j.copsyc.2018.12.024

Tomov, M. S., Truong, V. Q., Hundia, R. A., & Gershman, S. J. (2020). Dissociable neural correlates of uncertainty underlie different exploration strategies. *Nature communications*, *11*(1), 2371. https://doi.org/10.1038/s41467-020-15766-z

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Tsujimoto, S., Genovesio, A., & Wise, S. P. (2010). Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nature neuroscience*, *13*(1), 120–126. https://doi.org/10.1038/nn.2453

Tsujimoto, S., Genovesio, A., & Wise, S. P. (2011). Frontal pole cortex: Encoding ends at the end of the endbrain. *Trends in cognitive sciences*, *15*(4), 169–176. https://doi.org/10.1016/j.tics.2011.02.001

Tsujimoto, S., Genovesio, A., & Wise, S. P. (2012). Neuronal activity during a cued strategy task: Comparison of dorsolateral, orbital, and polar prefrontal cortex. *Journal of Neuroscience*, *32*(32), 11017–11031. https://doi.org/10.1523/JNEUROSCI.1230-12.2012

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 195. https://doi.org/10.1037/0096-1523.35.1.195

Vallat, R. (2018). Pingouin: Statistics in python. *J. Open Source Softw.*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

Van Gaal, S., & Lamme, V. A. (2012). Unconscious high-level information processing: Implication for neurobiological theories of consciousness. *Neuroscientist*, *18*, 287–301. https://doi.org/10.1177/1073858411404079

Van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. (2008). Frontal cortex mediates unconsciously triggered inhibitory control. *Journal of Neuroscience*, *28*(32), 8053–8062. https : / / doi . org / 10 . 1523 / JNEUROSCI.1278-08.2008

Van Gaal, S., Ridderinkhof, K. R., Scholte, H. S., & Lamme, V. A. (2010). Unconscious activation of the prefrontal no-go network. *Journal of neuroscience*, *30*(11), 4143–4150. https://doi.org/10.1523/JNEUROSCI.2992-09.2010

von Economo, C. F., & Koskinas, G. N. (1925). *Die cytoarchitektonik der hirnrinde des erwachsenen menschen*. J. Springer.

Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *Neuroscientist*, *20*, 150–159. https://doi.org/10.1177/1073858413494269

Wager, T. D., Jonides, J., & Reading, S. (2004). Neuroimaging studies of shifting attention: A meta-analysis. *Neuroimage*, *22*(4), 1679–1693. https://doi.org/10.1016/j.neuroimage.2004.03.052

Wajnberg, E., Fauvergue, X., & Pons, O. (2000). Patch leaving decision rules and the marginal value theorem: An experimental analysis and a simulation model. *Behavioral Ecology*, *11*(6), 577–586. https://doi.org/10.1093/beheco/11.6.577

Weidner, R., Pollmann, S., Müller, H. J., & von Cramon, D. Y. (2002). Top-down controlled visual dimension weighting: An event-related fmri study. *Cerebral Cortex*, *12*(3), 318–328. https://doi.org/10.1093/cercor/12.3.318

Wiens, S. (2007). Concepts of visual consciousness and their measurement. *Advances in Cognitive Psychology*, *3*, 349–359. https://doi.org/10.2478/v10053-008-0035-y

Wilke, A. (2006). *Evolved responses to an uncertain world* (Doctoral dissertation). https://doi.org/10.17169/refubium-7570

Wilke, A., Hutchinson, J. M., Todd, P. M., & Czienskowski, U. (2009). Fishing for the right words: Decision rules for human foraging behavior in internal search tasks. *Cognitive Science*, *33*, 497–529. https://doi.org/10.1111/j.1551-6709.2009.01020.x

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074. https://doi.org/10.1037/a0038199

Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision research*, *34*(9), 1187–1195. https://doi.org/10.1016/0042-6989(94)90300-X

Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? foraging rules in human visual search. *Journal of Vision*, *13*. https://doi.org/10.1167/13.3.10

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fmri data. *Neuroimage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Worsley, K. J. (2001). Statistical analysis of activation images. *Functional MRI: An introduction to methods*, *14*(1), 251–270.

Zajkowski, W. K., Kossut, M., & Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *Elife*, *6*, e27430. https://doi.org/10.7554/eLife.27430

Zhang, X., & Fang, F. (2012). Object-based attention guided by an invisible object. *Experimental brain research*, *223*, 397–404. https://doi.org/10.1007/s00221-012-3268-4

Zhang, Z., & Yuan, K.-H. (2018). *Practical statistical power analysis using webpower and r*. Isdsa Press.

Zinchenko, A., Conci, M., Taylor, P. C., Müller, H. J., & Geyer, T. (2018). Taking attention out of context: Frontopolar transcranial magnetic stimulation abolishes the formation of new context memories in visual search. *Journal of Cognitive Neuroscience*, *31*, 442–452. https://doi.org/10.1162/jocn_a_01358

A Experiment 1: Supplementary material

A.1 Model selection in the LMM analysis of RT data

To determine the final random effect structure of the LMM used to fit the RT data, we conducted likelihood ratio tests (Crainiceanu & Ruppert, 2004). First, we defined a model with visual awareness (AL1 - 3) and the change of the stimulus orientation (switch versus repeat) as the two fixed effects, we additionally added an interaction term of these two fixed effects and defined a by-subjects random intercept to account for non-independency of the data in the repeated measure within design, i.e., for baseline differences in RTs across subjects. This resulted in the following basic model: RT $\sim$ *awareness* + *switch* + *awareness:switch* + (1 | *subject*). This model was compared to a second model containing an additional by-subject random slope for awareness to model differing responses to the main factor awareness since subjects may cope differently with very low stimulus visibility. This by-subject random slope model was thus defined as follows: RT $\sim$ *awareness* + *switch* + *awareness:switch* + (1 + *awareness* | *subject*). Finally we defined a third alternative model with a more complex random effect structure by additionally entering a second by-subject random slope of switch to model differing response due to changes in the stimulus orientation (RT $\sim$ *awareness* + *switch* + *awareness:switch* + (1 + *awareness* + *switch* | *subject*)) as well as a model containing a by-subject random slope for switch only (RT $\sim$ *awareness* + *switch* + *awareness:switch* + (1 + *switch* | *subject*)).

Making use of the likelihood ratio test as implemented in the anova function of the R-package *lme4* (Bates et al., 2014) showed for the weighted RT data that the model containing the additional by-subject random slope for awareness significantly improved the overall model fit compared to the basic model containing only a by-subject random intercept, χ^2 (5, N = 13) = 18.525, p = .002. Next, we compared the model with a by-subject random slope for awareness with a third model containing the additional by-subject random slope for switch. Here the likelihood ratio test suggested that the model with the additional by-subject random slope for switch did not better fit the RT data than the more parsimonious model containing the by-subject random slope for awareness only, χ^2 (4, N = 13) = 1.533, p = .821. Similarly comparing the model with the by-subject random slope for switch only with the basic by-subject random intercept model did not improve the overall fit, χ^2 (2, N = 13) = 0.061, p = .970. Therefore we selected the model with a by-subject random intercept and a by-subject random slope for awareness (RT $\sim$ *awareness* + *switch* + *awareness:switch* + (1 + *awareness* | *subject*)) for hypothesis testing. In a previous study on unconscious response priming that also used a variant of the PAS (Ramsøy & Overgaard, 2004), a similar mixed model with an identical random effect structure was defined to analyze the unbalanced RT data (Avneon & Lamy, 2018).

We repeated the procedure outlined above also for the average RT model. Here the likelihood ratio test similarly showed that a model including a by-awareness random slope showed a better fit than the basic model with only a by-subject random intercept, χ^2 (5, N = 13) = 20.95, p < .001. Again, a model with an additional by-subject random slope for switch did not improve the overall fit compared to the

model with only a by-subject random slope for awareness, $\chi^2(4, N = 13) = 2.303$, $p = .680$. The same was true when comparing a model with a by-subject random slope for switch to the basic by-subject random intercept model, $\chi^2\ (2, N = 13) = 0.051$, $p = .975$. Hence also for the average RT model we chose the model with a by-subject random intercept and a by-subject random slope for awareness for final data fitting.

A.2 Detailed trial information

The median and the range of the number of trials for the average switch, the weighted switch, and the repeat condition for each level of subjective awareness (awareness level (AL) 1-3) are reported in the supplementary Table 12. Violin plots of the group distributions of the number of trials for each condition are depicted in Figure 19. In total 4.3% of all trials, we obtained for the weighted switch (i.e., orientation changes away from the frequent tilt) and 5.6% for the repeat condition rated as fully unaware (AL1). 7.4% of all trials were weighted switch trials with residual awareness (AL2) and 12.6% were AL2 repeat trials. 8.5% of the trials were weighted switch trials rated as almost fully aware (AL3), while 10.9% of all trials were AL3 repeat trials. The majority of trials were obtained for the average switch condition with 10.4% of all trials rated as fully unaware, 20.5% with residual awareness, and 19.8% rated as almost fully aware.

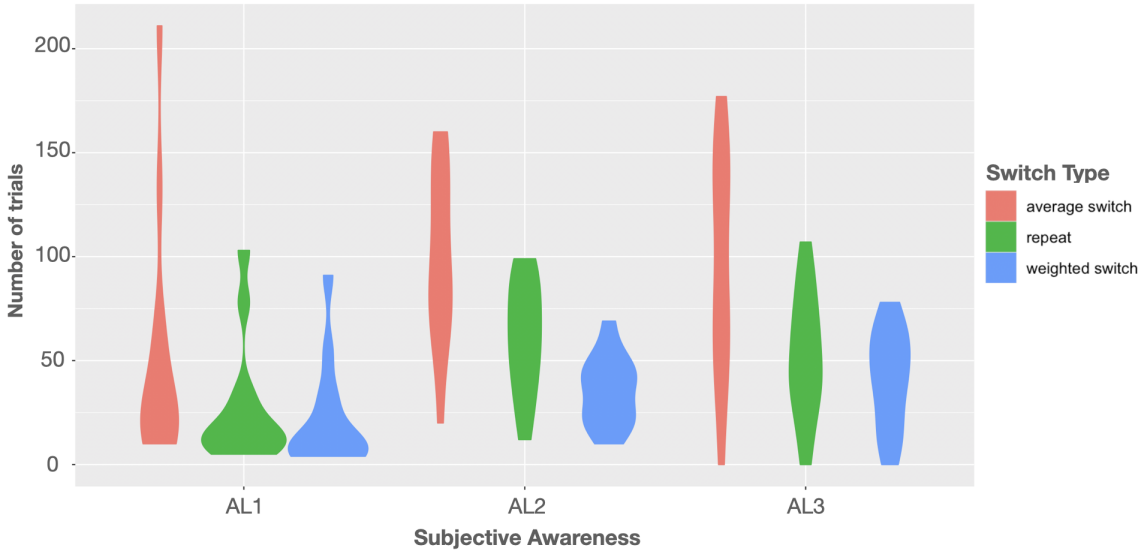


Figure 19: Violin plots show the number of trials for the average switch, the weighted switch, and the repeat condition as a function of subjective awareness (AL 1-3). Violin plots use density curves to depict distributions of numeric data. The width corresponds with the approximate frequency of data points in each region. The lower and upper limits of each plot is determined by the minimum and maximum value.

Table 12: *Medians and range (R = maximum - minimum) of the number of trials obtained for the switch and repeat conditions for each level of subjective awareness*

	AL1		AL2		AL3	
	Median	R	Median	R	Median	R
weighted switch	10	87	39	59	40	78
average switch	26	201	90	140	83	177
repeat	14	98	59	87	47	107

Note: AL1 = awareness level 1, (i.e. PAS rating = 1), AL2 = awareness level 2, AL3 = awareness level 3. PAS = perceptual awareness scale used to assess the subjective level of visual awareness.

A.3 Control analyses

The descriptive data clearly showed uneven proportions of trial numbers across conditions with more trials for the partially aware and almost fully aware switch and repeat condition. Thus, the means that we calculated based on sometimes only a few trials for the fully unconscious condition could have been more strongly affected by outliers and could be less reliable estimates of the true values so that the observed effect could reflect an artefact elicited by noisy data. To rule out this possibility, we already stated in the manuscript that we first repeated the analysis using a 2 SD and 2.5 SD cutoff to see if the results would be preserved with a more rigid cutoff, which was indeed the case. In addition, we conducted a control analysis in which we matched the number of trials by randomly sampling from AL2 and AL3 trials the same amount as AL1 trials and used these matched random selections per subject to repeat the mixed model analysis. The results were the same as before: we found a significant fixed effect of *switch*, $F(1,33.982) = 4.481$, $p = .042$, and awareness, $F(2, 11.706) = 13.336$, $p < .001$. Yet, paired comparisons again showed only for AL1 trials that RTs in repeat trials were significantly faster compared RTs in switch trials, $t(33.13) = -2.374$, $p = .024$, 95% CI [-187.5 ms, -14.4 ms], but not in AL2, $p = .953$, nor in AL2 trials, $p = .1753$. This observation clearly contradicts the possibility that the switch effect occurred randomly due to a low number of trials. We also tested the direct relation between switch costs obtained in the weighted RT model and the number of trials aiming to prove that the effect did not depend on the number of trials obtained for individual subjects. Since this analysis required the verification of the absence of an association, we used a Bayesian based linear mixed model. First, we constructed a model with a by-subject random intercept only indicating that all variance in the observed switch costs was explained by interindividual differences only. The alternative model additionally included the number of trials obtained for switch trials for each level of awareness as a single fixed effect. Next, using the R-package *BayesFactor* and its *lmBF* function (Morey & Rouder, 2018), we calculated Bayes factors for each model and divided these two factors to obtain a BF_{10} that would favor either the null model or the model that included the number of trials as a predictor. This analysis resulted in a $BF_{10} = 0.472$ with anecdotal evidence rather in favor of the null model (Quintana & Williams, 2018) which suggests an association

between the observed switch effect in AL1 trials for the weighted RT model and the amount of trials available for analysis in this condition was rather unlikely.

Next, we report accuracies as well as the level of subjective awareness of those trials that preceded an orientation change (i.e., "pretarget" trial information). In total 8.6% of all preceding trials were rated as subjectively fully unaware in which participants gave a correct response in the orientation discrimination task. Incorrect responses were given in 9.2% of all preceding trials rated as subjectively fully unaware. In 27.6% of all trials prior to an orientation change in which volunteers reported residual awareness of the stimulus orientation (AL2), correct responses were given, while in 7.5% of these preceding AL2 trials subjects made incorrect responses. In 29.4% of trials rated as almost fully aware (AL3) volunteers correctly discriminated the target orientation prior to a switch. Only 4% of these preceding AL3 trials were error trials. Finally, 9.7% of the trials were rated as fully aware and correct answers were given and only in 0.4% of these AL4 trials volunteers gave incorrect responses. Note that AL4 trials were not included in the analyses due to the very low number of trials. Table 13 summarizes the mean number of trials for correctly and incorrectly performed trials preceding an orientation change for each level of awareness.

Table 13: *Number of trials (mean ± standard error of the mean) for correct and error trials prior to an orientation change for each level of subjective awareness*

	AL1	AL2	AL3	AL4
correct	23 ±7	76 ±10	81 ±12	27 ±7
error	25 ±10	21 ±2	11 ±3	1 ±0.4

A.4 Accuracies and signal detection analysis data

The ability to correctly discriminate the target orientations (vertical versus non-vertical) served as the objective measure of visual awareness of the target orientation. As reported in the manuscript, we chose signal detection theory measures instead of simple accuracies (1-errors) to analyze volunteers' task performance in the orientation discrimination task. This we did because the unbalanced design and the frequency differences of the three target orientations would have made it difficult to determine the theoretical chance level for statistical testing of the accuracies on group level. The theoretical chance level of the sensitivity (A'), however, remains unaffected by frequency differences and uneven proportions of data across conditions so that the discrimination performance on group level could be readily tested against chance using this measure. Moreover, accuracies confound the effect of sensitivity and bias on behavior, while they can be measured separately using signal detection theory (e.g., Lynn & Barret, 2014). Still, to provide a full picture of the data here, we report accuracies (1- error rates) for each condition and for the weighted

and average data set.

Descriptive means of accuracies revealed an increase in task performance with increasing subjective awareness with only marginal differences between switch and repeat trials: in fully unconscious trials (AL1) the average switch model, comprising all switch trials and repeat trials, the mean accuracy was 55.93 ± 24.55% (SD) in repeat and 53.05 ± 18.97% in switch trials. In trials with residual visual awareness (AL2) the accuracy was 73.03 ± 13.29% in repeat and 77.10 ± 10.40% in switch trials. In trials rated as almost fully aware the mean accuracy in repeat trials was 90.39 ± 8.9% and in switch trials 88.56 ± 7.06%.

A similar pattern of behavioral performance was obtained for the weighted switch model in which the switch condition contained only orientation changes away from the heavily weighted tilt: the mean accuracy in fully unaware (AL1) weighted switch trials was 48.04 ± 17.85. For AL2 trials we found a mean accuracy of 77.97 ± 8.5% in switch trials. Again, highest accuracies were found in almost fully aware weighted switch trials (AL3) with a mean accuracy of 86.93 ± 7.01%. Note that the repeat trials in this weighted switch model were the same as in the average model. Figure 20 a) depicts boxplots of accuracies on group level for average switch and repeat trials for each level of awareness. Figure 20 b) shows accuracies on group level for the weighted switch and repeat trials for each level of awareness.

In addition to the accuracy data the average rates of hits (H), false alarms (FA), correct rejections (CR), and misses (M) for each level of subjective awareness are reported in Table 14 as well as the mean number of hit, false alarm, miss, correct rejection trials (Table 15).

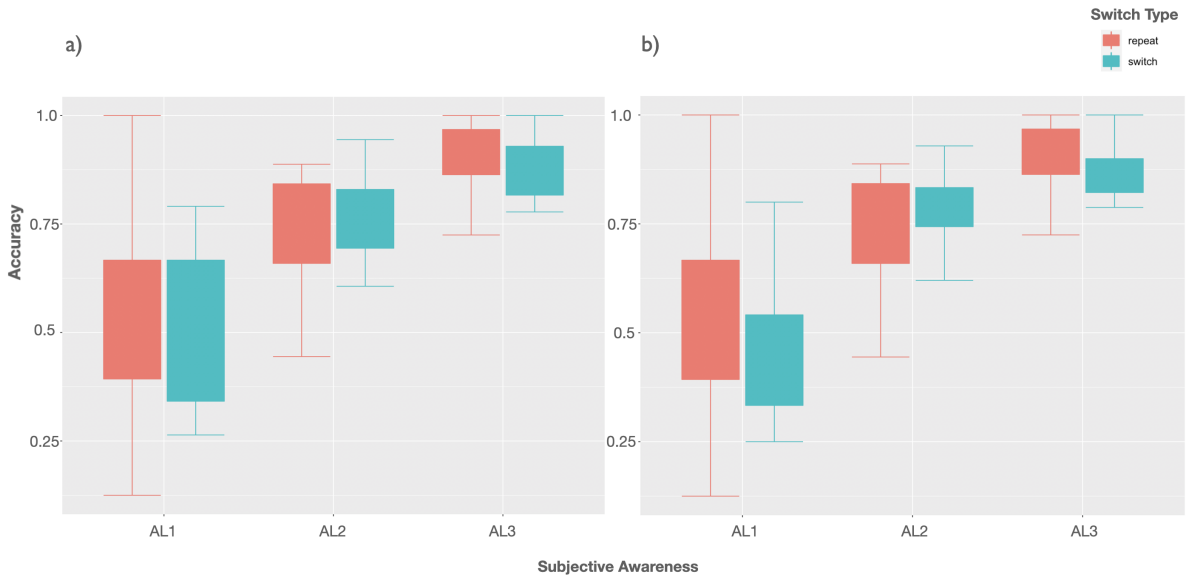


Figure 20: Accuracies (1-error rate) for the weighted switch, and the repeat condition in a), as well as for the average switch condition b) as a function of subjective awareness (AL 1-3).

Table 14: *Confusion matrix of the signal detection analysis*

Response	**Stimulus**	
	Tilted Gabor AL1	Vertical Gabor AL1
non-vertical	H = 0.579 ± 0.066	FA = 0.538 ± 0.066
vertical	M = 0.421 ± 0.066	CR = 0.461 ± 0.066
	Tilted Gabor AL2	Vertical Gabor AL2
non-vertical	H = 0.788 ± 0.029	FA = 0.261 ± 0.040
vertical	M = 0.212 ± 0.029	CR = 0.739 ± 0.040
	Tilted Gabor AL3	Vertical Gabor AL3
non-vertical	H = 0.849 ± 0.037	FA = 0.134 ± 0.038
vertical	M = 0.151 ± 0.037	CR = 0.866 ± 0.038

Note: H = Hit rate = hits/(hits + false alarms); FA = False alarm rate = false alarms/(hits + false alarms); M = Misses rate = misses/(misses + correct rejections); CR = Correct rejection rate = correct rejections/(misses + correct rejections). The matrix shows the averaged rates (M) and standard errors (±SE) of hits, false alarms, correct rejections and misses for each level of subjective awareness on group level.Mean values were obtained by calculating the rates for right- and left-weighted blocks separately and consequently averaging these rates across all blocks and subjects.

Table 15: *Average number of trials (M ± SE) of hits, misses, false alarms, and correct rejections across left- and right-weighted blocks*

	H (correct non-vertical)	M (incorrect non-vertical)	FA (incorrect vertical)	CR (correct vertical)
AL1	26.3 ±8.0	25.8 ±9.6	15.0 ±7.0	13.2 ±5.5
AL2	90.0 ±11.8	24.7 ±4.6	9.9 ±1.4	35.6 ±5.9
AL3	84.9 ±14.3	12.6 ±3.8	4.2 ±1.5	48.8 ±8.1

Note: H = Hit rate = hits/(hits + false alarms); FA = False alarm rate = false alarms/(hits + false alarms); M = Misses rate = misses/(misses + correct rejections); CR = Correct rejection rate = correct rejections/(misses + correct rejections).

B Experiment 2: Supplementary material

B.1 Analysis of low-level stimulus contrasts

When participants performed the task the luminance contrast of the stimulus in the upcoming trial was manipulated based on the current awareness rating to obtain a balanced amount of trials for each level of subjective awareness: if the current stimulus was reported as invisible, the contrast of the following grating was increased while it was decreased following ratings indicating residual or almost full visual awareness of the stimulus. On the downside this online manipulation could have introduced differences in the low-level physical intensities (i.e., luminance contrasts) of the gratings between the levels of subjective awareness and within these levels over time. Then this manipulation could mean a confound especially when the three levels of subjective awareness were compared.

To test this possible confound we again used a linear mixed model (LMM). The luminance contrast between stimulus and background (0-1) was defined as the dependent variable and the subjective awareness (AL1-AL3) as well as the number of Blocks (1-10) served as fixed effects. To allow for possible heteroscedasticity with respect to levels of the two fixed effect factors, subjective awareness and time (i.e., number of experimental blocks) we entered by-subject random slopes for the two. This resulted in a final model defined as *luminance* $\sim$ *awareness* + *block* + *awareness*:*block* + (1 + *awareness* + *block* | *sub*). Residual plots did not suggest deviations from homoscedasticity or normality. The estimated luminance values appeared to be affected by changes in the level of visual awareness indicated $F(2, 186.03) = 6.547$, $p = .002$. The post-hoc tests with Bonferroni correction indicated that the mean stimulus contrast of 0.0601 ± 0.0139 in AL1 trials was on average 0.0096 ± 0.003 lower compared to the mean contrast of 0.0697 ± 0.0153 in AL2 trials, $t(24.6) = -3.226$, $p = .011$, 95% CI (-0.0172, -0.00195) and on average -0.02121 ± 0.00318 lower compared to the mean contrast value of 0.0813 ± 0.0152 in AL3 trials, $t(13.3) = -6.677$, $p < .0001$, 95% CI (-0.0299, -0.01252). Finally, the contrast values in AL2 trials were on average -0.01162 ± 0.00263 significantly lower than those contrast values of AL3 trials, $t(36.9) = -4.422$, $p < .001$, 95% CI (-0.0182, -0.00503). Nor the fixed effect of *block* was significant, nor the interaction between the *awareness* and *block*, with $F(1,13) = 1.558$, and $F(2,372.32) = 0.106$, respectively. Taken together this analysis showed an increase of the stimulus contrast with increasing visual awareness. However, there were no significant fluctuations of the contrast values within the three awareness conditions. Figure 21 shows the contrast values for the three awareness levels as a function of block (i.e., time).

We are confident that these differences in the luminance contrast between AL1-AL3 are of no concern with respect to the searchlight analysis since we did not contrast between the levels of visual awareness and within these levels the luminance was constant, i.e., the control analysis showed no evidence for fluctuations over time meaning that contrast values were stable within the three awareness conditions. Regarding our univariate analysis, however, we contrasted awareness levels thereby comparing trials in which the stimulus intensity varied. Yet, it has been shown that

an increasing luminance contrast increases the BOLD signal at least in V1 (e.g., Chen et al., 2021; Goodyear & Menon, 1998). Thus, our findings of increased activity in the invisible condition as well as the increased BOLD response following invisible target changes are both difficult to explain with the differences in the physical stimulus properties because we found higher activity for the lower-contrast stimuli. This observation is better explained by an attention mechanism rather than by a difference in the luminance contrast.

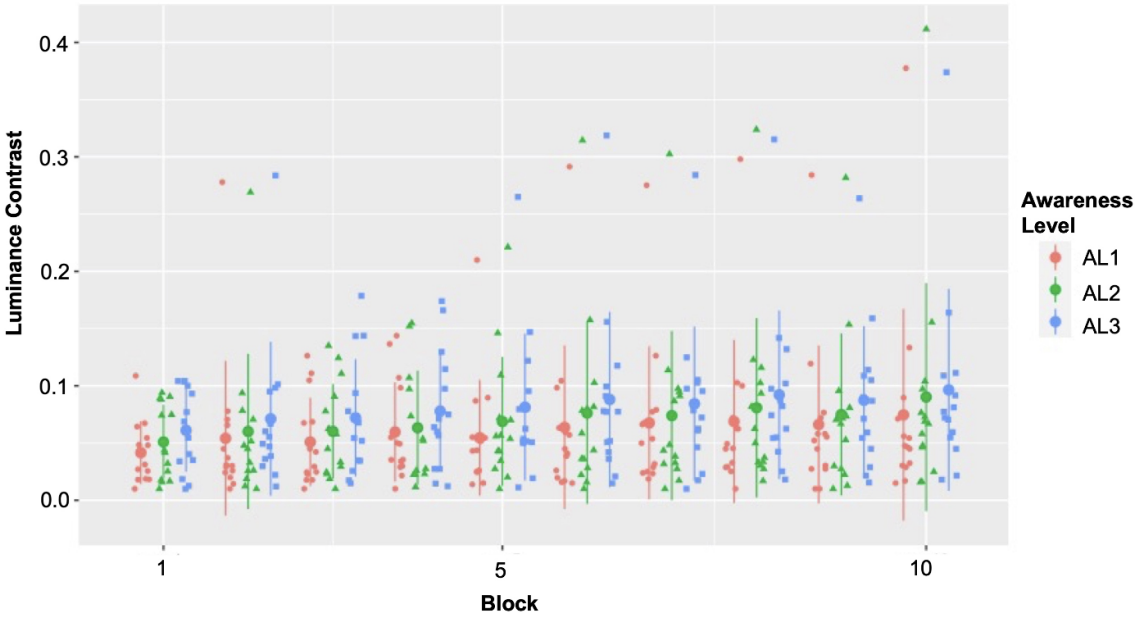


Figure 21: Luminance contrast values of the Gabor patch as a function of time for each of the three levels of awareness (AL1-AL3). Whereas the stimulus contrast remained stable over time, it increased with increasing visual awareness

B.2 Results of the ROI-based searchlight analysis using a 6 mm searchlight radius (SLR)

To test whether the informative clusters have a similar appearance across different searchlight radii (Etzel et al., 2013), we repeated the multivariate analysis using a 6 mm SLR in addition to the 9 mm SLR. Again, we used only left-versus right-tilted Gabors and included only trials in which the target was not perceived consciously (AL1) for decoding. No other parameters were changed therefore all methodological details match those reported in the manuscript (see the section 4.3.5). Most importantly also in this analysis we found clusters on the group level with significant searchlight centers ($p_{\text{Cluster}} < .05$) located in the right angular (AG), the superior parietal gyrus as well as the rLOC. Again, consistent the previous searchlight analysis we also observed clusters in the rMFG as well as in the right precentral gyrus. All together the clusters had a mean decoding accuracy of 65.9 ± 2.8%. The dot plot showing the decoding accuracies of all clusters that survived the significance testing by bootstrapping on group level the ROI is shown in Figure

22 (box on the right). The accuracies mapped on the MNI 152 template brain are depicted in Figure 22 on the left.

We additionally carried out a whole-brain searchlight analysis using a 6 mm searchlight radius (number of non-zero voxels = 221432). The results confirmed those obtained of the previous whole-brain searchlight using a 9 mm sphere diameter reported in the manuscript, showing clusters with significant searchlight centers ($p < .05$) at matching locations bilaterally in medial and lateral FPC, MFG, IFJ, TPJ, precuneus, and IPS, as well as in visual cortex including the lingual gyrus, cuneus, and occipital pole. Again, we found group clusters with local signals distinguishing between left- and right-tilted orientations in anterior cingulate gyrus, bilaterally in the temporal pole and OFC. The mean decoding accuracy across all clusters on group level with significant searchlight centers was 71.3 ± 3.5%. Figure 23 depicts the accuracy maps of those clusters surviving the two-step permutation test on group level for the 6 mm SLR.

Taken together these findings confirm the results of the prior analysis using a 9 mm SLR in that those regions that had shown an increased BOLD response in response to invisible target changes also represented the orientation information of the invisible targets. Among these areas were the right FPC and MFG as well as parietal regions partially comprising the right TPJ, such as the right AG and right SPL.

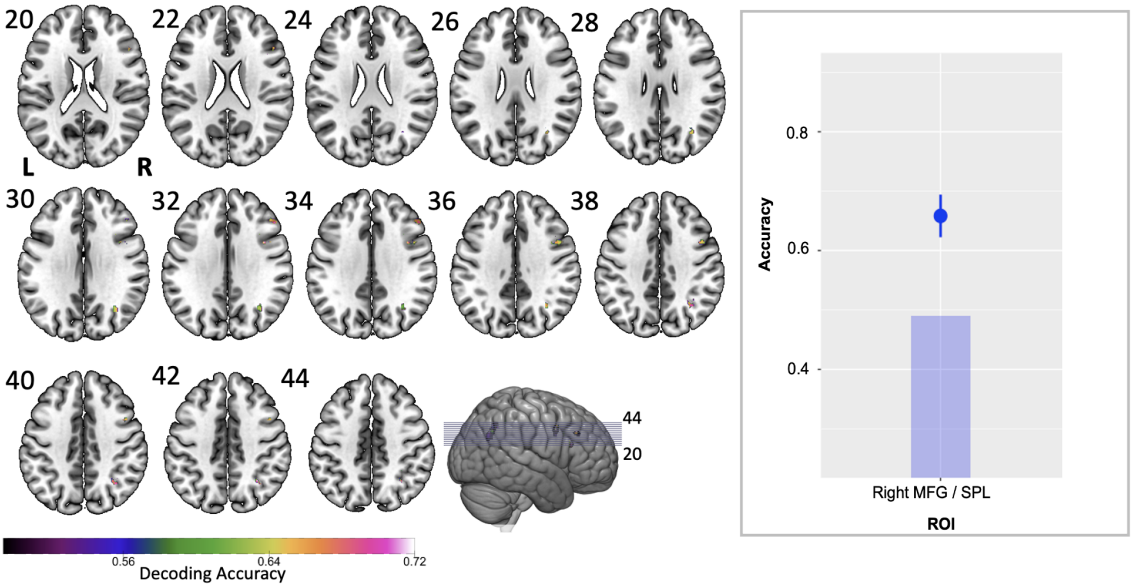


Figure 22: Decoding accuracies for the ROI-based searchlight analysis using a 6 mm radius mapped on a MNI 152 standard brain. Only those clusters with significant searchlight centers with accuracies exceeding the 99th percentile (49.5%) of the null distribution are shown with locations in right MFG, PCG, right AG, SPL, and LOC.

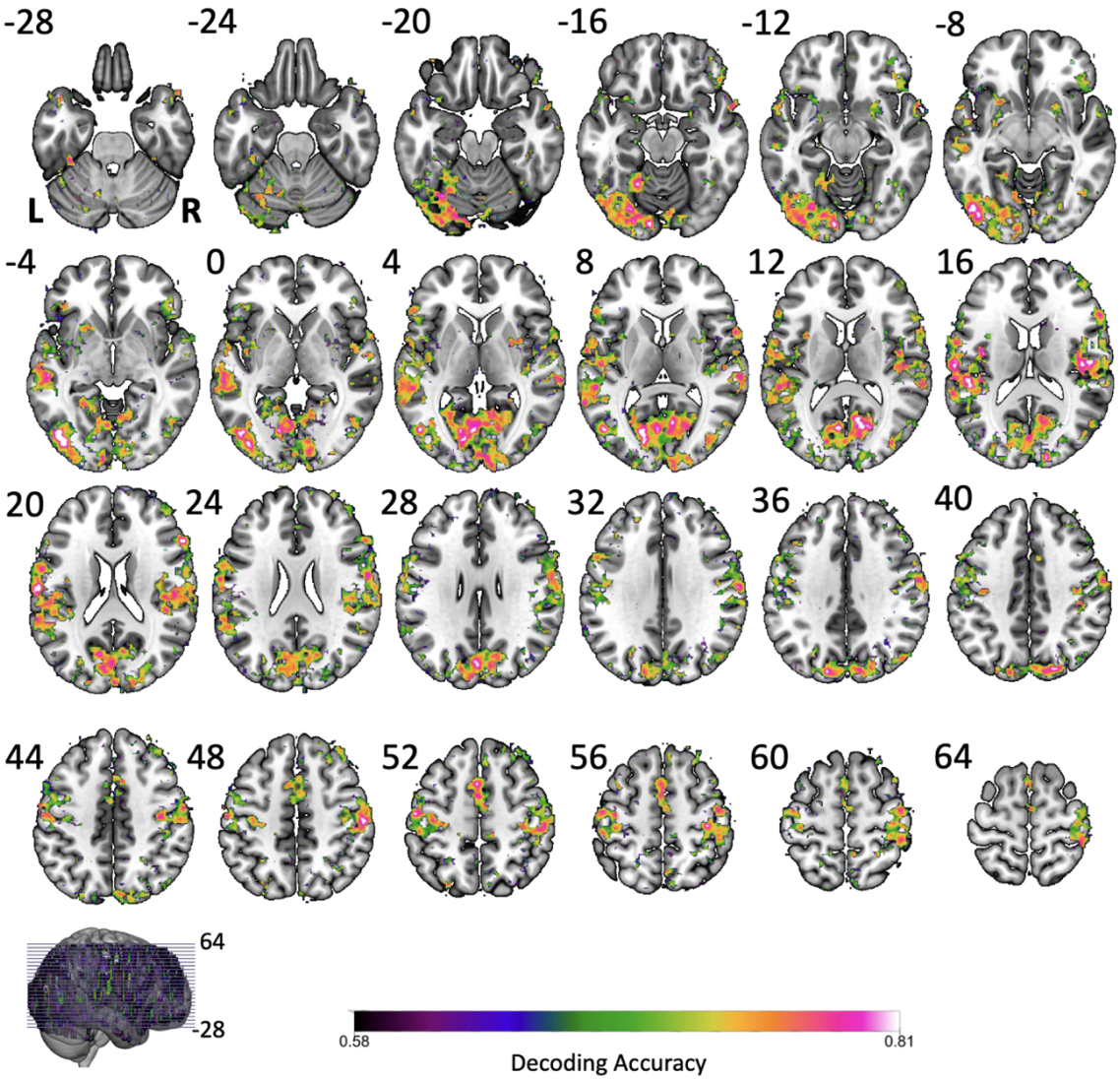


Figure 23: Decoding accuracies obtained in the whole-brain searchlight analysis using a 6 mm searchlight radius mapped on a MNI 152 standard brain, mostly replicating the findings of the 9mm SLR analysis.(Only clusters with above chance accuracies are shown, 99th percentile = 58.8%).