Neural Dynamics in Crossmodal Attention: Investigating a Modality-Weighting Account in Visuo-Tactile Search

Jan Nasemann

Graduate School of
Systemic Neurosciences
LMU Munich

Dissertation at the
Graduate School of Systemic Neurosciences
Ludwig-Maximilians-Universität München

1st February 2022

Supervisor
Prof. Dr. Zhuanghua Shi
Department of Experimental Psychology
Ludwig-Maximilians-Universität München

First Reviewer:	Prof. Dr. Zhuanghua Shi
Second Reviewer:	Prof. Dr. Hermann J. Müller
External Reviewer	Assoc. Prof. Dr. Anna Grubert

Date of Submission:	01.02.2022
Date of Defense:	08.09.2022

Acknowledgments

Without the many people in my life, I would only be half a person and scientist. First, I would like to thank my supervisors for their incredible and extraordinary support: Prof. Zhuanghua Shi, Prof. Hermann Müller, and PD Thomas Töllner. I am very grateful for everything I have learned from you! Second, I would also like to thank my TAC-Team, in particular Prof. Kathrin Koch and Prof. Paul Sauseng, for turning my scientific exercises and doubts into wonderful discussions. Third, a special thanks goes of course to the entire GSN Faculty and Staff, and Stefanie Bosse in particular, for giving me the opportunity to join the PhD program and being always there for me. Another thanks goes to my dear colleagues, I will have fond memories of all our wonderful conversations. Finally, I would like to thank my parents Bettina and Klaus and especially my beloved partner Juliana for her outstanding support and trust in me.

Table of Contents

List of Figures

Nomenclature

AEP	Auditory Evoked Potential
CCN	Central Contralateral Negativity (or N140cc)
CDA	Contralateral Delay Activity
CMA	Computational Model of Attention
DWA	Dimension-Weighting Account
EEG	Electroencephalography
ER	Error Rate(s)
ERL	Event-Related Lateralizations
ERP	Event-Related Potential
FEF	Frontal Eye Field
fMRI	functional Magnetic Resonance Imaging
IPL	Inferior Parietal Lobe
IPS	Intraparietal Sulcus
LGN	Lateral Geniculate Nucleus
LIP	Lateral Intraparietal Area
MSE	Modality-Shift Effect
MWA	Modality-Weighting Account
PCN	Posterior Contralateral Negativity (or N2pc)
P*d*	Distractor Positivity
PET	Positron Emission Tomography
PPC	Posterior Parietal Cortex
RG	Redundancy Gain(s)
RM	Repeated Measures
RMI	Race-Model Inequality
RSE	Redundant-Signals Effect
RT	Reaction Time(s)
SC	Superior Colliculi
SEP	Somatosensory Evoked Potential
SNR	Signal-to-Noise Ratio
TPJ	Temporo-Parietal Junction
VEP	Visual Evoked Potential
WTA	Winner-Takes-All

Summary

This thesis examines the functional architecture of the preattentive stage that precedes the deployment of selective attention within a visuo-tactile search paradigm. By combining the empirical evidence from visual and crossmodal search, it proposes a “modality-weighting account” (MWA) for crossmodal search that extends the “dimension-weighting account” (DWA) from visual search. To answer the main research question of whether multisensory attentional resources derive from a common or separate pool(s), two electrophysiological markers for visual and tactile attention (PCN & CCN) are analyzed regarding modality-specific and supramodal mechanisms.

Chapter 1 (*General Introduction*) gives an in-depth overview of the scientific background and the evolution of empirical concepts in visual and crossmodal search. In visual search, selective attention is guided by the most salient or relevant activation of the “priority map”, which reflects the weighted integration of independent analyzers or dimensions (e.g., color or shape) and their distinct feature activations (e.g., a red circle among blue circles). While this preattentive hierarchy, i.e., the DWA, has often been confirmed in visual search, empirical evaluation of a similar crossmodal concept is missing. A testable scientific framework is presented based on comparable empirical evidence showing performance benefits for redundant combinations and substantial switch costs across visual dimensions and sensory modalities. While the preattentive levels of dimensions and modalities should be the same in the multisensory DWA (null hypothesis), the MWA assumes a functional dissociation between them and proposes a “supramodal priority map” (research hypothesis). Also, the electrophysiological evidence for preattentive processes, multisensory integration, as well as visual and tactile attention is presented. Further, the visual PCN and tactile CCN components are introduced, reflecting spatial filtering and the outcome of preattentive computations. Based on the five factors of attention, a set of superordinate research questions for the visuo-tactile paradigm is derived that can be analyzed in combination with EEG recordings.

Chapter 2 (*Redundancy Gains*) provides strong evidence for the MWA by showing significantly higher redundancy gains, i.e., accelerated performances and PCN onsets and increased PCN amplitudes, for crossmodal but not visual redundancies. Also, the study reveals a consistent pattern for preattentive coactivations. While the CCN reflects a stable somatosensory component for

tactile target attention (modality-specific), the PCN is adaptive to visual and crossmodal information and a good predictor for behavior.

Chapter 3 (*Spatial Attention*) extends the insights on redundancy gains by analyzing their spatial properties, i.e., comparing the slopes from quadratic fitting. Here, vision and touch can be described by two distinct spatial functions: external quadratic or egocentric flat. While visual redundancies remained within their spatial (focal and peripheral) limits, crossmodal redundancies exceeded them, revealing crossmodal bending in spatial attention (favors MWA). Although the PCN was generally amplified for crossmodal redundant targets (supramodal), their quadratic slope was unaffected (modality-specific). Finally, an eigenvector model for spatial properties is discussed.

Chapter 4 (*Intertrial Effects*) examines the hierarchy of switch costs, which are substantially higher between modalities than (visual) dimensions (favors MWA). Here, the reaction time delay for modality shifts (e.g., vision → touch) was accompanied by reduced amplitudes in PCN and CCN. This finding indicates a supramodal mechanism for attentional resources distributed across sensory modalities. Also, a decay function across multiple successive trials and a visual bias that promotes location shifts are presented.

Chapter 5 (*General Discussion*) combines and discusses the various findings. Based on the evidence, the MWA reflects the suitable preattentive hierarchy for crossmodal attention, confirming the central claim. Furthermore, a “supramodal priority network” is proposed that prioritizes relevant features and spatial locations across sensory modalities by activating distinct network branches and nodes. Finally, a novel “Dynamic Search” paradigm is briefly introduced that is suitable to investigate the follow-up questions and likewise deals with the challenges of attention research.

1 General Introduction

Searching for answers,
The questions, they sprout.
Thou art humble!

The world we live in is complex. We must constantly adapt and update our beliefs and knowledge to find our way around it. This necessity to evolve applies to us as humans as well as science. In science, gathering accurate information to draw the correct conclusions is essential; likewise, every scientific journey starts with an open question. Although "everyone knows what attention is" (James, 1890), this thesis wants to learn more about the specific preattentive architecture preceding selective attention that integrates sensory information across multiple modalities (e.g., vision, sound or touch), which is known as *crossmodal search*. First and foremost, the main question wants to know whether crossmodal attention derives from "common or separate pools of selective attention" (cf. Miller, 1982; Töllner et al., 2009), i.e., do we have common attentional resources across sensory modalities, or is there, for example, a particular visual attention or tactile attention each with its unique capacities?

To answer this rather global and abstract attention research question, it is essential to obtain an empirical framework of (crossmodal) selective attention based on its attributes, factors, previous theories, and research background. From that, a set of concise empirical hypotheses is derived, enabling to answer (or at least approach) the main question from various perspectives by combining evidence from behavior and neural dynamics in EEG recordings.

Origins of Selective Attention

Although attention appears ubiquitous, in the elder philosophy, there is often no explicit term for attention; it is either unconditionally assumed or simply neglected. Nevertheless, attention seems inevitable regarding higher cognitions such as consciousness or reasoning. Notably, there appears to be a pattern of philosophers like Descartes and Hegel to assume an 'immediate connection' of sensory perception to our conscious experience and reflections. Likewise, other philosophers like Aristotle and Kant endeavor to understand our mind and soul by sets of a priori categories, where the mixture of substance, quantity, quality, and relations enable our sensory experiences, presumably including attention. Another immanent concept of attention can be found in Heidegger's 'being-in-the-world', which bundles a limited section of the 'worldliness' into our existence, revealing core attributes of attention like selection and limitation. These thoughts on attention throughout philosophy are very scarce and merely reflect an opener to shift the focus to experimental psychology and the obtained evidence from many empirical studies, which detached from philosophy during the late nineteenth century. Finally, by overcoming behaviorism and plain schemes of perception and action (e.g., Berlyne, 1951), attention research prospered from the paradigm shift of the cognitive revolution in the mid-twentieth century (e.g., Neisser, 1967) by focusing on the internal and neural processes in-between our perceptions and actions.

Still, attention appears closely linked but different to higher cognitions and consciousness (Dehaene, 2014; Dehaene & Naccache, 2001; Edelman & Tononi, 2001; Eimer & Grubert, 2015; Koch & Tsuchiya, 2007, 2012; Posner, 1994; Woodman et al., 2007), "attention is necessary, but not sufficient for awareness" (M. A. Cohen et al., 2012). It is commonly agreed that attention reflects a transient phenomenon that occurs in-between our perception and awareness, activating its own resources within specific attention networks in our brain (Corbetta & Shulman, 2002; Gross et al., 2004; Posner & Dehaene, 1994; Posner & Petersen, 1990; Walz et al., 2014; Woodman & Luck, 2003). In a famous experiment about 'inattentional blindness' (Simons, 2000; Simons & Chabris, 1999), participants were instructed to count the dribbling of a basketball of one team. Fascinatingly, they overwhelmingly failed to recognize the wandering gorilla in plain sight while focusing on the task, revealing an unequivocal dissociation between perception and awareness.

Since the late 1950s, attention research focused on two major questions about attentional selection: 1) the timing (early vs. late) and 2) the type of process (parallel vs. serial) (Allport, 1971; Broadbent, 1958; Deutsch & Deutsch, 1963; Kahneman, 1973; LaBerge, 1973b). Generally, our brain functions are a sequence of cognitive stages, or so-called *brain states*, which alternate systematically over temporal snapshots; similar to throwing a ball from person to person, information is passed across various brain states (e.g., McClelland, 1979; Pashler, 1997). In our case, it is assumed that attention follows perception and takes over from it. Hence, the *when* of this transition is of key importance, i.e., the temporal locus of selective attention. While *late* selection only occurs when the integration of perceptual information is completed fully, *early* selection interrupts those perceptual processes by selecting relevant information and concurrently neglecting irrelevant information, leaving them unprocessed. This debate on *early* vs. *late* selection spanned several decades, adding empirical evidence for selective attention being either *late* and *passive,* waiting for the ball or perceptual information, or reaching *actively* for this ball from *early* on. Another perspective on this debate came from the 'perceptual load theory' (cf. Lavie, 1995), which argues that early or late selection depends on the perceptual load of the experimental paradigm. In a series of experiments with target letters in visual displays, Lavie (1995) manipulated the set size, the distance and compatibility of distractors, and the task itself, e.g., by adding go/no-go conditions of combined colors and shapes. Evidence for both early and late selection models was found depending on the perceptual load of visual displays and tasks, leading to a truce in the debate.

Alongside the timing of selective attention, another debate ignited on whether perceptual integration and attentional selection occur in a parallel or serial fashion (Allport, 1971; LaBerge, 1973b). Overall, targets within visual displays are composed of distinct features, which activate independent visual analyzers or so-called *dimensions*; for example, a green square or a left-tilted bar contains primary information about colors and forms or orientation and forms, respectively (Allport, 1971; Treisman, 1969). Generally, selective attention reflects a filter, bottleneck, or spotlight (Broadbent, 1958; Duncan, 1980; LaBerge, 1983; Treisman et al., 1983). Hence, both processes can be encountered during the preattentive stage, a fast, automatic, and parallel activation of independent visual analyzers and a slow, controlled, and serial combination of those activations, enabling attentional selection. This conceptualization has been summarized in the two-process theory (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) and the two-stage model of *Guided Search* (Wolfe, 1994; Wolfe et al., 1989).

Despite this brief overview of the origins of attention research, those original questions will be reencountered in some form or another during the quest to derive a suitable empirical framework and preattentive hierarchy for crossmodal search.

Attributes of Attention

So far, attention seems ubiquitous as it reflects the gateway between perception and action, but also higher cognitive functions, e.g., working memory (Baddeley, 1992). Yet, attention is a transient state, which has given researchers scientific impetus over many decades. Commonly, three attributes describe attention in more detail. First, attention is *selective,* as it acts like a filter, prioritizing some information while blocking other information (e.g., Broadbent, 1958; James, 1890). Second, attention is *limited*, a process with limited-capacity resources, often described as a bottleneck (e.g., Lavie, 1995; Neisser, 1967; Treisman, 1969). Like a prism retrieving a section of the light spectrum, attention selects a fraction of the vast sensory perception (s. Figure 1–1). Third, attention is *dynamic*; it reflects an emergent process over time, e.g., by attending sequentially to several spatial locations or objects (e.g., Logan, 1996; Vecera & Farah, 1994; Wolfe, 1994). These three attributes of attention make it much easier to develop empirical research questions. For example, how can we systematically manipulate the prioritization of visual targets? Is there a cost function of selective attention that indicates limited resources? Are there distinct spatial functions for visual and tactile attention?

Five Factors of Attention

Before presenting the empirical framework of crossmodal attention, the understanding of selective attention must be revised. To begin with, it is necessary to summarize the various perspectives on attention, deriving five distinct factors or entrance hubs that enable empirical examination (e.g., Logan, 1996; Treisman, 1969; Wolfe et al., 1989; Wolfe & Horowitz, 2017).

So far, selective attention primarily relied on sensory perception (e.g., vision) and its functional integration (for a good reason). Naturally, our senses, or windows to the world, rely on attention to process relevant information from external sources. This type and direction of processing is often called a 'bottom-up' process, as it originates from perceptual information. During this 'bottom-up' stream, selective attention is captured by the most salient information, e.g., seeing a red traffic light or hearing the sound of sirens. Saliency stands out from the surroundings; it describes a contrast within visual scenes or displays and enables us to listen to a specific voice in a noisy room during a cocktail party. Overall, 'bottom-up' saliency or prioritization combines different aspects; it can be object-based, location-based, or discrimination-based (Allport, 1971; Duncan, 1984; Logan, 1996; Vecera & Farah, 1994). While *objects* and *locations* can be considered independent factors (Logan, 1996; Maljkovic & Nakayama, 1994, 1996; Vecera & Farah, 1994), discrimination or saliency derives from their combined computation. Although object recognition and spatial decoding are commonly active during visual encoding, e.g., during feature-binding (Mangun, 1995; Treisman & Gelade, 1980), they appear coarsely separated in the visual cortex across the ventral stream (*what*) and dorsal stream (*where*), respectively (DeYoe & Van Essen, 1988; Goodale & Milner, 1992; Grill-Spector & Malach, 2004; Konen & Kastner, 2008; McIntosh et al., 1994; Mishkin et al., 1983). This neuro-anatomical dissociation for *objects* and *locations* originates from V1 and propagates parallel to the inferior temporal or posterior parietal lobe in a bottom-up, stimulus-driven, or feedforward manner. Conversely, both streams share reciprocal connections or combined activations across the inferior parietal lobe (IPL) and visual areas, e.g., V4, MT/V5 (Hayden & Gallant, 2005; Milner, 2017; Singh-Curry & Husain, 2009; J. Wang et al., 1999). Arguably, visual information about objects and locations joins or converges at some stage or brain area into a coherent percept (e.g., Hillyard & Münte, 1984; Treisman & Gelade, 1980) to achieve discrimination based on saliency. This neural mechanism also includes serial limitations where "attention is either at one location or another" (Wolfe, 1994, p. 209). This thesis will focus on both 'bottom-up' factors for selective attention, *objects* and *locations*, and their interaction (s. Chapters 2 & 3). Further, the composition of feature singletons (*object features*) and their spatial functions (*spatial locations*) will be analyzed across visual and tactile modalities (s. Figure 1–1).

A third factor of selective attention can be found in the immediate past. Whereas *objects* and *locations* indicate the actual and real-time processing of perceptual information, selective attention relies heavily on temporal dynamics and preceding information. In experimental psychology, this temporal derivative of selective attention is analyzed as the difference between consecutive trials (or visual displays). It is summarized under *intertrial effects* (e.g., Found & Müller, 1996; H. J. Müller et al., 1995), where the comparison of response rates, e.g., reaction times, can either facilitate, decelerate or remain the same between the current trial (*n*) and the previous trial (*n*–1). In search paradigms, target repetitions produce response benefits (e.g., red circle → red circle), whereas systematic target changes lead to substantial switch costs (e.g., red circle → blue square). The best-known example of this is probably the modality-shift effect, where attentional shifts between sensory modalities (e.g., vision → touch) lead to tremendous switch costs, i.e., delayed responses (Miles et al., 2011; Spence, Nicholls, et al., 2001; Töllner et al., 2009). These *intertrial effects* open another perspective on selective attention and its limited resources, providing time-sensitive markers of attentional deflections, e.g., repetition benefits or switch costs from most recent events. These 'bounded or occupied' attentional resources represent residual cortical activations, which are commonly described as *neural traces* or *attentional traces* (Alho et al., 1990; Fecteau & Munoz, 2003; Gondan et al., 2007; Mannuzza, 1980; Ruge & Naumann, 2006). Chapter 4 will give a detailed empirical description of how these *intertrial effects* systematically modulate in a visuo-tactile search paradigm (s. Figure 1–1).

Besides these three factors of selective attention, representing primarily automatic and stimulus-driven mechanisms based on current or previous salient sensory information, another critical factor is 'top-down' or goal-driven attention (Desimone & Duncan, 1995; Wolfe, 1994). Instead of being entirely at the whim of environmental saliency, we can voluntarily guide attention based on our goals and expectations using *attentional templates* (e.g., Duncan & Humphreys, 1989; Eimer, 2014; Fecteau, 2007). Finding a familiar face in a crowd is a famous example of such a template, which pre-activates basic and complex visual features of the familiar face, e.g., shape and size of the head, nose, or ears, colors of skin, hair, or eyebrows, spatial relations of eyes, nose, and mouth, and many more. In contrast to bottom-up saliency computations, which run through the visual cortex in a feedforward direction (V1 → V4), top-down templates (or visual cues) neurally pre-activate relevant features and visual areas in a feedback direction, e.g., in V4 or IT, improving selective attention for both specific objects and locations (Chelazzi et al., 1993; Desimone & Duncan, 1995; Luck et al., 1993). Further, the short-term activations of attentional templates from memory or cues are strongly associated with working memory and neural activity in frontal regions (Carlisle et al., 2011; Eimer, 2014; Houtkamp & Roelfsema, 2006; Olivers et al., 2011; Stokes et al., 2013; Warden & Miller, 2010). Notably, the top-down control in selective attention extends the cortical scope from the parietal and visual cortex toward (pre-)frontal regions, resembling the well-known fronto-parietal attention networks (e.g., Corbetta & Shulman, 2002; Posner & Dehaene, 1994). Although attentional templates are evidently active before and during search displays (Eimer & Kiss, 2010; Grubert & Eimer, 2018), their ability to (proactively) suppress bottom-up saliency appears either limited, e.g., the additional singleton paradigm (Theeuwes, 1991, 1992, 2010), or is an open issue of debate, e.g., signal-suppression account (Gaspelin et al., 2015; Gaspelin & Luck, 2018b; Sawaki & Luck, 2010). Even though bottom-up and top-down factors clearly oppose one another, the distinction between *attentional templates* and *intertrial effects* is not so clear as both rely on ‘prior history’ or priming (Fecteau, 2007; Lamy & Kristjánsson, 2013). This thesis aims to improve the discriminability of these two factors by deriving a concise set of empirical hypotheses (here in Chapter 1) that will be evaluated later on for empirical examination (Chapter 5).

Generally, there is joint agreement on the four factors of selective attention (e.g., Theeuwes, 2010; Wolfe, 1994), which can be envisioned in an oppositely or orthogonally arrangement of involved brain functions (s. Figure 1–1), e.g., dorsal and ventral streams, bottom-up feedforward and top-down feedback connections, or current and previous activations. While the scientific community does not question those factors, the focus and debate of attention research rely (once again) on the exact timeline of selective attention (similar to early and late selection models) and is about *when* those factors interact (or interfere) with one another, e.g., stimulus-driven (automatic) vs. goal-driven (guided or controlled). Of note, other factors also guide selective attention, as can be seen in Wolfe's (2021) updated version of *Guided Search* (GS 6.0), including *reward* or *scene syntax and semantics*. However, as this thesis incorporates a visuo-tactile search paradigm for crossmodal attention, there will be, by default, no empirical intersections to study those additional factors. Nonetheless, this thesis aims to incorporate those contemporary trends whenever applicable, as it strives to propose a general model of crossmodal selective attention (Chapter 5) based on the observed empirical evidence (Chapters 2–4).

The fifth factor is vigilance or *sustained attention* (e.g., Esterman et al., 2013; MacLean et al., 2009). While the four factors take a snapshot of the task and its conditions, *sustained attention* depicts the (attentional) time course of the same task. Commonly, sustained attention measures the trial-by-trial variation in performances, often called the vigilant decrement in continuous performance tests (Riccio et al., 2002; Rosenberg et al., 2013; See et al., 1995). Interestingly, the direct influence of vigilance on selective attention has often been omitted due to the equally balanced and averaged conditions and targets over the whole duration of the experiment, thus eradicating frequent and systematic attentional fluctuations. Nonetheless, the role of sustained attention cannot be underrated, as it seems directly associated with the activation of fronto-parietal attention networks (deBettencourt et al., 2015; Helfrich et al., 2018; Posner & Dehaene, 1994; Rosenberg et al., 2016). Further, activity in attention networks reflects the trial-by-trial variability and also rhythmicity of attentional states, i.e., being either 'in-the-zone' or 'out-of-the-zone' (Kucyi et al., 2017; Rosenberg et al., 2015) associated with specific brain rhythms and oscillations, e.g., theta and alpha frequencies (Helfrich et al., 2018; Michel et al., 2021). Altogether, the five-factor model combines two perspectives of task/condition effects and attentional fluctuations over time. Similar to *attentional templates*, the factor of *sustained attention* will be critically reviewed in some detail (here in Chapter 1). Also, its practical feasibility and empirical relevance will be discussed later (Chapter 5).

Finally, we derive the five-factor model of selective attention, which resembles the empirical framework of this thesis (s. Figure 1–1). As mentioned above, this model aligns with previous models that included similar or other factors, e.g., set size, reward, habituation, learning, or distractor handling (e.g., Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Found & Müller, 1996; Gaspelin et al., 2015; Logan, 1996; Mackworth, 1968; Treisman, 1969; Wolfe et al., 1989; Wolfe & Horowitz, 2017); and it by no means claims to be complete. Nonetheless, the five-factor model pursues the scientific purpose of enabling a general model of crossmodal selective attention. This thesis argues that a general attention model must combine those factors, similar to essential ingredients in a recipe. Ultimately, this attention model should enable a dynamic quantification of each factor across search tasks, similar to a multiple regression equation. However, before we can develop an empirical and statistical conceptualization like this, it takes a deeper dive into the details of those factors and their empirical evidence.

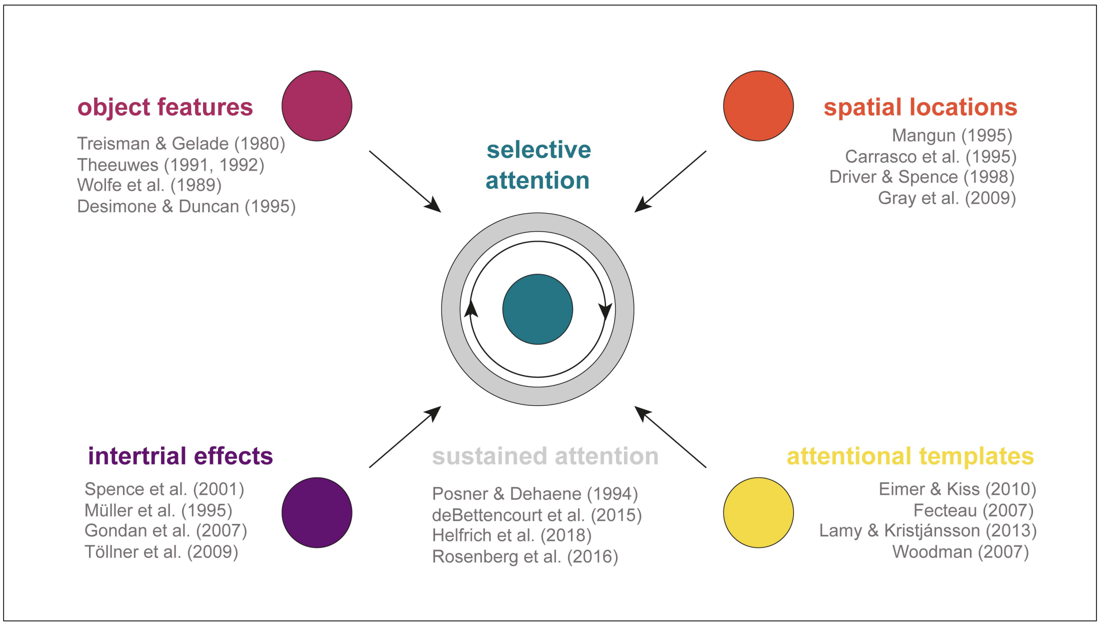

Figure 1–1. The five-factor model of attention is shown. Overall, selective attention (green), i.e., the deployment of attentional resources, modulates based on the impact of (1) *object features* (red) or feature activations, (2) *spatial locations* (orange), (3) *intertrial effects* (purple) or prior history, (4) *attentional templates*, and continuous processes of (5) *sustained attention*. Also, some relevant references for each factor are added. While object features and spatial locations reflect bottom-up driven processes in a feedforward direction, intertrial effects and attentional templates indicate feedback activity from previous activations (e.g., neural traces) or top-down control. Although the five-factor model is not exhaustive, its sparse description can be regarded as the skeleton of a computational model of attention.

Visual Search

Although evidence for the five factors of selective attention already depends on it, this section introduces visual search from experimental psychology. Before beginning to describe *displays*, *targets*, and various visual *task*s, it is noteworthy that the majority of early attention theories (e.g., Broadbent, 1958; Cherry, 1953; Deutsch & Deutsch, 1963; Treisman, 1960, 1969) were primarily based on auditory search, i.e., dichotic listening to voices or sounds. Excitingly, their research questions, concepts, and terms could be flexibly transferred to visual search (and also crossmodal search).

In their paper, Deutsch and Deutsch (1963) suggest that (selective) attention derives from perceptual and discriminatory mechanisms, which group and segregate (sensory) information. Further, their statement, “How such grouping or segregation takes place is a problem for perceptual theory …” (p. 83), resembles the core question of what happens between perception and attention. In a conceptual example, Deutsch and Deutsch (1963) collect a group of people (originally boys) from which they want to infer who is the tallest. Instead of comparing each person with one another, leading to an “enormous rate” of comparisons with increasing group size, the authors favor “a simpler and more economical way” to achieve their goal: collecting the people under a horizontal and adjustable board that measures ‘height’. Given that everyone stands up, the board would touch the tallest person like a threshold. Further, the board would move down when the tallest person leaves or up when a taller person enters the scene. This example refers to the mechanisms of selective attention. Here, attention capture is based on *weighting* the importance of (sensory) information, e.g., a person's height, or any other (physical) *dimension*, e.g., color or tone. Overall, the concept of attentional selection by Deutsch and Deutsch (1963), occurring after perceptual aggregation or dimension-based grouping, therefore *late*, already contains many essential aspects for the empirical framework of this thesis, being conceptually very similar to saliency summation models (cf. Koch & Ullman, 1985).

Another influential study is from Treisman (1969). In this study, the author reviews the evidence from various experiments of the 1960s, which include both auditory search (or selective listening) and visual search. Notably, visual search experiments were then conducted using the tachistoscopic technique, enabling the presentation of visual displays in the millisecond range (e.g., Lappin, 1967; Sperling, 1960).

In her study, Treisman (1969) proposes four selection types for attention: (i) output selection, (ii) input selection, (iii) analyzer selection, and (iv) test and target selection. Whereas (i) and (ii) refer to late and early selection models (e.g., Broadbent, 1958; Deutsch & Deutsch, 1963), where perceptual processes are either unrestricted or filtered, (iii) analyzer selection resembles an intermediate step between early and late selection. Those proposed analyzers are nothing less than the already introduced *dimensions*, being independent units or ‘collections’ of similar *features*, e.g., *color* (red, green, blue), *orientation* (left, right), or *shape* (bars, squares, circles). Further, independent analyzer units include ‘basic’ dimensions (e.g., color and orientation) but also more ‘complex’ ones, such as *letters* or *words*.

To better describe those selection processes, Treisman (1969) proposes a visual search task where the *display* includes “colored letters in different sizes and orientations” (p. 284). Here, participants must identify whether those (briefly presented) displays contain the letter ‘G’, i.e., the *target*, among irrelevant nontarget letters (e.g., ‘H’), by giving a positive (‘present’) or negative (‘absent’) response, respectively. Essentially, (visual) *targets* can be composed of features along (visual) dimensions, e.g., a red ‘G’ contains a *color* and a *letter*. On the contrary, a target cannot be ‘G’ and ‘H’ at the same time and location, representing a conflict within analyzer units. According to Treisman (1969), the knowledge of the *target*’s features enables a *test* within analyzer units, in *contrast* to other features, e.g., color-test: red vs. blue or yellow, or letter-test: ‘G’ vs. ‘H’, or ‘F’, thus describing (iv) test and target selection. Generally, this example of visual search (next to another of auditory search) describes attentional selection as a sequence of underlying processes, where each selection type (i – iv) can be associated with a specific stage or process within that pipeline, yielding an empirical framework. Arguably, *target* selection (red ‘G’) and (correct) response must undergo a systematic process of perceptual integration by activating relevant analyzers (e.g., color and letters) and by rejecting irrelevant analyzers (e.g., size and orientation), i.e., type (iii) selection. The feature discrimination (e.g., red vs. blue) occurs within those relevant analyzers (e.g., color), i.e., type (iv) selection.

Next, Treisman (1969) reviews this selection ‘architecture’ for divided and focused attention, i.e., to disentangle parallel or serial selection processes. While some evidence suggests independent (hence parallel) analyzers due to uncorrelated visual dimensions (Lappin, 1967), others favor serial discrimination between those dimensions (Egeth, 1966). Also, there is reasonable evidence that visual search occurs in a serial fashion, e.g., reaction time and the number of display items increase linearly, or responses become delayed for shorter interstimulus intervals (favoring a serial queue) (cf. Treisman, 1969). Similarly, focused attention on a single analyzer or specific analyzer sets seems unlikely due to interferences across relevant and irrelevant analyzers occurring within the perceptual or response stage or both. A famous example of analyzer interference is the Stroop test, where saying a word’s color conflicts with a color’s word, e.g., ‘green’ in yellow or ‘red’ in blue (Stroop, 1935). While in 1969, the evidence for (iii) analyzer selection appears still ‘equivocal’, feature-based (iv) target selection, on the other hand, consistently led to improved performances.

The special merit of this early study of Treisman (1969) is the proposition of ‘independent analyzers’ or dimensions (particularly for visual search) and her effort to explain the various stages and types of selection. Like Deutsch and Deutsch (1963), Treisman (1969) proposes a mechanism for selective attention in an ‘all-or-nothing fashion’ or as an adaptive ‘signal-to-noise ratio’, ultimately defining which features and dimensions capture attention. Interestingly, she implies that spatial locations (e.g., within a search display) are seized or perceptually registered in a ‘very early stage’, clearly before (iii) analyzers and (iv) targets.

With the advent of computer screens (e.g., CRT monitors), experimental psychology and visual search experienced a boost in scientific productivity from the 1980s onwards. The endless options to compose visual displays based on the number of items or set size, features, and locations led to various search tasks striving to decipher visual selective attention. Before we continue the evolution of our empirical framework, it is essential to introduce some basic terms and scientific milestones of visual search. As already mentioned, a (visual) *display* contains visual objects, e.g., *targets,* non-targets, cues, or distractors, defined by discriminative *features* on a homogenous background (e.g., black, gray, or white). Those *displays* are often presented relatively quickly (below 1 second or faster), and a *fixation cross* (or something similar) precedes them as a marker to align focal attention. Generally, a task consists of many displays organized in blocks and specific task conditions linked by an interstimulus interval (often blank). Besides display presentation, i.e., the sensory input, behavioral response describes another pillar of visual search, i.e., the processing output, obtaining the reaction time in milliseconds or accuracy in percentage from key presses. Sometimes error feedback is given by presenting visual and auditory signals, e.g., a beep tone.

Another common term of visual search is ‘pop-out’ search, which describes a salient ‘odd-one-out’ target or a ‘pop-out’ singleton among (mostly) homogenous nontargets, e.g., a green diamond among red diamonds, a white horizontal bar or black vertical bar among black horizontal bars (Luck & Hillyard, 1994a; Maljkovic & Nakayama, 1994). Crucially, the ‘pop-out’ target differs from the nontargets by a distinct *feature-contrast*, e.g., green vs. red, white vs. black, or vertical vs. horizontal. Crucially, those contrasts can occur systematically within and across visual dimensions, e.g., within color: red vs. green or across letter and shape: ‘G’ vs. circle (H. J. Müller et al., 1995). Again, it is important to emphasize that a *feature* can be any type of visual segregation, whereas a *dimension* contains a set of all similar features (e.g., colors). Obviously, ‘pop-out’ search is primarily driven by bottom-up saliency computations (Duncan & Humphreys, 1989; Koch & Ullman, 1985; Schreij et al., 2008). Although nontargets (e.g., green diamonds) and distractors (e.g., a red circle) are both task-irrelevant, distractors can also ‘pop-out’ by having a salient feature-contrast to the nontargets and target (e.g., green circle) that needs to be actively suppressed during the task, delaying response times as described in the ‘additional singleton paradigm’ (Theeuwes, 1992). Although this thesis will not analyze the impact of distractors in detail, it is nonetheless worth mentioning that both selection and suppression in attention research represent two sides of the same coin.

Another essential aspect of visual search is the task itself, designed to answer one or more concrete research questions. Similar to visual targets, the number of visual search tasks can be endless, yet we encounter a systematic and manageable variety of visual search paradigms. Arguably, visual search has three task classes: detection, localization, and identification, which themselves contain many other tasks and specific variations (cf. Töllner et al., 2012). In detection tasks, participants must indicate the presence or absence of *targets* (among nontargets) that occur at a predefined rate across trials (e.g., 20 %, 50 %, or 80 %). Responses are commonly collected by button (or key) presses after the visual display presentation, indicating whether a target was perceived as 'present' or 'absent'. A special case of detection is conjunction search. While detection is generally quite effortless, e.g., finding a letter among squares, the target in conjunction search combines the feature-contrasts of nontargets, e.g., finding a green 'T' among brown 'T's and green 'X's, thus increasing the difficulty of discrimination with increasing set size. The next section shows that conjunction search plays a critical role in the *feature-integration theory* and *Guided Search* (Treisman & Gelade, 1980; Wolfe et al., 1989). Different from detection tasks, localization task targets appear commonly in all trials. Here, participants are instructed to respond based on the target's spatial location, which can be 'left' or 'right', 'top' or 'bottom', and 'in' or 'out' of a display zone (Sauter et al., 2018; Wolber & Wascher, 2005). A combination of spatial locations and feature detection is the spatial cueing task, where visual displays are preceded by a visual cue that points to a particular spatial location influencing target detection (and selective attention) based on its validity (e.g., 75 % valid & 25 % invalid cues) (e.g., Eimer & Kiss, 2008; Folk et al., 1992; Folk & Remington, 1998; Wolber & Wascher, 2005).

The third visual search class is identification tasks, associated with a higher attentional load than detection and localization tasks, which 'only' ask for the presence and location of targets. In those tasks, participants must identify targets based on specific features, e.g., a target can be either a green or blue square among yellow square nontargets but also an 'M' or 'W' among vertical bars (Eimer, 1996). Crucially, participants had to discriminate those relevant target features (here, color or form) by pressing the respective buttons, e.g., left-hand: green or 'M' and right-hand: blue or 'W' (hence, the name discrimination task). Another common task is compound search, where participants initially must visually detect the target (e.g., a letter 'E' or a 90° rotated 'T') and, in the following step, report on the respective appearance (e.g., 'E' mirrored or not?, is the top of 'T' left or right?) by giving the correct button response (Deubel & Schneider, 1996; Zinchenko et al., 2020). Further examples can be found in additional singleton paradigms, where participants had to find a green diamond and identify whether the line within the diamond was horizontal or vertical (Theeuwes, 1992), or they had to find a tilted bar among vertical bars and indicate whether the gap in the target bar was on the bottom or top (Goschy et al., 2014; Sauter et al., 2018).

Closing the gap to our five-factor model of selective attention, visual search and its various tasks are well suited to examine (and answer) the multitude of specific research questions. For example, visual *objects* can be created by any combination of features (originating from distinct dimensions), allowing for systematic discrimination of contrast levels and bottom-up saliency in almost any task, e.g., the redundant-signals effect (Grubert et al., 2011; Krummenacher et al., 2002a; Töllner, Zehetleitner, Krummenacher, et al., 2011). Further, visual search can focus on *locations*, i.e., the examination of spatial attention, e.g., by using spatial cueing paradigms. The strong (and genuine) link between attention and spatial locations becomes evident by descriptions of attention as a 'spotlight', 'zoom lens', or 'gradient models' that indicate a relationship between the capacity of attention and the activity in the retinotopic visual cortex or receptive fields (Eriksen & St James, 1986; LaBerge, 1983; LaBerge & Brown, 1989; Luck et al., 1997; Mangun & Hillyard, 1988; N. G. Müller et al., 2003; Posner, 1980). Notably, the detection of target singletons becomes worse with increasing distance to the center of our focal attention, indicative of a spatial function of visual attention, which is known as the eccentricity effect (Carrasco et al., 1995; Staugaard et al., 2016; Wolfe et al., 1998).

Further, *intertrial effects* or prior history can be embedded by manipulating the display sequence (or target sequence) within and across blocks, enabling a systematic view of repetition benefits and switch costs for target objects and locations (Found & Müller, 1996; Rangelov et al., 2013; Talcott & Gaspelin, 2020; Wolfe, 2021). Also, top-down processes or *attentional templates* can be measured in visual search by giving certain objects and locations a high probability of appearance, i.e., making the target's presence predictable beyond the prior history in alignment with the task goal (Fecteau, 2007; Lamy & Kristjánsson, 2013). Crucially, top-down control can be observed in visual search as it selectively deploys attentional resources to relevant but not irrelevant visual inputs (Eimer & Kiss, 2010; Grubert et al., 2017). Finally, the fifth factor of *sustained attention* (seamlessly) aligns well with visual search tasks and can monitor attentional states ('in-the-zone' vs. 'out-of-the-zone') across trials, blocks, and sessions (Esterman et al., 2013; Kucyi et al., 2017; Rosenberg et al., 2015).

Overall, visual search evolved into a broad research field over the last decades, where the multitude of insights are summarized into well-known theories, e.g., theory of visual attention (TVA), perceptual load theory, guided search, or dimension-weighting account (DWA) (Bundesen, 1990, 1998; Found &

Müller, 1996; Huang & Pashler, 2007; Lavie, 1995; Lavie & Tsal, 1994; H. J. Müller et al., 1995; Wolfe, 1994; Wolfe et al., 1989). Visual search combines task designs and behavioral responses flexibly with physiological recordings such as eye-tracking, following processes of overt and focal attention along saccades and fixations, or controlling for involuntary saccades from center fixations in covert attention designs (Deubel & Schneider, 1996; Rayner, 1978, 2009; Theeuwes, 1992; Wollenberg et al., 2018). Also, visual search is compatible with brain imaging techniques such as EEG and fMRI, which enable the analysis of temporal and structural correlates of attention (e.g., Corbetta & Shulman, 2002; Kristjánsson et al., 2007; Leonards et al., 2000; Töllner, Zehetleitner, Gramann, et al., 2011; Töllner, Zehetleitner, Krummenacher, et al., 2011; Wei et al., 2009). Finally, visual search also profited from computational modeling and the signal detection theory (Green et al., 1966; Stanislaw & Todorov, 1999), providing an accurate description of perceptual, preattentive and attentional processes across the visual cortex, e.g., in saliency summation models (Itti & Koch, 2001; Koch & Ullman, 1985; Koene & Zhaoping, 2007; Lee et al., 1999).

In concluding this section, it can be stated that visual search is still quite an active field of experimental psychology, which is nowadays focusing on topics such as distractor handling (e.g., Gaspelin et al., 2015; B. Wang & Theeuwes, 2020), top-down control (e.g., Grubert & Eimer, 2018; Jenkins et al., 2018), reward (e.g., Hickey et al., 2010, 2014), probability cueing (e.g., Anderson et al., 2021; Goschy et al., 2014), and working memory (e.g., Gazzaley & Nobre, 2012; Oberauer, 2019; Olivers et al., 2011). Despite this impressive progress in studying visual selective attention, the ultimate goal for visual search must be the transition from the lab toward real-life settings and natural viewing (e.g., Võ & Wolfe, 2015), particularly with the advent of virtual and augmented reality. Similar to virtual reality, our actual (visual) reality, i.e., (visual) information processing, is, according to Broadbent (1956), a measurement of bits per second with a known limit and capacity of how much information we can concurrently perceive. Interestingly, this 'bit thought' links visual search to information theory (Borst & Theunissen, 1999; Shannon, 1948), where information is quantified as entropy or complexity (by bits and nits). Notably, the increase in complexity or similarity, i.e., too much or too less information, respectively, hinders the discrimination of visual displays or scenes and, thus, selective attention (Duncan & Humphreys, 1989). Hence, the optimal conditions to study visual search and selective attention derive from a balanced task set and design, including pertinent information. When the facts are straight, one can delve into unexplored territories.

Feature-Integration Theory

The following sections will collect the empirical framework's building blocks or elements. First, it is worth taking a deeper look into (visual) features, which were introduced above, besides the four types of selection (Treisman, 1969). In another scientific effort, Treisman and Gelade (1980) summarized the 'equivocal' evidence through a series of experiments into a "feature-integration theory of attention". The authors argue that early visual integration occurs automatically and in parallel through features (or visual contrasts) within the visual space. Yet, recognizing visual objects occurs later and serially as those features need to be coherently combined along the areas of the visual cortex. Essentially, the *feature-integration theory* states that the activation of specific features is parallel, while their 'binding' into objects must be serial and requires focused attention.

Again, Treisman and Gelade (1980) highlight that features belong to distinct dimensions or analyzers, e.g., colors or orientation, which are "separately analyzed by some functionally independent perceptual subsystem, and 'feature' to refer to a particular value on a dimension." (p. 98-99). The clear distinction between a parallel perceptual or preattentive stage that activates so-called *feature maps* and the subsequent serial attentional stage that identifies objects relates to the two-process theory (Shiffrin & Schneider, 1977). Yet, the interesting question remains to which extent those serial attentional processes can become parallel via unitization or practice (LaBerge, 1973a).

The empirical evidence for their feature-integration theory gave Treisman and Gelade (1980) by conducting a series of conjunction search experiments. Remember, in conjunction search, the target shares its features with all nontargets, while in disjunction or feature search, the target shares its features with only half of the nontargets (or less). Here, they showed that in conjunction search, i.e., finding a green 'T' among brown 'T's and green 'X's, reaction times revealed a significant linear slope that increased over set sizes (1, 5, 15, or 30 items in the visual display) for both positive or negative target events (present or absent). Interestingly, slopes in negative events were twice as high as positive ones, indicative of the extra time needed for an exhaustive search. In contrast, feature search (or disjunction search), finding a blue letter 'T' or 'X' or a brown or green 'S' among brown 'T's and green 'X's, revealed a flat (non-significant) slope, showing comparable reaction times across the various set sizes for positive target events. Hence, their finding suggests that the visual processing in conjunction search is serial and depends on set size (exhaustive and self-terminating search), taking around 60 ms per item, providing evidence against the unitization of distinct features that assumes an efficient search with flat slopes.

The same outcome of serial conjunction search persists for easier targets (e.g., finding a red ‘O’ among green ‘O’s and red ‘N’s) with equally distant items (e.g., embedded in a 4 x 4 matrix) and across the similarity range of targets and nontargets. Further, the authors argue that the prolonged search for conjunctions derives from the impaired spatial segregation of features and textures, as they are less discriminative than disjunctions. Also, Treisman and Gelade (1980) claim that (visual) spatial encoding and feature activation processes are distinct yet must coincide in conjunction search, opposing other accounts of attentional capture based on spatial orienting, e.g., spotlight or zoom lens (Eriksen & Hoffman, 1972; Posner, 1978).

Overall, the feature-integration theory proposes a preattentive mechanism that combines parallel detection with perceptual grouping (or segregation) based on feature activations before the spatially-bound deployment of focal attention that serially identifies visual items or objects. Although the statistical power within experiments is arguably a bit low (including only 6-8 participants), and the main conclusions will already be challenged in the next section, the feature-integration theory remains a landmark publication in visual search. Furthermore, the feature-integration theory can be regarded as a precursor of the idiomatic “binding problem”, where coherent (visual) percepts and higher cognitive processes originate from the neural synchrony of local feature activations (Damasio, 1989; C. M. Gray, 1999; Reynolds & Desimone, 1999; Singer, 1999; Singer & Gray, 1995; Treisman, 1996, 1998, 1999; Wolfe & Cave, 1999).

Saliency-Summation Models and Guided Search

Another perspective on visual selective attention stems from computational modeling, summarizing the various claims about spatial encoding and feature decomposition into a general framework (cf. Koch & Ullman, 1985). Agreeing on the two-stage model, the authors distinguish a “preattentive mode” and an “attentive mode” as collections of ‘early representations’ that selectively combine into a ‘central representation’. Again, elementary features (e.g., color, orientation, or moving direction) are processed rapidly and parallel with the spatial locations of the visual field, then follow the processes of focal attention and the serial recognition of complex objects.

Koch and Ullman (1985) introduce the concept of “topographical maps” to understand better the transition from the preattentive to the attentive stage. By referring to evidence from the research of spatial orienting and receptive fields (Eriksen & Hoffman, 1972; Goldberg & Wurtz, 1972), the authors argue that the distributed activations across *feature maps* must be encoded in a spatially coherent fashion, i.e., mapping each feature activation to distinct visual locations (or vice versa). Based on this (necessary) assumption, Koch and Ullman (1985) introduced the “saliency map” as another topographical map intermediate to early and central representations.

The saliency map summarizes loads of various feature maps (e.g., a red diamond among green circles) by registering them across the visual field, i.e., multiple features can co-exist in one location (e.g., a green tilted bar) or one feature can be linked to multiple locations (e.g., the color green). Further, the saliency map must incorporate a *weighting mechanism* to enable attentional capture, i.e., the transition to the non-topographic central representation. This mechanism yields a saliency summation that triggers the allocation of selective attention in a *winner-takes-all* routine that serially deploys visual selective attention to the location with maximum saliency (or highest overall feature-contrast) and not to multiple locations simultaneously (in line with Posner et al., 1980). Finally, Koch and Ullman (1985) argue that selection shifts must occur based on maximum changes in the saliency map due to constant re-weighting over time. This explicit description of an *attention shift* toward the most salient location is similar to the rough proposition of Deutsch and Deutsch (1963) of an adaptive board to measure a person’s height.

Besides their computational approach, Koch and Ullman (1985) also initiated the debate about the saliency map’s brain area and timing. By giving two options, the saliency map can be either located beyond the striate cortex, relying on the completion of early visual representation in V4 or MT, or it might be directly linked to V1 and the lateral geniculate nucleus (LGN), and thus preceding those feature activations. Notably, the authors underscore the 30-50 ms estimate for attention shifts (Bergen & Julesz, 1983) based on the physiological properties of neurons.

Although their ‘saliency-summation’ model is primarily built on bottom-up saliency that encodes elementary features (e.g., color and orientation) and locations, influencing many subsequent studies (Itti & Koch, 2001; Koene & Zhaoping, 2007; Lee et al., 1999), it nonetheless flexibly combines with the other factors of selective attention, such as top-down control and prior history, i.e., affecting the weighting routine due to priming and search priorities (e.g., Navalpakkam & Itti, 2007; Wolfe, 1994). This ‘multi-faceted’ character of saliency summation is similar to the visual cortex's hierarchical structure, combining feedforward and feedback connections (e.g., Hochstein & Ahissar, 2002; Van Essen & Maunsell, 1983).

In addition to ‘saliency-summation’, Wolfe and colleagues (1989) extend the ‘feature-integration theory’ by proposing a model of *Guided Search*, which incorporates an “attention map” (similar to the concept of a saliency map) after the parallel activation of feature maps. Again, the authors distinguish between parallel preattentive and serial attentive stages. In a series of conjunction search experiments, Wolfe and colleagues (1989) showed that a central claim of the feature-integration theory, i.e., expecting a linear slope for conjunctions over increasing set sizes, indicative of constant serial processing (scanning item by item), cannot be maintained. Instead of linear slopes of around 29 ms or 67 ms (including a significant 1:2 ratio) for present and absent targets, respectively, the encountered slopes of Wolfe and colleagues (1989) were much smaller (often non-significant) and ‘non-linear’. Significantly, those non-linear slopes, similar to logarithmic (or inverse quadratic) functions, rise substantially less over set sizes than linear slopes. This finding suggests that the purely serial search mode for conjunctions, as predicted by the ‘feature-integration theory’, is actually a hybrid search mode of both serial and parallel processes. Hence the conclusion of Wolfe and colleagues (1989) is that visual selective attention is also *guided* by the parallel activation of feature maps, at least to a certain amount, depending on how compatible the transmission from the preattentive to the attentive stage is.

Another intriguing finding of Wolfe et al. (1989) is the fact that stimulus saliency plays an enormous role in visual search concerning experimental outcomes and conclusions. Although the authors did not provide a detailed description of how salient feature loads are precisely integrated, like the *weighting mechanism* of Koch and Ullman (1985), they compared their conjunction search findings using a television monitor with bright colors, with an exact replication of the experiments in Treisman and Gelade (1980), which used the tachistoscopic technique with painted displays. Notably, Wolfe and colleagues (1989) replicated the steep (linear) slopes in the tachistoscope in difference to their shallow (nonlinear) slopes. However, they also showed that tachistoscopic slopes did depend on the saturation of colors, i.e., serial item search was about twice as fast for “high contrast” conditions (6 ms & 13 ms per item) in comparison to replicated “low contrast” conditions (11 ms & 24 ms per item). Overall, these results provide important evidence that stimulus saliency is a critical factor to consider (e.g., via saliency alignment of feature-contrasts), which is in accordance with other studies (Duncan & Humphreys, 1989; Itti & Koch, 2001; Nothdurft, 1993; Töllner, Zehetleitner, Gramann, et al., 2011).

The nonlinear evidence for attentional guidance of Wolfe et al. (1989), i.e., the logarithmic increase of conjunction search among set sizes, is very much in line with the proposed ‘winner-takes-all’ network of Koch and Ullman (1985) of local-to-global saliency comparisons of the visual field, reflecting a binary-search tree with a logarithmic time (or search) complexity. However, besides the stimulus-driven saliency, search performance for self-terminating tasks, e.g., rating the presence or absence of targets as in conjunction search, also depends on goal-driven factors such as task design. For example, a triple conjunction search (including size, color, and letter), e.g., finding a big red ‘O’ among small red ‘X’s, small green ‘O’s, and big green ‘X’s, reveals smaller slopes across set sizes due to the task-embedding of two feature contrasts as compared to regular conjunction search with only one feature contrast (Quinlan & Humphreys, 1987; Wolfe et al., 1989). On the other hand, triple and double conjunctions become comparable again when they both embed discriminations of only a single feature contrast between the target and nontargets. Also, search difficulty varies based on the number of included conjunction targets. In a simultaneous search for two targets (e.g., red ‘X’ or green ‘O’ among green ‘X’s and red ‘O’s), initial and strategic guidance towards a specific feature fails, e.g., neglecting all red stimuli when searching for a green ‘O’, which in turn produces the serial (linear) slopes, again (Nakayama & Silverman, 1986).

Wolfe et al.'s (1989) findings clearly suggest a modification to the original feature-integration theory. They argue that search efficiency and attentional selection depend on the amount of noise, i.e., the variation in stimulus saliency, between the parallel and serial stages. A high saliency of feature contrasts improves the signal-to-noise ratio, enabling attentional guidance through the ‘seamless’ transmission of preattentive computations. Subsequently, ‘feature-integration’ was modified due to empirical evidence, which formerly assumed a strict separation of parallel and serial processes.

For example, Treisman and Sato (1990) revised conjunction search as a response to the empirical exceptions from a purely serial search mode (e.g., McLeod et al., 1988; Pashler, 1987; Wolfe et al., 1989). Overall, they replicated those ‘exceptions’ of search slopes in their own experiments, favoring fast and (more) parallel conjunction search. Furthermore, the search speed for conjunctions relies on nontarget heterogeneity, i.e., display complexity; the more distinct nontargets were included, the more difficult the search. On the other hand, knowing a conjunction target in advance substantially improves the search compared to unknown targets, referring to involved mechanisms of top-down control, e.g., working memory. Notably, the feature-integration theory was updated by introducing an extra “master map” that combines the encoding of features and locations accordingly, that is, serial deployment of attention (Treisman, 1988; Treisman & Sato, 1990), similar to the functionality of the “saliency map” (Koch & Ullman, 1985) and “attention map” (Wolfe et al., 1989).

Based on their findings, Treisman and Sato (1990) excluded conjunction detectors and segregation assumptions due to additivity effects among features, which suggest the presence of both feature independence (evidence against detectors) but also feature summation (evidence against segregation). Hence, the authors propose a 'feature-inhibition' hypothesis, where irrelevant feature information of nontargets is suppressed in the master map, e.g., when searching a green 'T' among green 'X's and brown 'T's, the feature activation of the color brown and the letter 'X' can be (actively) inhibited, thus complementing target selection and search efficiency. Of course, this 'feature-inhibition' can only apply when the target is known and obviously fails for unknown targets. Further, the effectiveness of 'feature-inhibition' depends on target vs. nontarget discrimination. Finding a rotated 'T' among rotated 'L's is much more challenging than finding a pink 45° tilted bar target among green 45° bars or pink 135° bars. That is, 'feature-inhibition' is ineffective for similar edge comparisons in letters ('T' vs. 'L'), while for the latter color-orientation stimuli, it can flexibly choose one independent analyzer, e.g., green, that discards all green 45° bars by pooling them out. Again, the almost instantaneous exclusion of half of the non-targets via 'feature-inhibition' relates well to the shallow slopes and nonlinear inverse quadratic functions (of the binary search tree).

Wolfe (1994) summarized *Guided Search* into an updated version (GS2). As before, the author describes the activation of basic feature maps (e.g., orientation, color, or motion) during the preattentive stage, which altogether load onto an "activation map". In GS2, Wolfe (1994) explains in much more detail the underlying architecture. The author links stimulus-driven bottom-up processes of feature segregation to neural encoding patterns that ultimately translate signal strength, i.e., the differences of feature contrasts of signal detection, into 'categorical' features that interact with top-down processes such as task demand. More precisely, each 'categorical' feature represents a distribution of local and specialized firing neurons, including a variance or noise term similar to Gaussian distributions.

In addition to those feature distributions, Wolfe (1994) presents a linear model of parameters to forecast reaction times during visual search tasks. This formula includes a baseline (b), average time per item (x), set size (n), and standard deviation (σ), yielding a base estimate for behavioral responses, e.g., in feature or conjunction search. Arguably, this is a simplified formula, as it does not account for the display's complexity, i.e., the similarity and proximity of stimuli (see Humphreys & Müller, 1993; Koch & Ullman, 1985). However, the proposition of a 50 ms attention shift as average time per item is consistent with the 30-50 ms physiological estimate of Koch and Ullman (1985), although a bit high. At present, *Guided Search* is in its sixth version. GS6 includes additional factors of selective attention, e.g., prior history, reward, or scene syntax and semantics. Still, the “activation map”, which triggers attentional capture updates at a 50 ms rate about 20 times per second (or 20 Hz), yet object recognition (similar to the central representation) takes about 150 ms per item, about 3 “attention map” updates (or 7-8 Hz).

Fascinatingly, all three concepts of selective attention used a similar mechanism to explain the consistent mapping of features and locations, based on ‘saliency summation’, yet in different terms: “saliency map” (Koch & Ullman, 1985), “master map” (Treisman, 1988), and “attention map” or “activation map” (Wolfe, 1994; Wolfe et al., 1989). Another term was proposed by Fecteau and Munoz (2006): the “priority map”. The authors advocate a relabeling due to evidence from firing rates of neurons that respond to stimulus-driven salience and goal-driven relevance. Therefore, the term “saliency map” is not sufficient, as it does not express the role of dynamic ‘prioritization’. In summary, the attentional “priority map” aligns with the other terms that also proposed a bi-directional convergence of bottom-up and top-down processes in an implicit or explicit manner. This thesis uses the term “priority map” or “attentional priority map” to refer to the topographic map that guides selective attention. Besides a common agreement on a specific term, the most intriguing questions about the “priority map” are the *where* and *when*. The “Attentional Priority Map” section will examine these questions in greater depth.

Dimension-Weighting Account

As described in the previous sections, *features* are heterogenous, reflecting the variety of basic and complex properties of discriminative objects. This heterogeneity in visual search should be considered (Wolfe, 1994). So far, the primary research focus of the studies presented was to examine the two-stage model and its modes of selective attention, i.e., the transition of the parallel preattentive stage into the serial attention stage, and further how these stages guide the deployment of selective attention (Deutsch & Deutsch, 1963; Koch & Ullman, 1985; LaBerge, 1973b; Treisman & Gelade, 1980; Wolfe et al., 1989). While there is common agreement on topographic *feature maps* and the *priority map* in visual search, less advance was made to investigate the intricate relationships of feature activations, leaving this topic in an ‘equivocal’ state awaiting a systematic conception with empirical evidence.

The updated feature-integration theory exposed the necessity to organize the abundance of features or the ‘feature-space’, as it includes two competing mechanisms, i.e., ‘feature-activation’ and ‘feature-inhibition’, at the same level of preattentive computations (Treisman & Sato, 1990). Since features derive from the firing patterns of local neurons, ‘categorical’ features rely on adequate signal segregation (Wolfe, 1994). Hence, the more similar and spatially closer targets and nontargets become, the more activation and inhibition interfere with one another, seemingly, a conflict that impacts and delays visual search.

The presence of this conflict is quite vivid across many visual search studies. For example, finding a letter ‘L’ is much more difficult among ‘T’s than ‘X’s in simple texture discrimination (Bergen & Julesz, 1983), as ‘textons’, a letter’s building blocks, are more similar between ‘L’ and ‘T’ than ‘L’ and ‘X’. Another example is the ‘additional singleton paradigm’ (Theeuwes, 1991, 1992), revealing several interesting insights in a series of experiments. First and foremost, Theeuwes (1992) shows that the efficiency of target selection depends on the saliency of feature contrasts, consistent with previous findings. Finding a green circle among red circles (color-contrast) was around 60 ms faster in reaction times than finding a green circle among green diamonds (shape-contrast), indicative of an ‘asymmetric selectivity’ between color and shape. Further, distractor interference, i.e., the advent of an ‘additional singleton’ beside the target, also depends on the amount of saliency. Whenever the distractor is more salient than the target, the distractor automatically pulls attentional resources in a bottom-up manner, inducing a substantial delay in target selection with slower reaction times, e.g., a red-diamond distractor and a green-circle target among green diamonds. However, by (over-)correcting the saliency imbalance for color and shape, exchanging the high salient red color with a low salient yellow-green color, the effect patterns of the distractor bias reversed: a shape-contrast distractor (green diamond) produced higher response costs than a color-contrast distractor (yellow-green circle). The findings of Theeuwes (1992) suggest saliency trade-offs between visual dimensions, e.g., color and shape, that guide selective attention and distractor handling. Arguably, salient distractors cannot be ignored completely, and the ‘feature-inhibition’ of irrelevant distractors occurs on-the-fly.

Considering these findings and those of conjunction search, the evidence suggests that the performance during visual search depends on systematic variations in the task design (Bergen & Julesz, 1983; Theeuwes, 1992; Treisman & Sato, 1990; Wolfe, 1994). Hence, a systematic look at the heterogeneity and homogeneity of target and nontarget relations, i.e., their feature contrasts, is required, introducing another perspective besides pure saliency-summation. Interestingly, and maybe a bit expectedly, the ‘missing link’ awaits in the independent analyzer units or *visual dimensions*.

In advance to explain visual dimensions in greater detail, another study by Treisman (1988) highlights the relevance of this issue. Her study shows that feature search, i.e., the detection of disjunctive targets among nontargets, is flat for increasing set sizes and comparable for known and unknown targets but only for *within*-dimension feature-contrasts, e.g., a blue bar among green bars or a horizontal bar among vertical bars. Also, Treisman (1988) examines two substantial performance delays. While performances were fast for *within*-dimension targets and *across*-dimension heterogeneity in nontargets, processing delays occurred for mixed *across*-dimension targets and *within*-dimension heterogeneity nontargets, which empirically summarizes the above-stated conflict. Notably, the shift in scientific focus on *dimensions* relates back to Allport (1971) and the proposed analyzer selection (iii) by Treisman (1969).

With these findings in mind, Müller, Heller, and Ziegler (1995) conducted their experiments. Based on empirical evidence and conclusions (cf. Treisman, 1988; Wolfe, 1994), Müller and colleagues (1995) propose an extra layer of "dimension-specific saliency maps", or *dimension maps*, intermediate to *feature maps* and the *priority map*, as a functionally distinct processing step. Further, the authors argue that a dissociation for unknown "odd-one-out" targets that vary *within* or *across* visual dimensions proves that the existing preattentive hierarchy is incomplete.

Like Treisman (1988), Müller et al. (1995) replicated the indifferent and flat function across set sizes and the increased reaction time costs for unknown *cross*-dimension conditions of about 55-60 ms. While in the *within*-dimension condition, a target only varied in three orientations: a (small and gray) left-oriented, right-oriented, or horizontal gray line among small gray vertical lines; the *cross*-dimension condition, on the other hand, varied in size (large gray vertical line), color (small black vertical line), or orientation (small right-oriented gray line), again among small gray vertical lines. These extra costs in *cross*-dimension targets expose a gap in the preattentive architecture, as it cannot be explained by the 'global' and parallel integration of feature activations, where the "odd-one-out" should be equal across unknown targets. Instead, this finding of Müller and colleagues (1995) localizes a dimension-specific bias in the "early perceptual stage" (p. 7) that precedes the capture of selective attention.

In their second experiment, Müller and colleagues (1995) examined the interplay of relevant targets and irrelevant distractors by modulating the distractor either *within*- or *cross*-dimension. As feature search is not exhaustive as conjunction search is, the "odd-one-out" or "pop-out" target only needs to be detected but not identified. For this reason, positive and negative responses, indicating target presence or absence, respectively, are generally comparable.

However, adding a heterogeneous distractor, i.e., a task-irrelevant singleton that varies from nontargets, introduces an imbalance for positive and negative conditions, as the "odd-one-out" distractor produces a salient contrast that requires extra elimination effort to give a response. Hence, Müller and colleagues (1995) retrieved a significant delay for positive responses in both *within-* and *cross*-dimension (Δ 73 ms and Δ 29 ms, compared to the respective baseline in the first experiment), which was substantially higher in the *within*-dimension condition. This finding indicates that the (active) inhibition or feature segregation, i.e., omitting the task-irrelevant distractor, is more costly and, thus, less efficient for the *within*-dimension than the *cross*-dimension distractor. Again, this outcome advocates the functional relevance of visual dimensions. Yet, the authors admit that inhibiting a *cross*-dimension distractor is also incomplete, as a delay in positive responses, although smaller, persists.

To gain more insights into the impact of top-down control during the preattentive stage, Müller et al. (1995) varied the probabilities of their first experiment for each of the three targets in both conditions. Here, the right-oriented line occurred in 80% of target-present trials (*within-* and *cross*-dimension), and the other two targets were equally present in only 20% of target-present trials in both conditions. While no effect was found in the *within*-dimension condition, showing equal negative and positive responses, a massive effect appeared in the *cross*-dimension condition. As expected, the highly probable target was significantly faster than negative responses (Δ 41 ms) and low probable targets (Δ 111 ms), which also had much higher error rates (9 %).

Overall, the findings of Müller, Heller, and Ziegler (1995) revealed strong *cross*-dimension evidence that a preattentive hierarchy, which so far only consisted of *feature maps* and a *priority map*, is not sufficient and must take an extra computational level of *dimension maps* into account (s. Figure 1–2). Although the authors cannot completely resolve the issue of whether 'dimension elimination' in the second experiment is serial or parallel, their reasoning for *weight shifting* between visual dimensions seems plausible. Whereas *within*-dimension targets maintain the same dimension-weight across trials, in this case for orientation targets with comparable reaction times, *cross*-dimension targets, on the other hand, need to constantly adjust those weights due to dimension-shifts in the trial sequence from orientation to color to size, leading to substantial delays and reaction time costs. As before, in the 'saliency-summation' models, the *weighting mechanism* that integrates perceptual information is also preserved in this updated version, as the priority map combines the weights of dimension maps, which combine the weights of their respective feature maps.

Notably, *weight-shifting* between visual dimensions might elegantly explain the cross-dimensional costs by intertrial dependencies. Based on correspondence with Wolfe and a post-hoc analysis, Müller and colleagues (1995) argue that repetitions in target dimensions between consecutive trials (e.g., color → color from trial *n*–1 to trial *n*) were significantly faster (Δ 57 ms) than dimension-shifts (e.g., orientation → color). In contrast to other opinions that favor a purely stimulus-driven perceptual integration (Theeuwes, 1991, 1992), where salient stimuli like an additional singleton or distractor cannot be suppressed by top-down control, *dimension-weighting* appears to be a suitable explanation that includes both factors, stimulus-driven and goal-driven aspects. A highly expected target in one dimension (e.g., orientation) produces substantial response advantages compared to unexpected dimensions (e.g., color or size), indicating an unequal distribution of *weights* across visual dimensions based on (implicit or explicit) prioritization.

Also, Müller et al. (1995) claim that the *weights* (or importance) of visual dimensions cannot be ignored entirely; assigning a weight of zero to a specific dimension seems odd, as the (voluntary) exclusion of color information in daylight for healthy vision is sheer impossible. Hence, a 'known' distractor might always delay the attentional capture of a target whenever the distractor is more salient due to automatic adjustments in *dimension-weighting*, exceeding those weights of the prioritized target dimension, which in the next step must be actively suppressed. Notably, *dimension-weighting* is congruent with Theeuwes' (1991, 1992) findings, as only the high-salient distractor produced delays but not the low-salient distractor. In a recent study, Sauter and colleagues (2018) provided further evidence for *dimension-weighting* in a probability-cueing paradigm, showing that the distractor's dimension matters. While same-dimension distractors produce heavy interferences with the target (Δ 94 ms), the impact of different-dimension distractors is significantly lower (Δ 14 ms).

The proposition of a "dimension-weighting account" (DWA), introducing an intermediate layer of dimension-specific saliency maps or *dimension maps* into the preattentive hierarchy, is based on the evidence presented. Overall, the DWA combines bottom-up processes (features and locations), top-down control, and intertrial dependencies (four factors of selective attention) via the inherent weighting dynamics *within* and *across* visual dimensions. Further, the DWA entails a set of concise assumptions and research questions to bolster its functional relevance. Two essential claims of the DWA, which will be primarily addressed in this thesis, are the following: The detection (or localization) of pop-out singletons produces *intertrial effects*, i.e., a substantial amount of response costs due to *weight-shifting*, when the feature-contrast of targets switches *across* visual dimensions (e.g., color → orientation, or shape → color). Another claim suggests that target detection becomes faster by combining those task-relevant dimensions since the weights of multiple dimensions are summed, thus, improving the guidance of selective attention (e.g., color + shape or color + orientation). On the other hand, same-dimension shifts (e.g., blue → green) or combinations (blue and green targets) should be (more or less) comparable to same-feature repetitions (e.g., blue → blue) and combinations (two blue targets).

More investigations on the "dimension-weighting account" was done by Found and Müller (1996) by analyzing intertrial dependencies. In their experiments, the authors examined feature search: finding orientation (left- or right-tilted white bars) and color (blue or red vertical bars) targets among white vertical bars (nontargets) on a black background. In the first task, participants performed a detection task, where they had to respond to the presence or absence of targets (both conditions 50 %). The second experiment was an identification task, where participants had to indicate the target's dimension, either color or orientation, via button presses. Again, Found and Müller (1996) replicated the flat slopes for feature search across increasing set sizes (here 4, 9, 16), indicating a parallel search mode. Also, they showed behavioral advantages for color targets with faster response times and fewer errors than orientation targets. Overall, Found and Müller (1996) examined a consistent pattern of *intertrial facilitation* across color and orientation targets, comparing the current trial (*n*) as a function of the previous trial (*n*–1) with shift conditions. That is, same-dimension shifts (e.g., color → color) were significantly faster than different-dimension shifts (e.g., color → orientation, or vice versa) with a delta of 22–36 ms across experiments and visual dimensions. While for the detection task, no intertrial facilitation was found between same-feature (e.g., red → red) and different-feature (e.g., left-tilt → right-tilt) shifts, interestingly, the identification task revealed extra facilitation in color for same-feature repetitions (e.g., blue → blue) above different-feature shifts (e.g., red → blue). Overall, Found and Müller (1996) retrieved evidence favoring the "dimension-weighting account" as *switch costs* or *intertrial facilitation* derive primarily from dimensional shifts or repetitions, representing about two-thirds of the effect.

In another experiment, Found and Müller (1996) showed that counting three or four targets within a display also depends on whether target dimensions are homogeneous or heterogeneous. Counting four homogeneous targets (e.g., all color targets) is significantly faster than counting three homogeneous targets, as participants had to (extra) check for another heterogeneous target (e.g., one orientation target). Again, this finding supports a DWA as preattentive architecture as visual search becomes more efficient for homogeneous targets, adjusting the *weights* for only one dimension (e.g., color or orientation). In contrast, heterogeneous targets must update their dimension *weights* sequentially (e.g., color and orientation), which produces extra costs. As before, these *weights* accumulate onto the *priority map*, guiding visual selective attention to perform the task properly.

Essentially, the *dimension-weighting account* extends the concepts of previous theories, those of *feature maps* and an attentional *priority map*, with their empirical evidence for *dimension maps*. Crucially, the DWA embeds four factors of selective attention (except sustained attention), reflecting the dynamical assignment of dimension *weights* and their interplay that project onto the priority map regarding saliency, relevance, and prior history. With the DWA, the various findings in visual search and the continuous spectrum of task designs (e.g., as in feature and conjunction search from flat to steep slopes over set sizes) can be attributed to the efficiency of that weighted integration, producing either benefits or costs that reflect a more parallel or serial search mode, respectively (Fecteau & Munoz, 2006; Found & Müller, 1996; H. J. Müller et al., 1995; Treisman & Sato, 1990; Wolfe, 1994, 2021). Of note, the studies on DWA add the perspective of intertrial dependencies or prior history as a conceptual factor of short-term priming to attention research, becoming a productive approach over the following years (e.g., Becker et al., 2009; Eimer et al., 2010; Goschy et al., 2014; Krummenacher et al., 2009; Lamy & Kristjánsson, 2013; H. J. Müller et al., 2004).

The DWA provides answers to the 'equivocal' evidence of the various visual search experiments. For example, it appears reasonable that letter discrimination is more difficult for 'L's and 'T's than 'L's and 'X's, as horizontal and vertical 'textons' (or features) are identical but diagonally different for the latter (e.g., Bergen & Julesz, 1983). Also, it makes sense that letter discrimination is more challenging than feature search or conjunction search, as the letter's feature-contrast derives from only one, although more complex, *within* letter dimension, instead of feature-contrast(s) *across* dimensions, e.g., a green 45° bar among pink 135° bars, including color and orientation dimensions. Notably, the DWA favors the presence of dimension-specific saliency maps, i.e., independent analyzers, within the parallel preattentive mode, as dimension-wise contrasts primarily guide the rapid (or almost instantaneous) feature-contrast detection within a visual display, which Müller and colleagues describe as the "odd-one-out" or "that" response (Found & Müller, 1996; H. J. Müller et al., 1995). Ultimately, these topographic dimension loads from specialized analyzers are again spatiotopically combined in the attentional priority map, enabling the attentional capture of "that" difference over "there". In line with other studies, Found and Müller (1996) underscore the difference between detection or localization tasks, which only require a “that” and “where” response, and identification tasks, which additionally require a “what” response, revealing a dissociation for preattentive and attentive modes, respectively (Nothdurft, 1992; Sagi & Julesz, 1985).

The DWA has been a productive framework for examining the attentional processes during visual search over the recent decades (Gramann et al., 2007; e.g., Krummenacher et al., 2001a; Liesefeld & Müller, 2019). Despite the core assumption that visual dimensions or independent analyzers are organized within the same functional level of the preattentive architecture, their respective *weights* or influence on selective attention are somewhat asymmetric. First, dimension differences might simply result from saliency imbalance across dimensions, as shown by adjusted colors in the studies of Theeuwes (1992) and Wolfe et al. (1989). This reason might also explain the color benefits of Found and Müller (1996). Second, in line with Wolfe (1994) visual dimensions can be either basic (e.g., orientation, color, size, or motion) or more complex (e.g., shapes, letters, or words), resembling a hierarchical order of stimulus-driven feedforward and goal-driven feedback connections, representing the visual cortex with its dorsal and ventral streams (Hochstein & Ahissar, 2002; Van Essen & Maunsell, 1983). Third, the distribution of dimension maps within the visual cortex could reflect a distinctive temporal sequence of dimension loads onto the priority map, e.g., the integration of a color-contrast precedes that of a shape-contrast. Arguably, those ‘earlier’ and basic dimensions should either be anatomically closer to the priority map (e.g., in V1) or have shortcut connections.

Finally, while the DWA is well-suited for visual search, it remains an open issue whether it can be generalized to multisensory perception in “crossmodal search”, e.g., combining perceptual information from different modalities such as vision, touch, and sound. In case the DWA indeed applies to crossmodal search, the layer of distributed independent analyzers or dimension maps would simply expand across sensory modalities within the same hierarchical level. Generally, the crossmodal DWA claims a *flat* hierarchical organization of crossmodal dimensions with a supramodal priority map, i.e., a common pool of attentional resources. Hence, attentional orienting in visual and crossmodal search should be comparable without a substantial difference. Concluding, the DWA is a centerpiece for the scientific framework of this thesis, and its empirical evaluation for crossmodal search is one of the main objectives.

Crossmodal Search

So far, the scientific background and the two-stage model of selective attention, i.e., its preattentive and attentive modes of parallel and serial integration, have been introduced in the preceding sections with a focus on visual search and the preattentive hierarchy that culminates into the "dimension-weighting account" (DWA). Overalls, this preattentive hierarchy includes topographic layers of *feature maps*, *dimension maps*, and the *priority map*, which are connected by weighted projections. The attentional prioritization of features or dimensions occurs through the dynamic interplay of up-weighting or down-weighting processes, i.e., selective activation or inhibition. Similar to the other concepts of visual search (Koch & Ullman, 1985; Treisman & Sato, 1990; Wolfe, 1994), where the chronological order for feature maps and the priority map is not specified, the DWA architecture does not provide a clear statement as well (Found & Müller, 1996; H. J. Müller et al., 1995). Based on the empirical evidence, it can be argued that the activation of preattentive layers represents levels of 'attentional resolution', that is, the 'precision of attentional representation(s)'. While for detection or localization tasks, a dimension-specific contrast is sufficient to respond, identification (or feature-discrimination) tasks, on the other hand, require exact feature knowledge of the target and thus additional processing time (e.g., Found & Müller, 1996; H. J. Müller et al., 1995).

Of course, attention is not exclusively bound to vision, as humans are multisensory beings. In this context, it is worth emphasizing (again) that the early studies of selective attention primarily examined selective listening (e.g., Cherry, 1953; Treisman, 1960). Hence, attention's concept and empirical evolution appear independent of sensory modalities, indicating that selective attention reflects a *supramodal* capacity. But is that the case? Despite a deepened understanding of attention, this question cannot be answered by looking solely at individual modalities (e.g., vision or sound). It needs the interaction of those senses (e.g., vision and touch) to obtain a (more) complete picture of selective attention, which is known as *crossmodal search*. Overall, crossmodal search, i.e., "the adaptive attentional orienting to multisensory events", is well-suited to examine *supramodal* and *modality-specific* attention mechanisms. While a *supramodal* attention mechanism applies to all sensory modalities in the same way, *modality-specific* attention mechanisms would be different across sensory modalities, being unimodal or antagonistic to one another.

The distinction between *supramodal* and *modality-specific* attention mechanisms in crossmodal search is the central question of this thesis, which in the context of attention research, dates back over many decades. For example, Treisman (1969) reported better performances for divided attention that combines auditory and visual stimuli by summarizing the preliminary outcome of an unpublished pilot study. She further concludes: “While two inputs to the same modality must share the same analyzers throughout, the two inputs to different modalities would be at least partly analyzed by independent systems.” (p. 289). This statement complements the concept of attentional resources by adding the perspective of crossmodal search. Whereas dimensions or analyzer units are considered independent *within* visual search, they depend on (or share) the same pool of resources in visual selective attention. On the contrary, crossmodal search potentially combines independent pools of attentional resources *across* sensory modalities before the deployment of attention, proposing an extra layer of preattentive integration. Arguably, multisensory representations originate from the temporal coherence of neural decoding; they derive either from an overall and *supramodal* attentional priority map or the combination of *modality-specific* saliency maps (e.g., Macaluso et al., 2002a; Töllner et al., 2009). Concerning the preattentive hierarchy, this juxtaposition describes a fine but decisive distinction. It will be examined in greater detail after introducing crossmodal search with experimental designs, concepts, and insights.

There are plenty of studies in crossmodal search that combine multiple sensory modalities (e.g., Diederich & Colonius, 2004; Eimer & van Velzen, 2005; Macaluso et al., 2002b; Miller, 1982; Spence, Nicholls, et al., 2001; Spence & Driver, 2004). Among sensory modalities, the combinations of vision, sound, and touch proved particularly suitable for studying crossmodal selective attention, in contrast to odor and taste modalities, whose scope of application is rather restricted (e.g., Kauer & White, 2001; Laing & Glemarec, 1992). Like other sensory modalities, odor and taste primarily activate modality-specific brain regions, e.g., piriform or insular cortex, but also multimodal (or supramodal) areas, such as the far anterior insular cortex (Veldhuizen & Small, 2011). Although olfaction and taste contain mixtures of independent fragrances and ingredients, in line with the concept of dimensions, they primarily describe “chemosensory” processes (Mombaerts, 1999; Spence, Kettenmann, et al., 2000; Spence, McGlone, et al., 2001), which complicates the objectivity of the experimental designs. For example, it might be difficult for odor and taste experiments to retrieve conclusions about selective attention for several reasons. Although odor (and taste) detection, i.e., the onset of olfactory stimuli, is efficiently fast, the spatial propagation and stimulus offset in ‘odor displays’ depend on molecular movements (Hopfield, 1991), in difference to visual and auditory stimuli, with complete control of onset and offset timings. Further, the transition of olfactory sensations to selective attention and response decisions, e.g., the identification of an odor, might be more error-prone with an increased amount of variance, requiring a more explicit link to memories and 'expert knowledge'. Regardless, several fascinating attempts have been made to study multisensory interactions, including odor and taste, for example, in synesthesia (Deroy et al., 2013; Deroy & Spence, 2013; Spence et al., 2015; Thesen et al., 2004). However, pursuing a complete attention theory that combines all sensory modalities will be one of the most exciting challenges for future studies.

Whereas the evolution of a preattentive hierarchy, i.e., the proposition of an attentional priority map, can be genuinely attributed to the efforts of visual search, from the beginning, crossmodal search investigated the superordinate question of whether attentional resources, e.g., in a bimodal detection task, either derive from separate pools or a common pool, i.e., modality-specific or supramodal mechanisms, respectively (cf. Miller, 1982). Generally, crossmodal search describes perceptual integration as the activation of 'channels' from distinct sensory modalities that transmit 'signals'. Notably, by assuming independent channels, studies in crossmodal search are associated with divided attention (e.g., Miller, 1982; Treisman, 1969).

For example, Miller (1982) conducted a series of experiments, including a bimodal detection task containing visual and auditory stimuli, where participants had to respond to target-present trials while ignoring target-absent trials. Those tasks consisted of three targets: a centrally presented visual asterisk or letter 'X', an auditory 780 Hz tone, or their "redundant" visual-auditory combination. On the other hand, target-absent trials either showed nothing, a letter 'O', or a 360 Hz tone. Crucially, Miller (1982) consistently found a "redundant-signals effect" (Kinchla, 1974), retrieving substantially accelerated reaction times for the redundant signals, the visual-auditory target, in comparison to non-redundant or single-channel signals, i.e., a redundancy gain or delta of around 100 ms. This "redundant-signals effect" (RSE) even persisted for the comparison with the average of the respective faster non-redundant channels, e.g., visual or auditory signals (Δ 55 ms) (Biederman & Checkosky, 1970).

Overall, the RSE can be explained by two accounts, separate-activation, i.e., a *race-model* of two or more completely independent channels, or a co-activation account for combined integration. The reasoning behind both accounts is quite similar to the threshold criterion and *winner-takes-all* mechanism in visual search. While in separate-activation models, reaction times obey the assumption of independent and identical distributed random variables, redundancy gains simply reflect *statistical facilitation*, where a trial's winner, i.e., the faster channels, captures selective attention and triggers response behavior (Raab, 1962). Conversely, co-activation models propose an extra gain surpassing statistical facilitation's limit (or threshold), which arguably, can only be achieved by the interaction of those racers, which amplifies the overall signal strength, e.g., by activating supramodal brain regions.

The critical contribution of Miller (1982) was the *race-model inequality*, a test to evaluate whether separate-activation or co-activation models explain redundancy gains. The race-model inequality proposes a statistical limit by taking the sum of the cumulative density functions of non-redundant signals (e.g., auditory + visual). The test for *coactivations*, i.e., testing for violations of the race-model inequality, compares the reaction time percentiles (e.g., 5%, 10%, 15%, etc.) of cumulative density functions. In the case that the percentiles of redundant signals temporarily precede those of the additive limits marks a violation of the race-model inequality and favors the presence of coactivations beyond independent racers. Instead of rare occasions, Miller (1982) argues that "coactivation models may be the rule rather than the exception." (p. 269), as he found evidence for coactivations in other studies as well (Miller, 1978, 1981). Furthermore, Miller (1982) discusses the "locus of coactivation" within the timeline of selective attention, decision, and response behavior. In line with late selection models, Miller favors a post-perceptual and decision-related "locus of coactivation", after a sufficient level of (parallel and independent) perceptual processing.

Interestingly, although only in the Appendix section, Miller (1982) briefly presents an eight-parameter model to predict response behavior, i.e., the cumulative density functions of reaction times, based on linear integration of activation levels and equations of McClelland (1979). Generally, the sensory activation of visual, auditory, or redundant information influences the rate of response activation level that initiates response behavior, e.g., button pressing. Furthermore, Miller (1982) added parameters on trial-by-trial variability (or attentional parameters) to indicate the distribution of attentional resources and the sensitivity to attentional fluctuations among sensory modalities, and the variance of response rates. Arguably, this model achieved relatively good fits, yet, Miller did not continue to elaborate on this parametric model but rather on the race-model inequality and its various predictions. For example, violations of the race-model inequality are associated with coactivations and evidence accumulation processes that integrate combined sensory information over a prolonged time (Miller, 1986).

Again, it must be emphasized that despite favoring the decision process as the "locus of coactivation", Miller (1986) cannot provide evidence for this, being still an open quest. Nonetheless, models to explain the "redundant-signals effect" for crossmodal search gained momentum, yielding many exciting insights and extended concepts, e.g., superposition or interactive race models, to which reference is made but no detailed description (Blurton et al., 2014; Colonius, 1986; Diederich & Colonius, 1991; Gondan et al., 2010; Gondan & Minakata, 2016; Miller et al., 2001; Mordkoff & Miller, 1993; Mordkoff & Yantis, 1991, 1993; Schröter et al., 2007; Schwarz, 1989, 1994; Ulrich & Giray, 1986; Zehetleitner et al., 2015). Notably, the crossmodal concept of race-model inequality and the redundant-signals effect also influenced visual search by showing that violations, i.e., evidence for coactivations, primarily occur across but not within visual dimensions, which is in line with the assumptions of the dimension-weighting account (Feintuch & Cohen, 2002; Grubert et al., 2011; Koene & Zhaoping, 2007; Krummenacher et al., 2002a, 2002b, 2010, 2014; Töllner, Zehetleitner, Krummenacher, et al., 2011; Zehetleitner et al., 2008).

Fascinatingly, the *locus of coactivation* issue and debate have not been resolved (completely) until today. While some favor late selection accounts, i.e., the presence of coactivations in the post-selective and decision stage, or at least post-perceptually (Feintuch & Cohen, 2002; Miller, Beutinger, et al., 2009; Miller & Reynolds, 2003; Mordkoff & Yantis, 1991, 1993), others promote the preattentive stage as the origin of coactivations that precedes the capture of attention, i.e., the completion of priority map computations (Grubert et al., 2011; Koene & Zhaoping, 2007; Krummenacher et al., 2014; Töllner, Zehetleitner, Krummenacher, et al., 2011; Zehetleitner et al., 2009). However, neither side can completely exclude the other, which is made more explicit by the following two quotes: “We cannot rule out the possibility that attention can also facilitate perceptual processing. … Our findings strongly suggest that at least one important role of attention is postperceptual gating of information to high-level processes that deal with response execution.” (Feintuch & Cohen, 2002, p. 368), and “Although the latter [post-selective] cannot be definitely ruled out, there is a growing body of evidence that fits nicely with a salience [preattentive], but not a decisional (or motoric), origin of coactivation effects.” (Zehetleitner et al., 2009, p. 1755).

Regardless of those opposing perspectives, there seems to be agreement that the locus of coactivation must occur primarily within the preattentive and post-selective stages, namely after perceptual “low-level” processing and before the initiation of motor responses (Mordkoff et al., 1996; Schröger & Widmann, 1998). It also seems reasonable that there are multiple coactivation loci during the accumulation of information and attentional selection (Yantis & Johnston, 1990). The concept of multiple loci aligns with the cascade model of McClelland (1979), which proposes a sequence of activation levels for the various processing stages, e.g., sensory, decision, and response activation levels. As described above (in the parametric model of Miller, 1982), a level's rate (or load) depends on the summed (or linear) integration of its preceding level. Hence, each stage's projection can be associated with (additional) interactive or coactive processes. Since crossmodal search is primarily associated with processes of divided attention, it is essential to note that a core concept of visual search, the overall *attentional priority map*, is not explicitly considered, e.g., in the coactivation debate. Again, this refers to the superordinate research question of this thesis of whether multisensory attentional resources are supramodally or modality-specific organized and to gather evidence that favors either a global priority map or multiple local sensory priority maps. To resolve this question empirically, the conceptual mixture of visual and crossmodal search is paramount. It can only succeed by agreeing on a preattentive architecture that guides crossmodal attention and by comparing the dynamics of electrophysiological components as neural markers for selective attention.

Although intersensory facilitation in bimodal search tasks with performances of bimodal targets exceeding those of unimodal targets is sufficiently documented (e.g., Girard et al., 2013; Miller, 1982; Nickerson, 1973; Todd, 1912), further evidence of the additive effects across three sensory channels stems from Diederich and Colonius (2004). The authors tested visual, auditory, and tactile modalities in a simple detection task with varying stimulus intensity levels and stimulus-onset-asynchrony (SOA) conditions. Whereas the visual stimulus was constant, a 62.5 ms presented flash with 250 lux on the screen, three stimuli were included for the auditory 1000 Hz tone with different loudness (70 dB, 80 dB, 90 dB) via headphones, and also for the tactile 27.7 Hz vibration with different intensities (40 mV, 145 mV, 448 mV). Overall, Diederich and Colonius (2004) found evidence for trimodal facilitation above the detection of bimodal and unimodal targets by measuring the multisensory response enhancement (MRE) index, i.e., the percentage of redundancy gains in comparison to unimodal or bimodal baselines. Their findings suggest that redundancy gains increase by adding additional relevant channels, showing that trimodal visual-auditory-tactile signals exceed bimodal signals, which also exceed unimodal signals. Furthermore, the authors showed the relevance of physiological synchronicity, i.e., the optimal time lag between sensory stimuli (Raab, 1962), and presented evidence for inverse effectiveness, i.e., the nonlinear relation of stimuli's intensity (Meredith & Stein, 1986). Interestingly, multiple optimal trimodal combinations were found with an enhancement above 11 %, e.g., a low-vibrating tactile target that precedes the visual flash and the 70 dB auditory tone by 50 ms (MRE 13.1 %) or a middle-vibrating tactile target that precedes the visual flash and the 90 dB auditory tone by 30 ms and 80 ms, respectively (MRE 13.1 %).

In addition to enhancement measurements, Diederich and Colonius (2004) also examined race-models by adapting the Miller (1982) procedure for trimodal targets. By comparing the cumulative density functions of the trimodal (tactile-visual-auditory) target with those of the bimodal sums minus the recurrent unimodal target (e.g., tactile-visual + visual-auditory – visual), Diederich and Colonius (2004) consistently found violations of the (adapted) race-model inequality. Besides coactivations for bimodal targets, this finding further supports additional trimodal coactivations of neural encoding. Interestingly, the authors favor (or at least discuss the role of) the superior colliculus for multisensory encoding (Meredith & Stein, 1986). By taking into account that coactivations or race-model violations depend on the spatial proximity of multisensory events, i.e., the enhancements decrease with an increasing spatial distance between those crossmodal targets (e.g., Frens et al., 1995; Spence & Driver, 1997a), relates elegantly to the spatio-topographic encoding of the priority map in visual search. Hence, coactivation models implicitly seem to entail spatio-temporal decoding of multisensory events, thus boosting performance due to the coherent alignment of crossmodal signals.

In addition to those exciting results that support coactivation models, it must be clarified that the test for violations of the race-model inequality by assuming context-independent racers is a very conservative criterion (Miller, 2016), as the absence of violations does not preclude the presence of (context-dependent) coactivations per se (Liesefeld et al., 2017). Therefore, other indirect or less strict, i.e., context-sensitive, criteria can be utilized to examine coactive processing, such as the ‘maximum negative dependency’ (Colonius & Diederich, 2017) or by measuring redundant-signal repetitions and intertrial switch effects (Liesefeld et al., 2017). In contrast to the comparison of percentiles in the cumulative density functions, those criteria represent aggregated scores (e.g., mean reaction times). By assuming a dimension-weighting account, context dependency arises from limited attentional resources distributed as weights across (visual) dimensions and their dynamic adjustments during the task, i.e., increasing weights for feature repetitions and decreasing weights for feature changes. For example, redundancy costs, reflecting delayed responses of feature repetitions in redundant embeddings compared to exact (single-feature) repetitions, suggest the presence of interactive (or coactive) weight adjustments and thus can be regarded as evidence against independent racers (Liesefeld et al., 2017). The ‘maximum negative dependency’ describes a sample criterion for spike counts and reaction times. It compares the minima of oppositely sorted reaction times in non-redundant signals against the reaction times of redundant signals. Whenever redundant signals are substantially faster than the racer's minima, this favors coactivation models (Colonius & Diederich, 2017).

Like the additive effects of dimensions in visual search, combining sensory modalities in crossmodal search improves performances and produces redundancy gains (e.g., Diederich & Colonius, 2004; Forster et al., 2002; Grubert et al., 2011; Iacoboni & Zaidel, 2003; Krummenacher et al., 2002a; Töllner, Zehetleitner, Krummenacher, et al., 2011). Importantly, redundancy gains, i.e., violations of the race-model inequality, arise primarily for combinations across visual dimensions or sensory modalities but not for intradimensionally or intramodally redundant signals. Despite the similar findings in visual and crossmodal search, this also points to a scientific gap, as the intramodal redundant combinations derive only by one dimension, e.g., tactile pulses or visual flashes (cf. Forster et al., 2002). This thesis examines this gap by combining visual dimensions (e.g., color and shape) with sensory modalities (e.g., vision and touch) in crossmodal task designs. It tries to gather empirical evidence on whether an extended multisensory dimension-weighting account is suitable, assuming comparable dimensions and sensory modalities, or whether another level of modality-specific processing is needed to explain the data.

Obviously, the redundant-signals effect can be associated with the first factor of selective attention, as it combines the activations of *features* across dimensions and sensory modalities. However, as proposed by saliency-summation models, those feature activations need to be integrated in a spatially consistent manner, that is, by weighted projections onto the *priority map*. While the assumption of 'one' priority map in visual search is reasonable, it might be more complicated for crossmodal search, i.e., search with divided attention across sensory modalities. Arguably, the spatial decoding of *locations* is essential for sensory modalities – vision, touch, and audio. Whereas vision space represents a quadratic fit with an optimum in the fovea center that decreases toward the periphery, the touch space relies on an anatomical reference frame (Assumpção et al., 2018; Carrasco et al., 1995; Föcker et al., 2010; Heed & Röder, 2010; Staugaard et al., 2016; Teder-Sälejärvi et al., 1999). This empirical evidence suggests different organizations of spatial functions for vision and touch, with external and egocentric decoding, respectively. Further, this proposes modality-specific saliency maps within the visual cortex and somatosensory regions (Chambers, Stokes, et al., 2004; Forster et al., 2016). However, evidence on crossmodal space indicates that the egocentric touch space merges with the external vision space within the parietal region, e.g., intraparietal sulcus or temporo-parietal junction (Chambers et al., 2007; Downar et al., 2000; Driver & Spence, 1998; Eimer, Forster, et al., 2003; Eimer & Driver, 2000; Kanwisher & Wojciulik, 2000; Macaluso et al., 2000, 2002a; Spence, Pavani, et al., 2000; Spence & Santangelo, 2009), indicative of a supramodal mechanism that is embedded in the fronto-parietal attention networks and the personal space (Buschman & Kastner, 2015; Graziano & Cooke, 2006; Posner & Dehaene, 1994; J. F. Stein, 1989).

While features and locations represent the stimulus-driven factors of selective attention (feedforward factors), expectancy and prior history describe the goal-driven or priming factors (presumably feedback factors). In Spence, Nicholls, and Driver's (2001) experiment, the authors combine visual, auditory, and tactile stimuli in a localization task, i.e., the discrimination of left- and right-presented targets. Each trial presented either a visual red light, an auditory signal (90 dB white noise), or a tactile pulse (around 200 Hz) for 50 ms on the left- or right-located LED, speaker, or tactile stimulator on the index fingers, respectively. Overall, the experiment included four conditions. For each block (16 blocks total), participants were instructed to divide their attention equally across the three modalities or to focus on one of them, i.e., the expected modality with a 75 % occurrence rate. The findings of Spence and colleagues (2001) suggest that the target localization for the expected modality improves reaction times, generating substantial benefits for auditory (Δ 34 ms), visual (Δ 25 ms), and tactile (Δ 17 ms) signals about their baseline of equally divided attention. In contrast, unexpected targets produced sufficient costs (Δ 31-80 ms). Although quite compelling, this finding on endogenous expectancy only tells half of the story.

In the next step, Spence, Nicholls, and Driver (2001) showed that those benefits and costs primarily derive from exogenous intertrial relations, i.e., ipsimodal repetitions or crossmodal shifts, respectively. Interestingly, crossmodal shifts accounted for two-thirds of the total effect (compared to the sum of benefits and costs in milliseconds), revealing a vast “modality-shift effect” (MSE) in line with previous studies (cf. R. Cohen & Rist, 1992; Ferstl et al., 1994; Klein, 1977; Posner et al., 1976). Ultimately, the authors only looked at those crossmodal targets to test whether the relevance for (endogenous) expectancy persisted when correcting for the ipsimodal bias, i.e., the high occurrences of expected modality repetitions. Somewhat ‘unexpectedly’, the benefits in the expected modalities shrank (or even vanished) for auditory (Δ 10 ms), visual (Δ 10 ms), and tactile (Δ -6 ms) targets to non-significance. At the same time, the costs of the unexpected modalities remained highly significant. Finally, Spence, Nicholls, and Driver (2001) discuss attentional asymmetries across sensory modalities, where shifting attention away from touch towards the visual or auditory modality was most costly. Notably, their argumentation tallies with the above-described spatial functions, as these results suggest a genuine dissociation between external, visual or auditory, and internal or egocentric tactile modalities (by referring to Eimer &

Driver, 2000; Roland, 1982; J. F. Stein, 1989). Their explanation gains further support by showing comparable reaction times within the blocks of distributed attention, which makes a saliency imbalance between those modalities unlikely.

Overall, the study of Spence, Nicholls, and Driver (2001) is an excellent example of how the various factors of selective attention are intertwined, bottom-up saliency, top-down expectancy, and intertrial dependencies. It underscores the difficulty of obtaining effects of endogenous expectancy, i.e., top-down control, despite explicit instruction of participants. Furthermore, the authors strengthen the relevance of intertrial effects or exogenous priors for crossmodal search. The amount of switch costs in the modality-shift effect strongly indicates that attention resources are not independently distributed among sensory modalities, as it takes time to detach them. Generally, the study has a striking similarity with visual search and the dimension-weighting account (Found & Müller, 1996; H. J. Müller et al., 1995), as it deciphers the dynamic (re-)prioritization across trials, i.e., the distribution of resources or weights, similar to those of visual dimensions, yet for crossmodal search.

Evidence from other ‘modality-shift effect’ studies extends the understanding of modality-specific asymmetries and attentional engagement by analyzing the time-course between cue-target or target-target presentations (e.g., Miles et al., 2011; Rodway, 2005; Turatto et al., 2002, 2004). For example, Miles, Brown, and Poliakoff (2011) conducted a series of visuo-tactile search experiments similar to those of Spence et al. (2001). Crucially, the authors varied the ‘stimulus-onset asynchrony’ (SOA) (150 ms, 600 ms, & 1000 ms) and the ‘intertrial interval’ (1250 ms, 1600 ms, & 2000 ms) to investigate exogenous effects that influence modality encoding from preceding noninformative cues or targets. During the cue-target tasks, participants had to localize a visual or tactile target, a 200 ms red-light LED flash or a 100 Hz vibration (bone conductors), at the right hand’s index finger or little finger, respectively. Further, noninformative cues (orange LED or white noise vibrations) were presented in the middle of the hand’s back to avoid location-priming effects on the left-located index finger and the right-located little finger.

Miles et al. (2011) showed that the endogenous presence of a modality-shift effect (MSE) depends on attentional engagement with the preceding exogenous cue. Whereas no cue engagement (10 ms condition) produces solid benefits for visual target repetitions (but no tactile benefits), cue engagement (50 ms odd-ball condition) diminishes the MSE of the preceding target’s modality, except a late tactile repetition benefit (1000 ms condition). Counterintuitively, the (attended) exogenous cue (50 ms) produced reaction time costs for tactile cue-target repetitions in the 150 ms and 600 ms conditions, but reaction time benefits for the 1000 ms condition, congruent with the above endogenous target-target benefit.

Notably, no differences, neither costs nor benefits, were found in visual cue-target relations. This finding indicates that the exogenous cue effect is more substantial for touch but absent for vision. Although not discussed in too much detail by the authors, this finding indicates modality-specific “refractory periods” for vision and touch. Whereas vision can quickly (and almost independently) integrate new “same-modality” information (< 150 ms), tactile processing produces delays for “same-modality” cue-target combinations (< 600 ms).

Interestingly, the visual “refractory period” of attentional integration aligns with findings in visual search, where the impact or costs of various visual cues, producing a delay in reaction times, vanishes or flattens already after around 200-300 ms (Wolfe et al., 2004). However, congruent to the argumentation of Miles et al. (2011) of favoring a transition from exogenous to endogenous processes, evidence from other studies suggests that the timing within 100–300 ms for successive target presentations is especially vulnerable for attentional selection, i.e., producing a substantial delay in reaction times during matching tasks (Dalvit & Eimer, 2011; Jenkins et al., 2016). Notably, the matching of successive targets is most effective when both targets fall either within the same perceptual snapshot or cycle (< 100 ms), ‘percept-percept matching’, or when the preceding target completed the transition to working memory (> 300 ms), ‘image-percept matching’. Further, matching becomes less efficient during the stage of attentional selection (100–300 ms), as the new target enters before the old target exits, inducing competition or serial conflict for attentional resources.

Returning to the study of Miles et al. (2011), their findings suggest that the MSE is primarily driven by endogenous intertrial effects, producing switch costs (e.g., vision → touch) that diminish with an increasing intertrial interval in both vision and touch. Importantly, switching from touch to vision was less costly than vice versa, a modality-specific asymmetry, which once again demonstrates that visual processing is more effective than tactile processing, adding evidence to the claim of “visual dominance” (Klein, 1977; Posner et al., 1976). Although visual dominance across our sensory modalities seems intuitive, its scientific evaluation is more complicated. Similar to saliency in visual search targets, studies that promote visual dominance (or auditory dominance) in crossmodal search (Gondan et al., 2007; Miles et al., 2011; Posner et al., 1976; Spence, Nicholls, et al., 2001), often lack an adequate alignment (comparing the global mean of reaction times) or description of crossmodal stimuli.

Furthermore, Miles et al. (2011) show that the impact of the MSE extends over the two preceding trials, revealing additional benefits and costs for same-same and different-different modalities, which the authors link to the “neural trace theory” (Zubin, 1975). Congruent with the discussion in Miles et al. (2011), this finding indicates a dynamic weighting mechanism for crossmodal targets that pull attentional resources by activating modality-specific feature maps (Töllner et al., 2009). Essentially, weight adjustments during crossmodal search require time as they either decay or increase over successive trials, in accordance with the DWA in visual search (Found & Müller, 1996).

Despite evidence for endogenous modality-cueing, Miles et al. (2011) also found (as expected) spatial-cueing effects across the finger's locations. Interestingly, shifting from the index finger to the little finger or vice versa, i.e., the opposite location, during successive trials was substantially faster than repeated finger locations. Here, tactile location-shifts revealed more net benefits than visual location-shifts, while both modalities revealed a delay for repeated locations across stimulus-onset-asynchrony (SOA) conditions, indicative of 'inhibition of return' effects (Klein, 2000; Miles et al., 2011; Posner et al., 1985). Although Miles and colleagues (2011) present evidence for both modality-cueing and spatial-cueing across successive trials, they did not analyze both effects in combinations, e.g., repetitions vs. changes in targets and locations.

Similar to visual search, the endogenous top-down prioritization of targets is not questioned in crossmodal search as noninformative cues or distractors can be ignored sufficiently. Nonetheless, it remains challenging to distinguish between the effects of short-term priming and long-term templates (cf. Spence, Nicholls, et al., 2001). So far, it seems that knowing one or more crossmodal targets in advance fastens attentional selection by up-weighting (or pre-activating) target-specific feature maps for congruent targets, but delays for incongruent, unknown, or unexpected targets (e.g., Talsma et al., 2010). However, during the time-course of search tasks, those 'priors' dynamically adjust their weights based on the 'evidence' of the target's occurrence rates, especially in mixed block designs, reflecting statistical learning processes (e.g., Glicksohn & Cohen, 2013). Another perspective on goal-driven attention derives from 'reward' studies in visual search with macaque monkeys (Chelazzi et al., 1998). In this study, monkeys had to make saccades towards a specific image that matched with a preceding central cue image (e.g., flower-flower matching). Correct saccades were rewarded with a drop of juice. By recording the firing rates of neurons within distinct receptive fields in the inferior temporal cortex, the authors found a functional dissociation between spatially congruent ('good') and incongruent ('poor') firing rates of receptive fields that indicate continuous attentional engagement after 150-200 ms of array onset and 70-80 ms earlier than the target saccade.

The activation of topographic and relevant receptive fields resembles a matching (or spatial binding) for goal-driven and stimulus-driven information (cue-target matching) within the inferior temporal cortex. Interestingly, the 150-

200 ms timing aligns with the 150 ms duration of object recognition in *Guided Search* (Wolfe, 1994, 2021) and the timing of multisensory integration, where crossmodal interactions, i.e., additive or coactive effects, start at around 150-160 ms (Gondan et al., 2007; Talsma & Woldorff, 2005; Teder-Sälejärvi et al., 2002). Based on the presented empirical evidence, the time window from 150 ms onwards seems to play a critical role in visual and crossmodal search, as it combines the various factors of selective attention and embeds similar mechanisms of weighted integration. Furthermore, it reflects a temporal marker for transitioning from the preattentive parallel stage to the serial stage of attentional selection (and post-selective stage). To obtain a set of concise research questions, finalizing the empirical framework for crossmodal search will be essential by distinguishing modality-specific and dimension-specific encoding aspects, contrasting the extended dimension-weighting account.

Modality-Weighting Account

Although crossmodal attention based on modality-specific computations seems rather trivial given the presented empirical evidence and its genuine origin from divided attention research, it nonetheless becomes empirically relevant in the context of focused attention, where crossmodal information is combined and distinguished adequately, favoring the presence of supramodal mechanisms of attentional orienting across sensory modalities. Because crossmodal search studies 'only' differentiate between distinct processing stages or activation levels, e.g., sensory, decision, and response stages (McClelland, 1979), it misses out on a straightforward conceptualization and empirical evaluation of saliency summation and attentional prioritization like the dimension-weighting account (DWA) in visual search (Found & Müller, 1996). Therefore, the main concern of this thesis is the exact formulation of such a crossmodal account.

Following up on the crossmodal search findings, it is commonly agreed that crossmodal redundancies produce substantial performance benefits, while modality shifts create tremendous performance costs (e.g., Diederich & Colonius, 2004; Forster et al., 2002; Gondan et al., 2007; Miles et al., 2011; Miller, 1982; Spence, Nicholls, et al., 2001). As indicated earlier, the presence of 'weighted integration' can explain this empirical evidence reasonably well: combined weights in redundant signals enhance attentional selection by improving the signal-to-noise ratio, and costs for switching between sensory modalities arise from intersensory weight shifting. Overall, these findings demonstrate crossmodal dependencies that constitute performance behavior regardless of the underlying processing stages (e.g., preattentive or post-selective).

In another crossmodal search study that investigated the modality-shift effect, Töllner and colleagues (2009) conducted two visuo-tactile search experiments where participants had to discriminate either the target's modality (a visual LED flash or tactile vibration) or specific visual or tactile target features (red or green LED colors, 40 Hz or 100 Hz tactile vibrations) by pressing the respective foot pedals. Generally, the visuo-tactile search task was very similar to the experimental setup in Spence et al. (2001) but without an auditory condition. In the first experiment, the setup consisted of seven green-light LEDs and two tactile solenoids arranged symmetrically on the left and right sides and index fingers, respectively (three LEDs at each side and one in the center). White noise was played throughout the task. Each trial started with a 500 ms fixation cross at the center (the bottom of a computer screen) and was followed by a 200 ms target, either a visual green light or a tactile vibration. To respond to the target's modality, answers were given by pressing the respective foot pedal (e.g., left: vision & right: touch, or vice versa). In combination with a 1000 ms interstimulus interval and average visual (479 ms) and tactile (492 ms) responses, a trial lasted around 2000 ms.

Overall, Töllner et al. (2009) found two main effects and an interaction term in their first experiment for the factors of *modality shift* and *location shift*. Here, significant costs occurred for switching the modality between successive trials (511 ms), e.g., vision → touch, compared to modality repetitions (461 ms), e.g., vision → vision. Also, the authors reported significant benefits for repetitions in the target's location (481 ms), e.g., left → left, and a delay in location changes (490 ms), e.g., right → left. Interestingly, this finding of location repetition benefits contrasts those of Miles et al. (2011), who reported benefits for location switches during successive trials between the index finger and the little finger within the same hand. Despite a comparable task design and intertrial durations of target-target relations (roughly around 2000 ms), both experiments vary in that Töllner et al. (2009) analyzed the location switching between both hands combined for visual and tactile targets, while Miles et al. (2011) examined visual and tactile switching between fingers separately within the same hand, revealing a crossmodal interaction. Hence, this indicates a research gap that requires further examination.

To test their critical assumption that attentional switch costs, i.e., the presence of a modality-shift effect, already build up during the early preattentive stage that precedes attentional selection, Töllner et al. (2009) recorded the electrophysiological activity via electroencephalography (EEG) parallel to performances. Crucially, the authors wanted to examine whether a *supramodal* marker for attentional reorienting exists, indicating consistent switch costs for both visual and tactile modality shifts. By averaging EEG epochs, i.e., stimulus-locked segments, across frontal, central, and parietal electrode positions, they retrieved the *event-related potentials* (ERPs). They compared the amplitude of the early tactile (somatosensory) and visual component deflections within distinct time windows, namely the P1, N1, and N2 components. The authors found a significant increase in N1 amplitude for modality changes in vision and touch across those electrode positions relative to modality repetitions, indicative of a supramodal mechanism. At the same time, there were no findings in P1 and N2 or negligible interactions.

To ensure that this 'early' supramodal N1 effect, within 140-180 ms, originates during the perceptual and preattentive stage and not from a 'refractory' response-related bias of the preceding modality-different target, Töllner and colleagues (2009) conducted a second experiment by controlling the response patterns. Instead of simple sensory discrimination, participants had to distinguish visuo-tactile features: two visual LED colors, red and green, and two tactile vibration patterns, 40 Hz and 100 Hz. Importantly, response patterns were balanced by a visual and tactile feature for each pedal, e.g., a red-light or a 40 Hz target for left-pedal and a green-light or 100 Hz for right-pedal responses (or vice versa), eradicating a potential modality-specific response bias. Similar to the previous experiment, response patterns were counterbalanced and switched during the task's half-time break.

Again, Töllner et al. (2009) replicated the modality-shift effect (596 ms vs. 545 ms) for successive changes between visual and tactile targets. In contrast to their first experiment, they found a response difference between visual (546 ms) and tactile (595 ms) target detection, with the former being faster, yet both were seemingly slower than those in the first experiment (479 ms & 492 ms). Notably, this difference results from the adapted task design, indicating that 'modality-discrimination' (485 ms) is faster than exact 'feature-discrimination' (571 ms), reflecting 'rough' and 'precise' decoding modes. This delay for precise discrimination was especially the case for both tactile targets. While a simple tactile stimulation (experiment 1) could be detected relatively effortlessly, the tactile discrimination between 40 Hz and 100 Hz vibrations (experiment 2) was more difficult. Instead of a common position effect (experiment 1), Töllner et al. (2009) reported a three-way interaction for *modality*, *modality shift*, and *location shift*. Importantly, tactile target discrimination was fastest for dual repetitions: modality and location, e.g., left vibration → left vibration, whereas visual repetitions were comparable across location-shift conditions. Unfortunately, Töllner et al. (2009) did not analyze the four distinct features and their intramodal feature shifts separately (e.g., 40 Hz → 100 Hz, or red → green). Finally, the authors examined another factor of *response shift*. They reported interactive outcomes for repetitions and changes in responses, modalities, and locations, e.g., faster reaction times for repeated modalities and pedal responses.

The EEG findings in the second experiment were congruent with those in the first experiment, yet more detailed. Again, the somatosensory N1 sensitively depicted modality shifts (vision → touch) with increased amplitudes. Although the authors found an interaction between tactile modality shifts and location repetitions with an increased N1 (e.g., left visual target → left tactile target), crucially, the N1 was unaffected by response conditions. On the other hand, findings in the visual N1 were more restricted than before, as the enhanced negativity for modality shift (touch → vision) was mainly over frontal regions and, by declining, becoming absent over parietal regions. Ultimately, by finding no systematic interaction between *modality* (vision, touch) and *modality shift* (repetition, change), Töllner et al. (2009) interpreted their evidence, the presence of a 'modality-unspecific' anterior N1 effect, in terms of *supramodal mechanism* or marker that sensitively depicts modality shifts during successive trials already during the preattentive stage, contrasting response-related accounts (e.g., Mortier et al., 2005).

Besides these exciting findings, the relevance of Töllner et al.'s (2009) paper stems from the direct comparison with the visual search literature. Crucially, the authors reconciled their electrophysiological anterior N1 as a marker for "modality-shifts" with a similar, yet later, fronto-central '*transition* N2' component for visual "dimension-shifts" (e.g., red color target → left-tilted orientation target) (Gramann et al., 2007). Based on the preattentive hierarchy of the DWA (Found & Müller, 1996; H. J. Müller et al., 1995), Töllner et al. (2009) argue that the anterior N1 reflects a crossmodal search marker for "attentional weight shifting", and therefore, adheres the same 'weighting rules' as in visual search. This reasoning led the authors to propose a *modality-weighting account* (MWA), a crossmodal generalization and extension of the DWA. Remember, the DWA consists of layers of dimension maps, i.e., independent analyzer units, that combine specific activations from various feature maps. Finally, those activations get integrated onto an attentional priority map in a spatially consistent manner and ranked based on saliency and relevance, i.e., the interplay of attentional factors guiding the serial capture of selective attention.

The innovative idea of Töllner et al.'s (2009) MWA is the extension of the preattentive hierarchy by an extra layer of modality-specific computations that integrate the *weighted* signals of their underlying (modality-specific) dimensions (e.g., vision: color, orientation, etc.; touch: vibration, pressure, etc.). As before, attentional resources will be limited across these modality maps (crossmodal search), yet they likely exceed intramodal capacities (e.g., visual search). This issue relates to the fundamental (and opening) question of whether attentional resources share a common pool with a (*global*) limited capacity or whether each modality derives its own (*local*) resource limits, where the overall capacity equals the sum of separate pools (Miller, 1982; Treisman, 1969). Finally, these modality maps are linked to a topographic "supramodal master map" that combines the multisensory signals into an attention-guiding mapping of crossmodal space.

Although the MWA of Töllner et al. (2009) is 'merely' a conjecture from their anterior N1 findings, without 'unequivocal evidence', it nonetheless yields the last building block to obtain a complete scientific framework for empirical research. The 'co-existence' of a multisensory DWA and MWA, describing the underlying preattentive dynamics in crossmodal search, marks an open gap in attention research (s. Figure 1–2). The critical quest to favor an MWA, as depicted by Töllner et al. (2009), is to provide evidence for the additional layer of modality maps, which should substantially exceed the (subordinate layer of) dimension maps. By applying Occam's razor, the null hypothesis assumes no functional dissociation between modality-specific and dimension-specific processes, i.e., both thrive from same-level computations and weightings, which indicates that the extended (multisensory) DWA is sufficient. As shown later in the "Hypotheses and Limitations" section, this conceptual juxtaposition (MWA vs. DWA) leads to concise and superordinate research questions for each of the five attention factors, anticipating performance behavior as well as the neural dynamics of attentional correlates.

Besides the layer of modality maps, another decisive description of the MWA is the assumption of a *supramodal priority map* that combines the modality-specific spatial encoding of relevant sensory modalities. Regarding visuo-tactile search, this supramodal priority map fuses the spatial functions of external visual information and the egocentric tactile information (Assumpção et al., 2018; Carrasco et al., 1995) into coherent multimodal representations in crossmodal space. However, the empirical evaluation of a supramodal priority map is not trivial, as the variety of counter-accounts must be considered. For example, instead of a common, supramodal, and crossmodal priority map, it might already be sufficient to assume separate modality-specific priority maps (e.g., visual or tactile priority maps) interacting with each other after completing modality-specific saliency computations through crossmodal links. Whereas the outcome of 'multimodal representations' would be the same in both accounts, the neural functionality, however, is of decisive difference, describing crossmodal prioritization as either 'static' and 'unique' or 'distributed' within the frontoparietal attention network (Eimer et al., 2002; Eimer & Van Velzen, 2002; Spence & Driver, 2004).

Regarding the redundant-signal effect (RSE) and the modality-shift effect (MSE), crucially, the MWA assumes that the performance benefits from crossmodal redundant targets should exceed those of intramodal redundant ones that combine distinct dimensions, while the costs for modality shifts should exceed those of intermodal dimension shifts. Both outcomes favor the additional layer of modality maps above the layer of dimension maps. Interestingly, the empirical evidence for the RSE points to crossmodal links within parietal regions, presumably summing the weights of crossmodal activations, whereas the MSE in Töllner et al. (2009) suggests a supramodal involvement of frontal areas in the anterior N1 component for weight shifting between sensory modalities. Again, both brain regions represent characteristic attention network hubs and are further consistent with the findings of various studies (Chambers et al., 2007; Macaluso et al., 2002b; Man et al., 2015; Quinn et al., 2014) that found supramodal and invariant activations, for example in the temporo-parietal junction and the inferior frontal cortex during crossmodal search tasks. Likewise, there is agreement that unimodal processes, i.e., the processing of low-level sensory information, temporarily precede the activation of crossmodal links or the convergence to “higher association areas” (Quinn et al., 2014; Senkowski et al., 2008; Talsma & Woldorff, 2005; Töllner et al., 2009). Of note, multisensory integration or crossmodal binding occurs within 157-260 ms in the supramarginal gyrus and from 280 ms onwards from the middle frontal gyrus (cf. Quinn et al., 2014), falling within the 100-300 ms time window of attentional selection (e.g., Dalvit & Eimer, 2011; Jenkins et al., 2016).

Regarding the supramodal priority map, the anterior N1 within 140-180 ms seemingly disqualifies as a spatio-temporal candidate due to its early origin and spatial blindness, since spatial sensitivity posits a fundamental prerequisite of the priority map. Therefore, the anterior N1 indicates perceptual gating by updating (or re-adjusting) the weights of sensory channels as reflected by the amplitude increase for modality shifts. Furthermore, the anterior N1 reflects the ‘rough’ stage of ‘modality-discrimination’, consistent across both experiments, in contrast to the subsequent ‘precise’ stage of exact ‘feature-discrimination’ (experiment 2) that further relies on a higher ‘attentional resolution’, thus delaying reaction times. Therefore, the frontal N1 reflects a control mechanism for attentional orienting in crossmodal search but not a spatially-coherent representation of multisensory signals (Töllner et al., 2009).

Multisensory integration, i.e., the convergence of sensory information, occurs parallel to attention processes, spanning multiple stages with spatio-temporally

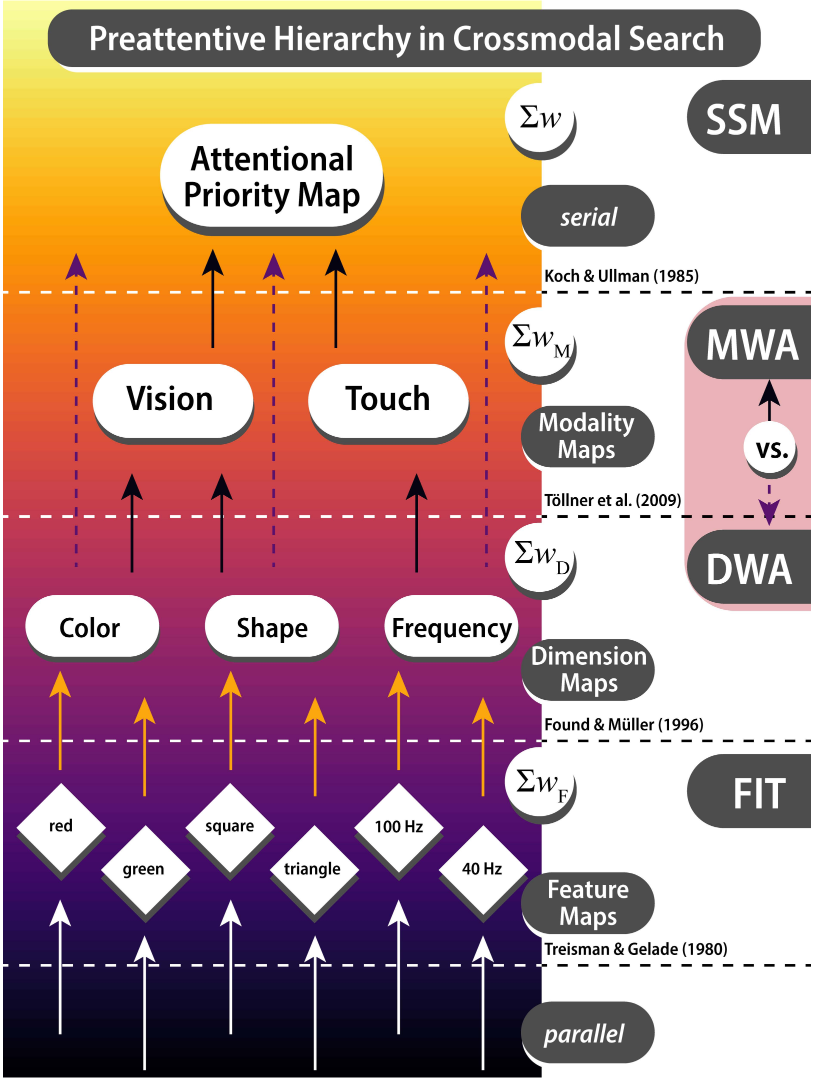

Figure 1–2. The preattentive hierarchy for crossmodal search is shown as scientific framework. The preattentive architecture includes a hierarchy of distinct topographic maps on multiple layers or levels. In visual search, this hierarchy combines the parallel activation of *feature maps*, as proposed by the feature-integration theory (FIT), within independent analyzer units or *dimension maps* (orange solid lines) that project their loads via weighted summation (Σw) in a spatially coherent fashion onto an *attentional priority map* (purple dashed lines). Based on saliency-summation models (SSM), the priority map guides selective attention serially to the most salient or relevant (display or scene) information, i.e., the highest feature-contrast based on a 'winner-takes-all' mechanism. The main research interest is to provide evidence for the preattentive hierarchy in crossmodal search. Here, it will be essential to examine whether a multisensory dimension-weighting account (DWA) already suffices or whether the preattentive hierarchy needs to be extended by another intermediate layer of *modality maps* (as proposed by Töllner et al., 2009), describing a modality-weighting account (MWA). Generally, the MWA proposes *modality-specific* saliency maps (e.g., for vision or touch) that merge their spatio-topographic loads onto a *supra-modal* priority map.

distinct properties (Man et al., 2015; Meyer & Damasio, 2009). Hence, it is no coincidence that the proposed modality-weighting account (Töllner et al., 2009) has similarities to other scientific concepts, such as the convergence-divergence-zone (CDZ) framework (Damasio, 1989; Meyer & Damasio, 2009), the global workspace theory (Baars, 1997; Dehaene et al., 1998; Shanahan & Baars, 2005), or transmodal nodes (Mesulam, 1998). In essence, all these models embed a hierarchical or network structure that includes forward and backward connections across sensory, memory, or attentional systems that are connected via heteromodal, transmodal, or association hubs, enabling higher cognitions (e.g., working memory) and subjective experiences (e.g., consciousness) as well as coordinating actions (e.g., motor responses).

Fascinatingly, these models (attempt to) describe the whole set of cognitive functions by more or less explicitly utilizing the term *maps* as distributed functional mechanisms and resources. For example, the CDZ framework distinguishes between perception and recall (or mental imagery) modes, reflecting the activation of relevant maps (e.g., feature binding) through perceptual forward integration or memory-based backward reconstruction. However, these models need more explicit descriptions of the underlying mechanisms despite their functional universality, which the MWA entails by incorporating the (explicit) weighting mechanism among layers of its preattentive hierarchy.

Last but not least, to enable the empirical comparability of both the DWA and MWA, it is vital to address the issue of 'search displays', which are generally quite different among visual and crossmodal search experiments. Whereas targets in visual search can 'pop-out' among many nontargets and differ by numerous contrasts in features and dimensions, i.e., complex displays from computer screen presentations (Luck, 2011), the displays in crossmodal search paradigms appear somewhat limited by including just one or a couple more LEDs, loudspeakers, and tactile stimulators (e.g., Gondan et al., 2007; Kanayama et al., 2012; Macaluso et al., 2002b; Miles et al., 2011; Spence, Nicholls, et al., 2001; Talsma & Woldorff, 2005; Töllner et al., 2009). Despite congruent findings, this 'display difference' between visual and crossmodal search makes a direct comparison seem a bit arbitrary. To overcome the 'display issue', this thesis introduces a novel visuo-tactile search paradigm by extending the number of tactile stimulators while guaranteeing visual search compatibility. Overall, the combination of visual projections and tactile vibrations within a joint (multisensory) screen enables the collocated presentation of visual and tactile targets and nontargets across a fixed number of locations (8 or 10), symmetrically arranged across the left and right hemifields and fingertips. Crucially, this expansion of crossmodal items (i.e., increasing the set size) enables a direct comparison between the intramodal dimension-specific dynamics of visual search and the modality-specific dynamics in crossmodal search. Using the visuo-tactile search paradigm, the empirical evaluation of dimension and modality maps is possible, favoring either an MWA or an extended DWA.

Electrophysiological Evidence

Having introduced the empirical framework (MWA vs. DWA), the primary focus of the thesis lies on the preattentive stage and the capture of selective attention as the outcome of attentional priority computations. Hence, it is essential to consider this time window in greater detail by examining the electrophysiological evidence. Of note, there are plenty of EEG findings parallel to the scientific evolution of visual and crossmodal search, similar to those of Töllner et al. (2009), whose findings should now be more intelligible, given the conceptual knowledge. Of course, other neuroimaging techniques, such as fMRI, PET, or TMS, also revealed various insights into crossmodal attention besides EEG findings, for example, establishing the structural nodes of the frontoparietal attention network (Brefczynski & DeYoe, 1999; Chambers et al., 2007; Chambers, Payne, et al., 2004; Chambers, Stokes, et al., 2004; Corbetta et al., 1993; deBettencourt et al., 2015; Downar et al., 2000; Fink et al., 1997; Kristjánsson et al., 2007; Kucyi et al., 2017; Macaluso et al., 2000, 2002a, 2002b; Man et al., 2015; Pollmann et al., 2000; Posner & Dehaene, 1994; Shafritz et al., 2002). Since EEG is the technique of choice for the following studies, those other techniques derive a supportive role in the quest for suitable EEG markers and the interpretation of their findings. Yet, fMRI will be crucial in the "Attentional Priority Map" section.

Generally, the EEG measures voltage changes over the scalp from many electrode channels with an incredible temporal resolution: the sampling rate, i.e., the number of measured time points per second, is often 1000, retrieving one voltage score per millisecond for each electrode. The scientific strength of the EEG reveals itself in combination with the exact timing of "task events", e.g., timestamps that indicate a visual display's onset (including targets, nontargets, and/or distractors) or button-press responses. Commonly, the EEG gets sliced into same-length segments or *epochs* that align on those task events, containing the various task conditions. The *event-related potential* (ERP) reflects the (condition-wise) average of those (selected) epochs across time points and electrodes (or electrode clusters). Crucially, the (average) ERP amplifies the signal-to-noise ratio (SNR) with increasing trials, i.e., the number of epochs; it strengthens the *signal* while canceling out the *noise*, revealing some stable "event-related components" with characteristic amplitude deflections and timings. The statistical evaluation is based on the *amplitude*, e.g., the positive maxima and negative minima, and the *latency*, e.g., the peak timing, of the subject-wise ERP components. Before analyzing (relevant) ERP components, a sequence of specific preprocessing steps is performed, such as re-referencing, high-/low-pass filtering, the removal of ocular and muscular artifacts, independent component analysis (ICA), and baseline correction to improve the SNR and ensure the validity of statistical findings (Bell & Sejnowski, 1995; Bigdely-Shamlo et al., 2015; Luck, 2014).

As indicated above in the previous sections, the main focus is on preattentive computations that precede and guide the capture of (crossmodal) selective attention. Therefore, choosing an adequate EEG component that indicates attentional processes for vision and touch is critical to establish empirical evidence favoring a preattentive hierarchy based on modality maps (MWA) or dimension maps (DWA), respectively. To obtain such neural markers, it is essential to consider the influence of the perceptual, preattentive, and attention stage on the various EEG components within 300 ms after stimulus onset.

By starting at the perceptual level, well-known visually evoked potentials or components are the P1, the first positive peak within 90-130 ms, and the N1, the first negative trough within 160-200 ms, which arise most clearly over lateral parieto-occipital regions (Eason, 1981; Hillyard & Münte, 1984; Mangun, 1995; Van Voorhis & Hillyard, 1977). Although the P1 and N1 are elicited consistently by sudden visual changes, indicating a (purely) perceptual bottom-up mechanism regardless of the underlying visual relevance, e.g., a visual target or distractor, nonetheless, those studies also revealed a mechanism for spatial attention that increases the contralateral P1 and N1 amplitudes for visual signals located in the attended visual hemifield. Yet, this ‘early’ attention impact was likely due to a preceding and relevant cue, initiating an attention shift towards the cued visual hemifield before the target’s appearance and, therefore (simply) resembles the cue-target validity (Mangun, 1995; Mangun et al., 1987; Mangun & Hillyard, 1987, 1991). Essentially, the ‘spotlight’ of attention is allocated in advance to the cued spatial location, amplifying the visually evoked P1 and N1 components in the visual cortex for the visual target.

Besides the visual-evoked potentials (VEPs), there are also auditory-evoked potentials (AEPs) and somatosensory-evoked potentials (SEPs), each indicating distinct spatio-temporal components for vision, audio, and touch that are linked to processes of perceptual integration (V. P. Clark et al., 1994; Cruccu et al., 2008; García-Larrea et al., 1992, 1995; Hillyard et al., 1973; Levit et al., 1973; Näätänen, 1988; Näätänen & Gaillard, 1983; Näätänen & Picton, 1987; Picton & Hillyard, 1974; Van Voorhis & Hillyard, 1977). Generally, those studies presented sensory information to the left or right visual hemifields, ears, or hands, and participants were instructed to either attend to the left and unattended to the right side or vice versa (blockwise or via cues). Concurrently, EEG signals were recorded, and the spatio-temporal deflections of early ERP components (e.g., P1, N1, P2) were analyzed and shown as topomaps (Driver & Spence, 1998; Eimer & Driver, 2001; Eimer & Schröger, 1998; e.g., Föcker et al., 2010). Of note, early ERP components for attended locations improved regardless of the preceding cue's modality (e.g., sound cues that precede visual stimuli) (e.g., Giard & Peronnet, 1999; McDonald et al., 2000).

Although the term "N1 component" is common in each sensory modality, its timing and spatial deflection differ genuinely between the auditory N1 (70-120 ms), visual N1 (160-200 ms), and somatosensory N1 (100-150 ms) (e.g., García-Larrea et al., 1995; Hillyard & Münte, 1984; Näätänen & Picton, 1987). Furthermore, while for auditory perceptual processing, i.e., the integration of auditory information, often an N1-P2 complex is mentioned, a 100 ms negative trough followed by a positive peak after around 180 ms (Näätänen & Picton, 1987; Picton & Hillyard, 1974; Woldorff & Hillyard, 1991), in contrast, the term P1-N1 complex is used in visual perceptual integration, indicating a positive peak (120 ms) that is succeeded by a negative trough (170 ms) (Gruber et al., 2005; Klimesch et al., 2004). Notably, auditory and somatosensory processing temporarily precedes visual integration based on the N1 timing, and some form of sensory dipole with opposite polarities occurs between auditory and visual processing (N1-P2 complex vs. P1-N1 complex).

The N1 component (and subsequent EEG components) are assumed to reflect the "activation of neural assemblies" (Picton & Hillyard, 1974, p. 197) within modality-specific brain regions, for example, the parieto-occipital visual cortex or the central somatosensory cortex. Since EEG signals derive from post-synaptic potentials (Luck, 2014), new insights from combined EEG and local field potentials suggest that the P1 component primarily derives from excitatory activity in the supragranular layer, while the N1 component reflects inhibitory activity (Bruyns-Haylett et al., 2017). This finding aligns well with previous studies on ongoing alpha oscillations, which (combined with theta oscillation) produce the P1-N1 complex, leading to the external and internal states for the P1 and N1, respectively (Gruber et al., 2005; Hanslmayr et al., 2007, 2011). Another takeaway from Bruyns-Haylett and colleagues' (2017) study is the 'Gaussian' description of EEG components. Similar to Luck (2014), the average ERP consists of independent and superimposed components (with Gaussian properties), i.e., additive post-synaptic potentials of supragranular, granular, and infragranular layers, which adequate experimental interventions and recordings can discern. Besides the component's maximum amplitude and latency, the authors also report the *full width at half maximum* (FWHM), a standard parameter in

Gaussians that depicts the width or duration at 50 % amplitude height, including an "onset" and "offset" timing.

To return to the functional role of the N1 regarding multisensory integration and selective attention, so far, based on the presented evidence, it reflects modality-specific and inhibitory processes. Hence, the various findings on the anterior N1 (140–180 ms) that show increased amplitudes for modality shifts (R. Cohen & Rist, 1992; Gondan et al., 2007; Töllner et al., 2009) might indicate a perceptual or preattentive coping mechanism towards 'unexpected events' (Spence, Nicholls, et al., 2001). In line with 'feature-inhibition' and the 'inhibition-timing hypothesis' (Klimesch et al., 2007; Treisman & Sato, 1990), the increased inhibitory N1 or higher alpha synchronization reflects additional efforts to improve the signal-to-noise ratio for target discrimination for the unexpected modality (e.g., touch), due to a bias of the preceding modality (e.g., vision), and reduced weighting precision in touch. Overall, it appears reasonable to assume that these early components, such as P1 and N1, reflect a modality-specific stage of perceptual integration that can be influenced (or superimposed) by the "processing negativity" (within 50–150 ms) linked endogenous attention and top-down control (e.g., prior history, cued locations, or anticipatory activity) (cf. Näätänen, 1982). However, as will be shown in the following paragraphs in a detailed study description of Talsma and Woldorff (2005), these components (sensory N1 and processing negativity) do not provide clear evidence for crossmodal attention, as they precede the stage multisensory integration and attentional capture. Therefore, they disqualify as suitable candidates for testing neural mechanisms across sensory modalities.

Based on empirical evidence, the dissociation between 'early' uni-modal and 'late' cross-modal perceptual processes appears adequate (e.g., De Meo et al., 2015; Teder-Sälejärvi et al., 2002), similar to the debate on 'early' and 'late' attentional selection (Pashler, 1997). In their study, Talsma and Woldorff (2005) conducted a detection task where participants had to respond to left or right visual white square wave grating), auditory (1600 Hz tone with 65 dB), or crossmodal targets, which were either attended or unattended, while concurrently measuring the EEG activity. Overall, the authors pursued to establish evidence for multisensory integration during the preattentive stage by comparing the temporal evolution and dissociation of EEG activity between the redundant auditory-visual (AV) signal and their non-redundant sum (A + V), similar to previous studies (Giard & Peronnet, 1999; B. E. Stein & Meredith, 1993; Wallace et al., 1993). To achieve this, Talsma and Woldorff (2005) excluded the 'anticipatory' activity, i.e., the endogenous processing negativity or contingent negative variation (CNV; Walter et al., 1964), by applying a 2 Hz high-pass filter, correcting for slow oscillatory drifts (Teder-Sälejärvi et al., 2002).

The authors retrieved three phases for multisensory integration from the difference in (AV) and (A + V): a frontal positivity within 100–140 ms that indicates enhancements from (redundant) multisensory integration, invariant for attended or unattended stimuli, and two centro-medial positivities within 180–220 ms and 320–360 ms. The factor attention, i.e., the difference between attended and unattended signals, was significantly enhanced within 140–300 ms, revealing a substantial negative deflection that peaks around 250 ms. Of note, this time window for attentional processes is congruent with previous suggestions of 100–300 ms (Dalvit & Eimer, 2011; Jenkins et al., 2016) and 200–300 ms (Palmer et al., 2019; Wolfe et al., 2004). Finally, multisensory integration and attention interacted within 160–200 ms, that is, the timing of the N1 component and preceding the 250 ms attention peak, and at a later stage between 320–420 ms.

Overall, the findings of Talsma and Woldorff (2005) show that multisensory integration occurs at a relatively late stage (around 180 ms), effectively distinguishing uni-modal perceptual and multisensory stages. Notably, the timing of multisensory integration aligns with the late visual N1 (160–200 ms), indicating temporal sensory segregation between late visual and earlier auditory (N1: 70–120 ms) and somatosensory (N1: 100–150 ms) processes of sensory integration. Although speculative, the interplay between the (external or excitatory) positive P1 and the (internal or inhibitory) negative N1 might reflect the (low-level) perceptual and preattentive binding mechanism to generate accurate representations (or mappings) of our sensory environment (Bruyns-Haylett et al., 2017; Hanslmayr et al., 2011; Klimesch et al., 2007). Support for this proposition comes from the presented studies as multisensory stimuli produce a frontal positivity within the P1 timing (100–140 ms), indicative of combined external signals (Talsma & Woldorff, 2005), and modality changes between successive trials are linked to an anterior N1 (140–180 ms) reflecting internal weight shifts (Gondan et al., 2007; Töllner et al., 2009). Again, it would be interesting to examine in greater detail whether the inhibitory N1, i.e., the enhancement in amplitude, reflects the internal weighting of relevant and irrelevant feature activations between and within sensory modalities.

However, since the multisensory integration only starts with the timing of the (late visual) N1 component, again, this disqualifies the N1 as a sensitive marker for crossmodal selective attention. Nonetheless, it is important to emphasize that the N1 is associated with early processes of attentional orienting, as it modulates contralaterally to the attended stimuli with increased amplitudes.

This adaptive sensitivity of the N1 points to the convergence of space-based and object-based computations into a coherent spatio-topographic mapping based on saliency and relevance, that is, the *attentional priority map* (Fecteau & Munoz, 2006; Koch & Ullman, 1985; Logan, 1996; Wolfe, 1994). Therefore, the N1 precedes the capture of spatial selective attention, likely reflecting a gateway between the (parallel) preattentive and (serial) attentive stage.

Before the finish line of selecting suitable neural markers begins, it should be noted that the sequence of ERP components within the selected time range of 100–300 ms can be regarded as a process of evidence accumulation, spanning across perceptual, preattentive, and attentive stages (or levels), including various ERP components, such as P1, N1, P2, N2, and even P3 (Driver & Spence, 1998; Eimer & Driver, 2001; Gaillard, 1976; Gramann et al., 2007; Levit et al., 1973; Luck & Hillyard, 1994a; Näätänen, 1982, 1988; Polich, 1986; Sams et al., 1983; Squires et al., 1976; Sutton et al., 1965). So far, a suitable neural marker should follow the preattentive stage, i.e., subsequent to the N1 component, and within the attention stage, e.g., around 250 ms. Further, a good candidate must entail spatial encoding due to the spatio-topographic aggregation of the priority map. Also, it should modulate sensitively to the five factors of selective attention (e.g., bottom-up saliency of features and locations, prior history, top-down control, and sustained attention). Finally, the neural dynamics of those markers should be similar and consistent for visual and tactile targets. By finding such neural markers, comparing the MWA and DWA should be possible by examining modality-specific and supramodal mechanisms based on a set of precise empirical hypotheses.

Posterior Contralateral Negativity

The above-stated catalog of requirements enables an explicit search for an adequate EEG component that indicates the deployment of selective attention based on the preceding preattentive computations. The N2pc is exactly such a component. It arises as posterior contralateral activity from pop-out targets in contrast to (pop-out) non-targets or homogenous displays within 200-300 ms after stimulus onset (cf. Luck & Hillyard, 1994a). Further, the N2pc, i.e., the target-related contralateral negative deflection, was present for feature contrasts in various visual dimensions, such as orientation, color, and size. Of note, Luck and Hillyard (1994a) found performance benefits, i.e., accelerated reaction times, for prior knowledge of the target's feature-contrast (e.g., a color difference), arising in the EEG components from 175 ms onwards (similar to Talsma & Woldorff, 2005), thus indicating an interaction of bottom-up and top-down processes during the stage of attentional selection (consistent with Treisman & Gelade, 1980; Wolfe et al., 1989). Interestingly, the presence of the N2pc is based on feature discrimination of target information against irrelevant non-targets or distractors, that is, it occurs within a ‘complex’ visual display that includes multiple (or at least one other) stimuli. In contrast, the N2pc is absent for target-only presentations. In another study by Luck and Hillyard (1994b), the authors describe the N2pc as a “spatial filtering process” that selects relevant contralateral target information while suppressing (or inhibiting) irrelevant non-targets or distractors within the visual scene.

In another study about the N2pc, Eimer (1996) obtained similar results for color and form targets among three or even one distractor (or non-target) item(s), producing a substantial and target-sensitive *posterior contralateral negativity*. This study included multiple visual discrimination tasks, where visual displays were presented for 150 ms. Participants had to discriminate color targets (red or blue), form targets (W or M), or word targets (‘RECHTS’ or ‘LINKS’) among respective nontargets. The novelty of this N2pc study was the statistical evaluation of N2pc amplitudes, which derived from the (mean average) difference of the ipsilateral from the contralateral amplitude within the N2 time range (220-300 ms). Similar to Luck and Hillyard (1994b), Eimer (1996) concluded that the N2pc reflects an “attentional filtering process” for the selection of relevant target information.

Fascinatingly, the “attentional selection” in the N2pc can be associated spatio-topographically with the preceding bilateral P1 component (80-120 ms), which time-wise refers to the initial firing rates of (retinotopic or position) receptive fields (cf. Chelazzi et al., 1998), i.e., spatio-topographic integration of the external visual scene. While the irrelevant receptive fields stop firing during the timing of the inhibitory visual N1, relevant receptive fields maintain engagement with the target, i.e., they continue firing based on saliency or relevance, bottom-up discrimination or top-down templates, respectively. Hence, the N2pc is the consequence or outcome of spatial prioritization and feature binding across visual dimensions, (likely) producing a negative deflection due to the continuous firing of contralateral receptive fields (Chelazzi et al., 1993; Luck et al., 1997). The distribution of (position) receptive fields is closely linked to the retina, including many smaller ones in the fovea and larger but fewer ones towards parafoveal and peripheral locations (e.g., Y. Chen et al., 2009; Sereno et al., 1994). This distribution pattern of positional receptive fields aligns with the eccentricity effect, i.e., a decay in visual detection performance from foveal toward peripheral regions (Carrasco et al., 1995; Staugaard et al., 2016; Wolfe et al., 1998). Also, the N2pc amplitude decreases with extending eccentricity (Schaffer et al., 2011) or remains high within parafoveal regions and then plunges peripherally (Papaioannou & Luck, 2020), providing further evidence that the N2pc reflects a spatial marker for attentional selection.

Another study stems from Wolber and Wascher (2005). Although the authors replicated the previous N2pc findings of selective “visuo-spatial processing” in a series of experiments, they also raised concerns regarding the naming convention. By analyzing the “N2pc” peak latency, Wolber and Wascher (2005) found a range from 197 ms to 325 ms for the asymmetric deflection of the N2pc. Hence, the term “N2pc” in reference to the timing of the N2 component might be a bit misleading, as the “visuo-spatial” component arises within the N1 to P3 timing; thus, names like N1pc, N2pc, or N3pc would be more consistent (yet more confusing) for the same underlying process (Verleger et al., 2012). Therefore, the authors endorsed the more general expression: *posterior contralateral negativity* (PCN). The term PCN is also the expression of choice for this thesis, yet PCN and N2pc refer to the same functional process of attended spatial target selection. Notably, Wolber and Wascher (2005) obtained the PCN component by taking the average difference between contra- minus ipsilateral electrode channels for left and right targets, respectively (e.g., the average of right: PO7–PO8 & left: PO8–PO7 signals). Further, the PCN latency was chosen subject-wise as the timing of the negative peak within a selected grand average time range, e.g., 150–300 ms or 200–350 ms.

The main findings of Wolber and Wascher (2005) revealed systematic differences in PCN latencies across their three experiments. For example, the PCN latency was delayed for less salient form targets and increasing set sizes of nontargets but faster for congruent color embeddings. Overall, the authors found highly significant positive correlations between the PCN latency and reaction time performances; the earlier the PCN peak latency, the faster the reaction time. Of note, instead of choosing the PCN electrodes in advance (e.g., the ‘canonical’ PO7 and PO8 channels), Wolber and Wascher (2005) selected the (symmetric) electrode pairs based on the grand average ‘maximum’ amplitude in N1 and PCN, deriving three different selections for each experiment: PO5/PO6 for *form* targets, P7/P8 for spatially cued *squares*, and PO7/PO8 for *color-form* targets (e.g., yellow circle, or red square).

Overall, the PCN established itself as a common neural marker for spatial target selection and attentional deployment in many studies, either as a direct measurement of systematic variations in PCN amplitude or latency (Kiss et al., 2008; Krummenacher et al., 2014; Lleras et al., 2008; Ruge & Naumann, 2006; Töllner et al., 2008, 2010; Woodman & Luck, 1999, 2003), or as an indirect measurement of asymmetric posterior deflections and/or the difference between attended and unattended target conditions within 150-350 ms (Driver & Spence, 1998; Gramann et al., 2007; Hillyard & Münte, 1984; Mangun, 1994; Mangun & Hillyard, 1988).

After demonstrating the functional relevance of the PCN by showing systematic variations in amplitude and latency, it is essential to continue the detailed description to obtain a complete holistic concept of the PCN component. Similar to the concept of event-related EEG components, the PCN deflection entails a (peak) *amplitude* and *latency*, reflecting magnitude or signal strength (y-axis) and timing (x-axis) parameters, respectively. Further, the PCN (as well as other ERP components) rely on a sufficient signal-to-noise ratio (SNR), i.e., maximizing the signal and minimizing the noise, which is ensured by a sufficiently large number of included trials and adequate preprocessing routines (e.g., artifact rejection procedures) (cf. Luck, 2014). To obtain a complete (PCN) component impression, further parameters can be derived, such as its temporal start and end, *onset* and *offset*, respectively, and the *mean amplitude*, the amplitude average within a specific time range (e.g., 225-275 ms). Whereas the selection of amplitude scores is precise, the choice of a 'standardized' onset or offset criterion is more relative. This issue can be scientifically resolved by utilizing the 'full width at half maximum' (FWHM) parameter of Gaussian distributions, the 50 % amplitude height, marking the *duration* (or width) from the onset to offset timing. Based on this description, the PCN (as well as other ERP components) can be regarded as a Gaussian curve (or the additive signal of Gaussian subcomponents from cortical layers) (Bruyns-Haylett et al., 2017; Luck, 2014). Essentially, the timing and magnitude parameters represent the speed and amount of the allocated attentional resources (s. Figure 1–3).

Another critical issue of the PCN addresses "noise susceptibility". Despite establishing a clear PCN description above, extracting those various parameters satisfactorily for each participant exposes a practical challenge. As shown, the PCN "occupies" a broad time range, i.e., 150–350 ms, in which attentional selection occurs. Given its temporal fluctuation across single trials, the PCN is less 'clocked' or 'time-locked' as, for example, the perceptual and event-driven visual N1. This considerable variation flattens or broadens the signal strength of the PCN already in the average signals within participants and accumulates in the grand average between participants. In combination with hemispheric asymmetry and the common EEG noise, the various noise factors are particularly influential in 'small' components, such as the PCN, posing a critical challenge to the extraction of correct timings.

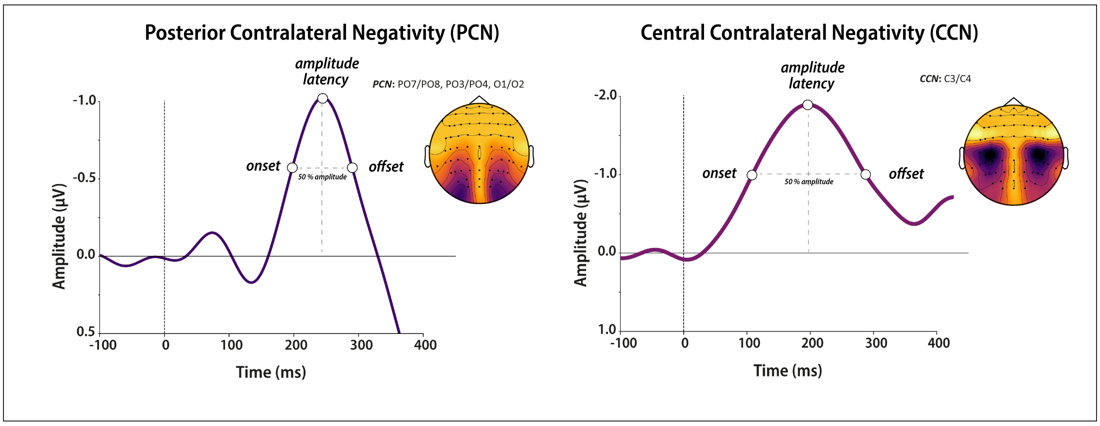

Figure 1–3. The Posterior Contralateral Negativity (PCN) and Central Contralateral Negativity (CCN) is shown on the left and right side with their temporal deflections and topomaps around the peak latency. *Left*. The *PCN* represents a neural marker for visual selective attention and spatial filtering with a negative deflection (150–350 ms) over parieto-occipital regions (e.g., PO7/PO8, PO3/PO4, O1/O2 electrodes). The difference wave of (attended) contralateral and (unattended) ipsilateral hemifields consists of a peak *amplitude* and three temporal parameters of peak *latency*, *onset*, and *offset*. Importantly, *onset* and *offset* derive from the 50% criterion of the peak *amplitude*, similar to the full-width-at-half-maximum (FWHM) criterion in Gaussians. *Right*. The CCN reflects a somatotopic marker for tactile spatial attention (140–340 ms) over central regions. Also, it can be described by peak *amplitude* and *latency*, and an *onset* or *offset*.

A noise or bias reduction technique is *jackknifing* (Miller et al., 1998; Miller, Ulrich, et al., 2009; Smulders, 2010; Ulrich & Miller, 2001). Instead of extracting each participant's latency (or amplitude) scores individually, the jackknife method computes a grand average for each participant by iteratively omitting the same participant. For example, in a sample with *n* participants, jackknifing will yield *n* averages each from *n*–1 participants. Those jackknife (grand) averages reduce the common and individual EEG noise tremendously, and adequate parameter extraction becomes feasible, i.e., the choice of onset, latency, or offset. However, since the obtained data reflect a ‘compressed’ solution, statistics need to either apply a correction of the test statistic (Miller et al., 1998; Ulrich & Miller, 2001) or convert the jackknife scores into individual estimates by the reversed jackknife transformation (Smulders, 2010). Regardless of the underlying approach – ‘normal’ or jackknife procedure – the outcome is (mathematically) identical. Yet, jackknifing is “more correct” as it disregards the noise and thus optimizes signal detection, given the requirement that a component like the PCN is (actually) present and observable.

Concerning the five factors of selective attention, the link with bottom-up *locations* and *features* seems evident since the PCN reflects a visuo-spatial marker for the attentional selection of target-relevant information. Importantly, the PCN is elicited for multiple (or any) visual dimensions: color, orientation, size, and form (or shape) (Eimer, 1996; Luck & Hillyard, 1994a; Wolber & Wascher, 2005). In agreement with these findings in visual search (and the assumptions of a DWA), the PCN modulates sensitively for visual redundant targets. For example, Töllner and colleagues (2011) found in a detection task with a 1:2 ratio in absent/present trials that visual redundant targets (orientation + color feature contrasts) elicited an earlier PCN latency and a higher PCN amplitude in comparison to non-redundant targets (orientation or color), except for same PCN amplitudes between the redundant and orientation target. Overall, the findings of Töllner et al. (2011) provided empirical evidence of an enhanced PCN due to cross-dimensional visual redundant targets. The authors conclude that the PCN latency and amplitude reflect "an index of the preattentive processing speed and … an index of the amount of attentional resource allocation" (Töllner, Zehetleitner, Krummenacher, et al., 2011, p. 139). Those redundancy gains in the PCN, i.e., the selective outcome of preattentive computations, were aligned with race-model violations in physical performances, promoting a preattentive 'locus of coactivation' instead of a post-selective origin, adding an electrophysiological view to the ongoing debate (Feintuch & Cohen, 2002; Krummenacher et al., 2002b; Zehetleitner et al., 2008).

Another study by Grubert and colleagues (2011) found similar redundancy gains in the PCN (or N2pc) component for color-shape targets – a red-diamond among green squares or a green-diamond among red squares – that exceeded single-dimensional red/green color and diamond shape targets in their baseline detection task, where participants had to respond to whether a color, a shape, or color-shape target was present. Of note, whereas Töllner et al. (2011) report the participant-wise PCN latency, the maximum negative deflection (within 150–350 ms), and the average of its ten surrounding time points as PCN amplitude, Grubert et al. (2011) report the participant-wise mean amplitude (average across 180-280 ms), and the *jackknifing* PCN onsets, i.e., the timing that exceeded an absolute threshold of –1μV. Yet both studies selected the PO7/PO8 electrode locations to examine PCN dynamics.

The findings by Grubert et al. (2011) revealed that PCN onsets were significantly earlier for the redundant target (171 ms) than for color (184 ms) and shape (196 ms) targets, which were comparable to one another. In addition to this baseline condition, the authors included a color-target and a shape-target condition to examine the influence of target prioritization on the PCN. For example, in the shape condition, participants had to respond to the presence of a shape feature contrast – a pure shape or redundant target – while ignoring pure color targets and absent trials. The same logic applied to the color-target condition. Although Grubert et al. (2011) found "reliable N2pc" for pop-out singletons, i.e., a significant negative deflection, interestingly, the (statistical) temporal benefits of redundant targets in PCN onsets vanished in the color-target (171 ms vs. 174 ms) and shape-target (176 ms vs. 190 ms) condition. Despite finding still a "reliable N2pc" in task-irrelevant color and shape pop-out singletons, critically, their PCN onsets and amplitudes were significantly delayed and attenuated in comparison to their task-relevant counterparts, respectively.

In summary, Grubert et al. (2011) demonstrated that the PCN component only partially (but not entirely) reflects bottom-up saliency processing of visual pop-out singletons by showing evidence for the impact of top-down control or task-driven prioritization based on task sets – the (pre-)activation of task-relevant target templates (see also Eimer & Kiss, 2010). Generally, both visual search studies and their observed PCN dynamics can be explained by the DWA. Whereas redundant signals combine their weights of dimension-specific activations for the guidance of attentional selection, which manifests in the faster and more enhanced PCN, single-target templates (actively) up-weight task-relevant dimensions (e.g., color-target) while concurrently down-weighting task-irrelevant dimensions (e.g., shape singletons), thus eradicating (saliency-driven) redundancy gains. Furthermore, the PCN amplitude refers to a limited capacity for the number of intradimensional targets (e.g., single vs. multiple colors). Whereas the PCN amplitude is larger and faster for a single color target, it becomes smaller and delayed for multiple color targets, indicative of a higher memory load in a letter-digit identification task (Grubert et al., 2016). This finding (as well) can be interpreted with the DWA, which assumes a limited capacity for each visual dimension (e.g., color). While a single feature (e.g., red) can utilize (almost) the full capacity (similar to focused attention), multiple features (e.g., blue and yellow) must share the same capacity (similar to divided attention). Overall, the PCN component reflects a visuo-spatial filtering process based on the preattentive summation of salient and relevant features across (and within) visual dimensions.

Also, the PCN modulates sensitively to dimensional changes or repetitions between successive trails. For example, Töllner et al. (2008) demonstrated earlier and enhanced PCN for dimension repetitions (color → color & shape → shape: -2.25 μV, 243 ms) in contrast to dimension changes (shape → color & color → shape: -1.95 μV, 251 ms), during an identification task, where participants had to select the target (e.g., a red circle or a blue square among blue circles) and respond to the orientation of its grating (horizontal or vertical).

The study of Eimer et al. (2010) also found a delay in PCN onsets for target (and/or 'distractor') 'swaps' in comparison to repetitions (242 ms vs. 188 ms). While participants had to respond to the diamond notch (top or bottom), task conditions were defined by color repetition or swaps between targets and distractors (actually nontargets) in two experiments. For example, when the previous trial consisted of a red-diamond target and three green-diamond nontargets, a ‘full repetition’ would reveal the same colors for the target and nontargets but in a different spatial configuration (e.g., red-diamond top-left → red-diamond bottom-right). In a ‘full change’, colors in target and nontarget were ‘swapped’, revealing a green-diamond target and three red-diamond nontargets. The second experiment extended this ‘swap’ paradigm by introducing a third color (blue), which also included partial repetitions (e.g., red-diamond target [same] but three blue-diamond nontargets [new]) and partial swaps (e.g., blue-diamond target [new] and three red-diamond nontargets [swap]). Of note, Eimer et al. (2010) computed the mean amplitude for an early (200–270 ms) and late (270–340 ms) time window. They calculated the jackknife PCN onset latencies with a relative threshold (instead of an absolute) by selecting the time point of 40% (jackknife) peak amplitude (see also Kiesel et al., 2008).

The main findings of Eimer et al. (2010) revealed a temporal dissociation in PCN; full and partial repetitions revealed accelerated onsets and higher amplitude during the ‘early’ time window than partial and full swaps with delayed onsets and higher amplitudes during the ‘late’ time window. Generally, the early vs. late amplitude effects can be interpreted as a temporal delay between repetitions and changes, that is, shifted PCN components (genuinely) produce amplitude differences between early (or onset) and late (or offset) time windows. Interestingly, the PCNs in Eimer et al. (2010) revealed a characteristic PCN *offset* pattern in their Figure 5, but without statistical evaluation. Here, the PCN offset appears fastest for full repetitions and heavily delayed for full swaps, whereas partial repetitions and swaps had comparable PCN offsets in-between.

While the analysis of PCN onsets and peak latencies is common, a scientific reflection on the functional role of PCN offsets is missing. Whereas the earlier timing scores seemingly reflect the origin and peak of (covert) visual target selection (and attentional deployment), the later PCN *offset* might indicate the completion of “target individuation” against nontargets within visual displays or scenes (see conference abstract of J. Foster et al., 2018). Likewise, the PCN *offset* could represent the transition toward a (central) object representation (Koch & Ullman, 1985) and post-selective processes, such as working memory, decisions, or response preparation (e.g., McClelland, 1979; Oberauer, 2019). Hence, PCN *onset* and *offset* (and *duration* width) might reveal a functional dissociation in the processing timeline, which (taken together) could better explain the increasing switch costs in reaction times. For example, performances in Eimer et al. (2010)

revealed the following pattern: full repetition < partial repetition < partial swap < full swap.

Generally, it is important to underscore that the PCN findings in Töllner et al. (2008) and Eimer et al. (2010), that is, the PCN delays (and lower amplitudes) for visual (dimension) changes are consistent with the amount of switch costs in behavioral response times. Whereas the findings by Töllner et al. (2008) align with the DWA, assuming prominent switch costs for cross-dimensional changes due to *weight shifting* (e.g., color → shape), the interpretation of Eimer et al.'s (2010) findings needs some consideration. Crucially, the task of Eimer et al. (2010) required the detection of a color pop-out singleton (e.g., red) in contrast to (three) same-color nontargets (e.g., green). Although intertrial changes only included 'feature-shifts' within the color dimension (e.g., red → green), which in terms of the DWA should be comparable, nonetheless, reaction times and PCN dynamics obtained substantial costs. This outcome can be explained by both a singleton-search mode ("odd-one-out" or dynamic templates) instead of a feature-search mode (static templates) and the permanent adjustments of intra-dimensional weights: the weight of 'feature-inhibition' and 'feature-selection' are (constantly) swapped between the target and nontargets (e.g., red ↔ green). Hence, the weight shift sum in a full swap is the same as a regular 'dimension-shift', thus in line with the DWA. Further, the dynamic color-swapping between target and nontargets introduces a higher (attentional) load, similar to the set size effect in conjunction search.

Of course, plenty of other studies provide further interesting evidence on the functional role of the PCN, yet going too much into the details would go beyond the scope. For example, reward expectation (Eckstein, 2011) seems to be linked with PCN dynamics – becoming earlier and larger for high-reward targets (Kiss et al., 2009; Wei & Ji, 2021). Nonetheless, some further (PCN) insights are given with a brief description.

For example, Zivony and colleagues (2018) underscored the genuine (uni-directional) link between PCN and attentional shifting by emphasizing that an observed PCN (indeed) indicates an attentional shift toward a (relevant) target (including the timing and amount of deployed attentional resources). In contrast, attention shifts can occur without a PCN (e.g., the observation of an irrelevant distractor) due to a lack of attentional engagement. Of note, Zivony et al. (2018) measured the PCN onset as the 15 % fractional peak latency, that is, the time point with 15 % peak amplitude (Kiesel et al., 2008; Luck, 2014). The requirement of attentional engagement for relevant targets is a reminder of the (default) PCN constraints – it measures lateral but not midline targets and is susceptible to noise due to temporal variation in the deployment of attention.

Another study worth mentioning is by Matusz and Eimer (2013). Interestingly, their experiments examined a (consistent) PCN onset and amplitude difference between visual and audiovisual task sets (for target-color singleton cues), being delayed and decreased for the latter. While in the visual task, only one attentional template was activated (e.g., a red bar), two bimodal templates were active during the audiovisual tasks (*visual*: red bar or small bar; *auditory*: high pitch or sound). Again, similar to focused and divided attention, Matusz and Eimer (2013) explained their findings by weighting differences in the visual (color or size) template based on unimodal or bimodal prioritization. While the visual task simply applies visual (template) matching, the audiovisual tasks require multimodal (template) matching alongside more distributed attentional resources. This indicates a difference in the perceptual load based on the number of concurrently active templates. Hence, the weights of the same visual singleton cues differ between unimodal and bimodal tasks, manifesting in the delayed and decreased PCN, suggesting a crossmodal involvement in the same modality-specific component. Overall, these findings align with the assumptions of the (multisensory) DWA and the MWA, which assume a reduced saliency for color singletons in the priority map due to other concurrently active and relevant templates in visual dimensions or sensory modalities. Crucially, the study of Matusz and Eimer (2013) provides evidence for a crossmodal influence on PCN expression. Yet the study cannot answer the critical question (of this thesis) whether the preattentive hierarchy needs to be extended for modality maps (above dimension maps), as it was not included in the task design. Nonetheless, the study of Matusz and Eimer (2013) points to a critical role of the superior temporal sulcus (STS) and the lateral intraparietal area (LIP) (McDonald et al., 2003; B. E. Stein & Stanford, 2008), which contain "multisensory spatial maps" and are relevant for "the control of space-based and feature-based visual attention" (Matusz & Eimer, 2013, p. 1007). The PCN depicts the activity of those areas.

Overall, the PCN is embedded within other lateralized components in the same spatial electrode cluster, such as the P*d* (distractor positivity) that modulates within 290–340 ms due to the attentional capture of distractor information (Feldmann-Wüstefeld et al., 2021; Hickey et al., 2009; Hilimire et al., 2012; Sawaki & Luck, 2010; Wyble et al., 2015), or subsequent memory-related processes in the SPCN (sustained posterior contralateral negativity) and CDA (contralateral delay activity) (Adam et al., 2018; K. Clark et al., 2015; Grubert et al., 2017; Ikkai et al., 2010; Jolicoeur et al., 2008; Luria et al., 2016; Mazza et al., 2007). Additional components are the stimulus- and response-related LRPs (lateralized readiness potentials), which indicate response preparation or execution, respectively, and are commonly measured over central (C3/C4) regions (Miller et al., 1998; Töllner, Zehetleitner, Krummenacher, et al., 2011; Ulrich & Miller, 2001).

Regarding the five factors of selective attention, it can be stated with the presented empirical evidence that four factors systematically affect the PCN component: bottom-up (i) *object features* and (ii) *spatial locations*, as well as (iii) *intertrial history,* and (iv) top-down *attentional templates*. Yet it remains an open issue whether the fifth factor (v) *sustained attention* or vigilance affects the PCN (based on a scarce scientific background), hypothetically the PCN reduces with increasing task fatigue and could sensitively depict fluctuations between 'in-the-zone' and 'out-of-the-zone' attentional states (Esterman et al., 2013; Kucyi et al., 2017; Rosenberg et al., 2015, 2016).

Overall, the PCN, as a visuo-spatial processing filter, is a suitable candidate (or marker) to summarize the underlying preattentive computations, i.e., the timing and amount of deployed attentional resources, similar to the *priority map*. Furthermore, the PCN succeeds the time window of multisensory integration (e.g., Talsma & Woldorff, 2005), adapting sensitively for crossmodal search (Matusz & Eimer, 2013), but (by definition) precedes post-selective stages (e.g., central object representation, decision, and response reparation). Therefore, the PCN is the perfect candidate to examine the functional difference between visual dimensions and (potentially) sensory modalities and to retrieve *modality-specific* and *supramodal* mechanisms that either favor an extended MWA or a multisensory DWA.

Central Contralateral Negativity

To fulfill the neural requirements for the visuo-tactile paradigm, a marker for tactile attention is (still) missing, similar to the PCN for visual attention. Fortunately, an adequate somatosensory component exists, the *central contralateral negativity* (CCN), also known as N140cc (Forster et al., 2016). Although the empirical evolution of the CCN component shows many parallels with the origins of the PCN (or N2pc), its corpus of evidence is (unfortunately) much smaller, likely due to feasibility constraints in tactile devices. Nonetheless, a brief chronological evolution of the CCN is presented in the following paragraphs.

In an 'early' study, García-Larrea et al. (1995) gave an introductory overview of the various findings of a somatosensory N1 component (SEP) within 100–150 ms. Here, the N1 is delayed due to higher attentional load, is higher for tactile target information, decays due to habituation, and vanishes for neutral stimuli. Further, García-Larrea et al. (1995) examined the “exogenous and endogenous components” of the somatosensory N1 by testing electrical stimulations (300 µs, 3–6 mA, series of 150–200 pulses) on two fingers of each hand (second and third finger) in neutral, García-Larrea et al., unattended, and attended conditions. The cognitive load varied across these task conditions. Participants had to relax in the neutral and ignore the tactile pulse series; they had to silently count the (less frequent) number of pulse series in either the unattended left or attended right hand. The authors found a dissociation between the early N120 and late N140 components. While the N120 reflects (modality-specific) tactile awareness, its amplitude increase (for tactile pulses) occurred regardless of the spatially attended left or right hand. In contrast, the N140 revealed a substantially higher amplitude over contralateral electrode locations for the attended right hand than the unattended left hand and vanished in the neutral condition. Overall, this finding of García-Larrea et al. (1995) indicates that tactile target detection is improved by (endogenously) deployed attentional resources. The highest amplitude effects were found for the attended right hand in C3, Cz, and C4 electrodes, significantly higher than in the unattended hand condition and earlier N140 latencies in C3 (6.5 ms). The ‘tactile attention effect’ for left (C3) and right (C4) somatosensory hemifields was interpreted by García-Larrea et al. (1995) with “callosal transfer times”, which potentially ensures the correct spatial (egocentric) decoding of right and left tactile stimulations.

The visuo-tactile search study by Eimer and Driver (2000) ‘replicated’ the N140 findings for attended (and relevant) tactile target information, originating from 140 ms onwards over central regions (C3, Cz, C4) in contrast to unattended (or ‘to-be-ignored’) tactile stimuli. In a similar experimental setting as in García-Larrea et al. (1995), participants had to vocally ‘judge’ whether a visual or tactile target occurred at the attended left or right side (or hand). All stimuli were presented for 200 ms. While (frequent) nontargets showed either a visual isoluminant green light across left or right LEDs or a continuous tactile vibration on the left or right index finger, (infrequent) targets included a 10 ms ‘gap’ after 95 ms in visual illumination and tactile vibration.

As expected, Eimer and Driver (2000) found substantial and characteristic negative difference waves (N*d* components) within 210–280 ms between attended and unattended visual and tactile targets over occipital and central regions, respectively. Although N*d* findings are derived from midline electrodes (Fz, Cz, Pz), they occur in the same time range as the PCN and reflect modality-specific spatial attention processes. Eimer and Driver (2000) found a clear attentional separation between vision and touch by comparing the Nd across the three task conditions, reflecting endogenous and independent spatial orienting to relevant (while ignoring irrelevant) targets. Whenever participants had to respond to visual targets in attended side presentations ('judge vision' condition), they found a visual N*d* but no tactile N*d*, and vice versa for the 'judge touch' condition. Yet, when both (visual and tactile) targets were relevant ('primary vision'), an early central effect (N*d*1) was observed for tactile targets within 140–200 ms, which Eimer and Driver (2000) interpreted as the activation of crossmodal links between both sensory modalities.

In another study by Eimer, Forster et al. (2004), the spatial orienting of the tactile N140 component is examined in greater detail. Again, a vast N140 amplitude gain for frequent tactile stimuli (strong vibrations) was found at the attended index fingers. Importantly, the N140 amplitude became more negative with increasing hand distance (near: 6 ms vs. far: 56 ms), which can be explained by the association of somatotopic and external space or "coordinate systems". Further, Eimer et al. (2004) also reported a subsequent "sustained enhanced negativity for tactile stimuli at attended locations" (p. 605) within 200–300 ms, i.e., the time window of selective attention, which interestingly was unaffected by the distance of hand position. In summary, these findings suggest an 'early' interaction between somatotopic (or anatomical) and external (or proprioceptive) during spatial registration (in N140), which seemingly vanishes during the later time window, and presumably indicates tactile attention processing.

So far, the findings between the visual N1 and the tactile N140 are congruent as both components reflect processes of spatial gating by showing enhanced negative amplitudes for attended sensory information. Similar to the visual N1, which precedes the PCN, the tactile N140 disqualifies as a tactile (target) marker as it modulates to (attended) tactile targets as well as tactile nontargets. Similarly, it seems conclusive to detect a sensitive marker for tactile attention in the subsequent lateralized divergence – the central contralateral negativity.

The somatosensory CCN, or N140cc, is a relatively novel EEG component. It was introduced and summarized by Forster et al. (2016) by studying the somatosensory 'body space'. In their experiments, six tactile stimulators (or solenoids) were either placed at the hand's fingertips ('hand task') or at the left and right shoulders, index fingers, and big toes ('body task'). In their detection tasks, two discriminative tactile stimuli patterns – 'tap' and 'buzz' – were presented for 603 ms across solenoids in each trial. Participants had to vocally respond to the presence of an assigned tactile target (counterbalanced 'tap' or 'buzz') at one solenoid, while the other five solenoids presented tactile nontargets ('tap' or 'buzz', respectively). The response behavior was comparable between tactile targets and task conditions. In the EEG, Forster et al. (2016) found a significant target-related central contralateral negativity (C3/C4 electrodes) within 140–340 ms with a maximum of around 220 ms, the N140cc at all body locations. Interestingly, the N140cc amplitude was more negative in the 'hand task' than in the 'body task'. Although tactile stimulation was quite long (603 ms), an N140cc indicates 'fast' tactile discrimination between both target–nontarget patterns (vanishing after 340 ms). Further, the N140cc was most prominent in fingers and became (significantly) reduced (but still relevant) in shoulders and then toes. Essentially, while the N140cc reflects the "allocation of attention to target locations" (p. 164), Forster et al. (2016) did not reveal evidence for the 'speed of tactile attention' as response rates were comparable across fingers and toes, despite substantial N140cc amplitude differences.

Arguably, the most interesting conclusion of Forster et al. (2016) of the N140cc is the link to the "somatosensory homunculus", where tactile sensitivity is maximized in the fingertips (cf. Sato et al., 1999). By linking the N140cc to the N2pc, which decays towards the periphery (Schaffer et al., 2011), the authors argue that the hands are the "fovea of touch" (p. 165) with the most negative N140cc amplitude. Hence, the N140cc can also be regarded as a "spatial filtering process", but instead of the external (visual) space, it decodes the egocentric (somatotopic) space topographically. Overall, the findings of Forster et al. (2016) evidently showed that the N140cc is related to tactile attention and tactile target discrimination ('tap' vs. 'buzz') within a more 'complex tactile display' (of six solenoids). Yet, the authors admit that the functional role of the N140cc, whether it reflects target selection or nontarget (distractor) suppression, is not fully clarified. Nonetheless, the study of Forster et al. (2016) reflects a landmark publication in the quest to understand tactile attention by providing a suitable EEG marker, such as the N140cc (s. Figure 1–3).

In agreement with this 'somatotopic decoding' scheme are insights from tactile learning in a contextual cueing task that assume an "anatomical reference frame" (Assumpção et al., 2018; S. Chen et al., 2020). Also, the N140cc modulates sensitively for concurrent target and distractor presentations; for example, it decreases for contralateral but not ipsilateral distractors, evidencing target selection (Gherri et al., 2021, 2022). Other recent studies underscore the target selection (instead of distractor suppression) role by reporting an N2cc (the same as N140cc) (Katus & Eimer, 2019). Furthermore, the (late) N140cc amplitude (230–310 ms) enhances with increasing hand distance (near vs. far), which suggests spatial interaction of proprioception and/or visual feedback (Ambron et al., 2018), as a contrary result as in Eimer et al. (2004). Again, to avoid confusion for the tactile attention component, the expression CCN (central contralateral negativity) will be used throughout the thesis (instead of N140cc or N2cc). Although the tactile CCN and visual PCN seem to have obviously much in common, there is a huge empirical gap between both components, as the CCN only recently gained momentum. While the PCN describes the 'amount' and 'speed' of (visual) attentional allocation, the CCN parameters seem to entail magnitude effects only (so far) via the evaluation of mean amplitudes, e.g., in early (140-240 ms) and late (240-340 ms) sections (Forster et al., 2016; Katus & Eimer, 2019). Therefore, the CCN reflects "tactile spatial filtering" and thus seems ideally suited to analyze tactile attention processes (just like the visual PCN). Before combining the empirical framework (DWA vs. MWA) and the *modality-specific* neural markers (PCN & CCN) into a set of concise research questions for the visuo-tactile search paradigm, it is imperative to consider the "attentional priority map" in greater detail by examining the crossmodal links between vision and touch.

Attentional Priority Map

By having introduced the empirical framework of preattentive computations and the neural markers of visuo-tactile attention, the *priority map* must now be considered. Consistent with the superior research question of "common or separate attention pools" (Miller, 1982) across sensory modalities, this section describes the *modality-specific* and *supramodal* mechanisms based on the spatio-temporal aspects of attentional selection.

The PCN and CCN components reflect the *modality-specific* saliency or spatial prioritization of relevant visual or tactile information. Hence, both components by themselves constitute visual and tactile *saliency maps* (or priority maps), summarizing the preattentive dynamics of weighted feature integration in a spatially consistent manner (Eimer, 1996; Forster et al., 2016; Luck & Hillyard, 1994a). So far, the evidence on PCN and CCN favors 'separate attention pools' for visual and tactile attention, respectively. Whether those 'separate pools' persist during visuo-tactile search (evidencing modality-specific priority maps) or combine through crossmodal links, reflecting 'common pool' dynamics (evidencing supramodal priority map) remains an open scientific issue. However, since multisensory integration precedes (and interacts) with attention processes (Talsma & Woldorff, 2005), it seems plausible that the lateralized deflections – the 'attentional divergence' between contra- and ipsilateral sides – in PCN and CCN would capture those crossmodal dynamics in their amplitude and/or latency expressions.

Generally, attention combines stimulus-driven and goal-driven (bottom-up and top-down) processes by activating a fronto-parietal network, the so-called attention network (Corbetta & Shulman, 2002; Petersen & Posner, 2012; Posner

& Dehaene, 1994; Posner & Petersen, 1990). Crucially, this attention network combines several relevant network hubs across frontal regions, like the frontal eye field (FEF), the ventral frontal cortex (VFC) or the dorsolateral prefrontal cortex (dlPFC), but also in parietal regions, such as the intraparietal sulcus (IPS) or superior parietal lobe (SPL), and the temporo-parietal junction (TPJ) (see Corbetta & Shulman, 2002; Petersen & Posner, 2012). Although the (fMRI) evidence on attention networks primarily derives from sustained attention studies, like the (gradual) continuous performance tasks (deBettencourt et al., 2015; Rosenberg et al., 2016), evidence from crossmodal paradigms promote the posterior parietal cortex as the area of interest for spatial orienting and multisensory activations (Chambers et al., 2007; Chambers, Payne, et al., 2004; Chambers, Stokes, et al., 2004; Macaluso et al., 2000, 2002a, 2002b).

Given the requirement of coherent spatial decoding, a supramodal *priority map*, as proposed by the MWA, has to combine the distinct spatio-topographic saliency computations across sensory modalities, e.g., within a specific brain region or via the establishment of crossmodal links between modality-specific areas during preattentive computation that precedes the capture of (crossmodal) attention.

For example, Chambers et al. (2004) reported two distinct time windows for spatial orienting during a visual search task – localizing a target with high-frequency grating – an early or 'fast' (90–120 ms) and a late or 'slow' (210–240 ms) timing, activating the (right) angular gyrus within the inferior parietal lobe (IPL). The authors argue that these activations for visual spatial attention derive from the "fast retinotectal pathway via the superior colliculus (SC), and from the slow geniculostriate pathway via the primary visual cortex." (p. 218). Notably, the "biphasic" timing of (visual) spatial attention by Chambers et al. (2004) is perfectly in line with the timing of P1 and PCN (or N2pc) components and their link to the firing of relevant receptive fields (e.g., Chelazzi et al., 1998; Luck et al., 1997; Mangun & Hillyard, 1988). The conclusion of Chambers et al. (2004) is very much in line with the 'attentional resolution' of the underlying saliency computations, e.g., the initial "where" of the "odd-one-out" ("that response"), followed by the feature identification ("what response") (H. J. Müller et al., 1995; H. J. Müller & Rabbitt, 1989; Sagi & Julesz, 1985). Whereas the early P1 reflects a "crude" spatial change detector that indicates that something is different in the (visual) scene or display (SC pathway), the late PCN timing enables the discrimination of "more complex visual information" (p. 218), that is feature binding from activations in the visual cortex, e.g., V1, V4, or MT (LGN pathway).

Interestingly, Chambers et al. (2004) describe the early 90–120 ms (P1) timing as "the disengagement of attention from its current location" (p. 218), likely due to the ‘new’ snapshot of visual information. Also, the P1 timing aligns with the initialization of saccade programming (Reichle & Reingold, 2013), overt and covert attention shifts in the frontal eye field (Buschman & Miller, 2009), and the ‘attentional divergence’ across receptive fields that amplify and reduce firings rates for relevant and irrelevant receptive fields, respectively (Chelazzi et al., 1993; Deubel & Schneider, 1996), promoting ‘attentional engagement’ in the later 210–240 ms timing (PCN) (Duncan & Humphreys, 1992; Zivony & Lamy, 2018).

Based on the evidence, it seems plausible that the computation of the visual priority map (or saliency summation) accumulates in two stages: the fast pathway registers the (external) spatial frame (e.g., visual display), while in the second stage, the slow pathway provides the content (e.g., Shulman et al., 2002). Essentially, both network pathways converge in the IPL, reflecting a full cycle of saliency summation (210–240 ms) that coalesce in the PCN's dynamics. So instead of promoting a specific brain region as the (visual) priority map, it is more reasonable to assume an interconnected “priority network” of relevant nodes or brain areas, such as SC, LGN, V1, V4, or MT (Fecteau & Munoz, 2006; Koch & Ullman, 1985; Koene & Zhaoping, 2007; Li, 2002; Meredith & Stein, 1986; Shim et al., 2013; Somers & Sheremata, 2013). Finally, the IPL or the lateral intraparietal area (LIP) measures the outcome of those “priority network” projections (Balan & Gottlieb, 2006; Bisley & Goldberg, 2010; Fecteau & Munoz, 2006) with deployed attentional engagement.

Similar to the visual priority map or the activity in the “attentional priority network” that is measured by the PCN over the posterior parietal cortex, the CCN, as shown in the previous section, reflects the somatotopic summation along the body schema – the “tactile priority map” (Assumpção et al., 2018; Eimer et al., 2004). However, despite the apparent distinction in spatial decoding, empirical evidence suggests that visual and tactile information, in combination with their external and egocentric spatial functions (Eimer, Forster, et al., 2003), merge in invariant areas of the posterior parietal cortex, indicating crossmodal links, e.g., in TPJ (Man et al., 2015; Quinn et al., 2014). For example, Macaluso et al. (2000) found that the “spatially congruent” presentation of combined visual and tactile information enhances the activity in the supramarginal gyrus of the IPL. In their follow-up studies with fMRI and PET, Macaluso et al. (2002a, 2002b) provided additional insights into the “supramodal network of spatial attention” which is embedded in the fronto-parietal attention network. Notably, the activation in IPS and TPJ represented crossmodal hubs of attentional (re-)orienting due to invalid cues, regardless of the target’s modality, being either a visual or tactile target. Furthermore, Macaluso et al. (2002a) refer to the modality-specific role of the somatosensory cortex for tactile information, being unaffected by visual information, while the parietal cortex and occipital visual cortex reveal supramodal activations for visual and tactile targets. Nonetheless, the authors conclude that the spatial specificity in occipital regions remains for vision, that is, external coordinates exceed (or dominate) the egocentric mapping (Eimer, Forster, et al., 2003; Spence, Shore, et al., 2001). Overall, crossmodal links in the (posterior) parietal cortex, e.g., in IPS and TPJ have been reported many times (Buschman & Kastner, 2015; Downar et al., 2000; Graziano & Cooke, 2006; Kanwisher & Wojciulik, 2000; Spence & Santangelo, 2009; J. F. Stein, 1989).

Overall, it seems that a “supramodal priority map” must be located within the attention network, especially within the parietal regions due to those crossmodal links. Again, it is unlikely to assume a specific brain area, but rather a “supramodal priority network”, whose neural activations and projections, similar to the “visual priority network” culminate over parietal regions. Further, the evidence clearly favors the ‘late’ stage, e.g., around 210–240 ms (Chambers, Payne, et al., 2004), thus answering the question of Treisman (1988) of whether the timing of the “master map” is early or late. Since this timing reflects the (spatio-topographic) attentional engagement for more complex (multi-)sensory objects, it also fulfills the requirement of stimulus-driven salience and goal-driven relevance (Corbetta & Shulman, 2002; Fecteau & Munoz, 2006) and occurs after (or parallel) to the parietal processes of multisensory integration (Talsma & Woldorff, 2005). Arguably, the efficiency of such a “supramodal attention network” depends on the coherence across relevant network hubs and their feedforward and feedback connections, establishing a ‘closed-loop system’ (deBettencourt et al., 2015; Desimone & Duncan, 1995).

Finally, the PCN moves into the center of interest as a suitable candidate for visual and crossmodal search by depicting the attentional engagement over parieto-occipital regions as a spatial marker for selective attention. Whereas the CCN is expected to be purely tactile and modality-specific, the neural dynamics of a supramodal PCN are expected to be amplified and exceed those of a (modality-specific) visual PCN, like a detector for preattentive coactivation.

Contemporary Research

Now that almost all the relevant empirical background has been outlined in-depth, this section aims to locate the topic of “crossmodal selective attention” within the broader landscape of “attention research” by looking at some other contemporary research topics. Indeed, the scientific quest to examine the preattentive hierarchy in (crossmodal) attention is fascinating (and arguably essential), yet it only reflects a rather small, specialized segment.

As mentioned, attention is a transitional stage between our perception and actions. Furthermore, it is "necessary but not sufficient" for the subjective experience of awareness and consciousness (Koch & Tsuchiya, 2012). Also, attention is central for other executive functions, such as working memory, but also short- and long-term memory, with a focus on (internal) object representations and other concrete or abstract information (Wolfe, 2021).

By having obtained the PCN as a sensitive marker for (visual) spatial selective attention (Eimer, 1996; Luck & Hillyard, 1994a; Wolber & Wascher, 2005), this enabled researcher to study the variety of (visual) attention paradigm with (almost) limitless research question, without knowing the exact preattentive hierarchy. For example, the topic of "distractor handling" has become more and more relevant over recent years (Carlisle & Nitka, 2019; Gaspelin et al., 2015, 2017; Kerzel & Burra, 2020; Kiss et al., 2012; Liesefeld & Müller, 2021; Sauter et al., 2019; van Moorselaar & Slagter, 2019; Won et al., 2019). Like a race-model, the parallel presentation of 'pop-out' targets and distractors (among nontargets) compete for attentional resources. The distractor can either capture attention before the target due to higher saliency or, if known and anticipated, it can be actively suppressed ('feature-inhibition') (Berggren & Eimer, 2018; Carlisle & Nitka, 2019; Sawaki & Luck, 2013). As briefly mentioned in the PCN section, the capture and active suppression of distractor information is measured by the P*d* component within 290–340 ms (Feldmann-Wüstefeld et al., 2021; Hickey et al., 2009; Hilimire et al., 2012; Sawaki & Luck, 2010; Wyble et al., 2015). The dynamic interplay between PCN and P*d* reveals a temporal sequence of attentional selection and suppression for targets and distractors, e.g., the PCN becomes delayed or reduced due to distractor interference, obtaining a sensitive template to examine attention shifts (Burra & Kerzel, 2013; Drisdelle & Eimer, 2021; Liesefeld et al., 2021).

The rekindled debate of how distractor handling plays out, either through bottom-up processes, top-down control, or intertrial history, goes hand in hand with the topic of "statistical learning" (Theeuwes, 2019). Essentially, successful distractor inhibition reflects a learning process based on spatial or dimension-specific regularities (Liesefeld & Müller, 2021; Theeuwes et al., 2022; van Moorselaar et al., 2020; B. Wang & Theeuwes, 2018). When a distractor occurs more frequently within a specific location, the spatio-topographic priority map learns this regularity. It decreases the location's weights (or relevance), i.e., the probability-cueing effect, which delays the target detection in the same location (Allenmark et al., 2019; Goschy et al., 2014; Zhang et al., 2019). Interestingly, the (statistical) learning of regularities in time and space can be generalized across (sensory) modalities (Frost et al., 2015, 2019).

Arguably, regularities produce an implicit or explicit rule or template during the task stored in “working memory”, another topic of interest. Although the concept of working memory dates back several decades (Baddeley, 1992, 2003), it spurred attention research since the last years (Grubert et al., 2016; Hanning et al., 2016; Liesefeld et al., 2020; Vogel & Machizawa, 2004; Woodman & Vogel, 2008; Zelinsky & Bisley, 2015). Crucially, (goal-driven) attentional engagement relies on so-called “attentional templates”, which are held in working memory, guiding our attention in an endogenous fashion, e.g., by flexibly activating relevant and discarding irrelevant templates in advance of the display presentation (Grubert & Eimer, 2018). Those templates consist of (a collection of) distinct feature representations like colors; they also can extend across sensory modalities (Grubert et al., 2016; Matusz & Eimer, 2013). The studies on attentional templates demonstrate the impact of top-down control on selective attention that varies depending on the task set (Wolfe, 2021). Similar to working memory, the capacity of attentional templates is limited (Olivers et al., 2011; Woodman et al., 2001).

To obtain electrophysiological insights from templates and working memory processes, those studies focused again on the PCN but also another component, the ‘contralateral delay activity’ (CDA), reflecting working memory processes within 300–900 ms (Carlisle et al., 2011; Ikkai et al., 2010; Luria et al., 2016). Interestingly, the origin of the CDA is also (primarily) located in the posterior parietal cortex; its amplitude indices the ‘cognitive load’ and increases with the number of relevant (task) items until a distinct limit and affects performances (Adam et al., 2018; Ikkai et al., 2010; Luria et al., 2016). While the CDA increases with the number of items or templates (working memory load), the PCN attenuates due to distributed resources among those templates (attention load), revealing a dynamic interplay between working memory and attention processes within parietal regions (Grubert et al., 2016; Heuer & Schubö, 2016). Overall, selective attention and working memory are intertwined based on their specific properties: while selective attention updates working memory, working memory controls the guidance of attention, e.g., during a task (for a review see Oberauer, 2019). Finally, the involvement of frontal (and parietal) regions during working memory activity underscores the importance of top-down control for attention during target selection and distractor suppression (Buschman et al., 2011; Gaspar & McDonald, 2014; Postle, 2006; Soto et al., 2006, 2007), which is closely tight to theta oscillations, e.g., the fronto-midline theta activity (Jensen & Tesche, 2002; Onton et al., 2005).

Another evolving corpus of studies deals with “sustained attention”; shifting the research perspective away from selection processes toward the temporal fluctuations of attentional resources during continuous task performance, e.g., vigilance decrements (Esterman et al., 2013; MacLean et al., 2009; See et al., 1995). Although sustained attention reflects the fifth factor of attention in this thesis and has been briefly introduced before, it shows promise as an evolving branch for future attention research, especially through the liaison with ongoing "brain oscillations" (Clayton et al., 2015; Fiebelkorn et al., 2018; Helfrich et al., 2018; van Es et al., 2022). While the electrophysiological evidence in selective attention primarily focuses on event-related lateralizations (e.g., PCN or CCN), research on sustained attention examines the oscillatory spectro-temporal mechanisms across hubs of the fronto-parietal attention network, like FEF and LIP. Evidence suggests that the communication within the attention network is coupled to the phase of theta oscillations (Helfrich et al., 2018) that indicates rhythmic perceptual sampling during attention processes (Bastos, Vezoli, Bosman, et al., 2015; Calderone et al., 2014).

Furthermore, Fiebelkorn et al. (2018) found a coupling between the 'good' theta phase with beta in FEF, suppressing attention shifts, and gamma in LIP, enhancing visual processing. On the contrary, the authors report another coupling between the 'poor' theta phase and the inhibitory processes of alpha. Notably, these attentional states resemble a dynamic oscillatory neural code of the interplay between selection and suppression, excitation or inhibition. Also, behavioral performances were related to the distinct states of 'good' and 'poor' theta coupling. Fascinatingly, Fiebelkorn et al. (2018) relate these distinct states to the computation within the priority map and derive the following conclusion: "By periodically attenuating the strongest peak in that priority map, the rhythmic properties of spatial attention promote active sampling (Schroeder et al., 2010), preventing spatial attention from remaining overly focused on a single location." (p. 850).

This quote describes the underlying weighting mechanism of attentional prioritization elegantly from an oscillatory perspective. Whereas the 'good' theta phase samples perceptual information (weight accumulation), the inhibitory alpha and 'poor' theta phase rhythmically reduce the importance (or weights) of prioritized but also irrelevant activations, thus parallelly enhancing selection while also triggering a new attention cycle. As the attention network and the "priority network" are essentially the same (Fecteau & Munoz, 2006; Fiebelkorn et al., 2018), these oscillatory patterns reflect a rhythmic "communication through coherence" between relevant network hubs, such as FEF and LIP (Bastos, Vezoli, & Fries, 2015; Fries, 2005). Interestingly, the 7–8 Hz theta frequency for attentional sampling (Fries, 2005) and the 14–30 Hz beta rhythm in FEF, are compatible with the 150 ms proposed pace of item recognition and the 30–50 ms timing of attention shifts (Koch & Ullman, 1985; Wolfe, 1994).

This intricate link between processes of selective and sustained attention marks an open scientific gap, which would be promising to close. For example, trial-by-trial variations or fluctuations in sustained attention can be measured by the "variance time-course" (Esterman et al., 2013; Esterman & Rothlein, 2019; Rosenberg et al., 2013, 2015) to classify "in-the-zone" or "out-of-the-zone" attentional states, which should be observable in the time-course of brain oscillations, e.g., theta and/or alpha frequency. Ultimately, the dynamic fluctuations of resources should manifest in the time-course of the PCN component; "in-the-zone" should accelerate and/or amplify the PCN. Also, P*d* should contain more alpha activity than PCN in a spectral decomposition, due to active suppression.

More evidence that underscores the relevance of brain oscillations derives from the fact that early ERP components such as P1 and N1, which are linked to perceptual integration, originate from the superposition of theta and alpha oscillations and their phase alignment during stimulus onset (Gruber et al., 2005; Klimesch et al., 2004). Also, prestimulus alpha activity predicts visual perception in attentional blink paradigms, showing that an internal (or inhibitory) state impairs the attentional orientation toward shortly presented external information (Busch et al., 2009; Hanslmayr et al., 2011; Slagter et al., 2016). Generally, it seems plausible to assume a "neural code" of coupled brain oscillations that enable cognitive states such as attention, working memory, or consciousness due to rhythmic processing and network communication (Friston et al., 2015; Lisman & Jensen, 2013; Mathewson et al., 2012; McLelland & VanRullen, 2016; Palva et al., 2010; VanRullen, 2016).

Given that network communication derives from feedforward and feedback connections, it represents a sequence (or hierarchy) of cognitive states along network hubs, e.g., areas in the visual cortex (e.g., Ahissar & Hochstein, 2004; Bastos, Vezoli, Bosman, et al., 2015). In "predictive coding" the brain creates a model, i.e., the internal representation of the environment, that is constantly updated by the matching (or synchronization) of top-down priors (or expectancies) and incoming sensory information (Alamia & VanRullen, 2019; Bressler & Richter, 2015; Hardstone et al., 2021; Rao & Ballard, 1999; Talsma, 2015). Overall, there are many ways in predictive coding to model cognitive functions (e.g., Spratling, 2017), like Bayesian priors (e.g., Aitchison & Lengyel, 2017), sensory uncertainty (e.g., van Bergen et al., 2015), or dynamic causal modeling (Bastos, Litvak, et al., 2015; Friston, 2012). Further, priors are measured (or estimated) by probability distributions (e.g., Gaussians), and the direction of information flow (feedforward or feedback) is assessed via Granger Causality (e.g., Bressler & Seth, 2011). Essentially, it is assumed that the brain tries to maximize evidence while minimizing entropy or uncertainty (Friston, 2012), a process that is described as the free-energy principle (Friston, 2009, 2010), in accordance with information theory (e.g., Borst & Theunissen, 1999).

Finally, other emerging topics are “neurofeedback” (Bagherzadeh et al., 2020; deBettencourt et al., 2015; Thibault et al., 2018) and the experimental transfer into “real-world settings” with the “co-registration of neuroimaging techniques” such as EEG and eye-tracking, e.g., to examine natural viewing or reading (Barczak et al., 2019; Coco et al., 2020; Dimigen et al., 2011; Kriegeskorte et al., 2008; Leszczynski & Schroeder, 2019; Võ & Wolfe, 2015). Altogether these ‘new’ topics rely on the profound scientific evidence from the previous topics to make significant contributions, including (crossmodal) selective attention, e.g., oscillatory synchronization. For example, neurofeedback requires adequate and feasible neural markers to track the ongoing (and fluctuating) activity of brain states (e.g., attention) and the knowledge of technical constraints, to improve performances by presenting target information in the optimal moment (e.g., ‘good’ theta phase). Concurrently, real-life settings are the natural consequence of decades of fundamental laboratory research, yet they require advanced calibration techniques to ensure the quality and validity of their findings.

Hypotheses and Limitations

Given the detailed and in-depth empirical description and introduction of the scientific background spanning over (many) decades of attention research, everything is now set to develop the main research hypotheses. Although the (ultimate) main objective is indeed to develop and derive a complete and concise *computational model of* (crossmodal) *attention* (CMA), likewise it is also clear that this cannot be achieved within one thesis, indicating both the first limitation but also the global and superior goal. Similar to other attempts that parameterize the functionality of attention processes and cognitive states by computational models (e.g., Bundesen, 1998; Friston, 2009; McClelland, 1979; Miller, 1982; Wolfe, 1994), it is reasonable to assume that processes of (crossmodal) selective attention can be described by a set of distinct and latent factors, containing specific loads (or weights) and interactions. Overall, this thesis focuses on five factors of attention referenced throughout the introduction in a confirmatory fashion: *object features*, *spatial locations*, *intertrial effects*, *attentional templates*, and *sustained attention* (s. Figure 1–1). This factor set does not claim to be exhaustive, yet it arguably depicts the skeleton structure of the underlying neural dynamics of selective attention. It contains the bidirectional streams of spatio-topographic feedforward (bottom-up) and feedback (top-down) processing and its temporal derivatives of the immediate past and resource fluctuations, which altogether assemble in the moment of attentional capture.

Of course, plenty of other factors or rules, such as reward or meaning (Wolfe & Horowitz, 2017), reflect additional functional or factorial embeddings for a CMA. However, it can be argued that a CMA includes primary and secondary factors. While primary factors describe the general architecture of an attention model, i.e., its functional skeleton, secondary factors unravel (sub-)classes within and across those primary factors. For example, while initiating a target template reflects a common process, the inherent or expected value of such a template is based on attribution and decision rules, e.g., via target ranking. Similarly, saliency represents a complex configuration that modulates by various variables, e.g., segregation or similarity (Duncan & Humphreys, 1989). These secondary factors can be considered categorical compounds of saliency and relevance that culminate in the priority map with their qualitative or quantitative expression (Fecteau & Munoz, 2006; Wolfe, 1994).

The main goal of a CMA is to quantify the impact or effect sizes of each factor (and their sub-factors) by linear combinations and interaction terms and to evaluate its prediction and time regression. This ultimate and global goal can only be achieved in several stages, i.e., a sequence of local goals. Besides the main research question of whether common or separate pools of attention exist across sensory modalities, this thesis's (additional) local goal is empirically examining those five (primary) attention factors. Since the deployment of crossmodal selective attention depends on the outcome of preattentive computations – the weighted prioritization and 'winner-takes-all' selection –, a scientific framework or preattentive architecture was introduced with testable statements and predictions, the *modality-weighting account*.

The modality-weighting account or MWA (proposed by Töllner et al., 2009) extends the *dimension-weighting account* or DWA (Found & Müller, 1996; H. J. Müller et al., 1995) from visual search toward crossmodal search. Overall, there is common agreement that preattentive computations derive from a hierarchy of multiple layers, whose activations accumulate onto the attentional priority map (Fecteau & Munoz, 2006; Koch & Ullman, 1985; Wolfe, 1994). Whereas the assumption of *dimension maps* reflecting independent analyzer units (e.g., color dimension) consisting of many *feature maps* (e.g., red, green, blue) is sufficient in visual search, it remains an open issue whether the same hierarchy is also suitable for crossmodal search, or whether it needs to be expanded with the additional layer of *modality maps*.

As the empirical background shows plenty of similar findings in visual and crossmodal search, for example, there are substantial redundancy gains (e.g., Forster et al., 2002; Krummenacher et al., 2002b; Miller, 1982; Töllner, Zehetleitner, Krummenacher, et al., 2011) or intertrial switch costs (e.g., Found & Müller, 1996; Krummenacher et al., 2009; Spence, Nicholls, et al., 2001; Töllner et al., 2009) for target combinations or changes across visual dimensions or sensory modalities, respectively. Despite this vast and compelling corpus of studies, the intricate question of whether dimensions and modalities derive from the same hierarchical level or rather different levels within the preattentive stage is still an open issue (s. Figure 1–2). Nevertheless, closing the gap is feasible by utilizing an adequate crossmodal search paradigm (here, visuo-tactile search) and focusing on the distinction between visual dimensions and sensory modalities across search tasks. Essentially, a decision or ruling that either favors an MWA (research hypothesis) or a DWA (null hypothesis) is applicable via the statistical evaluation of dimension- and modality-conditions.

Commonly, the outcome of task manipulations (or conditions) is measured by participant's performances that record behavioral variables such as error rates and reaction times. In addition to these standard dependent variables, which are indispensable for testing the hypotheses, this thesis also incorporates neural markers, i.e., PCN and CCN, as spatial detectors for the deployment of visual and tactile attention, respectively (Eimer, 1996; Forster et al., 2016; Gherri et al., 2021; Luck & Hillyard, 1994a; Wolber & Wascher, 2005).

The choice of these markers is based on electrophysiological evidence. As described in detail above, the PCN and CCN reflect spatio-topographic processes of target selection, i.e., the outcome (or summation) of the preceding preattentive computations, within their specific brain areas: visual or somatosensory cortex. Overall, the preattentive stage includes two distinct stages that precede the timing of attentional engagement with a maximum at 250 ms (Talsma & Woldorff, 2005). The first stage is the phase of *unimodal* perceptual integration that is tied to the modality-specific neural signatures of P1 and N1 components – VEP or SEP (e.g., García-Larrea et al., 1995; Mangun, 1995). While the P1 is considered to reflect excitatory perceptual processes, initiating firing in relevant receptive fields (Chelazzi et al., 1993, 1998) in the supragranular layer, the N1 component reflects inhibitory processes (Bruyns-Haylett et al., 2017; Gruber et al., 2005; Hanslmayr et al., 2007, 2011), likely improving the signal-to-nose ratio across unimodal activations. The second stage of multisensory integration occurs (arguably) after the peak timing of the N1 (tactile: N140 or visual: N180) around 180 ms (Talsma & Woldorff, 2005), and is associated with *crossmodal links* or network hubs (IPL, IPS, & TPJ) in the posterior parietal cortex (PPC) (Macaluso et al., 2002a, 2002b;

Man et al., 2015; Quinn et al., 2014). In line with these two stages of *unimodal* and *crossmodal* integration is also the timing of the two pathways of spatial attention. While the 'fast' (SC) pathway activates the 'spatial frame' (within P1 timing: 90–120 ms) during a first loop, the 'slow' pathway or second loop provides the (multisensory) 'content' of feature activation loads within the timing of spatial selective attention (210–240 ms) (Chambers, Payne, et al., 2004). The PCN and CCN reflect *modality-specific* components, whose neural dynamics (magnitude and/or timing expressions) modulate sensitively due to the location-based and feature-based prioritization (Ambron et al., 2018; Forster et al., 2016; Grubert et al., 2011; Katus & Eimer, 2019; Töllner et al., 2008; Töllner, Zehetleitner, Krummenacher, et al., 2011). Finally, the PCN should also be sensitive to detect *supramodal* activity due to crossmodal links in the PPC that combine the egocentric (or somatotopic) tactile information with the external visual information (Eimer, van Velzen, et al., 2003; Eimer & van Velzen, 2005; Macaluso et al., 2000; Man et al., 2015; Quinn et al., 2014).

Although it would be interesting to examine the full timeline of perceptual, preattentive, attentive, post-selective, and response stages to gain a complete picture of the underlying electrophysiological spatio-temporal components and patterns during crossmodal search, unfortunately (and understandably) though, this is out of the scope for this thesis. Of course, other components such as the N1, the CDA, or the lateralized readiness potentials (LRPs) (García-Larrea et al., 1995; Ikkai et al., 2010; Näätänen, 1982; Töllner, Zehetleitner, Krummenacher, et al., 2011; Ulrich & Miller, 2001) would undoubtedly contribute to a CMA, yet this voluntary restriction or limitation was made to ensure the quality of the primary research questions by focusing on behavior and the neural dynamics in PCN and CCN. Surely, those analyses can be made up in future studies.

Before presenting the superordinate research questions to test whether modality-weighting explains the preattentive modulations during crossmodal search in behavior and in PCN and/or CCN for each of the five factors, a brief description on the included (and non-included) EEG studies and the many pilot testings (including 8–12 participants each) is given.

Overall, the viuso-tactile search paradigm consisted of 8 or 10 symmetrically collocated tactile solenoids and visual fields on a transparent frame. Generally, participants had to place their fingertips on top while fixating the center (at around 50–55 cm distance). Further, crossmodal displays were presented for 250 ms and response were given by pressing two foot pedals. Three tactile pilot testings were conducted to derive a set of discriminable tactile features with distinct frequencies (e.g., 30 Hz, 40 Hz, 100 Hz, 150 Hz, 200 Hz). The amplitude of tactile frequencies was controlled and adjusted with an oscilloscope. For each

EEG experiment a series of 1–3 pilot testings were carried out to ensure optimal experimental conditions for the main experiment. For example, the redundancy experiment had to incorporate three feature-contrast (color, shape, & frequency) with a comparable saliency (or reaction time). To guarantee this, 3 pilots had to be executed by comparing a blue square with multiple colors and tactile frequencies. Also, some pilot testings did not qualify for the EEG, e.g., 'reverse redundancies': a stable target with changing nontargets, or were embedded in a colleague's project, e.g., 'additional singleton'.

For the EEG experiments, two EEG studies were included in the three main chapters examining redundancy gains (20 participants) and the modality-shift effect (18 participants). Furthermore, another EEG experiment was conducted for attentional templates (19 participants), which is not included as a unique main chapter (preliminary findings are presented in the General Discussion). Additionally, two other EEG studies were recorded: a redundancy-variation (14 participants) and a dynamic optimization paradigm (8 participants).

Finally, the moment arrived to present the set of superordinate research questions. Again, the main focus is to scientifically judge whether a common or separate pools across sensory modalities exist. This can be achieved by finding either *modality-specific* or *supramodal* mechanisms in crossmodal attention. While a supramodal mechanism reveals similar dynamics for the same task condition across sensory modalities (e.g., reduced amplitudes in PCN and CCN), a modality-specific mechanism may only apply to one modality (e.g., faster PCN onsets but no effect in CCN), or it is opposite between modalities. Remember, the other main goal is to provide empirical evidence for the presence of a MWA, therefore each question will (more or less) be related to this (e.g., favoring a MWA instead of DWA). Generally, these research question arise from the five factors of attention. Hence, each factor will be linked to 2–3 distinct research questions which derive from the introduced scientific background and its detailed descriptions.

(i) object features

As shown in visual and crossmodal search, substantial redundancy gains arise between visual dimensions and sensory modalities (Miller, 1982; Töllner, Zehetleitner, Krummenacher, et al., 2011). The open question in Chapter 2 is: (1) *Do crossmodal redundancies (vision + touch) outperform intramodal redundancies (color + shape)*? and: (2) *Do these crossmodal benefits show up in the PCN and/or CCN*? Also, (3) *is there evidence for a preattentive locus of coactivation*?, adding neural evidence for the ongoing debate (Feintuch & Cohen, 2002; Zehetleitner et al., 2009).

(ii) spatial locations

The second study (Chapter 3) extends the findings on redundancy gains by looking at interactions between the external visual and egocentric tactile spatial functions (Assumpção et al., 2018; Carrasco et al., 1995; Föcker et al., 2010; R. Gray et al., 2009; Staugaard et al., 2016; Wolfe et al., 1998). By analyzing processes of spatial attention in crossmodal search the following question(s) will be essential to examine: (4) *Are there specific spatial functions for vision and touch, that are quadratic and flat, respectively*? (5) *Do these modality-specific spatial functions also appear in visual PCN and/or tactile CCN*? The final question is more specific and linked to the slopes of quadratic fitting: (6) *Is there a systematic patten from quadratic fits in focal and peripheral slopes that explain intramodal and crossmodal benefits, e.g., limitation or amplification*?

(iii) intertrial effects

The third study (Chapter 4) examines the hierarchy of intertrial switch costs (Found & Müller, 1996; Spence, Nicholls, et al., 2001; Töllner et al., 2009). (7) *Is the shift between modalities (e.g., vision → touch) more costly than between visual dimensions (e.g., color → shape)*? (8) *Are these costs reflected by reduced amplitudes and delays in the PCN and/or CCN*? (Eimer et al., 2010; Töllner et al., 2008). Further, it will be interesting to investigate: (9) *How long do those switch costs, reflecting neural traces, persist across trials*? (e.g., Zylberberg et al., 2009).

Although the factors of **attentional templates** and **sustained attention** are not included as separate studies, nonetheless, some interesting research questions can be formulated:

(iv) attentional templates

For example, it would be interesting to gain a deeper understanding on how performances and PCN dynamics of a ‘stable’ template (e.g., a square target) modulate by combining it with another target template of varying similarity: e.g., a triangle target (same visual dimension), a color target (different visual dimension), and a tactile frequency target (different modality). (10) *Is there a* *functional template difference across shape targets which goes beyond priming effects, i.e., prior history*? (e.g., Lamy & Kristjánsson, 2013). Essentially, the square target should differ between intramodal and crossmodal templates pairs, as attention resources become more distributed across sensory modalities.

(v) sustained attention

There is a reasonable amount of empirical evidence promoting the functional role of brain oscillations within the fronto-parietal attention network, especially for theta and alpha oscillations (Fiebelkorn et al., 2018; Helfrich et al., 2018; Michel et al., 2021). Based on this scientific background, response fluctuations are closely linked to either “in-the-zone” or “out-of-the-zone” attention states (Esterman et al., 2013; Rosenberg et al., 2015, 2016) and should be related to brain oscillations and the PCN. (11) *Are there functionally relevant correlations between response fluctuations, brain oscillations, and the PCN*?

Overall, the established scientific framework that promotes modality-weighting in crossmodal search during the preattentive stage can be regarded as a consequence of decades of research. Although it displays only a section of attention research, nonetheless, the set of superordinate research questions (1–11) are a guide to examine the neural dynamics in crossmodal selective attention. The answers to these questions will be discussed in the final “General Discussion” chapter.

2

Redundancy Gains in Crossmodal Search

Tell me, how quickly
Does the world unveil itself?
Just wait an instant.

Hierarchy of Intra- and Crossmodal Redundancy Gains in Visuo-Tactile Search: Evidence from the PCN[1]

Jan Nasemann [1,2], Thomas Töllner [1], Hermann J. Müller [1,2], Zhuanghua Shi [1,2]

1 General and Experimental Psychology, Department of Psychology,
Ludwig-Maximilians-Universität München, Munich, Germany
2 Graduate School of Systemic Neurosciences,
Ludwig-Maximilians-Universität München, Planegg, Germany

---

[1] © 2023 MIT Press. This is the accepted version of the following article: Nasemann, J., Töllner, T., Müller, H. J., & Shi, Z. (2023). Hierarchy of Intra- and Cross-modal Redundancy Gains in Visuo-tactile Search: Evidence from the Posterior Contralateral Negativity. *Journal of Cognitive Neuroscience* 35(*4*), 1–28. It is published in final form at http://dx.doi.org/10.1162/jocn_a_01971. This article may be used for non-commercial purposes in accordance with the MIT Publication Agreement: https://direct.mit.edu/journals/pages/authors#author_posting

Abstract

Redundant combination of target features from separable dimensions can expedite visual search. The dimension-weighting account explains these "redundancy gains" by assuming that the attention-guiding priority map integrates the feature-contrast signals generated by targets within the respective dimensions. The present study investigated whether this hierarchical architecture is sufficient to explain the gains accruing from redundant targets defined by features in different modalities, or whether an additional level of modality-specific priority coding is necessary, as postulated by the modality-weighting account (MWA). To address this, we had observers perform a visuo-tactile search task in which targets popped out by a visual feature (color or shape) or a tactile feature (vibro-tactile frequency) as well as any combination of these features. The RT gains turned out larger for visuo-tactile versus visual redundant targets, as predicted by the MWA. In addition, we analyzed two lateralized event-related EEG components: the posterior (PCN) and central (CCN) contralateral negativities, which are associated with visual and tactile attentional selection, respectively. The CCN proved to be a stable somatosensory component, unaffected by cross-modal redundancies. In contrast, the PCN was sensitive to cross-modal redundancies, evidenced by earlier onsets and higher amplitudes, which could not be explained by linear superposition of the earlier CCN onto the later PCN. Moreover, linear mixed-effect modeling of the PCN amplitude and timing parameters accounted for approximately 25% of the behavioral RT variance. Together, these behavioral and PCN effects support the hierarchy of priority-signal computation assumed by the MWA.

Keywords: redundancy gains, PCN, CCN, crossmodal search,
dimension-weighting account (DWA), modality-weighting account (MWA)

Introduction

A fire alarm system includes both visual and auditory warning sub-systems, enabling people to make a fast escape decision in an emergency based on redundant, visual and/or auditory information. Facilitation by such redundant information is known as the redundant-signal effect or redundancy gains (Miller, 1982). In experimental settings, redundancy gains manifest in terms of both enhanced response speed and accuracy, as has been demonstrated in various search paradigms (Krummenacher et al., 2001b; Miller, 1982; Todd, 1912; Töllner, Zehetleitner, Krummenacher, et al., 2011).

Redundancy gains have been reported, for instance, for targets redundantly defined in multiple visual feature dimensions in visual-singleton, 'pop-out' search (e.g., color + shape, color + motion, shape + motion) (Feintuch & Cohen, 2002; Krummenacher et al., 2002a; Krummenacher & Müller, 2014; Mordkoff & Yantis, 1993; Töllner, Zehetleitner, Krummenacher, et al., 2011), or, respectively, in multiple sensory modalities in crossmodal search (e.g., vision + touch) (Diederich & Colonius, 2004; Forster et al., 2002; Miller, 1982). In crossmodal search, redundancy gains are also referred to as intersensory facilitation (Blurton et al., 2014; Forster et al., 2002; Meredith & Stein, 1986; Miller, 1986). Interestingly, Diederich and Colonius (2004) found detection performance to be better for trimodal targets, combining redundant information from vision, sound, and touch, compared with bimodal targets, for which performance was better than for unimodal targets.

Initially, redundancy gains were thought to arise as a natural consequence of statistical facilitation (Raab, 1962; Townsend & Ashby, 1983). In this view, redundant signals are conceived as independent 'racers' that are processed in parallel and compete with each other to trigger the (detection) response. In this race model, the response decision is made as soon as the faster racer reaches the decision criterion. Accordingly, the upper limit of the reaction-time performance is constrained by the faster of the two racers: the fastest responses cannot be faster than those produced by the faster racer. However, numerous studies have shown that the pure race-model cannot fully account for redundancy gains that exceeds the boundary of statistical facilitation (Colonius & Diederich, 2006; Feintuch & Cohen, 2002; Hagmann & Russo, 2016; Krummenacher et al., 2002a; Miller, 2016; Schwarz, 1989; Töllner, Zehetleitner, Krummenacher, et al., 2011), in particular, for redundancy gains that give rise to violations of the stringent criterion assumed in Miller's (1982) *race-model inequality*.

Thus, instead of independent racers, violations of the race-model inequality have been taken to reflect *coactivation* effects, that is: both signals combine to activate the (detection) response (Miller, 1982; Zehetleitner et al., 2009). It should be noted, though, that directly testing for coactivations by examining the race-model inequality is very conservative: the absence of violations does not preclude coactivations per se. Given this, other, direct and indirect measures – in particular, maximum negative dependency (Colonius & Diederich, 2017) or redundant-signal repetition (benefit) and switch (cost) effects (Liesefeld et al., 2017) – can be utilized in support of coactive processing. On the other hand, when the noise in the redundant stimuli is increased, the parallel race model, without the assumption of coactivation, can actually produce spurious violations of the race-model inequality (Otto & Mamassian, 2012). Interestingly, violations of the race-model inequality appear to be primarily linked to redundant targets that combine features across separable visual dimensions or across sensory modalities, while they are not observed for redundancies within dimensions or modalities (Forster et al., 2002; Krummenacher et al., 2002a).

The above-mentioned studies have typically examined redundancy gains either across dimensions within one sensory modality (e.g., visual search) or across modalities (e.g., crossmodal search). But it has rarely been addressed whether redundancy gains follow a specific hierarchy within a functional processing architecture that combines the two – dimension and modality – aspects. This issue is of particular importance for theoretical conceptions of the processing architecture underlying attentional selection in perception, specifically: do redundancy gains generated by multimodal signals arise at the same preattentive processing stage as redundancy gains produced by (separable) intramodal signals, or do they reflect processing at a hierarchically higher stage? This was the focal issue investigated in the present study, by comparing and contrasting two accounts: the *dimension-weighting account* (e.g., Found & Müller, 1996; H. J. Müller et al., 1995, 2003) and the *modality-weighting account* (Töllner et al., 2009). – To better understand the architectures envisaged by the two accounts, we consider them in some detail in turn.

Dimension-weighting: A common assumption of theories of attentional selection in visual search, including the dimension-weighting account (DWA), is that search-guiding perceptual information is extracted and integrated via the summation of saliency signals onto an overall-saliency, or 'attentional-priority', map (Fecteau & Munoz, 2006; Koch & Ullman, 1985; Lee et al., 1999; Wolfe et al., 1989). In more detail, at the pre-attentive stage, object features are coded by specific analyzer units forming separable feature maps (e.g., for the various colors, orientations, motion directions, etc.; first-order feature coding in terms of Gaspelin & Luck, (2018a). Of note, a feature activity responding to an item in its receptive field is modulated by the features of its neighbor items. Specifically, the feature activity is higher when the features of the items in its surround differ significantly from the item being detected (e.g., Knierim & van Essen, 1992; Nothdurft, 2000), due to a process called iso-feature suppression (Li, 2002), effectively reflecting *feature contrast*. Some theories, including the DWA (Found & Müller, 1996; Itti & Koch, 2001; H. J. Müller et al., 1995; Wolfe, 1998), assume that the feature-contrast signals are combined into *dimension-specific saliency maps* (e.g., for color, orientation, motion, etc.; second-order feature coding in terms of Gaspelin & Luck, 2018), and the dimension-specific saliency signals, in turn, are integrated into the supra-dimensional *attentional-priority map* (Fecteau & Munoz, 2006; Ferrante et al., 2018), which guides the allocation of focal attention. The DWA additionally assumes that the integration of the dimension-specific saliency signals occurs in a weighted fashion, that is: in a competitive process, greater weight is assigned to saliency signals from 'pertinent' dimensions and correspondingly less weight to non-pertinent dimensions – where the weights are modulated by both inter-trial history and top-down set (H. J. Müller et al., 2003). Thus, for instance, if color, rather than orientation, is target-defining on consecutive trials, the weight for color signals increases and that for orientation signals decreases, as evidenced by color repetition benefits (irrespective of whether a particular color feature is repeated or not; Found & Müller, 1996). Also, if a target is defined by dual features from different dimensions (such as color and orientation), there is a redundancy gain due to the parallel coactive processing of signals from the two dimensions (Krummenacher et al., 2001b, 2002a). However, as shown by Krummenacher et al. (2001b), the size of the gain can be influenced by the distribution of weights assigned to each dimension. For example, if the system has recently encountered targets that rely heavily on color information, a redundant target that is defined both color and shape will be mainly detected based on its color information because color has been given a higher weight in the recent processing history. The shape information will contribute less to the detection of the target because its weight is lower. If multiple redundant targets are encountered in a row, the weights assigned to the different dimensions become balanced, which allows for optimal coactivation to occur. This means that a second redundant target is more likely to be processed coactively than a first one (this is called the redundant-target repetition benefit). However, if the next target is defined by one feature (a non-redundantly, e.g., color), it will be detected less efficiently, showing a redundant- to non-redundant-target switch cost, and it occurs because the balanced weight distribution that was established by the previous redundant target is non-optimal for processing a single feature. Redundancy gains are strongest and violate the race-model inequality when the signals being processed come from a single target located in the same place

(Krummenacher et al., 2002). When the signals are separated by one or more other objects (i.e., they come from two distinct targets), the redundancy gains and violation of the race-model inequality disappear. This finding is consistent with the idea that cross-dimensional signal integration occurs in a location-specific manner, as proposed by virtually all 'saliency-summation' theories starting from Koch and Ullman (1985).

An extension of the DWA (which was developed to account for attentional guidance in *visual* search) that takes other modalities, such as that of touch, into account would assume that guidance signals generated in modalities other than vision are treated at the same hierarchical level as visual signals, effectively just adding other stimulus dimensions. Thus, for instance, in the tactile modality, vibrotactile stimuli would be coded in terms of their vibration frequency (frequency dimension) and amplitude (strength dimension). As illustrated on the right side of Figure 2–1, in the extended DWA scheme, the frequency dimension would just be another dimension (vibrotactile frequency-saliency map) at the same hierarchical level as the visual dimensions of color (color-saliency map), shape (shape-saliency map), etc., with the supra-dimensional priority map integrating the saliency signals from all these dimensions in a weighted fashion. Consequently, for computing the attentional overall-priority map, it would not matter whether saliency signals are coded in the same or different modalities. Accordingly, this scheme makes one critical prediction: redundancy gains should *not* differ between redundant targets defined within the same modality (e.g., visual search: color + shape) and targets defined across different modalities (e.g., visuo-tactile search: color + frequency), provided that the saliency of the respective (color, shape, and frequency) target-defining features is comparable (see e.g., Krummenacher et al., 2001b).

Modality-weighting: An alternative conceptualization, which does predict a difference, was proposed by Töllner et al. (2009) to account for their finding, in a crossmodal search scenario, of a modality-shift effect (i.e., slowed responses to targets in a given – visual or, respectively, tactile – modality on modality-switch vs. repetition trials): their modality-weighting account (MWA). In contrast to the DWA, the MWA assumes an extra layer of modality maps (see left side of Figure 2–1), above the (modality-specific) dimension maps and below the overall-priority map, where the modality maps integrate dimension-specific saliency signals into modality-specific saliency maps, which are then combined across modalities into the attentional-priority map. Some empirical support in line with this scheme is provided by consistent findings of modality-shift costs (Spence, Nicholls, et al., 2001; Töllner et al., 2009), as well as crossmodal redundancy gains (Diederich & Colonius, 2004; Forster et al., 2002; Iacoboni & Zaidel, 2003; Miller, 1982) –

indicating that *multisensory* integration follows similar rules of weighting and prioritization as *multidimensional* integration in visual search. Importantly, on the assumption that there is a relatively separate weight resource for each modality – so that up- or down-weighting of any (say: the visual) modality does not substantially influence the weight available for the other (say: tactile) modality –, the MWA does predict that the gains for crossmodally redundant (e.g., visuo-tactile) targets are larger than those for intramodally redundant (e.g., visual) targets.

Thus, a critical test to decide between the alternative architectures envisaged by DWA and the MWA – in particular, whether it is necessary to assume the additional layer of modality-specific saliency maps (MWA) – would be to directly compare the redundancy gains generated by crossmodal targets with those produced by intramodal targets. As, to our knowledge, there is no prior work providing conclusive evidence to answer this question, the present study was designed to allow this critical comparison by introducing a crossmodal search paradigm that combined target information from the color and shape dimensions in vision and the (vibro-tactile) frequency dimension in touch; thus, besides non-redundant (color C, shape S, frequency F) targets, there were both visually redundant (color+shape C+S) and crossmodally redundant (color+frequency C+F, shape+frequency S+F, color+shape+frequency C+S+F) targets, allowing direct comparison of the redundancy gains between the two types of redundant targets. Critically for this comparison, the baseline, bottom-up feature contrasts have to be (reasonably well) equated between the three types of non-redundant target (i.e., color, shape, and frequency) signals; if one was more bottom-up salient than another, detection of redundantly defined targets would be driven (mainly) by the more salient target feature, working against the establishment of redundancy effects. Thus, following Zehetleitner et al. (2013), we matched the three types of target feature introduced in the main experiment in terms of their bottom-up saliency by ensuring that they produced comparable reaction times and error rates in pilot experiments (in which only the three non-redundant types of target – C, S, and F – were presented in separate trial blocks). Crucially, in the main experiment, we then examined whether the redundancy gains in the response times are similar or larger for cross- vs. intramodal combinations of these (matched) target features, arguing in favor of either the 'flat' architecture of the DWA or the 'hierarchical' architecture of the MWA.

In addition to examining behavioral performance, we also recorded the EEG to evaluate the neural dynamics of the different types of redundancy gains. While some authors have argued that redundancy gains arise post-selectively, that is, after the deployment of attention (Feintuch & Cohen, 2002; Miller, Beutinger, et al., 2009), others have provided strong behavioral evidence that, at least in search paradigms (with multiple, target and non-target, stimuli), redundancy gains arise also pre-attentively, in the computation of attentional selection priority (Krummenacher et al., 2001b, 2002a; Zehetleitner et al., 2009) – in line with the DWA and MWA.

A good electrophysiological candidate component for attentional selection in the visual domain is the posterior contralateral negativity (PCN, also referred to as N2pc): a lateralized negative deflection (150–350 ms) over parieto-occipital regions (Ansorge & Heumann, 2006; Eimer, 1996; Luck & Hillyard, 1994b; Töllner, Zehetleitner, Krummenacher, et al., 2011; Wolber & Wascher, 2005; Woodman & Luck, 1999). The PCN seems to be a suitable indicator for attentional selection by indicating both the speed of pre-attentive computations and the amount of allocated attention. Examining this component in a visual search paradigm, Töllner et al. (2011) established an earlier and more negative PCN for visually redundant (color + orientation) targets relative to non-redundant (color, orientation) targets, whereas EEG signals reflecting post-selective target processing showed no additional effects beyond the PCN and were overall comparable for redundant and non-redundant targets – providing strong support for a pre-attentive origin of the behavioral redundancy gains.

In tactile search, there is also a component linked to tactile selective attention (in tasks that require finding a tactile target among tactile distractors), namely, the central contralateral negativity (CCN, also referred to as N140cc): an event-related lateralization with a negative deflection (140–340 ms), which specifically indicates tactile target information among tactile distractors, similar to the PCN, but over central regions (Eimer et al., 2004; Forster et al., 2016).

Accordingly, the purpose of the present study was to investigate to what extent redundancy gains can be explained by these components (PCN, CCN) that lead up to attentional target selection in visuo-tactile search. Interestingly, the temporal evolution of both components, the CCN (starting around 140 ms) and PCN (from 170 onwards) (Forster et al., 2016; Hopf et al., 2004; Kiesel et al., 2008), mirror the modality-specific temporal dissociation in the N1, with the tactile N1 (140 ms) over central regions (Eimer et al., 2004; Eimer & Driver, 2000; García-Larrea et al., 1995) preceding the visual N1 (170–220 ms) over parieto-occipital regions (Mangun, 1995; Novitskiy et al., 2011). Given the difference in the emergence of the two components, the processing of touch signals would temporally precede that of visual signals – rendering the CCN unlikely to be responsive to crossmodal, visuo-tactile redundancies. In contrast, the PCN, originating in ventral visual areas (Hopf et al., 2000; Tay et al., 2022), would be more likely to reflect the processing of visuo-tactile redundancies in the guidance of crossmodal search, as it emerges spatially close to supramodal brain regions in the posterior parietal cortex, such as the intraparietal sulcus or temporo-parietal junction, that combine visuo-tactile information (Chambers et al., 2007; Downar et al., 2000; Macaluso et al., 2002b).

In summary, the PCN appears to be the most suited component to examine the pre-attentive hierarchy assumed by the MWA (vs. the DWA), by looking for gains in its temporal evolution and/or magnitude of expression in response to crossmodally redundant (color+frequency C+F, shape+frequency S+F, color+shape+frequency C+S+F) vs. visually redundant (color+shape C+S) search targets. For triple (crossmodally) redundant targets (C+S+F), we expected to find additional gains (Diederich & Colonius, 2004; Wolfe et al., 1989). Importantly, in examining for these crossmodal redundancy gains, we controlled for potential confounds such as linear superposition and volume conductance that may arise from the temporal lag of the PCN relative to the CCN. Last but not least, we expected the PCN to be predictive of behavioral, reaction-time performance. Altogether, our aim was to contribute to a better understanding of the functional architecture of crossmodal attention, by investigating key predictions from the MWA (see Figure 2–1).

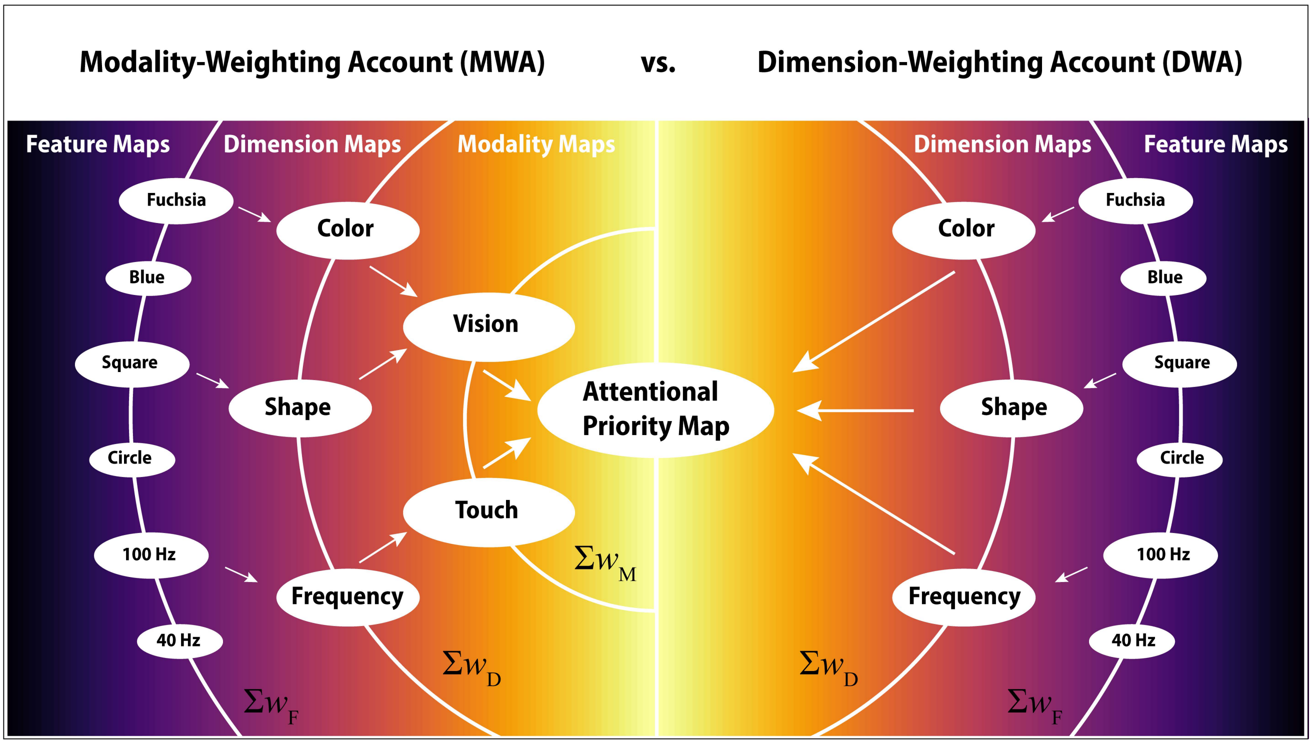

Figure 2–1. Functional architectures of the MWA (left) and DWA (right). In general, activations across hierarchical layers (spatiotopically organized maps) converge onto the attention-guiding overall-priority map, where each hierarchical layer integrates (or sums) signals from the respectively lower layer, in a weighted fashion (*w*; where the total weight is limited). Although both accounts assume that focal attention is allocated (in a winner-take-all process) to the location signaling the highest attentional priority, they differ in their intermediate structures between the entry-level maps of feature analyzers (F) and the attentional-priority map. The DWA (right) parsimoniously assumes only one intermediate layer of dimension-specific feature-contrast maps (D) representing the (bottom–up) saliency of display items within the respective feature dimension; the output of these dimension maps then converges directly, in a weighted fashion, onto the overall-priority map. In contrast, the MWA (left) extends the DWA by introducing an additional layer of modality-specific saliency maps (M), which integrate, in a weighted fashion, the dimension-specific saliency signals within a given modality and output, again in a weighted fashion, to the overall-priority map. The question of which account applies better can be tested empirically by assessing the pattern of redundancy gains in singleton search (RG; i.e., the gains from defining the target redundantly, by multiple features, vs. nonredundantly, by a single feature). Provided that all features are equated in their bottom–up saliency, the DWA predicts that the redundancy gains should be equal for any target–feature combinations across dimension maps (e.g., color + shape vs. shape + frequency). In contrast, the MWA predicts that cross-modal redundancies (e.g., color + frequency) produce greater gains than intramodal (e.g., purely visual) redundancies (color + shape). The entry-level feature-analyzer maps depicted in the figure code for the features of the targets (fuchsia, square, 100 Hz) and distractors (blue, circle, 40 Hz) used in the visuo-tactile search task of the present study; accordingly, there are two separable feature dimensions in the visual modality and one in the tactile modality.

Methods

Participants. 20 participants (6 females, mean age: 24.9, standard deviation: 2.8 years) were recruited for the experiment. They gave their informed consent prior to the experiment and were paid 9 €/hour for their participation. All had normal or corrected-to-normal vision, including normal color vision (tested using the EnChroma color-blind test); also, they all had (self-reported) normal tactile sensitivity in their fingertips. The inclusion criterion for the main task was achieving a tactile detection accuracy above 80% (see next subsection for details), which ruled out three participants who failed to reach this criterion. Accordingly, 17 participants were included in the further analyses. The sample size is in line with recent visual, tactile and multisensory neuroimaging studies, including 16-20 participants (e.g., Ansorge & Heumann, 2006; Busse et al., 2005; Chambers et al., 2007; Forster et al., 2016; Keller et al., 2017; Man et al., 2015; Talsma et al., 2005). Further, this sample size is also supported by assuming medium-to-large effect sizes (generalized η^2 or amount of variance explained) in EEG parameters for repeated-measures ANOVAs of targets and categories ($\eta^2 > .10$, *power* = .90, $a = .05$). The Ethics Committee of the Department of Psychology (LMU) gave their approval for this study.

Stimuli and Procedure. In the crossmodal display, the visual and tactile items were presented simultaneously via Matlab Psychophysics Toolbox using a projector (Optomo) and solenoid actuators (Dancer Design, diameter 1.8 cm), the latter linked to a MOTU system and a 10-channel amplifier. As illustrated in Figure 2–2, the two types of stimuli appeared collocated at 2 x 5 locations – symmetrically arranged along the fingertips of each hand, and separated by a central fixation cross –, with the visual items displayed just above the actuators.

Participants were seated comfortably looking down onto a semitransparent plexiglass table surface (table height: 84 cm; surface: 70 x 60 cm, tilted 60° relative to the upright observer) in front of them, which served as display screen (screen-to-eye distance: ~55 cm), with their fingertips resting softly on top of the actuators. On average, the effective display area subtended 38.1° x 12.5° of visual angle; visual items (color-filled-in squares or circles) were 3.1° in size (side length or, respectively, diameter) and isoluminant (~36 cd/m^2), with a center-to-center separation of the visual items to tactile solenoids of ~2.1° and an inter-item separation between visual and tactile target locations of ~1.9°. Participants wore (in-ear) headphones playing pink background noise (which is smoother and so experienced as more pleasant than white noise) and earmuffs (3M Peltor) to mask any potential auditory cues coming from the actuator vibrations or other external sources, with the loudness of the noise adjusted for each participant.

The task required detection of a singleton 'pop-out' target: participants had to discern whether a target was present or absent, responding by pressing one or the other of two foot pedals when any odd-one-out (target) item was present or absent among homogenous distractor (or non-target) items: blue circles paired with a 40-Hz vibration. Pop-out targets differed from distractors by feature contrast in *color* (C: fuchsia), *shape* (S: square), or (actuator-vibration) *frequency* (F: 100 Hz) or their respective redundant combinations (CS, CF, SF, CSF). The mapping of the foot pedal to the 'target-present' vs. '-absent' response was counterbalanced across participants and switched half-way through the experiment (after 9 trial blocks); to avoid a carry-over effect induced by response switching, participants performed an additional training block (50 trials, same targets) to ensure high accuracy levels (> 80 %) after response remapping.

Importantly, the distractors and the seven targets were iso-luminant (in the visual modality) and of similar vibration amplitudes (in the tactile modality). Based on the MWA, targets can be grouped into four *categories* (or contrast groups): non-redundant targets (NR: C, S, F), visually (dual-) redundant targets (VR: CS), crossmodally (dual-) redundant targets (CR: CF, SF), and triple-redundant targets (TR: CSF) (see Figure 2–2). To ensure a common baseline in non-redundant targets (C, S, F), a series of pilot studies[1] were conducted to select, and introduce in the main experiment, target feature-contrast values for color (fuchsia), shape (square), and frequency (100 Hz) that produced comparable response times and so can be assumed to be of similar bottom-up saliency (cf. Zehetleitner et al., 2013).

The experiment consisted of 1800 trials, presented across 18 blocks of 100 trials each. Within a block, each of the seven targets were presented once at each of the 10 (visuo-tactile) locations, yielding 70 target-present trials (i.e., each target appeared 180 times in total). In addition, there were 30 target-absent trials per block. The reasons for introducing 70% target-present (and only 30% target-absent) trials was that the main focus of the study was on target-generated redundancy gains and to ensure a sufficient signal-to-noise ratio for the corresponding PCN and CCN components, while also avoiding adverse effects of fatigue (a balanced present:absent ratio would have increased the trial number by a factor of 1.4; see also previous studies of redundancy gains in search tasks that used similar, unequal ratios, e.g., Töllner et al., 2011). To control for specific cross-trial target sequences (i.e., inter-trial transitions), we applied de Bruijn sequencing (Aguirre et al., 2011) to ensure that all transitions were equally likely across the 1800 experimental trials.

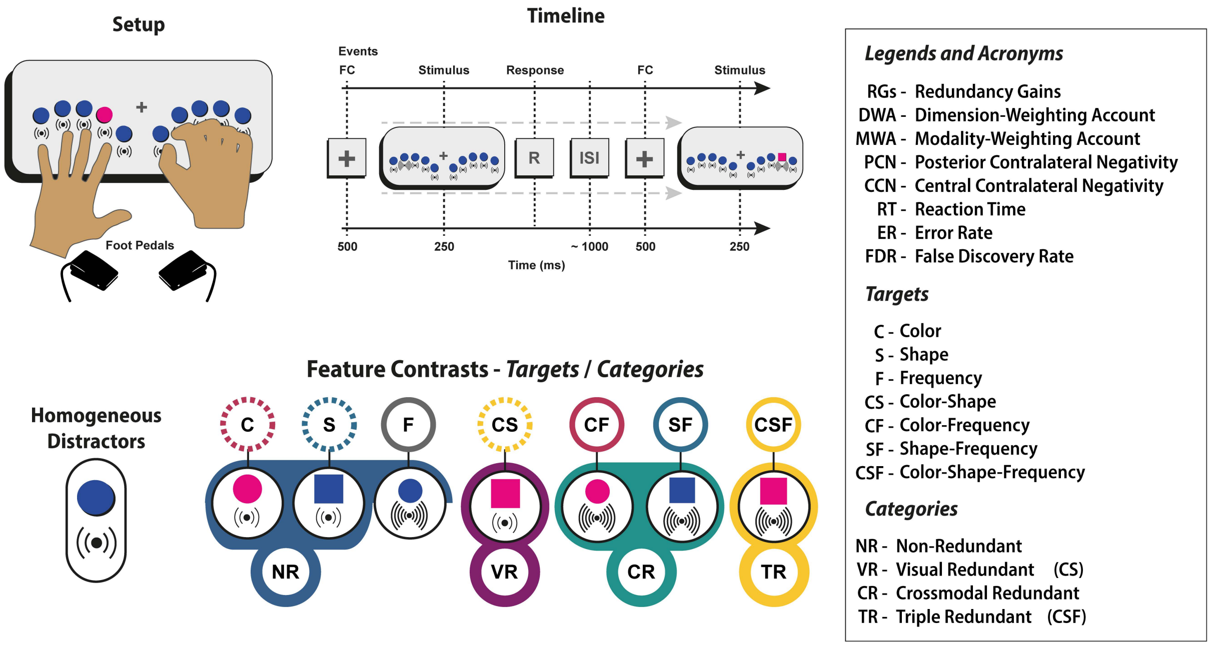

Figure 2–2. Illustration of the experimental setup including visual (projector-generated) and tactile (actuator-generated) stimuli. Participants placed their fingertips softly on the tactile actuators and gave responses via foot pedals. The timeline of the trials in the detection task consisted of a fixation-cross (FC), a visuo-tactile stimulus array, responses (R) via foot pedals, and an interstimulus interval (ISI). In case of wrong answers, error feedback via headphones (beep tone) was played. The durations of these events are depicted underneath in milliseconds (msec). Generally, distractor or nontarget (blue circle + 40 Hz) items were presented at each location, except for one location on target-present trials. Overall, there were seven possible *targets*, defined by feature-contrast(s) in color (fuchsia), shape (square), and frequency (100 Hz), and their respective combinations (C, S, F, CS, CF, SF, CSF). According to the MWA, assuming a hierarchy of redundancy gains, these targets were combined into *categories* of nonredundant (NR), visually redundant (VR), cross-modally redundant (CR), and triple redundant (TR) targets. For assessing behavioral performance, NR consisted of C, S, and F targets; for examining the PCN, which generally describes a visual component, NR included C and S targets, whereas F targets were examined separately. On the right panel, a list of relevant acronyms and legends for targets and categories is shown.

Participants were instructed to fixate at the center and to respond to target presence/absence as fast and accurately as possible. Each trial started with a 500-ms fixation cross, followed by the visuo-tactile stimulus array (with a target present or absent), which was briefly presented for 250 ms. Afterwards, participants had to respond by pressing a foot pedal. If the response was wrong, error feedback was presented in the form of a 300-ms warning 'beep' (330 Hz) via the headphones. The next trial began after an inter-trial interval randomly selected between 900 to 1100 ms (see Figure 2–2).

One or two days prior to the formal experiment, participants underwent a training session practicing the tactile detection task in 6 to 8 blocks of 100 trials each. That is, only tactile, 100-Hz (F) targets were present among homogeneous, 40-Hz distractors (with the visual array consisting of color-and shape-homogeneous blue circles), with a target being present in 50% of the trials. Participants had to reach a final level of above 80% correct responses on all, target-present and -absent, trials (in the last practice blocks) in order to be admitted to the main experiment (this ruled out three participants, whom we then replaced by three other participants). The same threshold (> 80% correct) was also applied to tactile target-detection performance in the main experiment, which led to the exclusion of three participants, although they had previously passed the training.

Performance Analysis. We analyzed both *error rates* and mean *reaction times.* Since the pop-out task was generally quite easy, for reaction times (mean = 597 ms, standard error = 21 ms) only correct-response trials ranging between 250 and 1200 ms were included (on average, 2% of the trials were excluded). For statistical analysis, we conducted repeated-measures ANOVAs, post-hoc testing with Bonferroni correction and repeated-measures correlations (Bakdash & Marusich, 2017). As effect-size parameter, the generalized eta-squared values (η^2) are given (Lakens, 2013), violations of sphericity are corrected with Greenhouse-Geisser, and 95% within-subject confidence intervals are shown and reported (Loftus & Masson, 1994). Overall, we analyzed performance for the seven targets (C, S, F, CS, CF, SF, CSF), as well as the four MWA-based categories (NR, VR, CR, TR). Moreover, we examined for coactivations by redundant targets both *directly* by looking at violations of the race-model inequality (after removing twins of error responses from the distribution of correct reaction times) by the first and, respectively, the second occurrence of two sequential redundant targets (Colonius et al., 2017; Krummenacher et al., 2001b; Miller, 1982; Ulrich et al., 2007), and *indirectly* through repetition effects and redundancy gains (Liesefeld et al., 2017). Furthermore, we compared the responses to redundant targets and, respectively, redundant-target categories with the minimum of their oppositely sorted feature contrasts, known a the ‘maximum negative dependency’ (Colonius & Diederich, 2017), using one-tailed independent t-tests and bootstrapping procedures (with 10.000 repetitions). Statistics were performed using R.

EEG Recordings and Preprocessing. We recorded the EEG using 64 Ag/AgCl active electrodes placed in the 10–20 system (Brain Products™) with a 1000-Hz sampling rate. During the task, we kept impedances below 5 kΩ. Further, we used Python and MNE (Gramfort et al., 2013, 2014) for EEG preprocessing and analysis and R for statistics. The preprocessing steps included a bandpass filter (0.1–70 Hz) and a notch filter (50 Hz) to eliminate line noise. We then used ICA (extended infomax) to separate ICA components of eye blinks and saccades and corrected the EEG recordings by applying the inverse transformation of the ICA and removing ICA components associated with ocular artifacts.[2] Subsequently, the EEG signals were re-referenced by mastoids (TP9, TP10) and a 40-Hz lowpass filter was applied.

For the main analysis, we applied an extra 12-Hz low-pass filter with zero-phase shift, to obtain 'cleaner' PCN and CCN signals, which are generally susceptible to noise (Miller et al., 1998; Ulrich & Miller, 2001). Further, this filtering includes the alpha and theta frequency bands, which both are linked to selective attention (J. J. Foster et al., 2017; Keller et al., 2017; Michel et al., 2021). To rule out possible filter effects, such as temporal smearing (Luck, 2014), we used paired t-tests to compare both (12-Hz vs. 40-Hz) signals. As regards behavioral criteria, only correct-response trials with reaction times within 250–1200 ms were selected for epoching the EEG, including a 200-ms baseline interval prior to onset of the search array. Next, we applied mean baseline correction and averaged the waveforms (ERPs) for left- and right-side presented *targets* and, respectively, target *categories*. Following this, we computed the lateralized signals (CCN, PCN) by subtracting the ipsilateral from the contralateral waveforms. For computing the CCN, we selected the difference waves for electrode positions C3/C4 (Eimer et al., 2004; Forster et al., 2016); and for the PCN, we combined PO7/PO8, PO3/PO4, and O1/O2 to improve signal-to-noise ratio (Feldmann-Wüstefeld et al., 2015; Töllner, Zehetleitner, Krummenacher, et al., 2011). Since the PCN primarily indicates visual selective attention, only color (C) and shape (S) targets were subsumed under the non-redundant category (NR), while purely tactile (F) targets were considered separately.

EEG Analysis. For the PCN (150–350 ms) and CCN (140–340 ms), we analyzed latencies and amplitudes with a zero-phase 20-ms moving average and applied the jackknife procedure (Kiesel et al., 2008; Miller et al., 1998; Ulrich & Miller, 2001). The peak *amplitude* and peak *latency* were selected by taking the negative peak, and the *onset* and *offset* by applying the 50% criterion from the peak amplitude. Next, we applied the inverse transformation of jackknife scores (Smulders, 2010); while this is equivalent to other corrections of test statistics (Ulrich & Miller, 2001), it has the additional advantage of examining correlations and linear mixed-effect models. Further, we also extracted the raw, non-jackknife amplitude from the peak minima. For the statistics, we applied the same repeated-measures ANOVAs and post-hoc testing – though, instead of using Bonferroni correction (which is very conservative), we used the false-discovery rate (FDR) for post-hoc comparisons (Benjamini & Hochberg, 1995). To control for a potential bias due to unbalanced categories VR (CS) and CR (CF, SF), we additionally compared the means for *color* (CS, CF), *shape* (CS, SF), and *frequency* (CF, SF) in behavioral performance and the EEG signals. To establish the presence of crossmodal benefits which we expected to find in the PCN in particular, a potential confound must be addressed, namely, that of earlier tactile CCN activity simply leading to a linear superposition of visuo-tactile information. Hence, we also compared the (PCN) onset timings for redundant targets and, respectively, target categories (e.g., SF, or TR) with their linear 'feature sums', that is, the summation of the respective non-redundant PCN signals (e.g., S+F, or NR+F). Crucially, finding faster PCN onsets by redundant signals compared to their linear feature sums would not only argue against the early tactile confound, but also be indicative of the presence of coactivations at the preattentive (attentional-priority computation) stage, reflecting some form of neural race-model violation. Finally, we evaluated linear mixed-effect models (LMM) using common information criteria (AIC, BIC), to make a statement about the predictability (marginalized R^2) of behavioral performance from the PCN dynamics.

Results

Performance. Analysis of behavioral performance (see Figure 2–3) focused on the error rates (ERs) and mean reaction times (RTs). ANOVAs with the single-factor *Target* revealed a prominent main effect in both ER (F (6, 96) = 34.91, p < .001, η^2 = 0.596) and RT (F (6, 96) = 39.24, p < .001, η^2 = 0.243): redundant targets (CS, CF, SF, CSF) were overall detected significantly more accurately and faster than non-redundant targets (C, S, F) (Bonferroni post-hoc tests, ps < .014). Importantly, the ERs and RTs were comparable among the non-redundant targets (ps > .085), indicating that these were well equated in terms of their physical saliency. For the four redundant targets, the ERs were very low overall (< 3 %), with fewer errors for CSF than CS targets (p = .005). In contrast, the RTs exhibited a clear hierarchy of redundancy gains as expected on the modality-weighting account (ps < .036): RTs became progressively faster from CS (549 ms) through CF (526 ms) and SF (521 ms) to CSF (508 ms) targets. The mean target-absent RT (658 ms) was the slowest.

With the RT (*target*) effects being in line with the modality-weighting account (MWA), we proceeded with an analysis based on our a-priori target *categories*. Single-factor *Category* ANOVAs also yielded main effects in ERs (F (3, 48) = 49.38, p < .001, η^2 = 0.609) and RTs (F (3, 48) = 128.02, p < .001, η^2 = 0.224). More errors were made with targets in the non-redundant (NR) category compared to the three redundant (VR, CR, TR) categories (ps < .001); again, responses were less accurate for the visual compared to the triple redundant category (*p* = .001). As for RT performance, NR targets were responded to significantly slower than redundant targets; for the latter, RTs became increasingly faster from VR through CR to TR targets (*p*s < .002). As both error rates and reaction times exhibit similar patterns of redundancy gains, we additionally applied repeated-measures correlations, which yielded a strong positive link between both variables in both the *target-* (*r* (101) = 0.669, *p* < .001) and *category*-based (*r* (50) = 0.806, *p* < .001) analysis. That is, faster responses were also more accurate, ruling out speed-accuracy trade-offs.

Additionally, we compared *colors* (CS, CF targets), *shapes* (CS, SF targets), and *frequencies* (CF, SF targets), to address possible concerns about the design imbalance between the VR and CR categories (one combination in VR vs. two combinations in CR); note that this comparison effectively isolates the contribution of the color, shape, and frequency signals to the redundant-target gain. The results revealed reaction times to redundant targets involving *frequency* (524 ms) to be significantly faster (*p*s < .003) compared to those involving *color* (538 ms) and *shape* (535 ms), which were comparable. In the error rates (which were overall near ceiling), there was a difference only between redundant *frequency* (2.1%) and *color* targets (2.8%) (*p* = .009), for both of which the ERs were comparable to *shape* targets (2.2%). – Thus, the findings, particularly in the RTs, are clearly in line with the MWA by showing significantly greater gains for crossmodal redundancies (see Appendix Figure 2–A1).

The mean RTs for redundant targets (e.g., SF) showed significantly fast compared to the non-redundant targets (e.g., S + F), but this redundant gain did not consistently meet the criteria for coactivation defined by the race-model inequality (Miller, 1982). This was only the case for SF vs. S+F targets, but not for any other redundant targets (all *p*s > .080). Since this test is very conservative, we also looked for indirect evidence of coactivations in terms of the existence of cross-trial repetition and redundancy-cost effects (cf. Liesefeld et al., 2017). Indeed, we found benefits for single-feature repetitions in the non-redundant category (i.e., repetition vs. change of single-feature target across trials *n–1* and *n*: 580 ms vs. 597 ms, *t*(16) = 3.18, *p* = .006), and double-feature repetitions in the redundant categories (i.e., double-feature repetition vs. single-feature repetition and complete feature change across trials *n–1* and *n*: 512 ms vs. 524 ms and, respectively 529 ms, *t*(16) > 4.10, *p*s < .002). Furthermore, single-feature repetitions led to slower RTs after redundant embeddings (585 ms) than after non-redundant presentation (566 ms) (*t*(16) = -3.50, *p* = .003), which is indicative of a bias due to redundant activations.

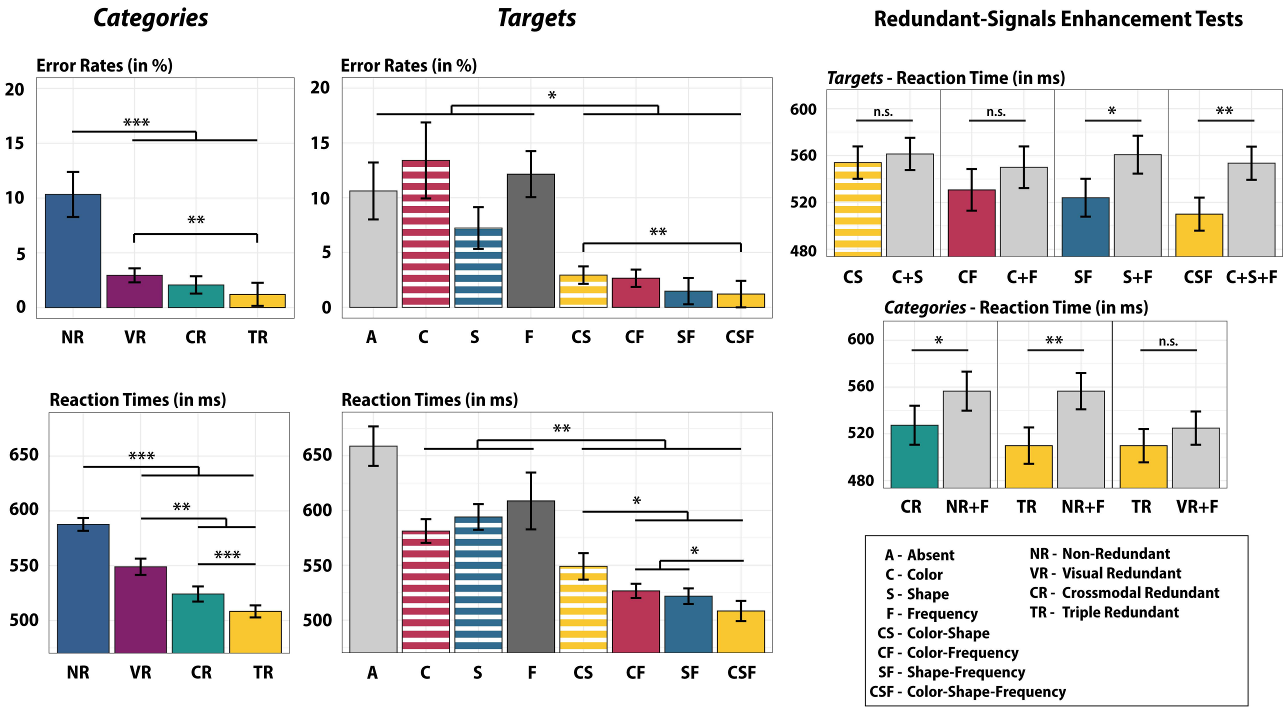

Figure 2–3. The error-rate (ER) and RT results are presented for categories (NR, VR, CR, TR; left panels) and targets (A = target absent, C, S, F, CS, CF, SF, CSF; middle panels). Horizontal lines and asterisks indicate statistically reliable differences from Bonferroni post hoc comparisons (*p < .05, **p < .01, ***p < .001). Of note, the ERs and RTs were comparable for C, S, and F targets, meeting our common-baseline criterion. The right shows the redundant-signals enhancement tests (cf. Colonius & Diederich, 2017) for coactive (vs. the minimum of their negatively dependent RTs) processing of the multiple features, or feature categories, defining redundant targets. Prominent redundancy gains were found in ERs and RTs (marked by *). Whereas in ERs, TR (CSF) was more accurate than VR (CS), and the RTs exhibited a clear hierarchy of redundant gains with benefits increasing from VR (CS) through CR (CF, SF) to TR (CSF), as expected on the MWA.

Finally, the redundant-signals enhancement test (cf. Colonius & Diederich, 2017), also known as the maximum negative dependency test, was used to compare the mean RTs of redundant targets (e.g., SF) to the minimum of the oppositely sorted mean RTs of the non-redundant targets (e.g., S + F). This test is used to determine if coactivations are presented and is less conservative than the race-model inequality test (Miller, 1982). The results of this test showed that coactivations were presented for SF vs. S+F, CSF vs. C+S+F, CR vs. NR+F, and TR vs. NR+F (*ps* < .037, one-tailed), but not for CS vs. C+S, CF vs. C+F, and TR vs. VR+F (*ps* > .15, one-tailed) (see the right-hand panel in Figure 2–3, and Appendix Table 2–A2).

Before turning to the PCN and CCN results, it is important to note that both EEG signals, whether filtered with 12-Hz or 40-Hz, were fully comparable, that is: they showed no differences whatsoever across latency scores (PCN: *t*s(16) <

$|1.55|$, *p*s > .141, CCN: $t(16) < |1.42|$, *p*s > 0.176), thus revealing no effects of temporal smearing Although the amplitudes were different (PCN: $t(16) > |2.35|$, *p*s < .032, CCN: $t(16) > |2.47|$, *p*s < .025), being lower for the 12-Hz signals, this result is rather trivial, since filtering in general converges to the ensemble mean as a limit process while preserving the systematic effect (this can be also seen from the exact-same result in mean amplitudes for both 12-Hz and 40-Hz signals; see Table 1 and Appendix Table 1). Further, in a comparison of (inverse) jackknife vs. (raw) no-jackknife amplitudes, we found the same effect of reduced amplitudes as in the previous comparison (12-Hz vs. 40-Hz). To ensure that both amplitudes within a participant are modulated in the same fashion, we applied repeated-measures correlations, which revealed a high and significant relationship (targets: *r* (84) > .870, *p*s < .001, CI = [.705; .882]; categories: *r* (50) > .893, ps < .001, CI = [.763; .941]), indicative of similar properties across amplitudes. Accordingly, jackknifing can be considered a form of filtering. Given this, the report below focuses on the 12-Hz signals and jackknife scores (a detailed description for 40-Hz targets and categories can be found in Appendix Table 2–A1 and Appendix Figures 2–A2 & 2–A3).

CCN. We found a prominent CCN over central regions (C3/C4) for tactile target signals (F, CF, SF, CSF). Compared to solely visual (i.e., non-tactile) target signals (C, S, CS), amplitudes for the tactile targets were significantly more negative (-1.94 µV vs. -0.78 µV, *p*s < .016). Among the tactile targets (F, CF, SF, CSF), there were no significant differences in CCN in either onsets or amplitudes (all *p*s > .29); hence, neither the CCN onsets (105 to 125 ms) nor the amplitudes (-1.89 to -2.01 µV) were sensitive to crossmodal redundancies (see Figure 2–4). Although the CCN latency analysis yielded a significant main effect (F (3, 48) = 4.81, p < 0.001, $\eta^2 = 0.122$), subsequent post-hoc comparisons, including false-discovery rate (FDR) correction, revealed F to be marginally earlier than CF, SF, and CSF (*p*s = .052). However, no main effects in CCN latency were present in 12-Hz categories, and in 40-Hz targets and categories (*p*s > .187). To examine for the presence of a late visual confound in the CCN for crossmodal (visuo-tactile) targets, we compared the mean amplitudes within the 300–400-ms window. While this revealed a marginal main effect in both *target*- and *category*-based analyses (*p*s = .077, η^2s < 0.038), the effect sizes were small and there were no differences in post-hoc testing (*p*s > .14). – Thus, the CCN reflects a consistent somatosensory component starting around 110 ms post stimulus onset and peaking around 200 ms with -2.00 µV, which is unaffected by crossmodal, visuo-tactile redundancies.

Table 2–1. PCN Results Presented Target- (Upper Part) and, Respectively, Category-dependent (Lower Part) for Onset, Latency, Offset, and Jackknife and Raw Peak Amplitude, and 200- to 300-msec Mean Amplitude (Means and, in Parentheses, Associated Within-subject Confidence Intervals)

PCN	*C*	*S*	*F*	*CS*	*CF*	*SF*	*CSF*	*F Value*	η^2
Onset (msec)	225 (12)	242 (28)	121 (28)	219 (19)	195 (52)	167 (24)	190 (13)	4.08*	.173
Latency (msec)	252 (13)	275 (12)	229 (40)	239 (9)	248 (8)	231 (21)	232 (9)	7.51**	.251
Offset (msec)	287 (33)	301 (12)	281 (110)	256 (22)	278 (10)	290 (18)	265 (10)	3.20*	.103
Peak amplitude (μV)	−0.58 (0.23)	−0.42 (0.30)	−0.57 (0.37)	−0.35 (0.40)	−1.04 (0.33)	−0.92 (0.24)	−1.20 (0.37)	5.2**	.106
Mean amplitude (μV)	−0.33 (0.22)	−0.24 (0.23)	−0.43 (0.85)	0.00 (0.38)	−0.71 (0.27)	−0.75 (0.21)	−0.72 (0.35)	5.31**	.151
Raw peak amplitude (μV)	−1.11 (0.23)	−0.86 (0.22)	−0.88 (0.39)	−0.81 (0.28)	−1.36 (0.25)	−1.29 (0.20)	−1.59 (0.30)	6.04**	.125
Raw mean amplitude (μV)	−0.33 (0.27)	−0.24 (0.23)	−0.43 (0.33)	0.00 (0.38)	−0.71 (0.27)	−0.75 (0.21)	−0.72 (0.35)	5.31**	.151
								df (5, 80)	

PCN	*NR*	*VR*	*CR*	*TR*	*F Value*	η^2
Onset (msec)	227 (13)	219 (19)	173 (24)	190 (12)	8.65**	.276
Latency (msec)	263 (16)	239 (9)	243 (6)	231 (10)	6.08**	.164
Offset (msec)	298 (20)	256 (20)	283 (8)	265 (10)	6.11**	.168
Peak amplitude (μV)	−0.46 (0.19)	−0.35 (0.40)	−0.96 (0.23)	−1.20 (0.38)	7.05**	.124
Mean amplitude (μV)	−0.28 (0.16)	0.00 (0.39)	−0.73 (0.16)	−0.72 (0.36)	6.37*	.172
Raw peak amplitude (μV)	−0.84 (0.18)	−0.81 (0.27)	−1.17 (0.20)	−1.59 (0.30)	9.20**	.159
Raw mean amplitude (μV)	−0.28 (0.16)	0.00 (0.39)	−0.73 (0.16)	−0.72 (0.36)	6.38*	.172
					df (3, 48)	

Furthermore, the results of the repeated-measures ANOVAs are reported as *F* tests, with the effect sizes (generalized η^2) and degrees of freedom (*df*) in targets (5, 80) and, respectively, target categories: (3, 48). In addition, the descriptive result for the tactile-only target (F) is presented (in gray) but omitted from statistics.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

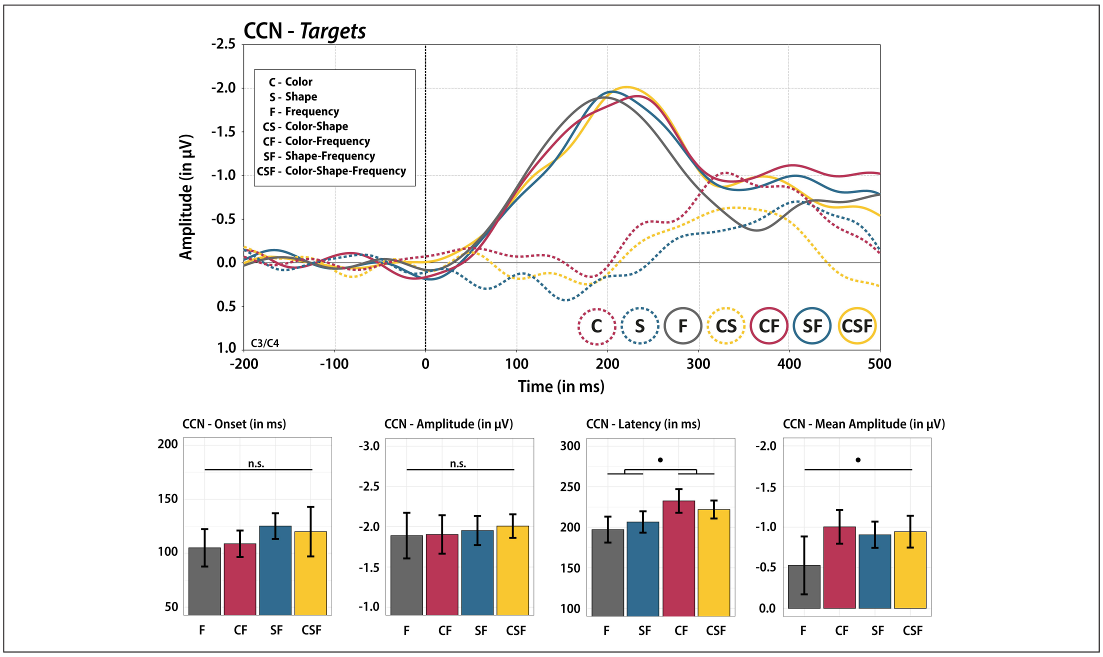

Figure 2–4. The CCN is shown over central regions (C3/C4) for the seven targets (C, S, F, CS, CF, SF, and CSF). Naturally, only targets including a tactile signal (F, CF, SF, CSF) elicited a prominent CCN, in comparison with solely visual targets (C, S, CS). Although the CCN was comparable in onsets and amplitudes (nonsignificant, n.s., comparisons) among the F, CF, SF, and CSF targets, there were at best marginal differences in latency and 300- to 400-msec mean amplitude in post hoc comparisons (FDR; • $p < 0.1$). Thus, the tactile CCN reflects a stable somatosensory component that is not sensitive to cross-modal, visuo-tactile redundancies.

PCN. Typically, the PCN has only been examined in visual search paradigms, so that it is unclear whether it also plays a role for tactile-only information. For this reason, we omitted the solely tactile target (F) from the subsequent PCN analyses and included only color (C) and shape (S) as non-redundant (NR) pop-out signals. But we analyzed the tactile PCN (F) separately. Altogether, all PCN variables – *onset, amplitude, latency,* and *offset* – showed consistent main effects in both *target*- and *category*-based analyses. A full statistical summary of the major findings in PCN is provided in Table 1.

To start with, the PCN *onsets* were modulated congruently across categories and targets. Analyzed *category*-based, crossmodal redundant targets (CR, TR) elicited earlier PCN onsets than solely visual and visual redundant targets (NR, VR) (all *ps* < .027, see Figure 2–5 and Table 2–1). Examined *target*-based, this crossmodal advantage was also evident in earlier PCN onsets for SF and CSF in comparison to C, S, and CS (all *ps* < .03). No crossmodal benefit was apparent for CF, due to a larger variance (see Figure 2–6).

Similarly, the PCN *amplitudes* were significantly more negative in the CR and TR than in the NR and VR *categories* (*p*s < .024). Again, this pattern was consistent when examined *target*-based: amplitudes were generally more negative for crossmodally redundant targets (CF, SF, CSF) than for visual targets (C, S, CS) (*p*s < .046), except for the SF vs. C post-hoc comparison (*p* = .141). Furthermore, findings in PCN amplitudes were mirrored by findings in 200–300-ms mean amplitudes. Again, crossmodal categories (CR, TR) and targets (CF, SF, CSF) had consistently higher amplitudes than NR and VR (*p*s < .053) and C, S, and CS (*p*s < .059), respectively. As mentioned above, the findings and patterns between jackknife amplitudes and raw (no-jackknife) amplitudes were statistically congruent (see Table 1).

This pattern indicates that the PCN is sensitive to crossmodal redundancies in terms of both onset (onsets occurring earlier) and amplitude (amplitudes becoming more negative). However, what remains open is whether these findings indicate true crossmodal gains (reflecting ‘coactivations’) by visuo-tactile targets, rather than simply deriving from the superposition of the earlier (tactile) CCN. We address this issue in the following sections. Of note, no gains in either onset or amplitude were found for purely visual redundancies (see Figures 2–5 and 2–6, and Table 2–1).

As for the PCN *latencies*, these were generally shorter for redundant (VR, CR, TR) than for non-redundant (NR) target *categories* (*p*s < .041), with CR being marginally shorter than NR (*p* = .065), but longer than TR (*p* = .041). This pattern is indicative of PCN latency being sensitive to redundancies. While the pattern of redundancy gains in PCN latencies was similar when examined *target*-based (i.e., with shorter latencies for CS, CF, SF, and CSF vs. C and S), the effect was significant only in comparison to S targets (*p*s < .013), but not reliable in comparison to C targets (*p*s > .059). The latter may reflect a dissociation in feature-contrast computation between color and shape targets (shapes being integrated wholes of more elemental features), while it may also be attributable to a lower PCN signal-to-noise ratio for *targets* in comparison to *categories* (see Figure 2–6 & Table 2–1).

Finally, the PCN *offsets* occurred earlier in VR and TR compared to CR and NR target *categories* (*p*s < .028), pointing to a benefit for visual redundancies. Examined *target*-based, the pattern was again somewhat more complicated (likely for the same reasons as with the PCN latencies): offsets occurred generally (though in some cases just marginally) earlier for visual redundant targets (CS, CSF) (*p*s < .069) compared to S, CF, and SF targets, though not in comparison with color-only (C) targets (*p*s > .372), and the offsets were comparable among C, S, CF, and SF targets (*p*s > .051).

In summary, crossmodal redundancies (CR, TR) produced benefits in terms of earlier PCN *onsets* and larger (negative) PCN *amplitudes*. Further, redundancies (VR, CR, TR) generally gave rise to shorter PCN *latencies*, and visual redundancies (VR, TR) appeared to elicit earlier PCN *offsets*. Given these distinct patterns and temporal effects for visual and crossmodal redundancies, the PCN – like the behavioral results (see above) – favors the MWA (see Figure 2–5 & Table 2–1).

In addition, the separate analysis of the frequency target (F) revealed evidence for a purely tactile PCN: the frequency target, too, elicited a prominent negative deflection over extrastriate regions (see Table 2–1 & Figure 2–6). This tactile PCN, however, occurred much earlier compared to the other targets, having an onset at around 121 ms. It plateaued around 229 ms with -0.57 µV and faded out by 281 ms. Given these properties, the tactile PCN is likely to reflect a wave propagation from the CCN (see above).

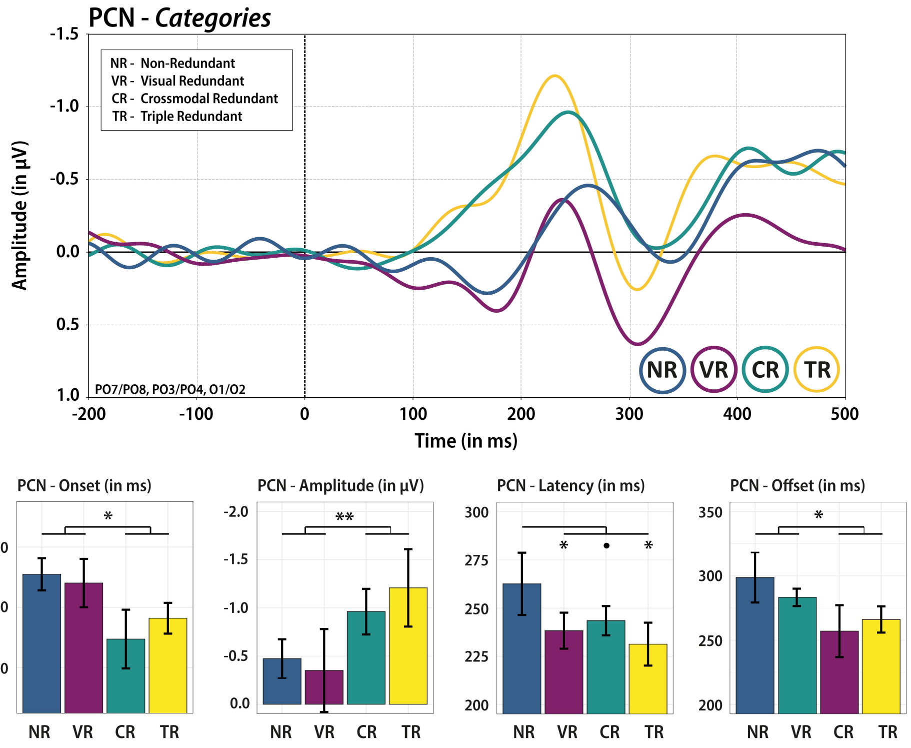

Figure 2–5. The PCN results are shown for categories (NR, VR, CR, TR) over parieto-occipital regions (PO7/ PO8, PO3/ PO4, O1/O2). On the top, grand averages of the PCN signals (< 12 Hz) are shown as negative deflections, which are most prominent within 200–300 msec post stimulus onset. On the bottom, bar plots are shown with FDR post hoc comparisons (• p < .1, *p < .05, **p < .01). Benefits were found for cross-modal redundancies (CR, TR) in both PCN onsets and PCN amplitudes. Furthermore, PCN latencies were generally shorter for redundancies (VR, CR, TR), with only a marginal advantage for CR vs. NR, whereas PCN offsets showed benefits for categories with visual-redundant information (VR, TR).

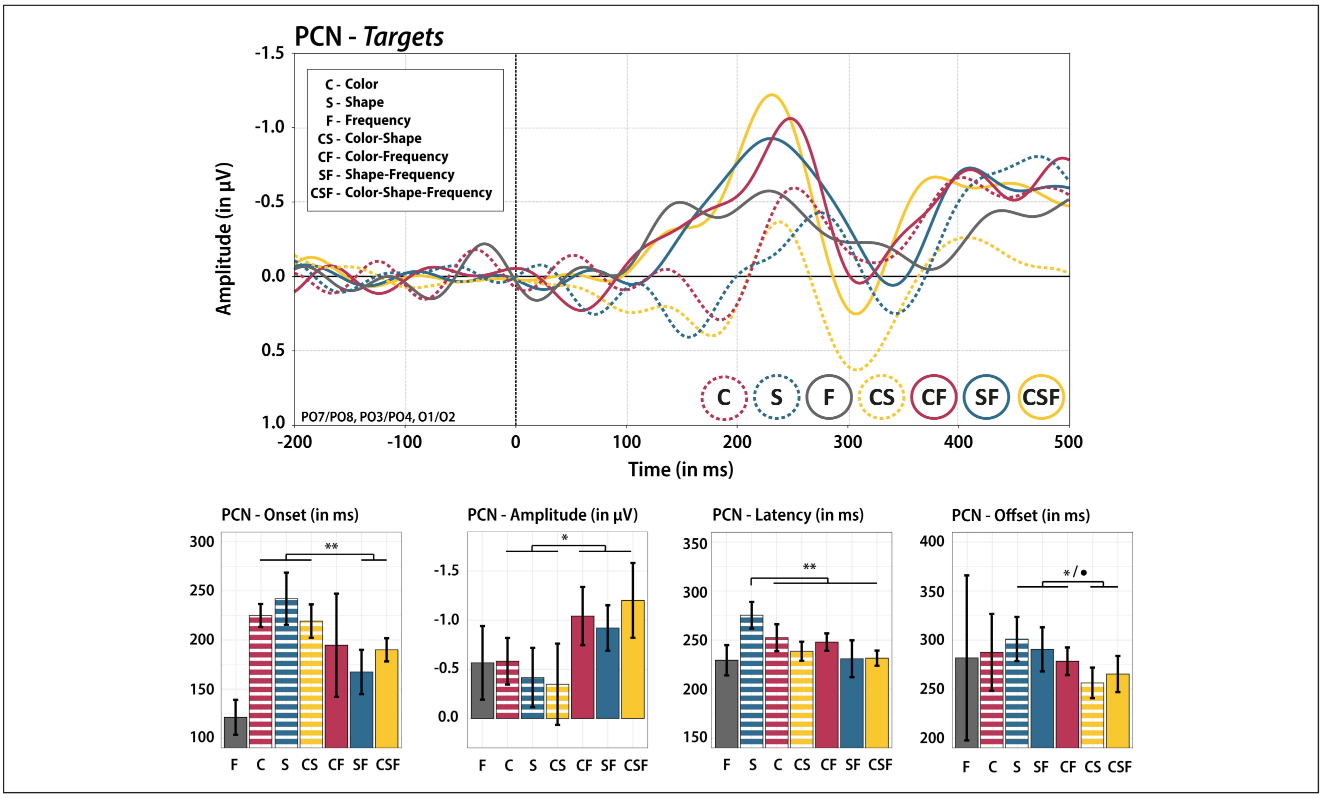

Figure 2–6. The PCN results are shown for targets (C, S, F, CS, CF, SF, CSF) over parieto-occipital regions (PO7/ PO8, PO3/ PO4, O1/O2). PCN grand averages (< 12 Hz) are shown (top) for targets, including a purely tactile PCN (F, gray line). Furthermore, bar plots are shown for onset, amplitude, latency, and offset (bottom), along with within-subject confidence intervals. Here, prominent findings in post hoc comparisons (using FDR correction) are indicated by asterisks (• $p < .1$, *$p < .05$, **$p < .01$). For targets, onsets and amplitudes revealed cross-modal benefits and redundancy gains, occurring earlier in CF, SF, and CSF compared with C, S, and CS. Redundant targets (CS, CF, SF, CSF) had shorter latencies compared with S, but not C. Finally, the PCN offsets occurred consistently earlier for targets including visual redundant information (i.e., CS and CSF), but not for sole-color (C) targets

Analogous to behavioral performance (see above), in a supplementary analysis, we also compared the effects of *color* (CS, CF), *shape* (CS, SF), and *frequency* (CF, SF) signals in redundant targets with regard to the CCN and PCN. In the CCN, only targets including a frequency signal produced a prominent negative deflection, underpinning the tactile sensitivity over central regions. For the PCN, frequency (174 ms), shape (203 ms), and color (213 ms) signals differed significantly in the onsets (*p*s < .041), while the amplitudes were more negative with frequency (-0.96 µV) compared to color (-0.69 µV) and shape (-0.61 µV) signals (*p*s < .023); and the PCN offsets occurred earlier with color (269 ms) and shape (270 ms) compared to frequency (283 ms) signals (*p*s < .061). No significant differences were found in the PCN latencies. These findings complement those from the main analysis by showing crossmodal (frequency) benefits in the PCN onsets and amplitudes, and a visual (color and shape) effect in the PCN offsets. –

Of note, there was a slight PCN onset advantage for redundant-shape (CS, SF) over redundant-color (CS, CF) targets (203 vs. 213 ms), with both exhibiting comparable PCN latencies and offsets (see above). Accordingly, the fact that in the *target*-based analysis, the timing of PCN was not reliably expedited by redundant-color vs. sole-color targets (see above) is unlikely attributable to fundamental differences in feature-contrast computation between color and shape signals, thus favoring an explanation in terms of reduced signal-to-noise ratios (and/or more conservative post-hoc testing in the *target-* as compared to the *category*-based analysis) (see Appendix Figure 2–A1).

Neural Race vs. Coactivation. In a further, critical analysis, we examined whether the crossmodal dynamics in the PCN derive purely by linear superposition of early (tactile) activation reflected in the CCN on the later (visual) PCN, or, alternatively, whether the crossmodal gains in the PCN onsets can be attributed to coactivations by visuo-tactile targets. To this end, we compared the PCN onsets for redundant *Targets* and, respectively, target *Categories* and the respective linear sums of the non-redundant feature contrasts contained in them (e.g., CF vs. C + F). While we did not find temporal benefits in the PCN onsets for CS and CF targets compared to their respective feature sums ($p > .131$) (indicative of linear superposition), for SF vs. S+F and CSF vs. C+S+F (target-based comparisons) and for CR vs. NR+F and TR vs. VR+F (category-based comparisons) we established significantly earlier PCN onsets ($p < .04$). Overall, the PCN onsets were consistently faster than their feature sums (see Figure 2–7 and Appendix Table 2–A2). Of note, finding these consistent temporal benefits in PCN onsets aligns nicely with the coactivation analysis of the reaction times above, which indicates coactivations for the very same combinations (see Figure 2–3). Thus, although linear superposition clearly plays a role (witness the summed-feature neural signals reflecting quite similar PCN patterns as crossmodal redundant signals), it fails to explain the whole effect pattern. Instead, we take our finding of evidence for coactivations in the PCN onsets to argue in favor of crossmodal signals combining to expedite the coding of attentional selection priorities.

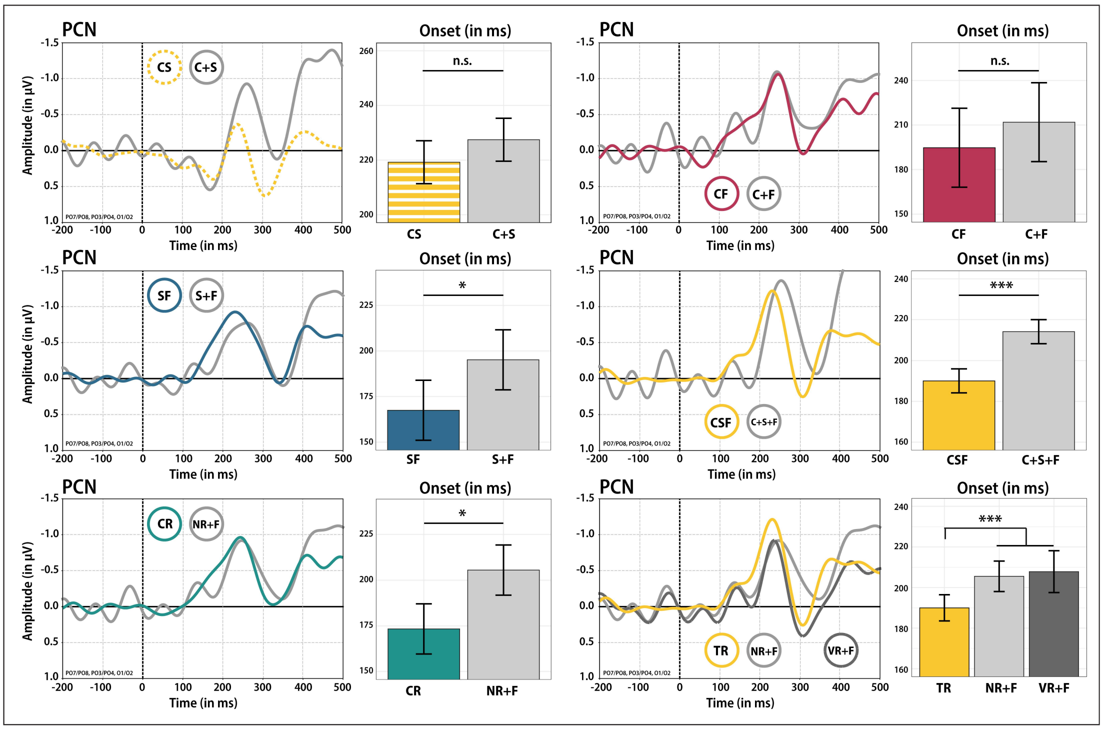

Figure 2–7. PCN onsets for redundant targets and target categories in comparison with their respective feature sums. PCN curves are shown for redundant signals (colored and dashed lines) and the linear summation of the respective features combined in them (gray; e.g., SF vs. S + F). Bar plots (with within-subject confidence intervals) present the results from one-tailed independent t tests of the mean PCN onsets, being either significant (*p < .05, **p < .01, ***p < .001) or nonsignificant (ns). In brief, PCN onsets were expedited above the (temporal) limit of their linear superposition in SF, CSF, CR, and TR, indicative of neural coactive processing of these cross-modal targets or target categories.

Finally, Figure 2–8 depicts the temporal evolution, from onset through peak latency to offset, of the topomaps for each type of *Target*, with the activation map for a given target scaled to half-to-full amplitude range (µV). Descriptively, visual targets elicited a prominent PCN over parieto-occipital regions (see C, S, and CS), with the PCN for color (C) appearing more prominent over occipital regions and that for shape (S) over parietal regions, and both combing with visual redundant color-shape (CS) targets. Generally, visual-target-related activity appeared to propagate towards parieto-central regions. Targets involving tactile information (F, CF, SF, CSF) produced a prominent CCN component over central regions. Of note, the CCN for purely tactile targets (F) was not observed to (strongly) propagate towards occipital regions, while visual and tactile generators appeared to combine for crossmodal targets. Finally, for CF targets, the two neural generators – a central tactile CCN and an occipital visual PCN – are quite clearly discernible, whereas for SF and CSF targets the topomaps show no clear-cut separation. Interestingly, for redundant and particularly crossmodal targets (CS, CF, SF, CSF), the visual and tactile activations in the topomaps appear to combine, which might be interpreted in terms of linear superposition or, alternatively, crossmodal interactions (i.e., coactivation). Although these observations are purely descriptive, we consider it interesting for future studies to examine these spatio-temporal patterns of the event-related lateralizations in greater detail.

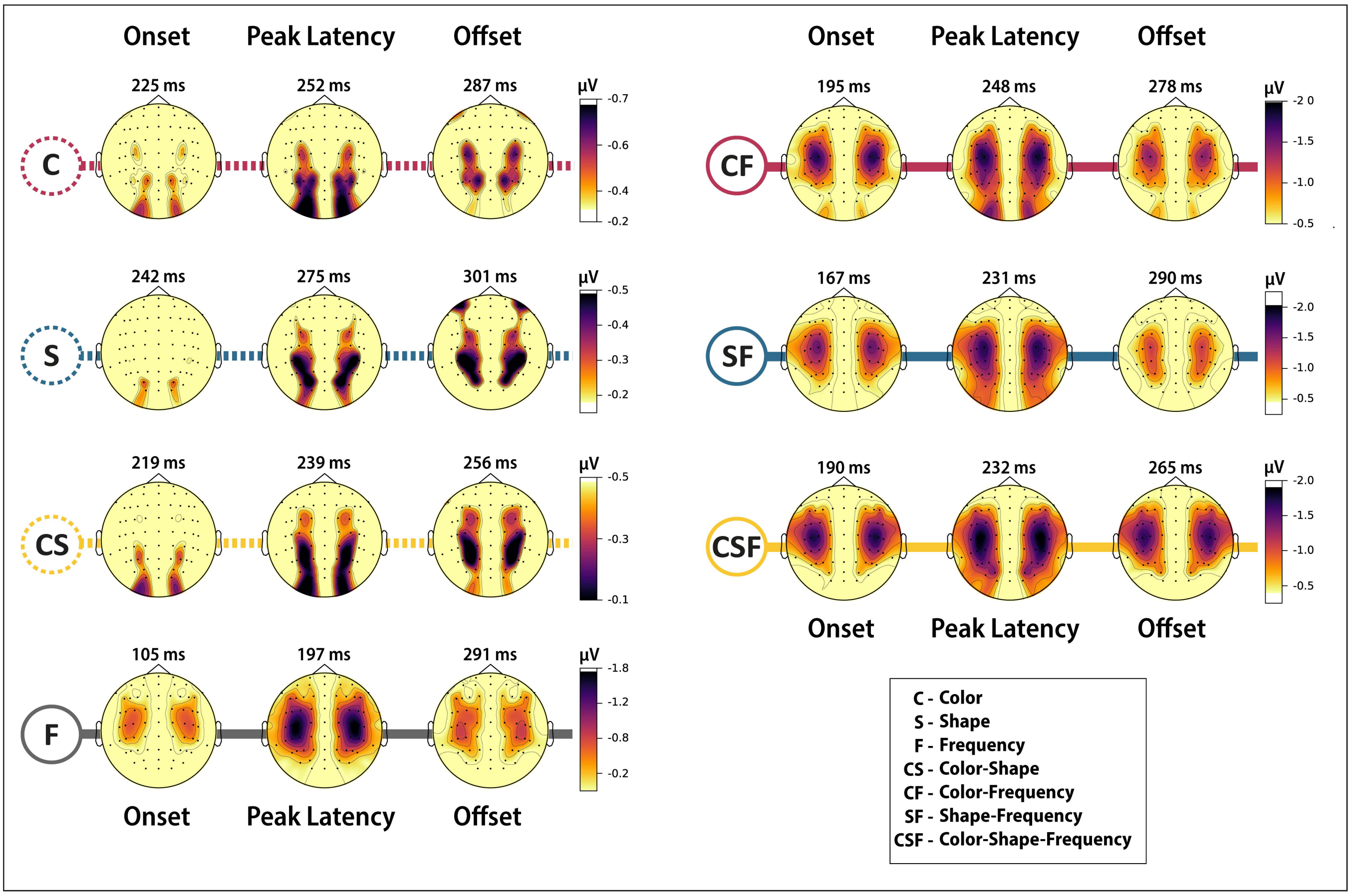

Figure 2–8. Temporal evolution of the topomaps for the seven targets, from onset through peak latency to offset, within the range of half-amplitude (in the PCN) to (above) peak-amplitude (PCN or CCN, depending on which amplitude is more prominent). Overall, a visual PCN (C, S, CS), activating parieto-occipital regions can be discerned from a tactile CCN (F) over central regions.

Linear Mixed-Effect Models. Finding similar patterns of hierarchical redundancy gains in reaction times and PCN dynamics, pointing to crossmodal coactivations in attentional-priority coding, we went on to examine (in a final analysis) how well the PCN itself predicts behavioral (RT) performance. To this end, we applied linear mixed-effect modeling (LMM), by stepwise including PCN *onset* (ms), *amplitude* (μV), *latency* (ms), and *offset* (ms). To control for the different scales (ms, μV), we applied z-normalization. Generally, this procedure was applicable due to the inverse transformation of jackknife scores, yielding individual estimates. Importantly, we omitted the tactile target (F) from the LMM analysis, since it produced a prominent bias owing to its unique PCN characteristics (having a very early onset and a loose offset). Examined *target*-based, we found three prominent predictors of RT performance, in PCN onset, amplitude, and latency (*p*s < .001), while offset contributed at only marginally (*p* < .083) – with the dynamics in the PCN explaining 19.8 % of the variance in the reaction times in total. The *category*-based model performed even better, explaining 25.5% of the RT variance, by revealing PCN onset, amplitude, and offset as prominent predictors (*p*s < .001) (see Appendix Table 2–A3). – In summary, the holistic dynamic in the PCN, reflecting a hierarchy in perceptual and preattentive processing in line with the MWA, turned out to be a good predictor of performance.

Discussion

Previous studies have provided evidence of redundancy gains in visual (Krummenacher et al., 2001b; Töllner, Zehetleitner, Krummenacher, et al., 2011) and, respectively, multisensory target detection (Diederich & Colonius, 2004; Forster et al., 2002; Miller, 1982). Thus far, however, the gains deriving from cross-dimensional (within the same modality) and, respectively, crossmodal redundant target definition have been studied separately. The present study combined both aspects within the same, crossmodal search paradigm in order to examine whether redundant-signal integration across (the visual color and shape) dimensions and across (the visuo-tactile) modalities operate at the same or distinct hierarchical levels in the functional architecture of search guidance.

The theoretical framing of the study was provided by Töllner and colleagues (2009) in their "modality-weighting account" (MWA) of crossmodal search, which is an extension of the "dimension-weighting account" (DWA; e.g., Found & Müller, 1996) of visual search. Both accounts assume a distinct hierarchy of perceptual and preattentive signal coding that precedes attentional selection, where information from separable lower-level mechanisms of feature (-contrast) coding is integrated, in a weighted fashion, onto on overall-saliency map which guides the allocation of focal attention, more recently referred to as 'attentional-priority' map (Fecteau & Munoz, 2006). However, the architectures assumed by the two accounts differ in one crucial aspect: whereas the DWA would consider dimensional (feature-contrast) maps to be sufficient to account for crossmodal signal integration, the MWA assumes an extra, modality-specific level of saliency signaling, that is: modality-specific saliency maps, whose signals are then integrated, (again) in a weighted fashion, onto a crossmodal attentional-priority map (see Figure 2–1). Accordingly, the main focus of the present study was on the necessity of assuming this extra hierarchical level in (crossmodal) search guidance. To achieve this, a crucial prerequisite had to be met: we had to, and did, ensure that the non-redundantly defined targets (C, S, F) were of comparable physical saliency (as assessed by the time taken to detect them with comparable accuracy), so that all (separable) target signals had an equal chance to influence detection when they were combined in redundantly defined targets. Further, to strengthen the validity of our findings, we conducted both *target*- and *category*-based analyses following the hierarchical classification deriving from the MWA (see Figure 2–2).

In behavioral performance, we found striking benefits for redundant targets in both error rates (ERs) and reaction times (RTs). Since the ERs were generally quite low (near ceiling) for redundant targets (2–3%), they would have been expected to be less sensitive to depicting the hierarchy of redundancy gains. Yet, finding fewer errors for triple compared to visual redundant targets supports previous reports of ‘pop-out’ target detection becoming more accurate by generating an additional feature-contrast signal (in another, crossmodal dimension) (Diederich & Colonius, 2004; Wolfe et al., 1989). For the search RTs, we found increasing redundancy gains from purely (dual) visual (VR) through (dual) crossmodal (CR) to triple (TR) redundant targets, as predicted by the MWA. Crucially, CR targets generated extra gains relative VR targets even given that the respective target signals combined together in the various redundant-target definitions were well equated in their baseline feature-contrast levels; also, we were able to rule out a potential bias from imbalanced feature combinations, by showing that performance modulations by redundant color (CS, CF), shape (CS, SF), and frequency (CF, SF) signals showed exactly the same pattern, namely: favoring crossmodal combinations (see Figure 2–3 & Appendix Figure 2–A1). These effect patterns argue strongly in favor of the MWA. Additionally, finding TR to exceed CR suggests that redundancy gains may not yet have reached their limit; that is, performance might potentially increase further (asymptotically) by adding additional feature contrasts. Overall, this pattern was evident in both *target*- and *category*-based analyses, without confounding speed-accuracy trade-offs: performance accuracy and speed were positively correlated.

While the pattern of behavioral redundancy gains strongly favors the MWA, the RT gains breach the bounds of the race-model inequality only for SF targets; for all other redundant targets, there was no strong evidence supporting coactivations. Nonetheless, not finding significant violations of the race-model inequality does not preclude coactivations, as the test is rather conservative. Interestingly, indirect evidence of coactivations was provided by the presence of significant cross-trial repetition-benefit and redundancy-cost effects (Liesefeld et al., 2017; Miller, 2016), which can be attributed to dynamic adjustments of the integration weights associated with particular target feature signals (Krummenacher et al., 2001b). While feature repetitions on consecutive trials were generally associated with benefits, costs arose when the critical feature of a non-redundant target was included in a redundant target on the preceding trial. Accordingly, the biases we found among the various target feature signals argue against the assumption of independent racers. This is consistent with the results of another coactivation test developed by Colonius and Diederich (2017), which assumes maximum negative dependency across feature contrasts as criterion instead of an independent race. When we applied this test to the present RT data, we found evidence of coactivations (i.e., significant breaches of the maximum negative dependency threshold) by redundant targets or target categories for SF, CSF, CR, and TR (while tests of the race-model inequality had only revealed a violation for SF), but not for CS and CF. Along with the evidence from cross-trial repetition-benefit and cross-trial redundancy-cost effects, these findings support the idea of visuo-tactile coactivation (rather than mere statistical facilitation) in the computation of attentional-selection priorities (see Figure 2–3).

Thus far, we have considered the behavioral evidence favoring the MWA in accounting for the pattern of redundancy gains in crossmodal search. Of note, though, the MWA provides a mere functional model of the hierarchy of signal-coding and -integration stages that precede the final stage of attentional target selection, without specifying the 'where' and 'when' of the critical distinction between dimension- and modality-specific saliency maps in neural terms. To achieve this, we recorded the EEG in our study and, in the analyses, focused on the PCN and CCN: two lateralized components thought to reflect the transition from pre-attentive signal processing to attentional selection.

In terms of the CCN, our results revealed that it is a stable somatosensory component over central regions that is sensitive to vibro-tactile target information, similar to previous findings (Eimer & Driver, 2000; Forster et al., 2016; Katus et al., 2015). However, importantly, we found no redundancy gains for targets defined by a combination of a tactile and one or two visual features: both the CCN onset timing and amplitude were comparable among the various tactile (F) and visuo-tactile (CF, SF, CSF) targets conditions (see Figure 2–4). The CCN appears to be modality-specific for touch: it emerges at approximately 110 ms and temporally precedes the PCN, thereby precluding crossmodal interactions. We only observed a weak late (within 300–400 ms) crossmodal influence on the (tactile) CCN from relatively 'delayed' (visual) PCN activity projected onto central regions in the CCN latency (main effect only with 12-Hz targets) and CCN mean amplitudes, likely because this component is impacted by latency jitter. More research is needed to confirm whether the CCN is truly insensitive (uninfluenced by linear superposition or spreading activation) to crossmodal activity, especially in its later stage, as it is a relative 'new' ERL component. Although we counterbalanced motor-response preparation during half-time break, it would be of interest in future work to examine the relationship between the CCN and these processes in more detail, as the sLRP (i.e., stimulus-locked lateralized readiness potential) starting around 250 ms (Miller et al., 1998; Töllner, Zehetleitner, Krummenacher, et al., 2011; Ulrich & Miller, 2001) may also be a source of redundancy gains (Feintuch & Cohen, 2002; Miller, Beutinger, et al., 2009; Töllner, Zehetleitner, Krummenacher, et al., 2011). Recall that we introduced only one vibro-tactile feature (F) in the present study, so that it cannot be ruled out that the CCN is sensitive to tactile redundancies, that is, combinations of multiple tactile features.

Regarding the PCN, we found prominent and consistent effects in both target- and category-based analyses. In particular, there was a crossmodal effect in both PCN *onset* and *amplitude*: with CR and TR targets, the PCN was elicited much earlier and was more negative compared to VR and NR targets. Further, the PCN *latencies* were generally shorter for redundant (including VR) targets, and the PCN *offsets* occurred earlier with visual redundancies (VR, TR). While the enhanced PCN amplitude only indicates a crossmodal amplification, there were two temporal effects within the time course of the PCN: earlier *onsets* with crossmodal redundancies, and earlier *offsets* with visual redundancies.

Our novel finding that the PCN is sensitive to crossmodal, visuo-tactile information expands its previously known visual scope (Eimer, 1996; Töllner, Zehetleitner, Krummenacher, et al., 2011; Woodman & Luck, 1999) to include some form of multisensory signal integration (in contrast to the CCN, which we found to be influenced only by tactile information). To confirm this interpretation, we examined an alternative explanation based on the distinct time courses of the earlier CCN and the later PCN, namely, that the crossmodal benefits simply derive from the overlay or linear superposition of these independent neural generators. By comparing the PCN onsets for redundant targets and categories with their respective feature sums, we found clear temporal benefits for *crossmodal* targets (in particular: SF and CSF) and categories (CR and TR) that exceeded the limit of their linear superposition, supporting the operation of neural crossmodal coactivations at the stage of attentional-priority computation[3]. It's worth noting that the pattern of coactivations driving the PCN onsets exactly mirrored the pattern revealed by the redundant-signal enhancement test for the RTs.

Prior studies suggest that extrastriate regions like the intraparietal sulcus (IPS) and temporo-parietal junction (TPJ) are involved not only in modality-specific, but also supramodal activations within the time window of the PCN – thus likely playing a role in crossmodal attention (Chambers et al., 2007; Chambers, Stokes, et al., 2004; Downar et al., 2000; Macaluso et al., 2000, 2002a, 2002b). Accordingly, the PCN originating within parietal regions (Hopf et al., 2000; Tay et al., 2022) might be a suitable candidate reflecting these crossmodal processes. By positively establishing crossmodal gains in the timing (onset) and amplitude of the PCN, we can interpret our primary effects in terms of the hierarchical functional architecture assumed by the MWA. In particular, the fact that there are substantial crossmodal benefits in PCN onsets would argue in favor of an architecture assuming 'modality maps', rather than just 'dimension maps'. Thus, the crossmodal (and likely coactive) PCN modulation provides a strong piece of evidence (in addition to that from behavioral performance) that the MWA provides an apt account of crossmodal redundancy gains.

Of note, we also found a PCN elicited by purely tactile (F) targets (see Figure 2–6). This tactile PCN appeared slightly delayed and reduced relative to the central CCN, which might indicate spreading activations or traveling waves (e.g., Klimesch et al., 2007; Nunez et al., 2001; Schack et al., 2003; A. von Stein & Sarnthein, 2000). However, since the tactile PCN occurred earlier compared to all visual and visuo-tactile PCNs (122 ms vs. 167–242 ms), the tactile PCN seems to fall into different time frames of sensory processing and visual integration (Bastos, Vezoli, & Fries, 2015; Bastos, Vezoli, Bosman, et al., 2015; Fries, 2005). Interestingly, we did not find differences in the PCN *onsets* for visual redundant (VR) vs. non-redundant (NR) targets. Our non-finding may be owing to the fact that, in our crossmodal search paradigm, crossmodal PCN effects superseded the underlying visual effects, particularly in the *onsets* and *amplitudes*. Note that there was still a temporal effect of visual redundancies in the PCN *latencies* and, most prominently, in the *offsets*. Thus, taken together, the crossmodal and visual effects point to a temporal evolution of the PCN, which starts being sensitive to crossmodal redundant targets and only later to visual redundancies. In terms of the MWA, this suggests that crossmodal targets generate an amplified and earlier PCN which exceeds visual redundancies, thus boosting selective attention. An alternative explanation why PCN *offsets* were sensitive to visual redundancies may derive from the notion that, besides the PCN, visual attention is linked to another component: the so-called distractor positivity (Berggren & Eimer, 2018; Hilimire et al., 2009, 2010), which runs parallel to the PCN, peaking within 290–340 ms, and is thought to be indicative of active (visual) distractor suppression. Accordingly, the offset benefit for visual redundancies (VR, TR) might derive from a more efficient feature-contrast discrimination between redundant targets and distractors (fuchsia square [target] vs. blue circles [distractors]) as compared to non-redundant targets (color: fuchsia circle [target] vs. blue circles [distractors]; shape: blue square [target] vs. blue circles [distractors]). Furthermore, this visual effect might also account for the noise in, or seeming absence of, the tactile PCN offset (see Table 2–A1 and Figure 2–6), as the visual discrimination cannot come into play with purely tactile targets (in which case all visual items are uniform blue circles). Finally, the dynamics of visual discrimination in the PCN offset can be linked to visual 'dimension' maps (or a supradimensional visual 'modality' map), as it appears to be modality-specific for vision.

The CCN and PCN have distinct onsets, with the CCN occurring before the PCN by at least 60 ms. This difference in timing is similar to the modality-specific timing of the somatosensory N1 and the visual N1 (Eimer et al., 2004; Eimer & Driver, 2000; García-Larrea et al., 1995; Mangun, 1995; Novitskiy et al., 2011). While ERPs (sensory-perceptual processing) and ERLs (attentional processing) provide different information about neural activity, the temporal profile of the CCN, PCN, and N1 all appear to be consistent with each other (CCN < PCN, and tactile N1 < visual N1). The PCN might reflect a stage of multisensory integration that begins around 180 ms in crossmodal sensitive parietal regions (Busse et al., 2005; Man et al., 2015; Quinn et al., 2014; Talsma & Woldorff, 2005).

The present findings, and the insights gained from them, raise a number of interesting questions for future research. In particular, it would be interesting to determine the (asymptotic) limit of redundancy gains by either adding more redundant target signals within a given modality and/or more modalities and fitting exponential approximations. Further, and arguably the most interesting question concerns whether the PCN reflects a suitable neural marker enabling the investigation of coactivation in redundancy gains for the preattentive stage, contributing to the ongoing debate at which stage (pre-attentive vs. post-selective) redundancy gains occur (cf. Feintuch & Cohen, 2002; Zehetleitner et al., 2009). Although our study provides consistent evidence in favor of (preattentive) coactivations within the PCN onsets, it remains crucial to corroborate this finding across various EEG paradigms. Finally, going beyond the clear demonstration of crossmodal redundant benefits (over and above visual redundant benefits) in the present study, it remains crucial to show that the capacity of crossmodal attention exceeds the limits of unisensory attention (e.g., visual attention).

We conclude that the PCN component reflects processes of crossmodal attentional selection; in particular, it is indicative of an attention-guiding (i.e., preattentive) stage of multisensory coactive signal integration, consistent with the MWA: the PCN exhibits the most prominent and early effects with crossmodally redundant targets, followed later on by modality-specific effects with visual redundant targets. As conceived by the MWA, the output of this stage is a map of crossmodal (visuo-tactile) attentional-priority signals. That is, the activations achieved by the stimuli on this map are inherently ‘feature-blind’ (e.g., Wolfe, 2021): they no longer carry information about the specific visual and tactile features that constitute the stimuli in the display array. Accordingly, focused attention is required to extract the features of selected items (by backtracking to the feature-coding levels) and combine them into integrated, multisensory objects (cf. Treisman & Gelade, 1980). The PCN is a good predictor of reaction times, explaining around 25% of the variability in reaction times in a category-based analysis and around 20% in a target-based analysis. This supports the idea that the preattentional processes – of ‘predictive coding’ under conditions of uncertainty – play an important role in conscious perceptual decisions and decision-making (van Bergen et al., 2015; van Bergen & Jehee, 2019). The parietal PCN plays a significant role in regulation of attentional selection across all modalities. However, the PCN appears to only reflect processes of crossmodal integration when it is combined with visual processing.

Appendix

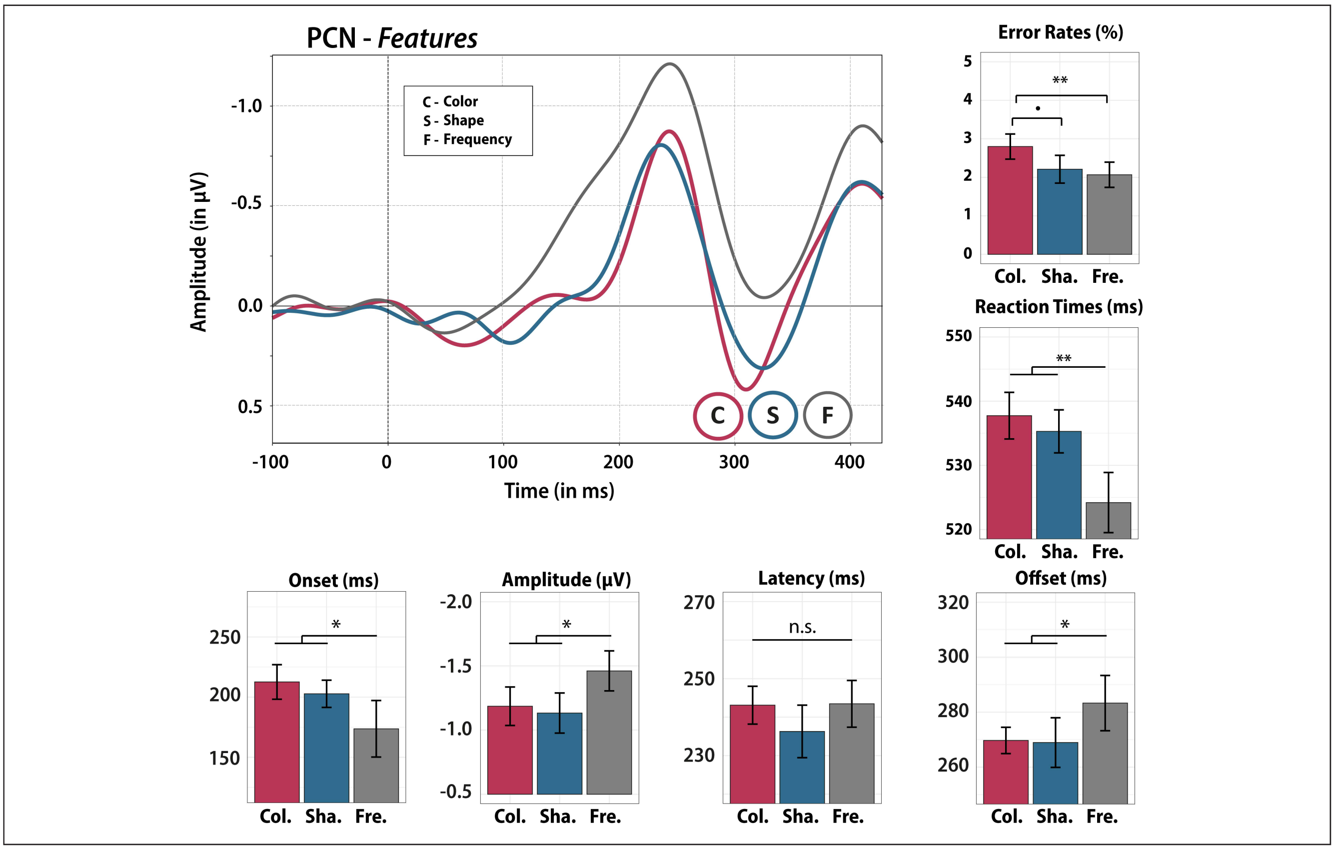

Figure 2–A1. The PCN results are shown for *features* in redundant *color* (CS, CF), *shape* (CS, SF), and *frequency* (CF, SF) over parieto-occipital regions (PO7/ PO8, PO3/ PO4, O1/O2). PCN grand averages (< 12 Hz) are shown (top). Furthermore, bar plots for ERs and RTs are shown on the right panel, and PCN dynamics in *onset*, *amplitude*, *latency*, and *offset* (bottom), along with within-subject confidence intervals. Here, prominent findings in post hoc comparisons (using Bonferroni and FDR correction, respectively) are indicated by asterisks (• p < .1, *p < .05, **p < .01). For *frequency*, RTs and onsets were faster and amplitudes are more negative; whereas PCN latencies were comparable, and PCN offsets were earlier for *color* and *shape*.

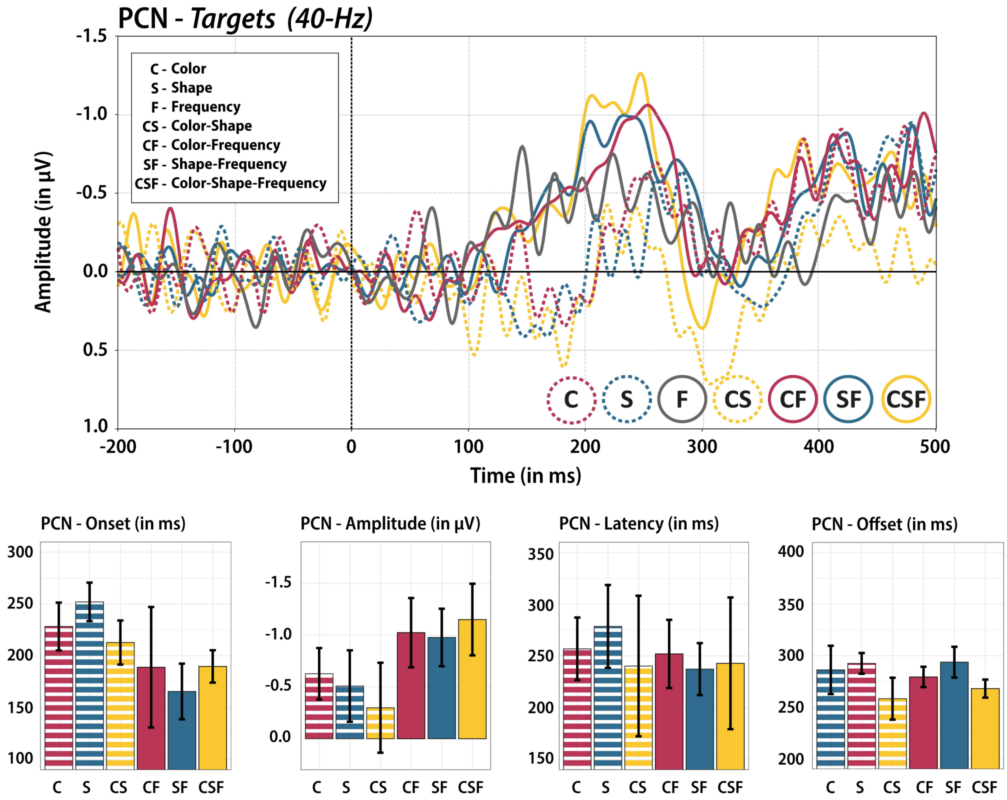

Figure 2–A2. The PCN results are shown for targets (C, S, F, CS, CF, SF, CSF) over parieto-occipital regions (PO7/ PO8, PO3/ PO4, O1/O2). PCN grand averages (40 Hz) are shown (top) for targets, including a purely tactile PCN (F, gray line), and features. Furthermore, bar plots are shown for *onset*, *amplitude*, *latency*, and *offset* (bottom), along with within-subject confidence intervals. From post hoc comparisons, onsets and amplitudes generally revealed cross-modal benefits and redundancy gains, occurring earlier and being more negative for SF and CSF compared with C, S, and CS. No difference was found in PCN latency (likely because of high variances), and PCN offsets occurred earlier for CS and CSF compared with S, CF, and SF.

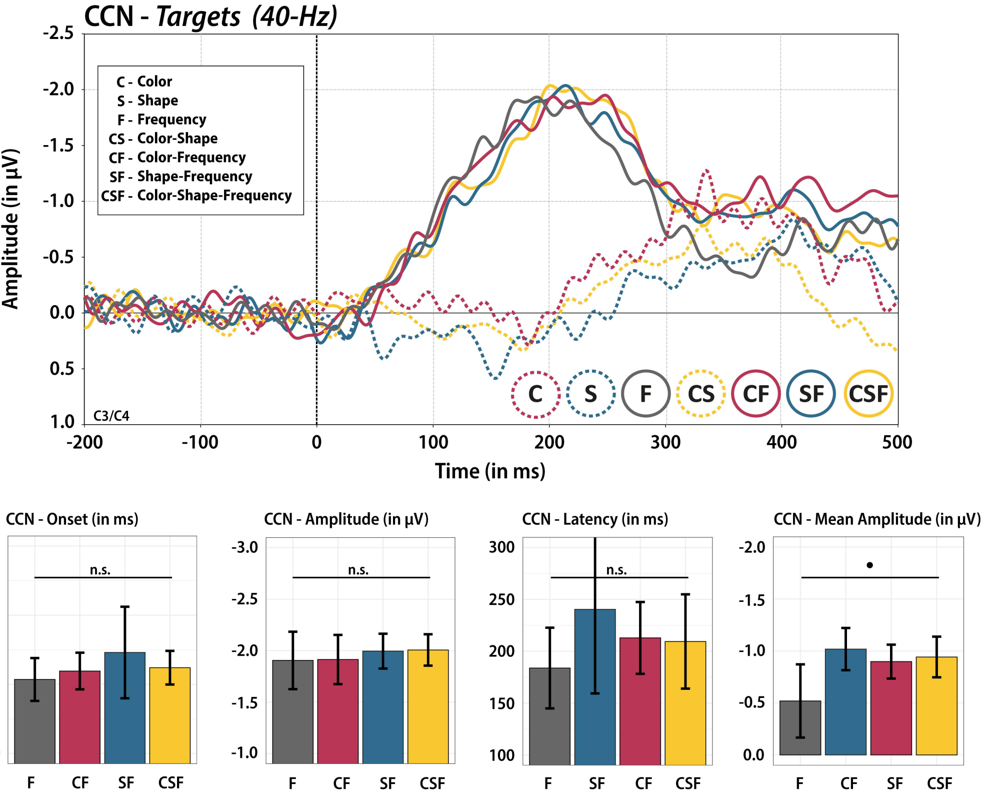

Figure 2–A3. The 40-Hz CCN is shown over central regions (C3/C4) for the seven targets (C, S, F, CS, CF, SF, and CSF). Naturally, only targets including a tactile signal (F, CF, SF, CSF) elicited a prominent CCN, in comparison with solely visual targets (C, S, CS). Whereas the CCN was comparable in onsets, amplitudes, and latencies (nonsignificant, n.s., comparisons) among the F, CF, SF, and CSF targets, there were at best marginal differences in 300- to 400-msec mean amplitude in post hoc comparisons (FDR; • $p < .1$). Thus, the tactile CCN reflects a stable somatosensory component that is not sensitive to cross-modal, visuo-tactile redundancies.

Table 2–A1. Forty-hertz PCN Results Presented Target- (Upper Part) and, Respectively, Category-dependent (Lower Part), for Onset, Latency, Offset, and Jackknife and Raw Peak Amplitude, and 200- to 300-msec Mean Amplitude (Means and, in Parentheses, Associated Within-subject Confidence Intervals)

PCN	*C*	*S*	*F*	*CS*	*CF*	*SF*	*CSF*	*F Value*	η^2
Onset (msec)	228 (24)	252 (19)	148 (163)	213 (22)	189 (60)	166 (28)	190 (16)	4.22*	.171
Latency (msec)	257 (31)	278 (41)	205 (269)	240 (70)	252 (34)	237 (26)	243 (66)	0.44	.021
Offset (msec)	286 (24)	293 (10)	253 (277)	258 (21)	279 (10)	294 (15)	268 (9)	3.38*	.121
Peak amplitude (μV)	−0.63 (0.26)	−0.51 (0.35)	−0.62 (0.40)	−0.30 (0.45)	−1.02 (0.34)	−0.98 (0.29)	−1.15 (0.36)	4.10*	.093
Mean amplitude (μV)	−0.33 (0.23)	−0.25 (0.23)	−0.43 (0.27)	0.00 (0.39)	−0.70 (0.27)	−0.76 (0.20)	−0.71 (0.20)	5.04**	.147
Raw peak amplitude (μV)	−1.37 (0.29)	−1.16 (0.23)	−1.17 (0.33)	−1.02 (0.31)	−1.57 (0.27)	−1.53 (0.21)	−1.74 (0.32)	4.17*	.096
Raw mean amplitude (μV)	−0.33 (0.23)	−0.25 (0.23)	−0.43 (0.27)	0.00 (0.39)	−0.70 (0.27)	−0.76 (0.20)	−0.71 (0.20)	5.05**	.147
								df (5, 80)	

PCN	*NR*	*VR*	*CR*	*TR*	*F Value*	η^2
Onset (msec)	235 (31)	213 (15)	172 (38)	189 (12)	4.53*	.176
Latency (msec)	262 (32)	240 (64)	241 (25)	242 (71)	0.16	.007
Offset (msec)	291 (10)	258 (20)	283 (8)	268 (10)	5.57*	.172
Peak amplitude (μV)	−0.53 (0.16)	−0.30 (0.42)	−0.97 (0.23)	−1.15 (0.40)	6.25**	.118
Mean amplitude (μV)	−0.30 (0.17)	0.00 (0.40)	−0.73 (0.16)	−0.71 (0.37)	6.01*	.167
Raw peak amplitude (μV)	−1.07 (0.20)	−1.02 (0.31)	−1.34 (0.20)	−1.74 (0.32)	6.59**	.132
Raw mean amplitude (μV)	−0.30 (0.17)	0.00 (0.40)	−0.73 (0.16)	−0.71 (0.37)	6.03*	.167
					df (3, 48)	

Furthermore, the results of the repeated-measures ANOVAs are reported as *F* tests, with the effect sizes (generalized η^2), with degrees of freedom (*df*) in targets: (5, 80) and categories: (3, 48). In addition, the descriptive result for the tactile-only target (F) is presented (in gray) but omitted for statistics.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 2–A2. Results for the Redundant-signals Enhancement Tests in Behavior's RTs (in msec) Are Shown at the Top, Including Mean and Standard Deviation (in Brackets) for Redundant Targets and Their Respective Feature-sum (e.g., SF vs. S + F). Further One-tailed t Statistic and p Values, and Degrees of Freedom (df)

RTs	*Redundant Target*	*Feature-Sum*	*t Statistic*	*p*
CS vs. C + S	554 (52)	561 (24)	−0.56	.292
CF vs. C + F	531 (59)	550 (31)	−1.91	.126
SF vs. S + F	524 (50)	561 (31)	−2.53	.011*
CSF vs. C + S + F	510 (50)	553 (22)	−3.26	.002**
CR vs. NR + F	527 (54)	556 (31)	−1.91	.037*
TR vs. NR + F	510 (50)	556 (31)	−3.2	.002**
TR vs. VR + F	510 (50)	525 (27)	−1.07	.15
			df (16)	

12-Hz PCN Onsets	*Redundant Target*	*Feature-Sum*	*t Statistic*	*p*
CS vs. C + S	219 (8)	227 (8)	−1.165	.131
CF vs. C + F	195 (27)	212 (27)	−0.719	.241
SF vs. S + F	167 (16)	195 (16)	−1.869	.04*
CSF vs. C + S + F	190 (6)	214 (6)	−4.557	<.001***
CR vs. NR + F	173 (14)	206 (14)	−2.686	.016*
TR vs. NR + F	190 (8)	206 (14)	−3.749	<.001***
TR vs. VR + F	190 (8)	208 (8)	−2.639	.009**
			df (16)	

40-Hz PCN Onsets	*Redundant Target*	*Feature-Sum*	*t Statistic*	*p*
CS vs. C + S	213 (14)	234 (14)	−1.796	.091
CF vs. C + F	189 (34)	213 (34)	−0.796	.438
SF vs. S + F	165 (19)	199 (19)	−2.012	.061*
CSF vs. C + S + F	190 (7)	214 (7)	−4.013	.001**
CR vs. NR + F	172 (22)	207 (22)	−1.829	.086*
TR vs. NR + F	190 (7)	207 (7)	−2.764	.007**
TR vs. VR + F	190 (9)	208 (9)	−2.881	.036*
			df (16)	

Similarly, 12-Hz and 40-Hz PCN onsets (in msec) are shown with mean, within-subject confidence intervals in brackets, *T* statistic, and *p* values.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 2–A3. Results of the Target- and, Respectively, Category-based LMMs

	PCN	*AIC*	*BIC*	*R^2*	*Estimate*	*SE*	*df*	*t Statistic*	*p*	*95% CI Bounds* *Lower*	*Upper*
Targets	(Intercept)				0	0.19	81	0	1	−0.37	0.37
	Onset	207	218	0.07	0.17	0.06	81	2.89	.005	0.06	0.28
	Amplitude	200	213	0.129	0.25	0.07	81	3.67	<.001	0.12	0.39
	Latency	189	205	0.175	0.16	0.06	81	2.53	.013	0.04	0.29
	Offset	188	206	0.198	0.11	0.07	81	1.75	.084	−0.01	0.24
Categories	(Intercept)				0	0.21	47	0	1	−0.4	0.4
	Onset	147	156	0.088	0.24	0.07	47	3.31	.002	0.1	0.38
	Amplitude	145	157	0.126	0.31	0.09	47	3.43	.001	0.13	0.48
	Latency	141	154	0.175	−0.05	0.11	47	−0.5	.618	−0.26	0.15
	Offset	132	148	0.255	0.35	0.1	47	3.38	.001	0.15	0.55

AIC, BIC, and (marginalized) R^2 report the change in the models' explanatory power when stepwise including the PCN onset, amplitude, latency, and offset parameters. Furthermore, the summary of these four predictors is shown with its (standardized) estimate, standard error (*SE*), degrees of freedom (*df*), *t*-statistic, *p* value, and 95% confidence interval.

Acknowledgments

We thank Hans Colonius for recommending the redundant-signals enhancement test.

Reprint requests should be sent to Jan Nasemann, Allgemeine und Experimentelle Psychologie, Department Psychologie, LMU München, Leopoldstr. 13, D-80802 Munich, Germany, or via e-mail: Jan.Nasemann@psy.lmu.de.

Data Availability Statement

Our data will be made available upon request via email to the lead author (e-mail address: Jan.Nasemann@psy.lmu.de).

Funding Information

This research was supported in part by a German Science Foundation [Deutsche Forschungsgemeinschaft] (https:// dx.doi.org/10.13039/501100001659), grant numbers: TO940/1-1, TO940/1-2 awarded to T. T., H. J. M., and Z. S.

Notes

1 To ensure comparable reaction times (RTs, ms) and error rates (ERs, %) for the non-redundant – *color*, *shape*, and *frequency* – *feature contrasts* introduced in the main experiment, we conducted two pilot studies testing targets (blockwise) in simple (target-present vs. -absent) detection tasks using the same experimental setup. Overall, visual stimuli were kept isoluminant (mean luminance of ~36 cd/m^2) and the amplitudes of the tactile target and distractor frequencies were similar. In an initial pilot study (N=8 observers), two (blue: RGB [0 0 255]) visual target *shapes* among blue circle distractors and six tactile target-distractor frequency combinations (in Hz) were tested, which yielded the following ERs and RTs: *square*: 6%, 570 ms; *triangle*: 14%, 596 ms; *100-30*: 7%, 521 ms; *150-30*: 10%, 549 ms; *200-30*: 14%, 531 ms; *100-40*: 9%, 567 ms; 150-40: 13%, 531 ms; 200-40: 23%, 582 ms. From these targets, the blue *square* and the *100-40* frequency pair (target: 100 Hz, distractor: 40 Hz) were selected as feature contrasts for *shape* and *frequency* targets, as these produced the most similar performance measures (ERs, p = .057; RTs, p = .93). In the second pilot study (N=10 observers), we included five colors from the red-to-blue spectrum: *square*: 5%, 547 ms; *100-40*: 10%, 618 ms; *color1* [106 0 115]: 5%, 519 ms; *color2* [104 0 124]: 4%, 530 ms; *fuchsia* [101 0 135]: 6%, 567 ms; *color4* [98 0 147]: 8%, 568 ms ; *color5* [94 0 162]: 10%, 558 ms). Of these, *square* (shape), *fuchsia* (color), and *100-40* (vibro-tactile frequency) were most comparable in terms of ERs (ps > .054) and RTs (ps > .102) – that is, in terms of behavioral measures of bottom-up saliency (cf. Zehetleitner et al., 2013). These three *feature contrasts* were the introduced as basic target features in the main experiment.

2 To confirm that our findings genuinely reflect processes of covert attentional orienting rather than overt saccadic activity, we analyzed the percentage of saccades in epochs that occurred during the critical 150–350-ms period, which includes the CCN and PCN time window. Since we did not directly record the horizontal eye movements, we examined for saccadic activity in left (F9, AF7, FT9) and right (F10, AF8, FT10) electrode channels (which are common locations for the corneo-retinal dipole), using a 20-90 Hz band-pass filter (Keren et al., 2010) and Hilbert transform to obtain the envelope power, before applying the inverse ICA transformation. Given reduced ocular signals at the 'frontal' channels, we classified saccadic activity as bilateral activity exceeding a fixed threshold criterion of 10 µV, instead of 20 or 30 µV for the horizontal EOG channels (cf. Drisdelle & Eimer, 2021; van Moorselaar et al., 2020). This procedure revealed saccadic activity to be present in an average of 4.85% (SD: 3.72%) of epochs, with a range of 0.3–12.1% across participants. Given this, and the fact that saccadic events were randomly distributed across our experimental conditions, our reported PCN (and CCN) findings can be taken to reflect processes of covert (rather than overt) attention.

3 However, it is important to note that our failure to find significant temporal benefits for redundant CS and CF targets (over and above their feature sums) does not exclude the possibility of an explanation in terms of linear superposition for such targets.

3

Spatial Attention in Crossmodal Search

And thus, while searching,
What and where becomes
that over there.

Crossmodal Bending in Spatial Attention

Jan Nasemann [1,2], Hermann J. Müller [1,2], Zhuanghua Shi [1,2]

1 General and Experimental Psychology, Department of Psychology,
Ludwig-Maximilians-Universität München, Munich, Germany
2 Graduate School of Systemic Neurosciences,
Ludwig-Maximilians-Universität München, Planegg, Germany

Abstract

During bottom-up saliency summation, selective attention elegantly combines feature-based and location-based information. However, these processes differ between sensory modalities. While visual attention is optimal in focal or foveal areas, it decreases toward the periphery; known as eccentricity effect. On the contrary, tactile attention follows the body-schema and is equally distributed across fingers. This study examines the spatial properties of redundancy gains by analyzing the spatial slopes ($\beta 0$, $\beta 1$, $\beta 2$) from quadratic fitting for external and quadratic vision and egocentric or somatotopic (constant) touch. As expected, significant quadratic fits were found for all targets containing visual information, and a flat function for purely tactile targets. Notably, the visual redundant color-shape target remained within the spatial limits of focal ($\beta 0$) color and peripheral ($\beta 2$) shape, while crossmodal redundant targets exceeded those limits, yet the visual pattern of spatial benefits in focal color and peripheral shape was repeated. Further, event-related lateralizations revealed a modality-specific pattern in $\beta 2$ slopes – showing a quadratic visual PCN and a constant tactile CCN – in line with external and egocentric properties. Also, there was a supramodal PCN mechanism with crossmodal amplification in $\beta 0$. The findings are discussed, and a spatial vector space is proposed with feature-contrast eigenvectors.

Introduction

Selective attention acts like a music conductor; it hits the tunes of our perceptual orchestration by prioritizing relevant features and locations – the *what* and *where* (e.g., Duncan, 1984; Logan, 1996; Wolfe, 1994). Although feature activations and spatial locations represent distinct processes in visual search (Bundesen, 1990; Hayden & Gallant, 2005; Luck et al., 2000; Luck & Hillyard, 1994a; Mangun & Hillyard, 1988), they both converge consistently during saliency summation and weighted integration onto the spatio-topographic *priority map* (Fecteau & Munoz, 2006; Found & Müller, 1996; Itti & Koch, 2001; Koch & Ullman, 1985; Lee et al., 1999; Mangun, 1995; Treisman & Gelade, 1980; Wolfe, 1994), and share common regions within the parietal cortex (Brefczynski & DeYoe, 1999; Corbetta & Shulman, 1998; Fink et al., 1997; Hopf & Mangun, 2000; Kanwisher & Wojciulik, 2000; McIntosh et al., 1994; Sereno et al., 2001).

In visual search, studies have shown that visual attention operates best in focal areas and declines toward the periphery as shown by decelerated reaction times; a phenomenon known as the *eccentricity effect* (Carrasco et al., 1995; Hansen et al., 2009; Jonas et al., 1992; Larson & Loschky, 2009; Staugaard et al., 2016; Wolfe et al., 1998). Further, the eccentricity effect is linked to the distribution of receptive fields being dense in focal regions and sparse in the periphery (e.g., Hansen et al., 2009), which can be described by linear or quadratic slopes (Carrasco et al., 1995; Föcker et al., 2010; R. Gray et al., 2009; Wolfe et al., 1998).

Given that we live in a multisensory world, spatial attention does not rely purely on visual information alone but also others, such as auditory or tactile information (Assumpção et al., 2018; Föcker et al., 2010; Heed & Röder, 2010; Holmes & Spence, 2004; Röder et al., 1999; Spence & McGlone, 2001; Teder-Sälejärvi et al., 1999). In contrast to vision, which represents an *external* (cortical) mapping, spatial orienting in touch is initially *egocentric* or somatotopic, reflecting an anatomical reference frame that follows the body schema, being most sensitive to the hands, the "'fovea of touch" (Assumpção et al., 2018; Eimer, Forster, et al., 2003; Forster et al., 2016; Holmes & Spence, 2004; Medina & Coslett, 2010), and subsequently merged with (visual) external coordinates into a crossmodal space (Driver & Spence, 1998; Eimer, Forster, et al., 2003; Graziano & Cooke, 2006; Kanwisher & Wojciulik, 2000; Spence, Pavani, et al., 2000). Although vision and touch activate distinct *modality-specific* brain regions for spatial integration within parieto-occipital and central regions (Chambers et al., 2007; Chambers, Payne, et al., 2004; Chambers, Stokes, et al., 2004; Eimer & Driver, 2000; Forster et al., 2016; García-Larrea et al., 1995), there is also evidence for *supramodal* activations in the intraparietal sulcus (IPS) and temporo-parietal junction (TPJ) (Diederich et al., 2003; Downar et al., 2000; Macaluso et al., 2000, 2002a, 2002b; Man et al., 2015; Quinn et al., 2014; Spence, Pavani, et al., 2000). Altogether these findings indicate a dual role for spatial attention in the parietal cortex, being either modality-specific in visual search or supramodal in crossmodal search (Chambers et al., 2007; Föcker et al., 2010; Spence & Driver, 1997a).

The crossmodal influence on spatial orienting has been demonstrated in several cueing paradigms, where preceding auditory or tactile cues (but not visual cues) improved the peripheral sight in vision (R. Gray et al., 2009; Lu & Dosher, 2000; Rosli et al., 2008; Spence & Santangelo, 2009). For example, Gray et al. (2009) analyzed the reaction times for the separation in various cue-target combinations by comparing their slopes via quadratic fitting. Further, the authors reported steep slopes for visual cue-targets, replicating the eccentricity effect with a narrow and focal distribution of spatial attention, the slopes in crossmodal cue-targets (e.g., auditory or tactile cue) were significantly more flat with increasing cue-target separation, thus effectively broadening spatial attention.

Similar findings of spatial decline have been reported in early (perceptual) EEG components, such as the P1, N1, or N2. Essentially, these components had higher amplitudes for attended locations that declined toward (peripheral) unattended locations, indicative of spatial prioritization due to endogenous or exogenous expectancy (Driver & Spence, 1998; Eimer & Driver, 2001; Eimer & Schröger, 1998; Föcker et al., 2010; Mangun, 1995; Mangun et al., 1987; Mangun & Hillyard, 1987). However, to describe the spatial orienting in crossmodal attention more accurately, e.g., in a crossmodal detection task (s. Chapter 2), event-related lateralizations, such as the PCN (or N2pc) in visual attention and the CCN (or N140cc) in tactile attention, acting as a spatial filter, might be more adequate (Eimer, 1996; Forster et al., 2016; Luck & Hillyard, 1994b; Wolber & Wascher, 2005; Woodman & Luck, 1999). While the PCN amplitude decreases from the fovea to the periphery, similar to the visual eccentricity effect and the distribution of receptive fields (Papaioannou & Luck, 2020; Schaffer et al., 2011), the CCN amplitude in hands, the ‘fovea of touch’, is maximized and presumably equal across fingers (Forster et al., 2016).

Given this empirical background, this study extends the preceding study on redundancy gains in visuo-tactile search by examining the underlying spatial functions across the seven visual, tactile, and crossmodal targets. Overall, the previous study provided strong evidence in favor of a modality-weighting account (MWA) as a suitable preattentive hierarchy for crossmodal search by showing that the redundancy gains in crossmodal redundant targets substantially exceed those visual redundant ones. This finding of crossmodal benefits was backed by accelerated reaction times and the neural dynamics of the PCN, revealing earlier onsets and higher amplitudes. Furthermore, evidence for the presence of coactivation is provided by the test of 'maximum negative dependency' (Colonius & Diederich, 2017) for two crossmodal targets (shape-frequency & color-shape-frequency). While the CCN reflects a stable somatosensory component for tactile target information, the PCN over parietal regions sensitively depicted the pattern of redundancy gains and was revealed to be a good predictor for performances. Also, the PCN onsets mirrored the (behavioral) coactivation pattern perfectly by exceeding the threshold of linear superposition, providing strong evidence for a preattentive origin (or locus) of coactivations (for a detailed description see Chapter 2).

Despite the convincing evidence in the previous study, the main focus was on aggregated targets, i.e., averages across the ten locations for each target in the visuo-tactile paradigm. For this reason, an in-depth analysis of the underlying spatial functions is missing, reflecting a scientific gap that this study aims to close.

Critically, this study examines the spatial functions across visual, tactile, and crossmodal targets by comparing the respective slopes from quadratic fits. Based on the modality-specific spatial functions – *external* and quadratic in vision, but *egocentric* and flat in touch –, the detection of the seven targets, i.e., the feature-contrast combinations in color, shape, and frequency, should reveal systematic modulations in focal and peripheral slopes. Essentially, we expect to find a quadratic spatial function for visual targets (color, shape, & color-shape) with a focal optimum and a decay toward peripheral locations that manifests in decelerated reaction times and decreased PCN amplitudes. On the contrary, the tactile target (frequency) should reveal a flat spatial function with comparable reaction times and CCN amplitudes across finger locations. Finally, we expect to find substantial interactions between spatial slopes across redundant targets. While the visual redundant target (color-shape) should remain within the *modality-specific* limits of visual attention, crossmodal redundant targets (color-frequency, shape-frequency, color-shape-frequency) should exceed those visual and tactile limits, providing a *supramodal* mechanism for the underlying redundancy gains. Ultimately, this study aims to provide evidence of whether spatial interactions already occur within the preattentive stage of multisensory integration, e.g., via crossmodal links in the parietal cortex (Chambers et al., 2007; Macaluso et al., 2002b) in PCN, or rather reflect maintain their independent external and egocentric spatial orienting in PCN and CCN, respectively.

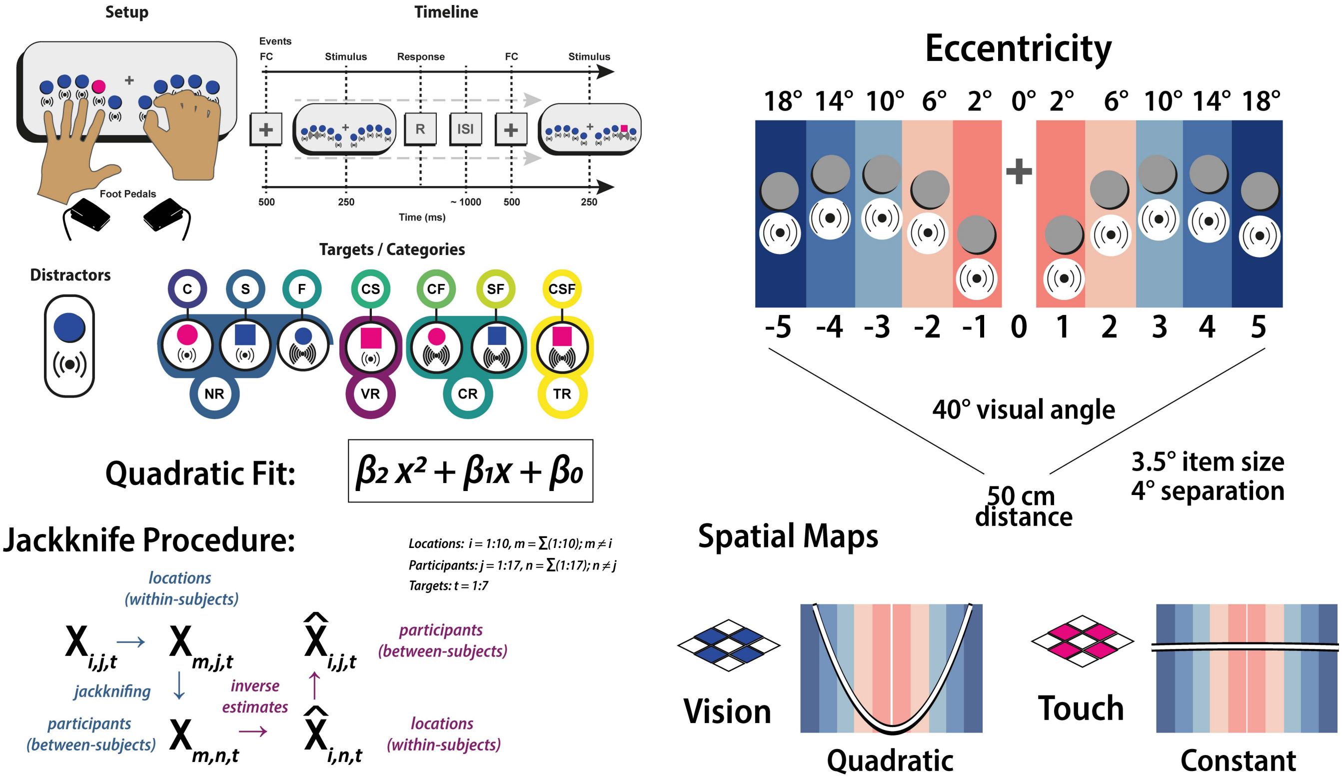

Figure 3–1. Here, the experimental setup is shown in the top left panel from the previous study on redundancy gains (for a detailed description s. Chapter 2). This study extends the findings on redundancy gains by analyzing the (spatial) slopes of quadratic solutions: $y = \beta_2 x^2 + \beta_1 x + \beta_0$. To analyze *targets* and *locations* sufficiently, a double jackknife procedure (bottom left panel) was conducted across (within-subject) *locations* and (between-subject) *targets* with subsequent double inverse transformation (Smulders, 2010). The top right panel shows the visuo-tactile display for each of the ten locations with their respective averaged eccentricity, i.e., visual angle, in degree (°) from central (focal) locations (red) to peripheral locations (blue). For the main analysis in behavior, PCN, and CCN, the symmetric locations were collapsed and mirrored. The bottom right panels shows the distinct spatial functions for vision and touch, being quadratic and constant respectively.

Methods

As mentioned above, this study (re-)analyzes and extends the previous findings by analyzing the spatial slopes. Therefore, everything remained the same (s. Chapter 2 Methods), e.g., participants, stimuli and procedure, EEG recordings, and preprocessing, except for the adjusted data analyses in performances and event-related lateralizations. Before we explain them in detail, it is important to address two important issues.

First, participants were seated in front of the visuo-tactile display, laying their fingertips on top of the tactile solenoids (s. Figure 3–1), while keeping their eyes fixated at the center with a distance of around 55 cm. Since no head fixation was applied, the distance of spatial locations varied during the task. Therefore, the visual angles can be described on average (40° display size, 3.5° item size, 4° inter-item separation), omitting within- and between-subject variations. Nonetheless, the validity is ensured by the fact that visual angle variations (e.g., head movements) reflect a (linear) noise term while preserving the constant external mapping of the ten locations.

Second, quadratic fitting was applied in line with other studies (R. Gray et al., 2009; Lu & Dosher, 2000) with the (common) formula: $y = \beta_2 x^2 + \beta_1 x + \beta_0$. By assuming a quadratic and symmetric function across spatial locations, importantly, we only expect the slopes of $\beta 2$ and $\beta 0$ to be relevant, reflecting the slope towards the periphery and the focal saddle point of central fixations. Although the formula contains a linear term with the $\beta 1$ slope, it (presumably) can be omitted since both spatial properties of the center and periphery are already covered by $\beta 2$ and $\beta 0$. Nonetheless, we evaluate the expected non-involvement of $\beta 1$ statistically by using one-sample t-tests.

Performance Analysis. Again, we only included correct trials having responses within 250–1200 ms, which on average excluded 2 % of trials as outliers. To test whether reaction times (RTs) and error rates (ERs) in the left and right spatial locations were congruent and symmetric, we applied paired t-tests for each of the seven targets. Initially, spatial locations were averaged across targets and participants, and symmetrically collapsed (s. Figure 3–1). Further, we applied repeated-measures (RM) ANOVAs, which include the factors of *targets* and *locations*. In general, violations of the sphericity assumption were controlled with Greenhouse-Geisser correction, and the generalized eta squared (η^2) is reported as effect size. Quadratic fitting was done by using the symmetric RTs for two solutions. We applied a *regular* fit, with $\beta 0$ as the saddle point in the center and $\beta 2$ as the gradient toward the periphery, and an *orthogonal* fit, where $\beta 0$ reflects the mean RT or centroid of the parabola. Further, the prominence of the slopes ($\beta 0$, $\beta 1$, $\beta 2$) was statistically evaluated by using RM ANOVAs and one-sample t-tests. To reveal modality-specific or supramodal patterns, the spatial slopes $\beta 2$ and $\beta 0$ were compared across targets, with a special focus on visual and crossmodal redundancies.

EEG Analysis. In order to ensure a sufficient signal-to-noise ratio (SNR), we applied a *within-subject* jackknife procedure across target locations. For each target location, we averaged the epochs of the other nine target locations and computed their specific ERLs (PCN and CCN). For the PCN, we included electrodes PO7/PO8, PO3/PO4, and O1/O2, while for CCN, C3/C4 was chosen. Then, we applied the common *between-subject* jackknife procedure across participants (Kiesel et al., 2008; Miller et al., 1998; Ulrich & Miller, 2001) and extracted the negative peak *amplitudes* in PCN and CCN within the time interval of 150–350 ms post-stimulus presentation. To further improve the SNR in ERLs, we collapsed the five symmetric target locations by averaging them. Afterwards, we applied a double inverse transformation of the *within-subject* (locations) and *between-subject* (participants) jackknife scores (Smulders, 2010). Similar to performances, we applied RM ANOVAs for *targets* and *locations*. Finally, quadratic fitting was done by mirroring (doubling) the symmetric locations for the amplitudes in PCN and CCN to reveal the slopes $\beta 0$, $\beta 1$, and $\beta 2$ as spatial properties. Afterwards, the slopes were again statistically analyzed by RM ANOVAs, and one-sample t-tests (s. Figure 3–1).

Results

Performances. Initially, we compared left and right locations across targets. For the majority, left and right locations across targets were comparable in both error rates (ERs) (ps > .118) and reaction times (RTs) (ps > .077). Further, some minor differences were found within targets, either in ERs or RTs. For CS, there was a bias on the first and left locations, showing slower RTs ($p = .019$), while ERs were comparable (p = .152). The exact same bias was found in SF, showing ERs to be different ($p = .025$) and RTs comparable ($p = .077$). Finally, in F, the tactile frequency, RTs were consistently faster for right locations than left ones ($p < .001$), showing benefits for the right hand. Yet, ERs for F were comparable (p = .146). Since ERs and RTs or both were consistent in the left and right hemifields across targets, it seems that spatial attention is broadly evenly distributed for left and right locations during the visuo-tactile detection task, which allows us to use the symmetric solutions.

In ERs, we found two significant main effects in targets (F (6, 96) = 37.11, $p < .001$, $\eta^2 > .250$) and locations (F (9, 144) = 16.37, $p < .001$, $\eta^2 > .168$), and also a prominent interaction (F (54, 864) > 7.08, $p < .001$, $\eta^2 > .226$). The same results in RM ANOVA were also present for RTs, showing main effects for targets (F (6, 96) = 39.25, $p < .001$, $\eta^2 > .455$) and locations (F (9, 144) > 52.50, $p < .001$, $\eta^2 > .384$), and a significant interaction (F (54, 864) > 7.99, $p < .001$, $\eta^2 > .190$). Generally, redundant targets (CS, CF, SF, CSF) were faster and more accurate than non-redundant targets (C, S, F) (ps < .001), which perfectly replicates our previous findings (s. 2.1). Further, an eccentricity effect for ERs and RTs was present across locations, showing best outcomes within center locations and decreasing reactions and accuracies towards the periphery (ps < .001). Finally, and most importantly, the interactions indicate that the spatial functions across targets differed (s. Figure 3–2).

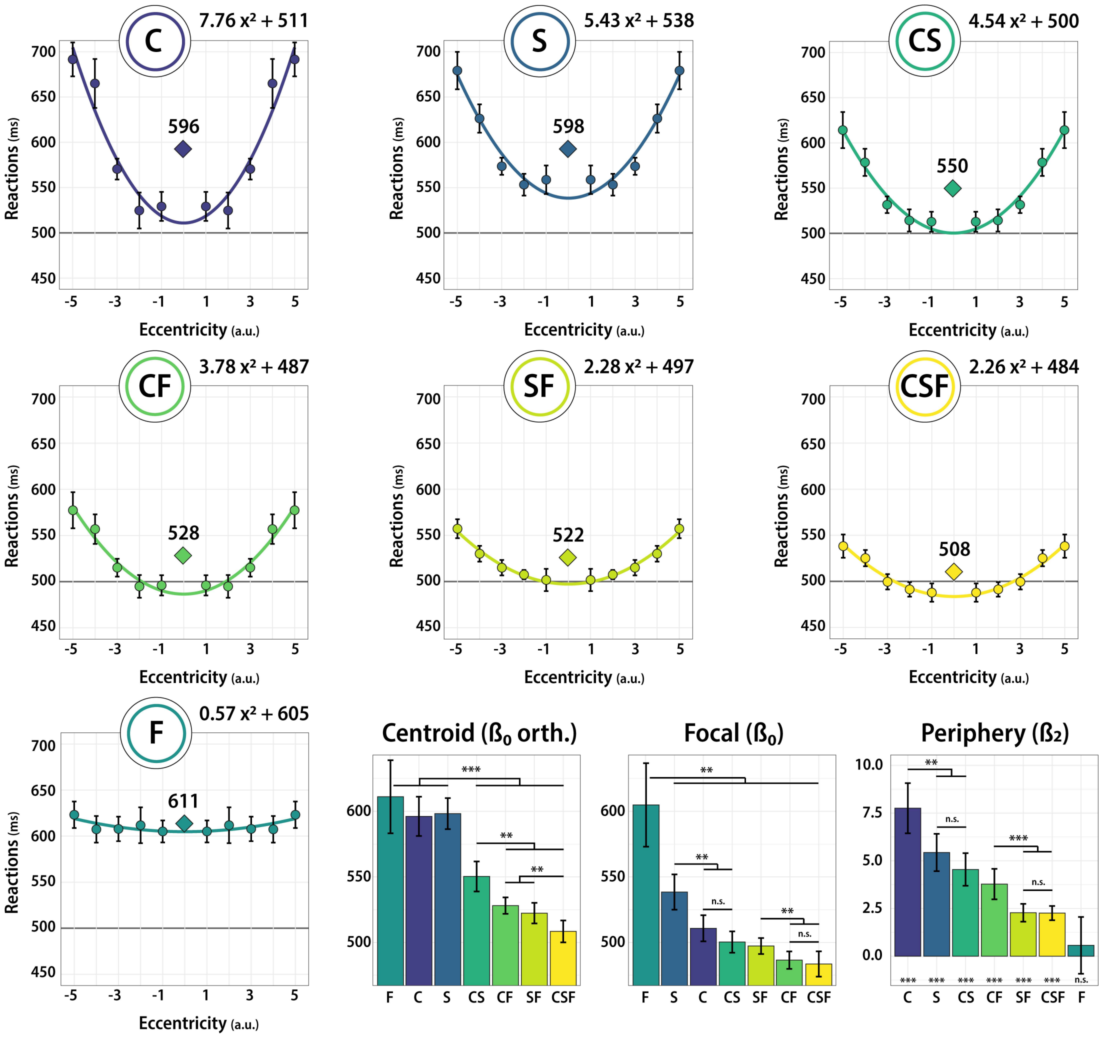

Figure 3–2. The performances in reaction times (in milliseconds) are shown for each of the seven *targets* (C, S, F, CS, CF, SF, & CSF) and symmetric *locations* (eccentricity with adjusted units, a.u.). Each target plot contains the quadratic fit for $y = \beta_2 x^2 + \beta_0$ on the top, shows the overall mean reaction time (center diamond) the location-specific reaction times (circles) including within-subject confidence intervals. The 'quadratic fit' curve is shown as lines. The statistical analysis of the distinct slopes is shown in bar plots at the bottom, significant differences are indicated by asterisks (*** $p < .001$, ** $p < .01$, $p < .05$). While the centroid or orthogonal $\beta 0$ replicates the findings on redundancy gains, the focal $\beta 0$ (saddle point) shows color benefits and the peripheral $\beta 2$ shows shape benefits in vision. Overall, visual targets (C, S, CS, CF, SF, & CSF) revealed substantial quadratic fits ($\beta 2$), whereas the purely tactile target (F) was flat.

By having the evidence from the RM ANOVA that the spatial functions across targets differ, we can now analyze the spatial properties and slopes ($\beta 2$,$\beta 1$,$\beta 0$) by fitting the quadratic function for symmetrically collapsed locations. Analyzing the *orthogonal* slope $\beta 0$ actually reveals the mean RTs or centroid of the parabola, and thus the main effect across targets is equivalently significant as above (F (6, 96) = 38.43, $p < .001$, $\eta^2 > .258$) (s. Chapter 2 Results), which further underlines the validity of quadratic fitting. Again, we found the same hierarchy of redundancy gains as before with increasing benefits in RTs across non-redundant (C, S, F), visual redundant (CS), crossmodal redundant (CF, SF), triple redundant (CSF) targets (*p*s < .013).

Another main effect was found for the *regular* $\beta 0$, indicating the saddle point of the parabola and the optimum of focal attention in the center (F (6, 96) = 37.36, $p < .001$, $\eta^2 > .315$). Interestingly, the saddle points in $\beta 0$ slopes deviated from the common hierarchy of redundancy gains across targets. Here, we found a dissociation between color and shape, showing slower RTs for S compared to C and CS (*p*s < .017) in focal areas, while both C and CS were comparable (p = .299). The same pattern of focal benefits for colors was repeated within crossmodal targets, showing CF and CSF to be consistent ($p = .697$), while both exceeded SF (*p*s < .030). Furthermore, SF was comparable to C and CS (*p*s > .697), while crossmodal benefits stayed apparent for all other comparisons (*p*s < .018). Finally, the saddle point of RTs in the tactile target F was significantly slower than all other targets (*p*s < .018). Generally, these findings in $\beta 0$ slopes indicate particular benefits for color in focal areas.

As expected, the linear slope $\beta 1$ was not relevant in both *regular* and *orthogonal* quadratic fits, since it was de facto zero in all targets ($|\beta 1| <$ 4E-14), showing no differences in one-sample t-tests (*p*s > .329), except for F, where the $\beta 1$ slope was marginally different to zero (p = .052). Hence, $\beta 1$ is not relevant.

The analysis of $\beta 2$ revealed another interesting pattern with a significant main effect across targets (F (6, 96) = 26.93, $p < 0.001$, $\eta^2 = 0.518$). Again, $\beta 2$ slopes indicate the spatial gradient towards the periphery. While $\beta 2$ was comparable between S and CS ($p = .277$), the $\beta 2$ slope in C was much steeper and significantly different from both (*p*s < .048). As before, the same pattern was repeated within crossmodal targets, showing $\beta 2$ in SF and CSF to be consistent ($p = .947$), but both were less steep than CF (*p*s < .007). Here it seems that peripheral performances appear to be better for shape information. Although we found overly benefit for crossmodal redundancies (*p*s < .058), interestingly, the $\beta 2$ slopes in S and CS were similar to CF ($p < .342$).

Finally, by using one-sample t-tests for the $\beta 2$ slopes, no difference was found in the solely tactile target F ($p = .204$), while the other six targets (C, S, CS, CF, SF, CSF), containing visual information, were significantly higher than zero (*p*s < .001). Therefore, the spatial function of tactile information in F seems to be flat or constant (egocentric), while the visual and crossmodal targets follow a quadratic function (external) (s. Figure 3–2).

Event-Related Lateralizations. For the PCN, similar to the analysis in redundancy gains (s. Chapter 2), we excluded F and analyzed it separately, due to its special role as a purely tactile signal. First, we conducted the same RM ANOVA as in performances. Importantly, in ERLs, we only included the five collapsed locations for the analysis to improve the signal-to-noise ratio in PCN and CCN signals. Two main effects were present in the PCN for the factor Targets (F (5, 80) = 4.88, $p < .001$, $\eta^2 = .127$) and Locations (F (4, 64) = 16.87, $p < .001$, $\eta^2 = .169$), without an interaction (F (20, 320) = 1.28, $p < .193$, $\eta^2 = .026$). Overall, PCN amplitudes were the most negative in crossmodal targets and central locations (s. Figure 3–3 & Figure 3–4). Interestingly, the RM ANOVA in the CCN revealed only one main effect in Targets (F (6, 96) = 28.44, $p < .001$, $\eta^2 = .503$), showing the same dissociation between visual (C, S, CS) and tactile (F, CF, SF, CSF) targets as in the redundancy gains' study (s. Chapter 2). Further, there was no effect in Locations (F (4, 64) = 1.25, $p = .295$, $\eta^2 = .005$) and no interaction (F (24, 384) = 0.85, $p = .659$, $\eta^2 = .019$). Actually, finding a significant effect in PCN locations but not in CCN locations, is in line with our assumptions, expecting vision to be quadratic and touch to be flat and constant. Furthermore, finding no interactions already suggests that the spatial functions in PCN and CCN are comparable across targets.

The analysis of spatial slopes ($\beta 2$, $\beta 1$, $\beta 0$) revealed further insights. In PCN, no difference was found in ß2 (F (5, 80) = 0.52, $p = .764$, $\eta^2 = .017$), suggesting that all six targets containing visual information (C, S, CS, CF, SF, CSF) have similar gradients towards peripheral locations. Further, a one-sample t-test revealed that the $\beta 2$ slopes of the same six targets were significantly different to zero (ps < .013), indicating good quadratic fits, while the $\beta 2$ slope in F, the tactile PCN, was comparable to zero ($p = .853$), and therefore, not quadratic but flat.

Again, the linear slope $\beta 1$was actually zero (< 1E-17) and insignificant (ps > 0.05) for the quadratic fit of PCN amplitudes. The *orthogonal* $\beta 0$ was significantly different across the six targets (F (5, 80) = 6.45, p < .001, $\eta^2 = .120$). Here we found the same crossmodal benefits in the PCN amplitude as before (s. Chapter 2), with more negative amplitudes in CF, SF, and CSF compared to C, S, and CS (ps < .054). A similar effect was also present in the *regular* $\beta 0$, which indicates the focal saddle point of PCN amplitudes (F (5, 80) = 6.14, $p < .001$, $\eta^2 = .119$). Here we found the same crossmodal benefits in amplitudes (ps < .059) as within the *orthogonal* $\beta 0$, except for the comparison between C and SF ($p = .185$).

In the CCN, ß2 was not different across targets (F (6, 96) = 1.20, $p = .313$, $\eta^2 = .055$) and overly not different from zero (ps > 0.255). However, gradients in SF and CSF were prominently different to zero (ps < 0.025), which indicates that a quadratic fit adds some improvement for these crossmodal targets in the CCN (s. Figure 3–5). The $\beta0$ slope was again zero and irrelevant (*p*s > .381). Finally, results in both $\beta0$ (*regular* and *orthogonal*) were very similar (F (6, 96) > 14.33, *p* < .001, η^2 > .315). The CCN amplitudes in tactile targets (F, CF, SF, CSF) were significantly more negative than purely visual ones (C, S, CS) (*p*s < .005).

In summary, these findings suggest that the ß2 slopes in PCN are indeed quadratic towards peripheral locations for visual and crossmodal targets, while $\beta2$ was flat for the tactile PCN (F). Further, we found crossmodal amplification in $\beta0$ slopes, which is in line with the findings in redundancy gains. The CCN, on the contrary, was overly flat in $\beta2$ slopes and thus CCN amplitudes were comparable across locations, except for some quadratic modulations in SF and CSF. Finally, both $\beta0$ slopes indicate higher tactile CCN amplitudes. Since no interactions were found and $\beta2$ were consistent, the spatial functions in PCN and CCN appear to be *modality-specific*, while the crossmodal amplification in PCN suggests the presence of a *supramodal* mechanism.

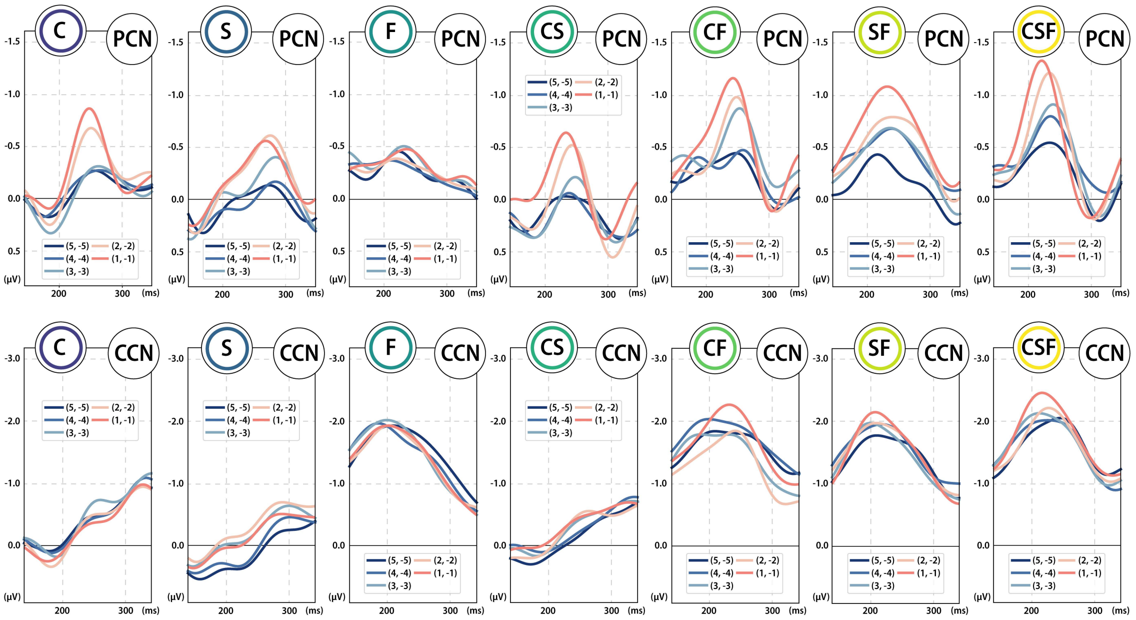

Figure 3–3. The PCN (top) and CCN (bottom) within 150–350 ms are shown for *targets* and symmetric *locations* (e.g., 1, -1) ranging from the periphery (blue) to the center (red). Notably, the CCN was only present for tactile targets (F, CF, SF, & CSF) and (more or less) constant across locations. The PCN was adaptive to visual information by showing enhanced PCN amplitudes for central locations in visual targets (C, S, CS, CF, SF, & CSF) that decay toward the periphery (replicating the eccentricity effect). The PCN of the purely tactile target (F) remained constant.

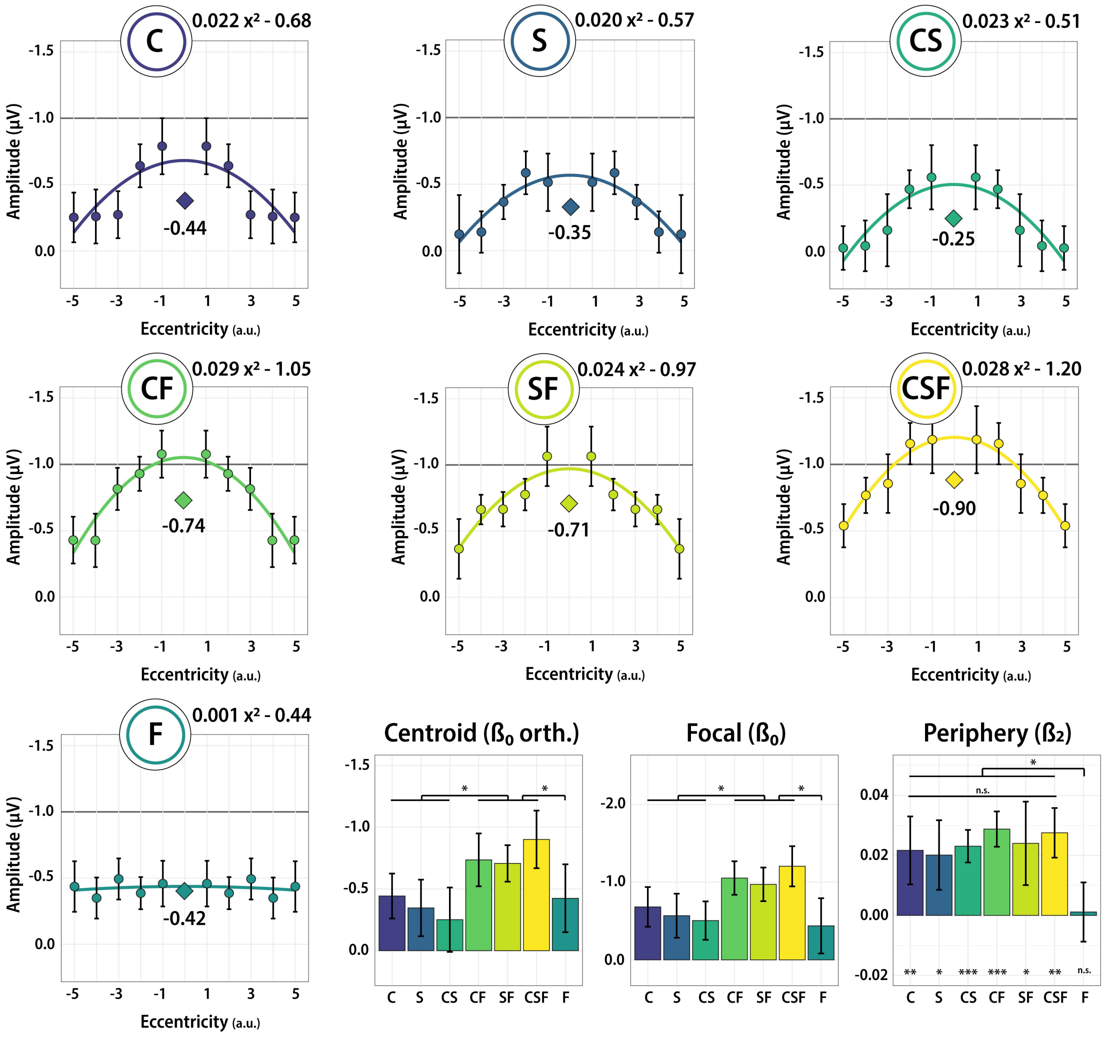

Figure 3–4. The PCN amplitudes (in µV) are shown for each of the seven *targets* (C, S, F, CS, CF, SF, & CSF) and symmetric *locations* (eccentricity with adjusted units, a.u.). Each target plot contains the quadratic fit for $y = \beta_2 x^2 + \beta_0$ on the top, shows the overall mean reaction time (center diamond) the location-specific reaction times (circles) including within-subject confidence intervals. The ‘quadratic fit’ curve is shown as lines. The statistical analysis of the distinct slopes is shown in bar plots at the bottom, significant differences are indicated by asterisks (*** $p < .001$, ** $p < .01$, $p < .05$). While the centroid or orthogonal $\beta 0$ and focal $\beta 0$ replicates the findings on redundancy gains, showing enhanced crossmodal PCN amplitudes, the peripheral $\beta 2$ shows significant yet comparable quadratic fits for all visual targets (C, S, CS, CF, SF, & CSF) and a flat $\beta 2$ for purely tactile targets (F).

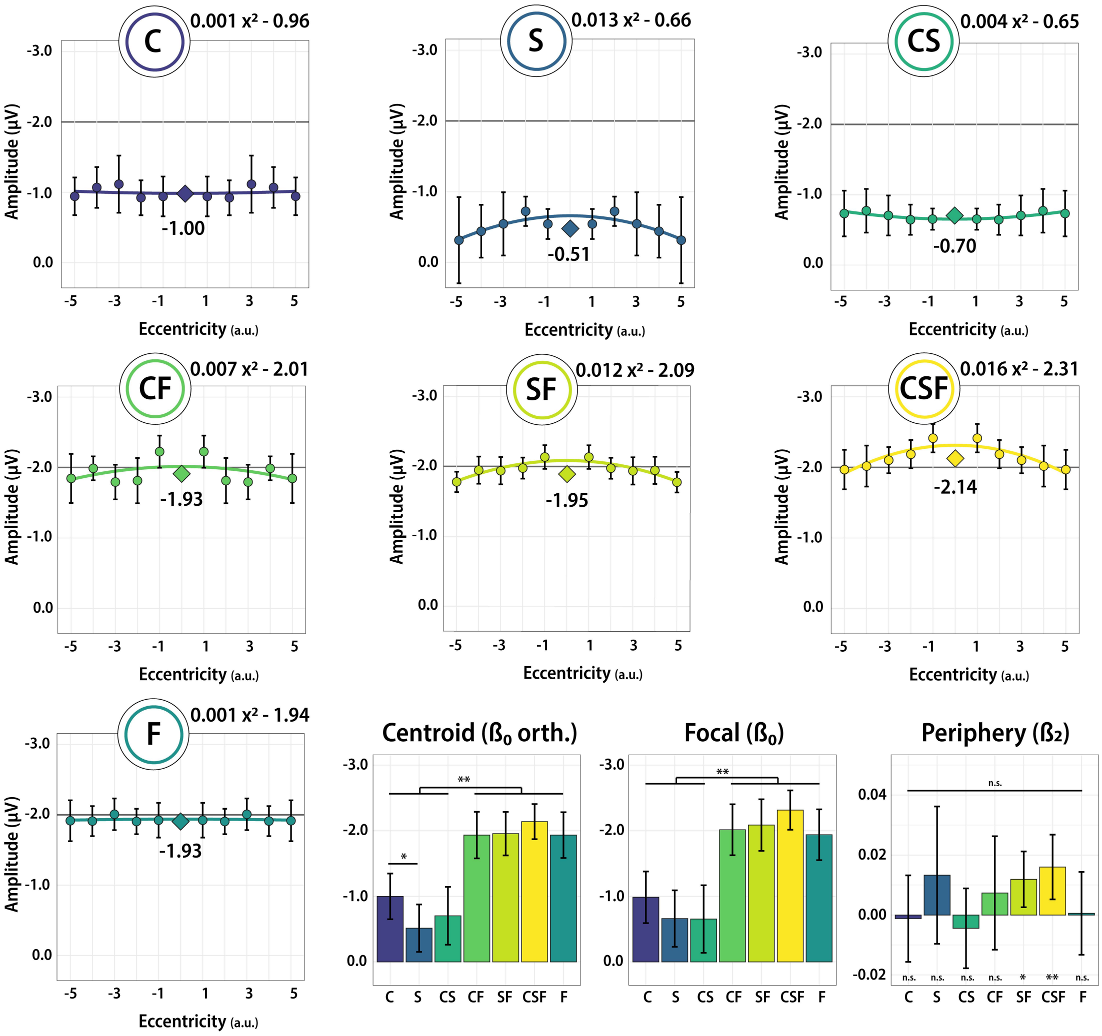

Figure 3–5. The CCN amplitudes (in µV) are shown for each of the seven *targets* (C, S, F, CS, CF, SF, & CSF) and symmetric *locations* (eccentricity with adjusted units, a.u.). Each target plot contains the quadratic fit for $y = \beta_2 x^2 + \beta_0$ on the top, shows the overall mean reaction time (center diamond) the location-specific reaction times (circles) including within-subject confidence intervals. The 'quadratic fit' curve is shown as lines. The statistical analysis of the distinct slopes is shown in bar plots at the bottom, significant differences are indicated by asterisks (*** $p < .001$, ** $p < .01$, $p < .05$). While the centroid or orthogonal $\beta0$ and focal $\beta0$ replicates the findings on redundancy gains, showing a stable somatosensory enhancement in CCN amplitudes for all tactile targets (F, CF, SF, & CSF), the peripheral $\beta2$ were non-significant and thus constant except for SF and CSF.

Discussion

The main purpose of this study was to extend our findings on redundancy gains by analyzing its spatial properties. To achieve this, quadratic fitting was applied to investigate the focal ($\beta0$) and peripheral slopes ($\beta2$) as parameters of spatial attention. In line with the eccentricity effect in visual search for external spatial mappings (Carrasco et al., 1995; Staugaard et al., 2016), we expected to observe (steep) quadratic slopes ($\beta 2$) in visual targets (R. Gray et al., 2009; Lu & Dosher, 2000) and a flat slopes ($\beta 2$) in purely tactile targets, due to somatotopic or egocentric mappings of the body schema across finger locations (Assumpção et al., 2018; Eimer et al., 2004; Forster et al., 2016; Holmes & Spence, 2004; Medina & Coslett, 2010). Furthermore, redundancy gains are expected to modulate systematically based on the spatial parameters ($\beta 0$, $\beta 2$). While visual redundancies (CS) should remain within the spatial limits of vision (C, S), crossmodal redundancies (CF, SF, CSF) should exceed those limits due to 'coactive' crossmodal spatial interactions. Finally, the examination of spatial slopes in the event-related lateralizations of PCN and CCN components, reflecting spatial filters for visual and tactile attention, are expected to depict the *external* and *egocentric* mapping, respectively. While spatial (slope) interactions in the PCN, the area of crossmodal links, can be regarded as evidence for crossmodal spatial integration in the preattentive stage (*supramodal* mechanism), comparable slopes across visual targets (C, S, CS, CF, SF, CSF) in the PCN and tactile targets (F, CF, SF, CSF) in the CCN would suggest independent spatial integration for vision and touch (*modality-specific* mechanism).

Overall, the reaction times and error rates between left and right locations were comparable across all six visual targets (C, S, CS, CF, SF, CSF), allowing us to collapse and combine symmetrical locations by excluding this spatial bias. In touch, the frequency target (F), however, we found faster reaction times in the right hand with comparable error rates in both hands. This difference might be linked to right-handedness, which likely improves tactile sensitivity in the dominant hand. Yet, tactile reaction times were consistent across the fingers within each hand, and thus indicate an egocentric and flat distribution of tactile spatial attention within the "fovea of touch" (Forster et al., 2016). Generally, these findings support the assumption that spatial attention is deployed broadly and symmetrically across the ten display locations during the visuo-tactile detection task.

Initially, the analysis of reaction times and error rates revealed significant main effects and an interaction for the factors Targets and Locations, indicating that the spatial functions differ across targets. While the main effect in Targets is identical to the hierarchy of redundancy gains (s. Chapter 2), differences across Locations in the six visual targets were in line with the eccentricity effect, revealing better performances in the centers that worsened toward the periphery (Carrasco et al., 1995; R. Gray et al., 2009; Staugaard et al., 2016). To examine those spatial properties in more detail, we compared the slopes $\beta 2$, $\beta 1$, and $\beta 0$ from quadratic fitting. Whereas the *orthogonal* solution of the centroid $\beta 0$ slope (reflecting the mean reaction times), again replicates the pattern of redundancy gains (s. Chapter 2), the *regular* quadratic solution revealed two interesting target patterns in focal saddle points ($\beta 0$) and the peripheral (quadratic) slope ($\beta 2$). As expected, the linear term of the $\beta 1$ slope was irrelevant.

The first spatial pattern was found in (*regular*) $\beta 0$ saddle points, indicating the reaction time optimum (the minimum) in focal areas, by showing a consistent benefit for color targets (C, CS) over the purely shape target (S). Despite a clear crossmodal enhancement in $\beta 0$, interestingly, the same pattern of color benefits was repeated within crossmodal targets, showing higher focal gains for CF and CSF than SF targets. So, it is quite clear that color perception exceeds shape perception in focal regions. Furthermore, these findings also suggest that visual redundancies (CS, CSF) in focal regions are strongly linked to color (C, CF) targets, since their $\beta 0$ slopes are similar (s. Figure 3–2). This finding indicates a modality-specific limit within vision, revealing no extra gain or benefit beyond color perception. On the contrary, we also found supramodal benefits in crossmodal redundancies (CF, SF, CSF), were the combination of vision (C, S, CS) and touch (F), always lead to significant improvements in focal regions ($\beta 0$ effect). Generally, visual information (color and/or shape) is very effective in focal areas and clearly exceeds tactile information (frequency), which refers to findings of visual dominance (Gondan et al., 2007; Klein, 1977; Posner et al., 1976; Spence, Shore, et al., 2001; Xia et al., 2018).

In contrast to color benefits in focal areas, the second spatial pattern revealed a $\beta 2$ effect for shapes, being less steep toward the periphery. Here, the peripheral slope ($\beta 2$) was steeper in C compared to S and CS, which both were comparable. Again, the same pattern was repeated in crossmodal targets, showing lower $\beta 2$ slopes in SF and CSF over CF. In summary, the findings in focal and peripheral slopes, now perfectly explain visual redundancy gains (CS & CSF), which arise due to both spatial properties. Interpreting both spatial patterns, it seems obvious that color promotes focal benefits, while shape promotes peripheral benefits.

The combination of color and shape information, obviously improves both spatial properties. Therefore, visual redundancy gains arise as the optimal combination of focal and peripheral properties. Furthermore, it is important to emphasize that spatial attention in visual redundant targets appear modality-specific; it remains within the limits of its respective feature contrasts (here color and shape). Finally, crossmodal targets (CF, SF, CSF) generally exceed these visual limits by revealing extra gains in both spatial slopes ($\beta 0$ and $\beta 2$), and thus indicate supramodal adjustments in crossmodal spatial attention. This supramodal boost in spatial properties provides empirical evidence for the presence of crossmodal bending in spatial attention. Yet, the same rules of visual optimization are again preserved in crossmodal targets, with improved focal color (CF) and peripheral shape (SF) that coalesce in CSF.

Overall, the spatial function in vision was sufficiently explained by quadratic fits, which is in line with the eccentricity effect and the external cortical mapping in receptive fields (Carrasco et al., 1995; Carrasco & Frieder, 1997; R. Gray et al., 2009; Harvey & Dumoulin, 2011; Yoshor et al., 2007). By having obtained distinct spatial properties in focal ($\beta 0$) and peripheral ($\beta 2$) slopes, theses findings extend the previous insights from redundancy gains, which now can also be explained by distinct patterns of spatial optimization. For touch, the solely tactile frequency (F) revealed a flat spatial function, in line with the assumption that spatial tactile attention is equally distributed across fingers and egocentric (Assumpção et al., 2018; Eimer, Forster, et al., 2003; Forster et al., 2002; Holmes & Spence, 2004; Medina & Coslett, 2010). For future studies, it would be interesting to continue the examination of distinct feature contrasts and their spatial functions across dimensions and sensory modalities on whether spatial limits in intramodal or crossmodal combinations persist or are exceeded, respectively. From the mapping of both spatial slopes, showing the relationship between the focal saddle point ($\beta 0$) and the peripheral slope ($\beta 2$) (s. Figure 3–6), some exciting questions arises. In the coordinate space, it seems that the spatial properties of crossmodal targets derive from scaling coefficients, which might reflect eigenvalue solutions and feature vectors. Furthermore, for each visual dimension there might be a specific set of lambdas (i.e., eigenvalues), which enables us to obtain accurate estimates about the interactions across features. Additionally, redundancy gains should only occur within the area of its feature vectors, while the intersection of feature vectors would refer to physical threshold or limits. Hence, our findings clearly suggest that spatial properties play a major role for explaining redundancy gains by revealing distinct characteristics, which should be considered.

To find out whether those spatial dynamics are also evidenced in event-related lateralizations, we analyzed both PCN and CCN, two neural markers of visual and tactile spatial attention, respectively. Overall, the main interest was to retrieve the same spatial function of reaction times, that is, a quadratic spatial function in PCN amplitudes for visual targets (C, S, CS, CF, SF, CSF) and a flat spatial function in CCN amplitudes for tactile targets (F, CF, SF, CSF).

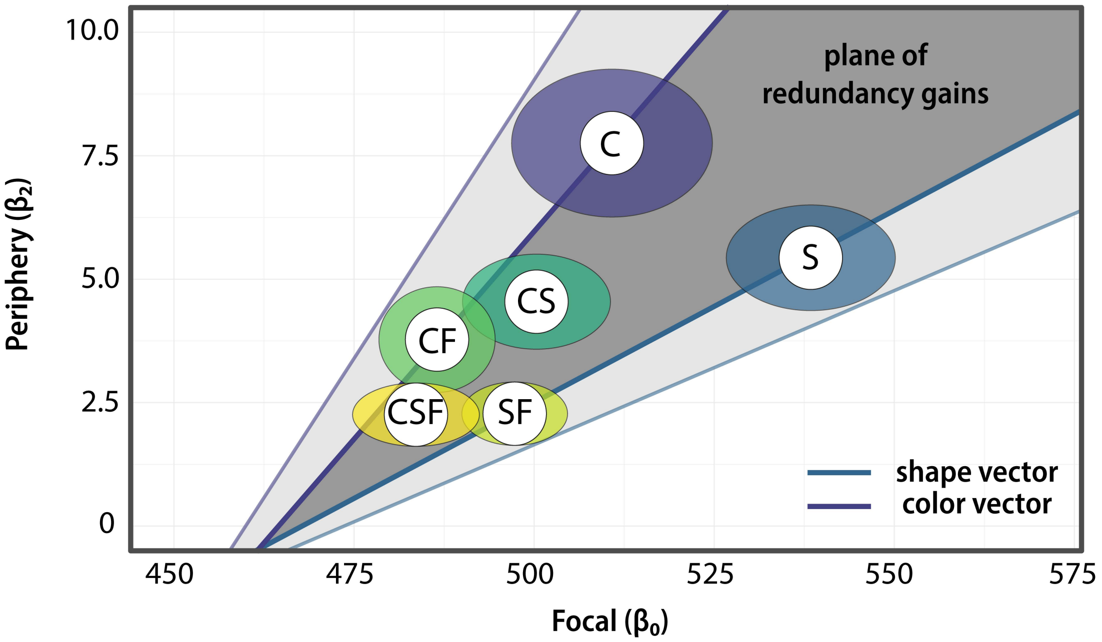

Figure 3–6. The two-dimensional spatial mapping of focal $\beta 0$ (x-axis) and peripheral $\beta 2$ (y-axis) is presented for the six visual targets (C, S, CS, CF, SF, & CSF). Importantly, color circles or ellipses indicate the within-subject confidence intervals of both spatial slopes ($\beta 0$, $\beta 2$). The feature vectors for shape (S, SF: blue) and color (C, CF: purple) were added as lines, creating a specific plane within their means (dark gray) and error rates (bright gray). Arguably, the plane indicates the area for redundancy gains, and the intersection between both feature vectors might refer to the physical threshold or limit (around 467 ms) of performances. Overall, this schema represents eigenvector solutions for visual and crossmodal feature-contrast.

Before we discuss our findings, it is important to discuss one critical issue. In order to retrieve locations-specific target activation in PCN and CCN with a sufficient signal-to-noise ratio, it was necessary to apply a ‘double jackknife procedure’, across targets locations and participants. Although from a mathematically point of view, applying (double) inverse transformation is unhesitating, being a set of linear transformation, from an empirical perspective, it is important to highlight that the obtained results (merely) reflect individual estimates (Smulders, 2010). Therefore, it will be essential in future methodological studies to evaluate the relationship between ‘real scores’ and ‘individual estimates’ more thoroughly to establish this technique as a common procedure, enabling a multitude of applications.

In this study, we found significant quadratic slopes ($\beta 2$) in the PCN for all targets containing visual information, but (overly) not in in the CCN and the solely tactile PCN (F) (s. Figures 3–4 & Figure 3–5). Interestingly though, the $\beta 2$ slopes of SF and CSF were significant in the CCN, the same targets that showed evidence coactivation in the previous study (s. Chapter 2). Of note, a similar quadratic slope was also present for shape (S) in the CCN, although not significant, due to high variance in the periphery. Nonetheless, this finding suggests that shape information in general might be the cause that influences the CCN (thus producing preattentive coactivation). In the PCN, no evidence was found for crossmodal bending, since the quadratic slopes of the six targets (C, S, CS, CF, SF, CSF) were comparable. This indicates that the $\beta 2$ slope appears *modality-specific* in the PCN for vision, likely representing the spatial firing patterns of receptive fields based on external coordinates (Carrasco et al., 1995; Luck et al., 1997).

For the other slopes, the *regular* and *orthogonal* $\beta 0$, saddle point and centroid, respectively, we replicated the hierarchy of redundancy gains (s. Chapter 2), by showing crossmodal benefits in PCN amplitudes, with more negative amplitudes (in CF, SF, & CSF), and a purely tactile CCN (in F, CF, SF, & CSF). Hence, we can argue that crossmodal targets lead to a *supramodal* amplification in PCN amplitudes, similar to effects in cortical magnification (Carrasco et al., 2003; Carrasco & Frieder, 1997; Staugaard et al., 2016).

Further, this explanation for the PCN is also in line with other studies, which suggest that the spatial egocentric mapping of tactile information is forwarded to parieto-occipital regions, where they combine with the external visual mapping (Driver & Spence, 1998; Eimer, Forster, et al., 2003; Spence, Pavani, et al., 2000).

This is also supported by the fact that the tactile PCN is constant, which indicates a common baseline across locations, and thus the superposition of visual and tactile mappings does not bias the peripheral gradient in $\beta 2$, and thus preserves the external mapping of visual spatial attention. This explanation is also supported by the fact that on the contrary, no visual influence was found in CCN amplitudes which remained comparable (F, SF, CF, CSF). Overall, the evidence favors modality-specific spatial functions across locations ($\beta 2$), being quadratic in visual PCN and flat in tactile CCN, and a supramodal amplification in PCN ($\beta 0$).

Another interesting aspect might be that quadratic fitting is not fully sufficient in PCN, since some locations appear quite distant to the curve of the quadratic function (e.g., 3rd location in C, or 4th location in CF). For colors, this can be explained by the fact that the visual angle is more closely linked to the fovea, and though spatial perception decreases steeper for locations outside an visual angle of 8–10°, which in our case occurs between 2nd and 3rd locations (Larson & Loschky, 2009). Therefore, it might be interesting to also apply higher polynomials to obtain additional spatial slopes of visual feature-contrasts.

In summary, we found strong evidence that redundancy gains can be described by spatial characteristics in vision and touch, being quadratic and flat, respectively. Furthermore, we found interactions in their focal and peripheral slopes ($\beta 0$, $\beta 2$), revealing focal gains for colors and peripheral gains in shapes. Whereas visual redundancies appear modality-specific, since its benefits remain within the limits of spatial properties, crossmodal redundancies, on the contrary, exceeded those modality-specific limitations, and thus indicate a supramodal interaction with crossmodal benefits. Although vision is highly efficient in focal areas, its lack towards the periphery seems to be effectively compensated by other senses (Solovey et al., 2015), which becomes apparent by our findings suggesting crossmodal bending of spatial attention. Additionally, finding the same modality-specific spatial functions in PCN and CCN, being again quadratic and flat respectively, underscores the differences in visual and tactile spatial integration, being external and egocentric, respectively. Yet crossmodal amplification in PCN amplitudes indicates supramodal benefits similar to cortical magnification of receptive fields. Therefore, the dynamics of spatial attention in PCN are both modality-specific and supramodal, which is in line with previous studies (Chambers et al., 2007; Föcker et al., 2010; Spence & Driver, 1997b). For future studies it would be interesting to investigate whether the mapping of spatial slopes and properties (focal, periphery) includes eigenvectors of features and their combinations. Finally, it seems promising to expand the analysis of crossmodal spatial attention on neural dynamics by looking into further components and wave complexes.

4

Intertrial Effects in Crossmodal Search

Fragile continuum
But yet a honey river
And grip goes under.

Reduced Amplitudes in PCN and CCN indicate a Common Mechanism for Modality-Shifts in Visuo-Tactile Search

Jan Nasemann [1,2], Thomas Töllner [1], Hermann J. Müller [1,2], Zhuanghua Shi [1,2]

1 General and Experimental Psychology, Department of Psychology,
Ludwig-Maximilians-Universität München, Munich, Germany
2 Graduate School of Systemic Neurosciences,
Ludwig-Maximilians-Universität München, Planegg, Germany

Abstract

Selective attention is the gateway between our perceptions and higher cognitive functions that modulates sensitively to bottom-up, top-down, and also intertrial effects, which emerge from the history of events. While repeated sensory information tends to improve our performances, shifts between sensory modalities (e.g., vision → touch) lead to prominent switch costs, known as the modality-shift effect (MSE). Generally, selective attention integrates perceptual and preattentive processes through saliency summation. In visual search, feature integration can be explained by the hierarchy of the "dimension- weighting account" (DWA), which combines layers for visual dimensions (e.g., color and shape), and their features (e.g., red and green). Here, switch costs arise between but not within dimensions. In this study, we extend the DWA to crossmodal attention using a visuo-tactile search paradigm. This "modality- weighting account" (MWA) assumes distinct and increasing switch costs for shifts between dimensions and modalities (MSE), which was confirmed by reaction times. Also, the cost for switching between modalities persisted over three successive trials, and a dissociation for location-shifts (left vs. right) was found between vision and touch. Further, we included EEG recordings and analyzed the latency and amplitude of visual (PCN) and tactile (CCN) lateralizations. Whereas latencies in the visual PCN indicate a modality-specific mechanism, being equally delayed for shifts in dimensions and modalities, a supramodal mechanism was found in amplitudes, which were prominently reduced for modality-shifts in both PCN and CCN. Interestingly, the PCN appears adaptive for vision and touch, indicating crossmodal attention to be linked to extrastriate regions (beyond linear superposition). In general, these findings are in favor of the MWA and its hierarchy, showing a delay or reduction of attention resources in PCN and CCN due to intertrial effects.

Introduction

For many decades, selective attention has been regarded as the gate that links perception and action in a coherent fashion (Allport, 1987; Berlyne, 1951; James, 1890; Pashler, 1997). In current theorizing, selective attention in perception is conceived as a limited-capacity resource that is allocated based on bottom-up (physical) stimulus saliency and top-down stimulus set or expectancies, where inter-trial effects reflecting the 'history' of recent selection events may be subsumed under the latter (Awh et al., 2012; Bundesen, 1990; Found & Müller, 1996; Maljkovic & Nakayama, 1994, 1996; H. J. Müller et al., 1995; Wolfe, 1994).

One type of inter-trial effect is the *modality-shift effect* (MSE), which refers to the (de-)prioritized processing of modality-specific information from multisensory (e.g., visuo-tactile) stimuli. In general, performance is facilitated when the task-critical, 'target' information is repeated within a given modality, whereas performance is impeded when the critical information changes from one modality to another (R. Cohen & Rist, 1992; Gondan et al., 2004; Jonides & Mack, 1984; Rodway, 2005; Spence, Nicholls, et al., 2001). That is, prioritizing a specific modality on (preceding) trial *n–1* is beneficial when this modality remains task-relevant on (current) trial *n* (repetition benefit), but costly when another modality becomes relevant on trial *n* (*switch cost*). Of note, though, shifting between modalities may be asymmetric. For instance, (Miles et al., 2011) found switching from vision to touch to be more costly than switching from touch to vision – an asymmetry attributed to visual dominance (Gondan et al., 2007; Klein, 1977; Posner et al., 1976; Spence, Shore, et al., 2001; Xia et al., 2018). In short, the MSE is a prominent inter-trial effect, which is evident across sensory modalities (R. Cohen & Rist, 1992; Rodway, 2005; Spence, Nicholls, et al., 2001; Turatto et al., 2002).

From a neurophysiological perspective, the MSE can be explained as follows. Initially, modality-specific sensory information activates primary cortices in the respective brain regions (e.g., Macaluso et al., 2002b), with the evidence accumulation being modulated by bottom-up (perceptual) and top-down (attentional) processes (e.g., De Meo et al., 2015; Macaluso et al., 2016; Talsma et al., 2010). Finally, processing of the information in a given (current) trial episode is influenced by 'neural imprints' (Fecteau & Munoz, 2003) from preceding trials, which reflect the cross-trial alternation of, and thus the activation balance between, the task-relevant sensory modalities. As a result, it takes less time to extract the response-relevant information when the target modality is repeated (because the relevant modality is 'pre-activated'), whereas shifting the modality causes a cost (because the residual activation of the preceding, now no longer relevant modality must be compensated for). Both components – the pre-activation (of the relevant modality) and residual activation (of the irrelevant modality) – have been termed *neural trace* (Gondan et al., 2007; Maljkovic & Nakayama, 2000; Mannuzza, 1980; Zubin, 1975; Zylberberg et al., 2009) or, respectively, *attentional trace* (Alho et al., 1990; Jaśkowski et al., 2002; Näätänen, 1990; Ruge & Naumann, 2006).

For a more principled understanding of the MSE – the observable end result of modality switching –, it is helpful to consider in greater detail the putative (neuro-) functional architecture that underlies it. According to 'saliency-summation' models (Itti & Koch, 2001; Koch & Ullman, 1985; Lee et al., 1999; Li, 1999, 2002; Wolfe, 1994; Wolfe et al., 1989), the allocation of selective attention is driven by an 'attentional priority map' (e.g., Fecteau & Munoz, 2006), that is, a spatiotopically organized representation of the (visual) scene whose units code the rank order in which objects at the various field locations should be selected for focal-attentional processing: the higher the activation of a unit, the greater the attentional priority of stimuli in its receptive field. In vision, the priority map integrates, or summates, signals derived from entry-level feature coding in the various visual dimensions (such as orientation, color, motion, etc.). Features that exhibit a higher contrast compared to their surround contribute stronger bottom-up 'saliency' signals to priority computation (i.e., through iso-feature suppression mechanisms; e.g., Knierim & van Essen, 1992; Li, 2002; Nothdurft, 2000), and specific features may be top-down enhanced if they match the attentional set (e.g., matching the templates of task-relevant targets held in visual working memory; e.g., (e.g., matching the templates of task-relevant targets held in visual working memory; e.g., Desimone & Duncan, 1995; Treisman & Sato, 1990; Wolfe, 1994). Once intermediate-level representation between the entry-level feature maps and the attentional-priority map has been postulated by "dimension-weighting account" (DWA; Found & Müller, 1996; H. J. Müller et al., 1995). This account assumes that feature-contrast signals are computed per dimension and may be represented in dimension-specific feature-contrast maps (e.g., for orientation contrast, color contrast, motion etc.; see also Itti & Koch, 2000). Critically, the signals in these maps are integrated by the attentional-priority map in a weighted fashion (e.g., color signals dimension may be assigned a greater weight than orientation signals), with the weights determined, in a competitive fashion, by both top-down set and inter-trial history (e.g., H. J. Müller et al., 2003). Concerning selection-history effects, the competitive weighting of dimension explains why dimension shifting (e.g., to a color target on trial *n* from an orientation target on trial *n*–1) causes performance costs compared to dimension repetitions (e.g., a color target on trial *n* following a color target on trial *n*–1), whereas feature changes (vs. repetitions) within a repeated dimension affect performance comparatively little. And it also explains why, with targets defined redundantly (i.e., targets that are salient) in multiple dimensions, signal integration across dimensions (by the priority map) is modulated by the relative weight that has accrued to the various dimensions as a result of inter-trial history: the degree to which two dimension-specific signals drive evidence accumulation by the priority-map units is scaled by the dimensional weights, and signal-summation, or 'co-activation', effects are optimal when the relative weights are within a comparable range (e.g., Krummenacher et al., 2001a, 2009, 2010; H. J. Müller et al., 2003, 2004; H. J. Müller & Krummenacher, 2006; Pollmann et al., 2000).

Going beyond the DWA (which was formulated to account for signal summation in the visual modality), Töllner and colleagues (2009) proposed a "modality-weighting account" (MWA) – essentially an extension and generalization of the DWA to crossmodal (search) task scenarios. According to the MWA, the priority map is an attention-guiding representation that integrates saliency signals across multiple sensory dimensions (such as vision, audition, touch, etc.). In architectural terms, it adds an extra layer of modality-specific maps situated between the intra-dimensional feature-contrast maps (assumed in the DWA) and the attentional-priority map (s. Figure 4–1), where there is competitive weighting of the (signals from the) various modalities determining their driving influence on signal integration by the priority map. Töllner et al. (2009) proposed this account based on their finding of a prominent MSE in a visuo-tactile discrimination task, that is: modality shifting (e.g., to a visual target on trial *n* from a tactile target on trial *n*–1) caused substantial performance costs compared to modality repetitions (e.g., a visual target on trial *n* following a visual target on trial *n*–1). In addition to assessing behavioral performance, Töllner et al. (2009) also recorded the EEG. Analysis of the EEG data provided evidence of the amplitude of a particular event-related potential (ERP) component, namely, the anterior N1, to indicate a supramodal mechanism of weight-shifting: whenever a modality-shift occurred the amplitude of the anterior N1 increased consistently across visual and tactile targets and experimental tasks. This finding is consistent with a variety of other studies that also found the N1 to be a sensitive component distinguishing between intra- and cross-modal shifts (Gondan et al., 2007; Levit et al., 1973; Squires et al., 1976; Sutton et al., 1965; Verleger & Cohen, 1978). Nevertheless, whether the mechanisms reflected in the N1 play indeed a central, 'supramodal' role in regulating the attentional priority of the various modalities is not entirely clear from the wider literature. For instance, some studies found only an auditory N1 effect (R. Cohen & Rist, 1992), with others even reporting an inverse effect for vision (Gondan et al., 2007). Also, a later component, the N2 has been shown to play a role in uni-modal visual search (Folstein & Van Petten, 2008), with shifts of the target-defining dimension producing an amplified N2 over fronto-central regions (Gramann et al., 2007). Overall, these inter-trial effects in early EEG components (N1, N2) can also be explained by a superimposed processing negativity being linked to early perceptual and pre-attentive coding processes (Näätänen, 1982; Näätänen & Picton, 1987; Talsma & Woldorff, 2005).

Exactly within the time window of these early ERP components falls the "posterior contralateral negativity" (PCN, also referred to as N2pc): a lateralized negative deflection at posterior electrodes (within 150–350 ms post stimulus) which is taken to be indicator of the allocation of visuo-spatial attention in the scene (Eimer, 1996; Luck & Hillyard, 1994a; Töllner, Zehetleitner, Krummenacher, et al., 2011; Wolber & Wascher, 2005; Woodman & Luck, 1999). In particular, in visual search, the timing and amplitude of the PCN have been linked to the speed of the attention allocation (i.e., the transition from pre-attentive to focal-attentional processing) and, respectively, the amount or 'engagement' of processing resources at the spatial focus of attention (cf. Zivony et al., 2018). Töllner et al. (2008) found a temporal delay in PCN peak latencies in a visual pop-out search task when the target-defining dimension switched from shape on trial *n*–1 to color on trial *n* (shape → color), or vice versa. Further, the PCN amplitudes were more negative for targets with repeated visual dimensions (e.g., color → color). Consistent with the DWA, no difference was found between exact target feature repetitions (e.g., green → green) and feature shifts (e.g., red → green). In another study using variations of Maljkovic and Nakayama's (1994) 'priming of pop-out' paradigm, Eimer et al. (2010) found expedited PCN onsets for target repetitions and delayed onsets for prominent changes (involving full swapping of the target and distractors features), by employing the jackknife procedure (Ulrich & Miller, 2001). However, amplitude effects derived merely from the differential onset timings, with more negative amplitudes for repetitions in the early PCN and for changes in the late PCN.

For tactile search – which likewise involves processes of attentional target selection (Spence & Gallace, 2007) –, an analogous component to the posterior contralateral negativity (PCN) has been observed over central regions: a "central contralateral negativity" (CCN or N140cc), which reflects the allocation of tactile attention (Eimer et al., 2004; Eimer & Driver, 2000; Forster et al., 2016). That is, the CCN is also a lateralized negative deflection (140–340 ms post stimulus), which builds up for tactile targets (but not tactile distractors), varying in amplitude for different stimulated body parts, with the most negative deflections observed for stimulation of the fingers. Thus, generally, lateralized negative deflections appear to reflect a common process of selective attention that integrates spatially-coded information (including sound, as revealed by an N2ac in anterior regions; Gamble & Luck, 2011).

Based on the evidence reviewed above, the present study employed a multi-modal, visuo-tactile search paradigm to re-examine, both behaviorally (examining reaction times and response accuracy) and electrophysiologically (analyzing the PCN and CCN as sensitive markers for visual and, respectively, tactile attention), the specific hierarchy of summative attentional-priority computation across feature, dimension, and modality levels as assumed by the MWA (see Figure 4–1). Critically, we focused on the pattern of target shift-cost effects to test whether the costs are indeed greater for modality shifts than for (intra-modal) dimension shifts, as expected on the MWA, which assumes an extra stage of modality weighting above a stage of intra-modal dimension weighting; or, alternatively, whether they are of comparable magnitude, as would be predicted by a parsimonious extension of the DWA to incorporate multi-modal search scenarios, which considers all stimulus dimensions equal in that a single (dimensional) weight resource is competitively shared among all dimension, whether they are of the same or of different modalities. In particular, on the MWA, we expected the switch costs to reflect the assumed hierarchy: they should be largest for modality-shifts (MSE), followed by (intra-modal) dimension shifts, and negligible for (intra-dimensional) feature shifts. In this study, these assumptions about intertrial effects are investigated empirically by using a visuo-tactile search paradigm. Furthermore, we analyzed the persistence of switch costs across successive trials, i.e., a decay function of the neural traces (Zylberberg et al., 2009). Also, we looked location-shift patterns between visual and tactile targets.

Besides examining the behavioral shift effects, we also analyzed the PCN and CCN components, expecting to find modulations in their timing and magnitude that reflect the hierarchy assumed by the MWA: they should become delayed and less negative with the (behavioral) increase in shift costs from the feature through the dimension to the modality level. Importantly, in contrast to higher amplitudes for modality shifts in the N1 and N2, indicative of a perceptual-coping mechanism (Gondan et al., 2007; Gramann et al., 2007; Töllner et al., 2009), less negative amplitudes in the PCN (as found in visual search (Eimer et al., 2010; Töllner et al., 2008) and CCN on target shift trials could be taken to reflect reduced allocation of attentional (weight) resources due to the carry-over, across trials, of neural traces from previously task-relevant relevant (i.e., pre-activated) dimensions or modalities. In contrast, target repetitions would ride on the same neural traces (accessing the same or similar information), thus making the deployment of attention more efficient – expressed in faster reaction times and earlier onsets and higher amplitudes of the lateralized components.

In summary, while there are many issues worth investigating in the context of inter-trial effects (such as visual dominance or the functional role of N1), the focus of the present study lies on the MWA as a multi-modal extension from the DWA. [insert location-shift and neural decay function here]

Accordingly, the main goal is to find a *supramodal* mechanism which combines processes of selective attention across the visual and tactile modalities (Farah et al., 1989; Lakatos et al., 2009; Spence, 2002). As outlined above, by

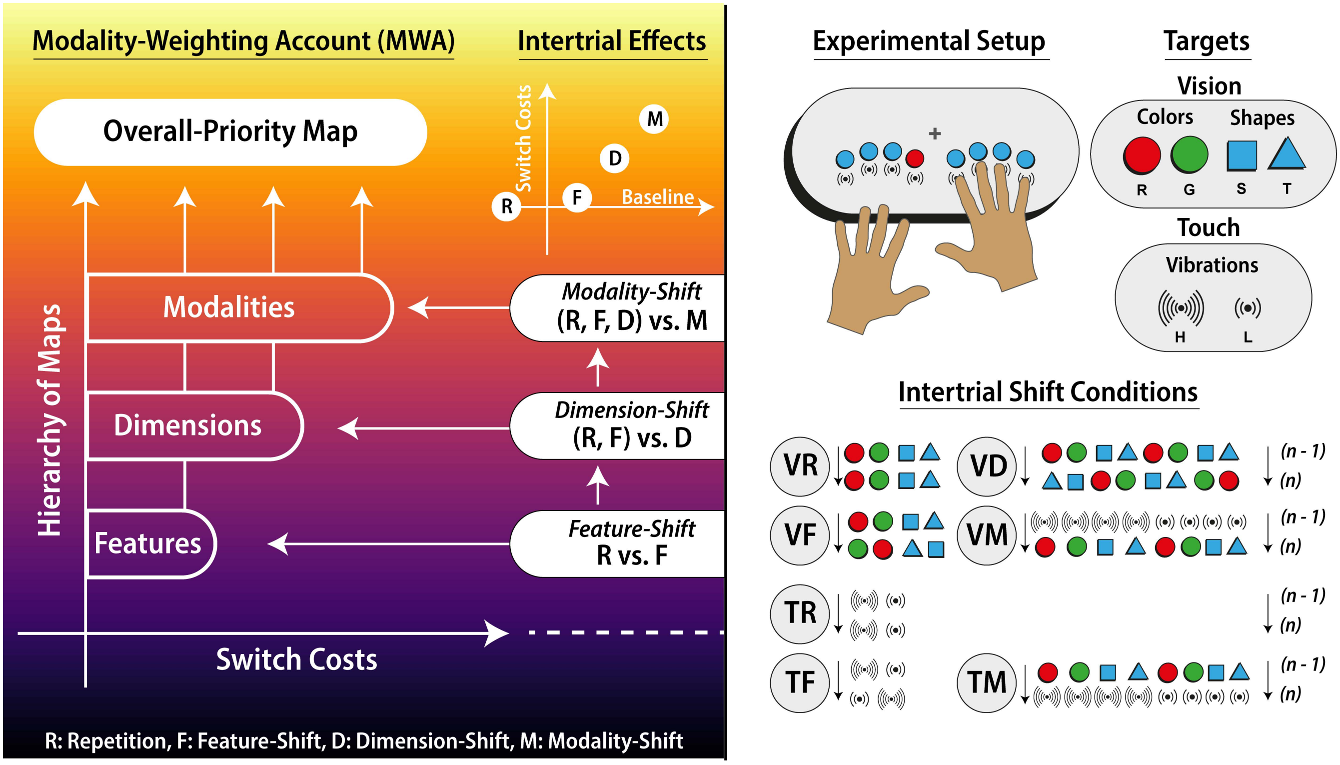

Figure 4–1. *Left Panel.* The modality-weighting account (MWA) is illustrated as a hierarchy of layers of spatiotopic maps, from feature through dimension to modality maps, that propagate their weighted signals upwards to an attention-guiding saliency (or 'attentional-priority') map. It is assumed that target switch costs increase with increasing hierarchy: (intra-dimensional) feature switching → (intra-modal) dimension switching → modality switching. For the analyses of intertrial effects, switch costs are derived by subtracting repetitions as baseline from the respective shifts (feature, dimension, or modality), or by conducting orthogonal contrasts within each layer of the MWA. *Right Panel.* The experimental setup is depicted, with the display items including collocated visual and tactile information. Participants were instructed to place their fingertips on the actuators, and to issue responses via foot pedals. Further, red circle (R), green circle (G), square (S), triangle (T), low (tactile) frequency (L, 20 Hz), high frequency (H, 55 Hz) were the six singleton targets in the visual and, respectively, tactile modalities, which popped out among homogeneous distractors (cyan circles, 200-Hz vibrations). The inter-trial target-shift conditions, i.e., the changes in the target definition from the previous (n–1) to the current trial (n), are illustrated (on the right) for all possible inter-trial conditions: visual target (VR = visual target, feature repetition, VF = visual target, feature change, VD = visual target, dimension change, VM = visual target, modality change) and tactile targets (TR, TF,

showing delayed latencies and reduced amplitudes for modality shifts, we anticipate the neural dynamics of the PCN and CCN to be indicative of such a common, cross-modal pool of attentional resources.

Methods

Participants. 18 participants were recruited for this study. All were right-handed and had normal or corrected-to-normal vision, and normal tactile sensitivity in their fingertips. Due to technical failure in EEG recording, one participant had to be excluded. Thus, 17 participants (mean age 25.9 years, four females) were included in the study. They gave their informed consent prior to the study and were reimbursed (9 €/h) for their participation. The study was approved by the Ethics Board of the Faculty of Pedagogics and Psychology at LMU Munich, Germany.

Stimuli and Procedure. The visuo-tactile setup (Figure 4–1) consisted of a transparent frame, with eight evenly spaced tactile actuators (Dancer Lab), and parallel visual projection from the backside (Optomo). Participants were seated in front of the setup, fixating their eyes at the center of the projection surface (eye-to-surface distance of approx. 57 cm), and placing their fingertips (excepting the thumbs) softly on top of the actuators. The actuators, delivering vibro-tactile stimulation, were arranged in a slightly curved fashion on each side, reflecting the lengths of the various fingers and so affording a comfortable hand/finger posture during the experiment. The visual items, projected onto the surface, were collocated with the actuators, to render the impression of an array of integrated visuo-tactile objects.

The experimental task was to localize any cross-modal (i.e., either a visual or a tactile) ‘pop-out’ target presented among seven homogeneous distractors on either the left or right side of the stimulus array display; participants indicated the side of the target by pressing the corresponding foot pedal. The pop-out item was defined by either a visual or a tactile feature contrast relative to distractors. In total, there were six targets: against the visual distractor context, composed of seven cyan circles, visual targets were defined in either the color dimension, being featurally either red or green, or in the shape dimension, being triangle- or square-shaped. Against the tactile distractor context, composed of 200-Hz vibrations, the tactile targets were defined in the (vibro-tactile) frequency dimension, being of a low, 20-Hz frequency or a high, 55-Hz frequency. The six targets were equally likely (and appeared equally likely at each of the 8 possible locations), with the target order across trials determined by De-Bruijn sequencing

(Aguirre et al., 2011). Whatever the target on a given trial, participants were instructed to produce a localization response to it as fast and accurately as possible. A trial started with the presentation of a central fixation cross for 500 ms, which was immediately followed by the search display consisting of a pop-out target at any position among the seven distractors. The search display was presented only briefly, for an exposure duration of 200 ms, to avoid eye movements. Next, the foot-pedal (left/right-localization) response was recorded. In case of an incorrect response, the visual feedback display 'Error!' was shown for 1000 ms. The inter-trial interval was about 1000 ms (varying within +/- 50 ms).

To balance the inter-trial conditions (which are in the focus of the present study), we applied De Bruijn sequencing (Aguirre et al., 2011), that is, a randomization procedure that ensured that all types of – to-be-examined – inter-trial target transitions from trial n–1 and to trial n were equally likely. In particular, we examined the following four types of inter-trial target *shifts*: *Repetitions* = *R* (e.g., square → square), *Feature shifts* = *F* (e.g., low → high), *Dimension shifts* = *D* (e.g., color → shape), and *Modality shift* = *M* (e.g., touch → vision) (s. Figure 4–1). In total, the experiment consisted of 728 trials, presented in 24 blocks of 72 trials each. Each target was repeated 288 times, equally distributed across all eight locations.

Performance Analysis. For behavioral performance, we examined the error rates (ER) and mean reaction times (RTs) in repeated-measures (RM) ANOVAs with the factors *Target* (*dimension*) and *Shift*. Throughout the Results, we report the generalized eta-square values (η^2, (Lakens, 2013) and depict within-subject confidence intervals (Loftus & Masson, 1994; Morey, 2008), unless indicated otherwise. Further, Greenhouse-Geisser correction was applied for violations of sphericity. For RTs, only correct responses within the time range 250–1200 ms were analyzed (i.e., response-error trials and trials with outlier RTs were excluded; see Results for details). For the analysis of *switch costs*, we took repetitions (R) of singleton target features (e.g. red → red) as baseline and subtracted this from feature shifts (F), dimension shifts (D), and modality shifts (M). For the *neural decay*, six two-way RM ANOVAs were conducted in reference for the preceding trials (*n*–1, *n*–2, *n*–3, *n*–4, *n*–5, & *n*–6), separately for *feature* (color, shape, & frequency) and *shift* (feature, dimension, modality). Again, shift conditions were subtracted from their respective baselines of exact repetitions. Also, *location-shift* between left and right hemifields and hands were analyzed across visual and tactile conditions, by utilizing RM ANOVAs.

EEG Recording and Analysis. The brain activity was recorded by 64 Ag-AgCl electrodes placed in the 10-20 system using BrainAmp and Brain Vision Recorder (BrainProducts™), at a sampling rate of 1000 Hz. During the experiment, impedances were kept below 5 kΩ. The EEG recordings were analyzed with Python and MNE (Gramfort et al., 2013, 2014), while statistics were conducted in R.

For the EEG preprocessing, initially, we applied a band-pass filter (0.1–70 Hz), including a notch filter to exclude line noise. Further, we conducted an extended infomax ICA and corrected eye blinks and saccades by inverse transformation of their respective independent components. Finally, the EEG signals were re-referenced by mastoid electrodes, and another 12-Hz low-pass filter was applied with zero-phase shift, in order to retrieve more interpretable signals further including alpha and theta oscillations which are both linked to selective attention (Keller et al., 2017; Michel et al., 2021).

Following this, events were segmented into 1-second epochs, including a 200-ms baseline interval. For the PCN and CCN, initially, epochs were averaged (ERPs) separately for left and right target locations, including an average baseline correction. In order to obtain a sufficient amount of EEG epochs (a high signal-to-noise ratio), here, we selected the shifts from the orthogonal contrasts within each hierarchical layer of the MWA (s. Figure 4–1), that is: in the first contrast, we combined EEG events (or epochs) for exact repetitions (high → high, triangle → triangle) and feature-shifts (e.g., red → green, low → high) for vision and touch respectively. In the second contrast (being only applicable for vision), we included all events (EEG epochs) for dimension-shifts (e.g., shape → color, color → shape) and "repeated dimensions" (e.g., color → color, shape → shape), subsuming both repetitions and feature-shifts. Finally, we analyzed modality-shifts (e.g., touch → vision, vision → touch) and modality repetitions (vision → vision, touch → touch), our third contrast, separately for each modality.

Finally, ipsilateral signals were subtracted from contralateral ones, where the PCN was averaged across extrastriate regions (PO7/PO8, PO3/PO4, P5/P6, O1/O2), and the CCN across central regions (C3/C4, C5/C6, C1/C2, CP3/CP4). To extract the amplitude and timing of these components, we applied the jackknife procedure for the PCN (200–340 ms) and CCN (140–340 ms) to derive the *amplitude* and *peak-latency* parameters; the timing of the *onset* and *offset* was chosen by the common 50% criterion (Kiesel et al., 2008; Miller et al., 1998; Stahl & Gibbons, 2004; Ulrich & Miller, 2001).

By nature, the PCN derives with a relatively low signal-to-noise ratio due to event-related lateralizations, making it difficult to determine the exact timing. To compensate for this circumstance, it is important to have a good signal-to-noise ratio by including many trials, and to apply the jackknife procedure. However, inaccuracies can still occur, since timing is closely associated with the amplitude criterion (50%). Further, which PCN time parameters are sensitive to experimental manipulations appears to vary across studies, being sometimes prominent in onsets (Eimer et al., 2010) and at other times in peak latency (Töllner et al., 2008; Töllner, Zehetleitner, Krummenacher, et al., 2011), although in each case an underlying temporal effect exists. For this reason, here, we introduce a more stable and 'novel' *latency* estimate, by simply averaging the three time points (onset, peak latency, offset). This step is intended to reveal temporal effects more sensitively, while also improving comparability.

For the statistics, jackknife amplitudes and latency scores were transformed into individual estimates, which approximate the real-timing scores, enabling statistical comparisons without further corrections (Smulders, 2010). Again, we applied RM ANOVAs, and paired t-tests for the contrasts in RT and PCN / CCN parameters for each level of the MWA. For the contrasts in vision, since trials are balanced, targets for colors and shapes are combined across shift conditions. Finally, we apply repeated measures correlations (Bakdash & Marusich, 2017) to uncover links between the PCN /CCN and behavior. The topomaps, which are presented in the results (s. Figures 4–6, 4–7, & 4–8) and show central and extrastriate activations of CCN and PCN, are averaged around 250 ms after stimulus onset.

Results

The main goal of this study is to investigate intertrial effects in a crossmodal search paradigm. Thereby, we assume that switch costs can be explained by the hierarchy of the MWA, where intertrial costs are negligible for *feature-shifts* (F), but increasingly prominent for *dimension-shifts* (D) and *modality-shifts* (M) (s. Figure4–1).

First, we report the error rates (ERs) and mean reaction times (RTs) for the six targets of the cross-modal localization task. Overall, the ERs were comparable within each of the two visual dimensions, color (red: 1.4%, green: 1.9%) and shape (squares: 2.7%, triangles: 3.8%) (*p*s < 0.133). For the vibro-tactile targets (low frequency: 14.4 %, high-frequency: 7.9 %, the ERs were higher compared to the visual targets and differed from one another ($p < 0.001$). Importantly, the ERs on inter-trial *shift* trials were overall comparable among the six targets (*p*s > 0.136), ruling out inter-trial biases in target accuracies. Comparing the ERs across dimensions (color, shape, vibration), an RM ANOVA revealed a significant Dimension effect (F (2, 32) = 38.72, $p < .001$, $\eta^2 = .542$): ERs were significantly higher (*p*s < 0.001) for vibro-tactile targets (10.6 %) compared to both color (1.7 %) and shape (3.3 %) targets, which were also different ($p = 0.036$) (s. Figure 4–2).

For the RT analysis, only correct trials were selected, and on average 2.1% of outliers (with RTs outside the 250–1200-ms window) were excluded. A target-based RM ANOVA revealed the Target main effect to be significant (F (5, 80) = 41.00, $p < .001$, $\eta^2 = .232$): RTs to color (red: 516 ms, green: 527 ms), shape (square: 564 ms, triangle: 568 ms), and vibro-tactile (low-frequency: 613 ms, high-frequency: 583 ms) target were all significantly different (*p*s < .001), apart from squares, triangles, and high-frequency vibrations, which were comparable to one another (*p*s < .072, *uncorr.*). A dimension-based RM ANOVA yielded a similar pattern (F (2, 32) = 41.58, $p < .001$, $\eta^2 = .216$), revealing color (521 ms), shape (566 ms), and vibro-tactile (597 ms) targets to have distinct processing times (*p*s < .006). The effects in RTs and ERs among dimensions are likely linked to bottom-up feature-contrast computations, which follow a hierarchy from color though shape to vibrations (s. Figure 4–2). Although there were some differences between target features (e.g., red and green), these appear to be largely consistent within their respective dimensions. Since the main focus of the present study lies on inter-trial target switch costs, we went on to examine for these effects omitting differences among dimensions in the analyses. Further, we only examined switch costs in the RTs (as there were no evident switch costs in the ERs).

Switch Costs. To start with, we conducted separate RM ANOVAs with the cross-trial target *Shift* (R = repetition, F = feature shift, D = dimension shift, M = modality shift) for visual (V) and tactile (T) target stimuli. For the former, there was a prominent main effect of shifts (F (3, 48) = 26.60, $p < .001$, $\eta^2 = .031$): the RTs for the various conditions were 532 ms (VR), 534 ms (VF), 542 ms (VD), and 555 ms (VM), that is, they increased in line with the hierarchy assumed in the MWA. The same pattern was also revealed for tactile stimuli: 578 ms (TR), 572 ms (TF), and 608 ms (TM) (F (2, 32) = 23.61, $p < .001$, $\eta^2 = .043$).

In order to take a more direct look at the shift effects and their underlying dynamics for each dimension, we calculated the switch costs (F, D, M) by subtracting repetitions (R) as a common baseline. As expected, no significant costs were found for (intra-dimension) feature-shifts (e.g., green → red) versus repetitions (e.g., red → red) in one-sample *t*-tests (t (16) < | 1.72 | *p*s > .258); that is, target changes at the feature level have little impact on performance, consistent with the MWA (See Supplementary Figure 4–A1 for a summary of the switch costs (ER, RT) for specific target features.) At the next level, (intra-modality) dimension shifts produced significant switch costs for both color targets (Δ 14 ms, t (16) = 4.75, $p < .001$) and shape targets (Δ 7 ms, t (16) = 2.65, $p = .018$).

Finally, modality-shifts produced costs for both color targets (Δ 24 ms, t (16) = 6.18, $p < .001$) and shape targets (Δ 21 ms, t (16) = 4.92, $p < .001$), with the modality-shift costs significantly exceeding the dimension-shift costs (ps < .002). Likewise, shifting the modality was costly for responding to tactile targets, whether they were preceded by color targets (Δ 31 ms, t (16) = 5.67, $p < .001$) or shape targets (Δ 28 ms, t (16) = 4.73, $p < .001$). This pattern, including the greater costs of shifts across modalities vs. shifts across dimensions within the same (here: exclusively the visual) modality is consistent with the key assumption – of a modality-specific priority-coding level – made by the MWA.

Of note, although the cross-modal shift costs (M) looked numerically somewhat larger for vibro-tactile targets than for visual targets (see Figure 4–2), they did not differ statistically (ps > .113), that is: we found no robust evidence of a bias towards the visual modality (or visual dominance), which would have produced significantly higher switch costs for touch vs. vision. The absence of such a bias is interesting especially when taking into account the fact that our design was skewed by default, with visual targets occurring twice as frequently as vibro-tactile targets. When applying a correction term for tactile vibrations (0.66), the modality-shift effects become even more similar (ps > 0.368). Given this, the present data provide little (and certainly no conclusive) evidence regarding the issue of visual dominance.

Location-Shift. In the location analysis, two significant main effects were found for ERs and RTs for the factor *Dimension* (F (2, 32) > 38.59, ps < .001, $\eta^2 > .524$), replicating the previous findings, and *Location* (F (1, 16) > 10.85, ps < .005, $\eta^2 > .014$), showing faster and more accurate response for right hand or hemifield locations instead of left locations. These accuracy benefits were most evidence in right hand tactile targets (s. Figure 4–3). The analysis of *location-shift* in vision and touch showed characteristic differences. Here, visual targets (VR: visual repetition, VF: visual feature-shift, VD: visual dimension-shift, VM: visual modality-shift) were significantly slower in RTs for repeated locations (F (1, 16) = 14.18, $p = .002$, $\eta^2 = .278$), the lack of an interaction (location-shift x switch costs) made it consistent across visual target ($p = .183$). In tactile targets (TR: tactile repetition, TF: tactile feature-shift, TM: tactile modality-shift), there was no main effect in *location-shift* (F (1, 16) = 3.72, $p = .071$, $\eta^2 = .140$). Yet, there was a significant disordinal interaction revealing TR and TF to be faster for repeated hands, while in TM the effect was reversed, tactile performance was faster when the preceding visual target occurred at the opposite side (s. Figure 4–4).

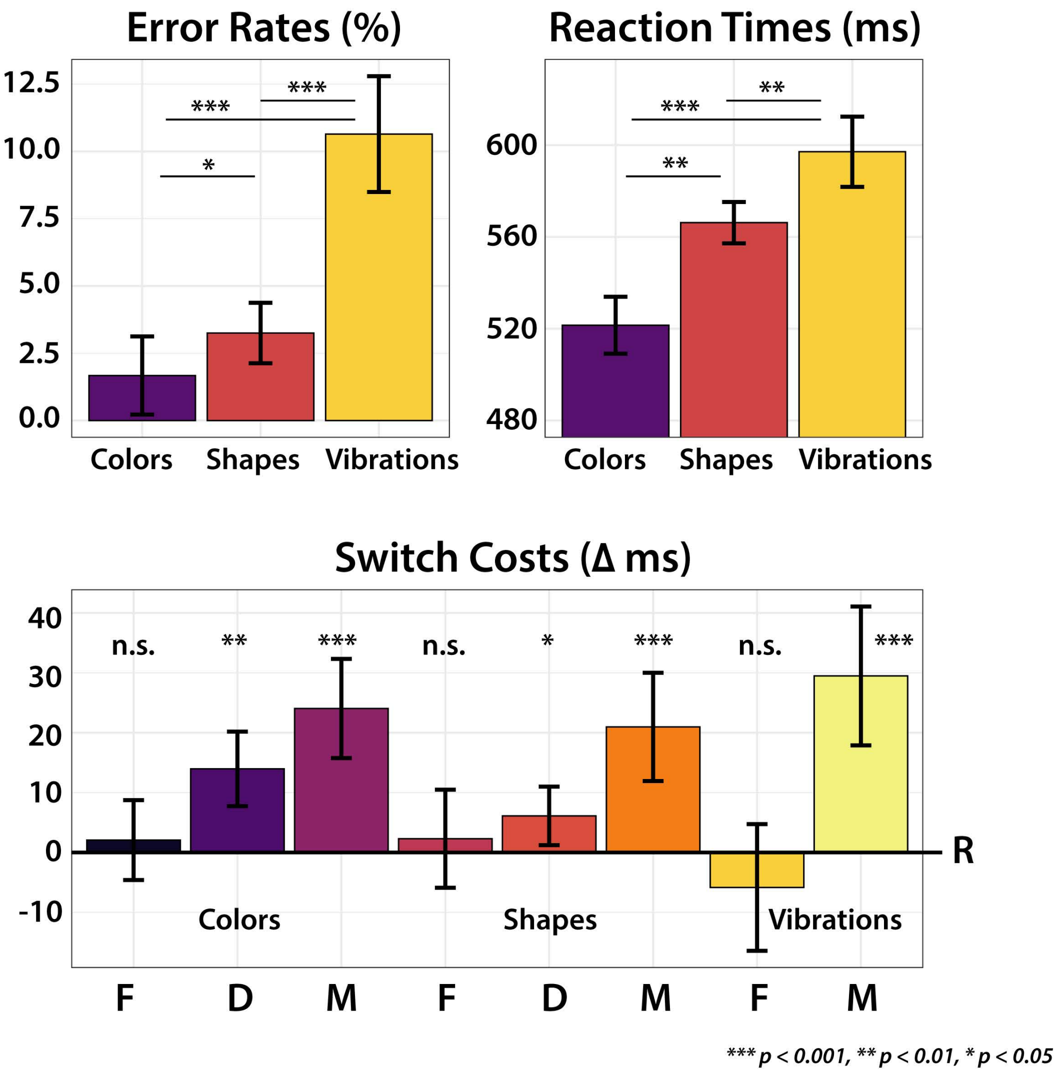

Figure 4–2. The upper panels present the error rates (ER, in percent) and reaction times (RT, in milliseconds) for color, shape, and vibro-tactile targets. The lower panel shows the switch costs separately for color, shape, and vibro-tactile targets, calculated by subtracting their respective feature-repetition (R) baseline from (intra-dimension) shifts in features (F), (intra-modality) shifts in dimensions (D), and shifts in the modality (M), respectively. For vibro-tactile targets, there could not be an intra-modality shift in D (as there was only one vibro-tactile dimension: frequency); further, M depicts the shift costs for vibro-tactile target averaged across shifts from (preceding) color and shape targets (which produced equivalent costs). Further, significance level is indicated by asterisks, or n.s. (non-significant), 95% confidence-intervals are added.

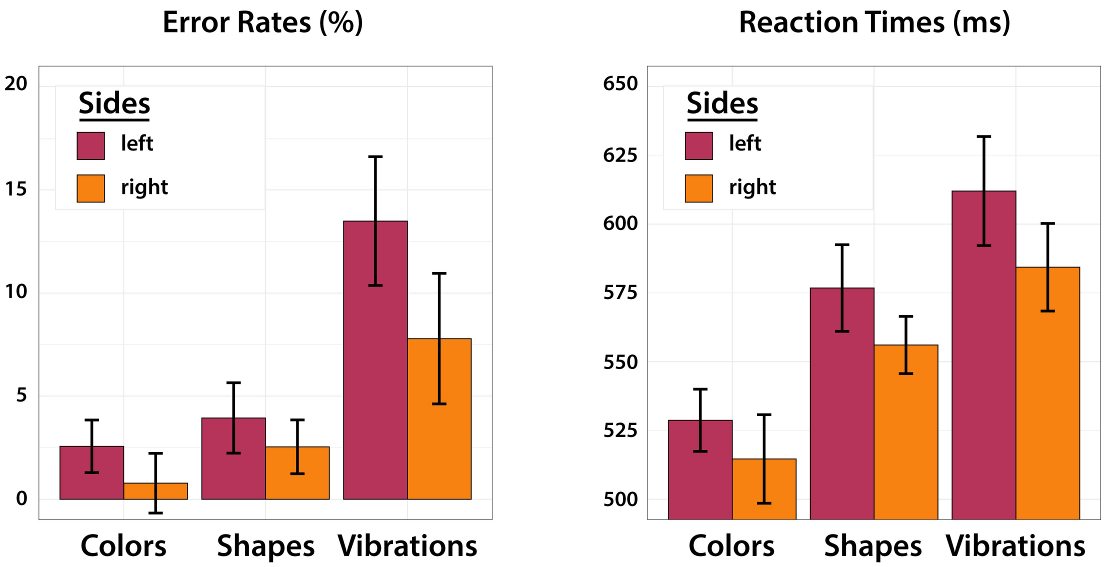

Figure 4–3. Error rates (in percentage) and reaction times (in milliseconds) are shown for left and right locations and sides across color, shape, and vibro-tactile targets. Overall, there was a systematic advantage for the right visual hemifield and hand, being significantly more accurate and faster than the left visual hemifield and hand.

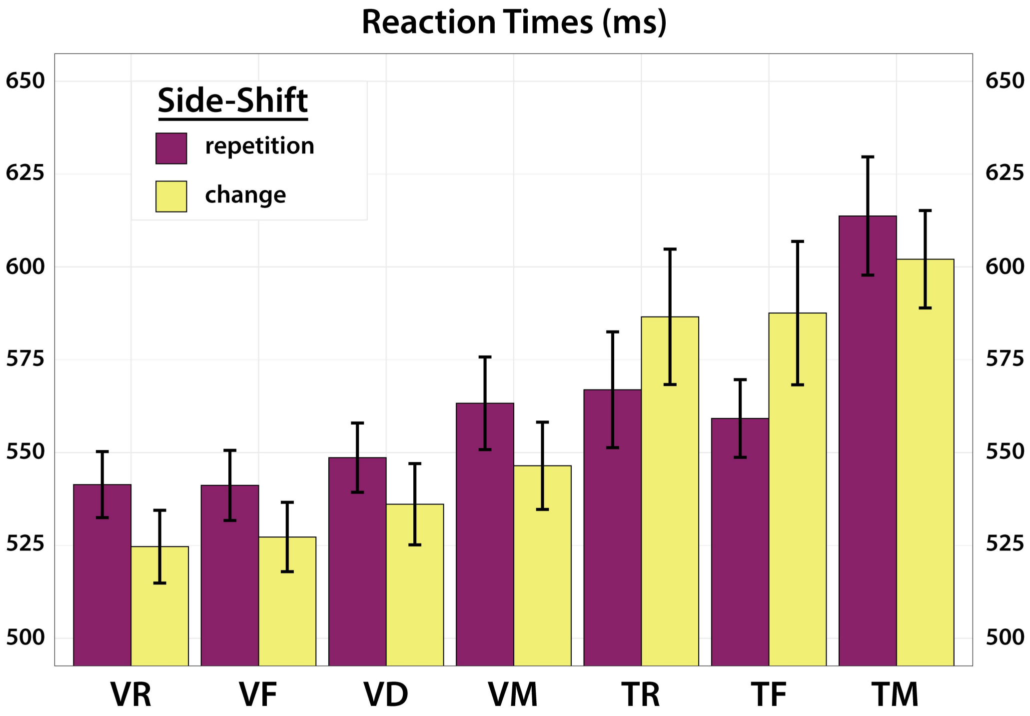

Figure 4–4. The *location-shift* analysis for repetition and change is presented for visual and tactile conditions is presented for reaction times (in milliseconds). Whereas visual targets were systematically faster for locations changes (e.g., left → right) in comparison to location repetitions (e.g., right → right) across all visual shift conditions (VR: repetition, VF: feature-shift, VD: dimension-shift, VM: modality-shift). On the contrary, tactile targets were faster for hand and tactile repetitions (TR, TF). Interestingly, a tactile modality-shift was better for changed locations (e.g., right → left) due to preceding visual information.

Neural Decay Function. The extended intertrial analysis found a convergence of switch costs, vanishing at *n*–4. Overall, the significant switch costs of modality-shifts persisted across preceding trials *n*–1 and *n*–2 across all dimensions (*p*s < .01), and for color and frequency in *n*–3 (*p*s < .05). Although dimension-shifts already vanished in *n*–2, feature-shifts in frequency and shape became substantial (*p*s < .05) (s. Figure 4–5).

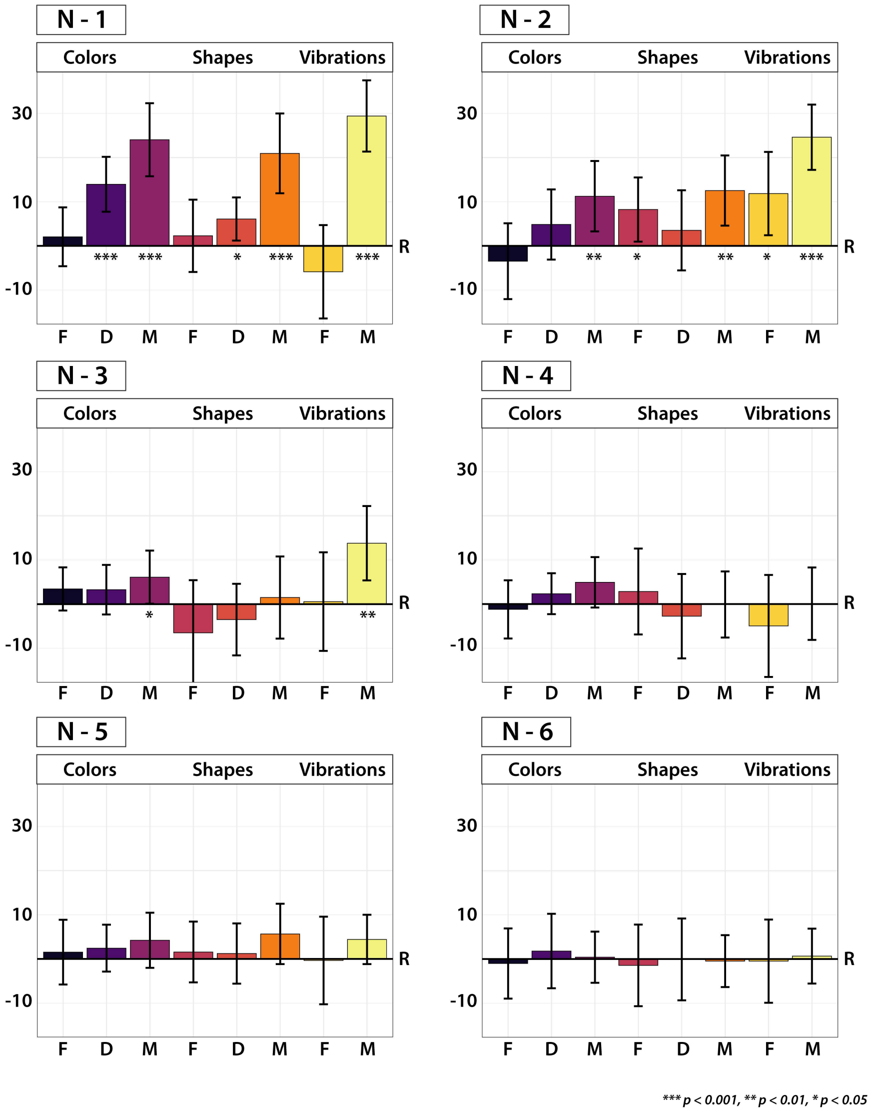

Figure 4–5. The *switch costs* (in milliseconds) are shown for the timeline of six preceding trials (*n*–1, *n*–2, *n*–3, *n*–4, *n*–5, & *n*–6). Again, the *shifts* for feature (F), dimension (D), and modality (M) are subtracted from identical repetitions for color, shape, and vibro-tactile targets, representing the delta of switch costs. The within-subject confidence intervals are shown, significant switch costs are indicated with asterisks.

Contrast 1: Feature Shifts. In line with the MWA, RTs were not significantly affected by (intra-dimension) feature shifts vs. repetitions (F vs. R), for either the visual or vibro-tactile targets (vision: Δ 2 ms, t (16) = 0.72 p = .481; touch: Δ -5 ms, t (16) = 1.17, p = .258). That is, no switch costs arise at the level of the feature maps. This finding is also reflected in the PCN and CCN for visual and vibro-tactile targets, respectively: neither the latencies nor the amplitudes of the respective components differed between F and R (ps > .163) – in line with the MWA (see Supplementary Figures 4–A2 & 4–A3).

Contrast 2: Dimension Shifts. For the contrast at the dimension level, the two shift types at the (visual) feature level were combined and compared to a (visual) dimension shift (F, R. vs. D). In RT, an (intra-modal) dimension shift produced significant costs over and above an (intra-dimension) feature shift or repetition (Δ 9 ms, t (16) = 6.26, p < .001). An analogous, temporal effect was found in both the PCN onsets (F (1, 16) = 7.90, p = .012, η^2 = .065), which were delayed for dimension shifts (223 ms) vs. dimension repetitions (213 ms), and the PCN *latencies* (F (1, 16) = 5.44, p = .033, η^2 = .041; 256 ms vs. 263 ms). No further effects were found in the PCN timing and amplitude (ps > .155) (see Figure 4–6).

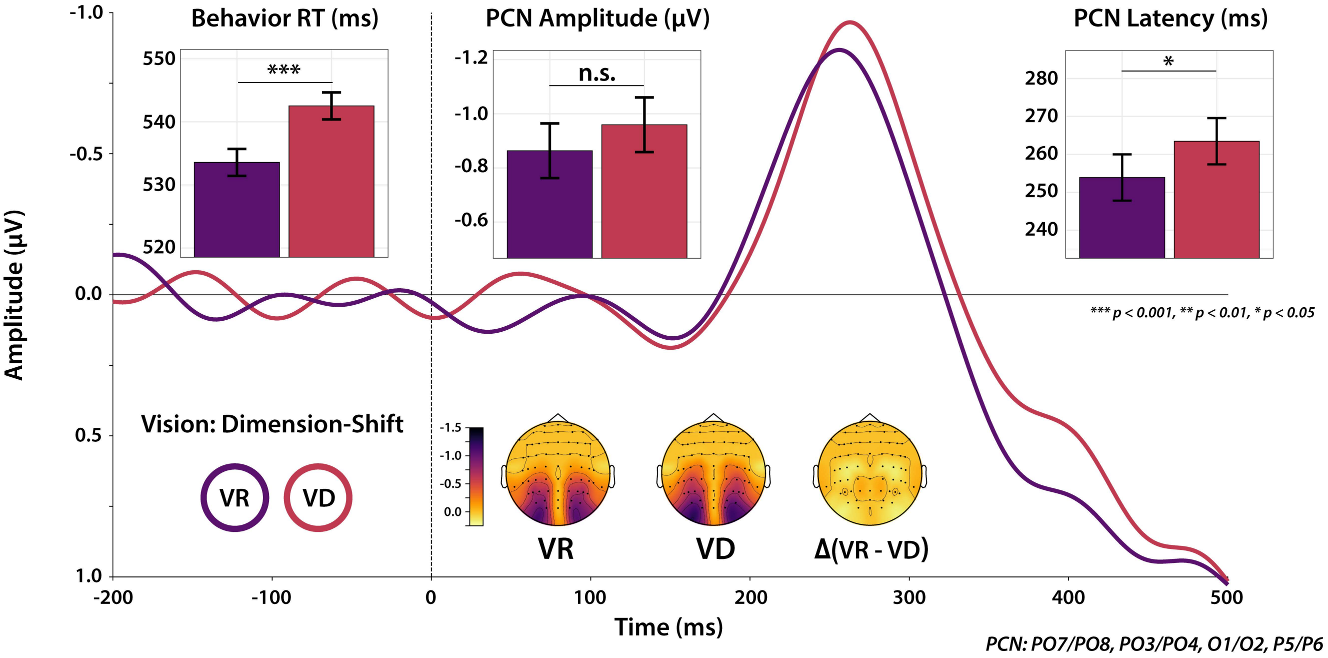

Figure 4–6. The PCN for visual targets as contrasts in the dimension layer is shown. VR (purple) denotes dimension repetitions (including both repetitions and feature shifts) and VD (fuchsia) dimension shifts. The PCN is presented in the center; the topomaps underneath (averaged activity in the time window 240–260 ms) depict the central and extrastriate activations. The bar plots present descriptives of the contrast in reaction times (RTs), PCN amplitude, and PCN latency, including within-subjects confidence intervals. Findings are either marked as non-significant (n.s.) or significant (indicated by asterisks). Overall, visual dimension-shifts produced a significant delay in reaction times and PCN *latency*, while PCN amplitude was

Contrast 3: Modality Shifts. Moving up the hierarchy of the MWA, comparing modality shifts vs. modality repetitions (R, F, D vs. M), there were significant RTs costs both for switching from vibro-tactile to visual targets (Δ 18 ms, t (16) = 6.11, $p < .001$) and for switching from visual to vibro-tactile targets (Δ 32 ms, t (16) = 6.24, $p < .001$). For visual targets, the PCN amplitude was significantly reduced (F (1, 16) = 12.63, $p < .002$, $\eta^2 = .023$) for modality shifts vs. intra-modal repetitions (-0.71 µV vs. -0.96 µV). And, while there was little difference in PCN onsets (p = .438; modality shifts vs. intra-modal repetitions: 222 ms vs. 218 ms), both the PCN latencies and the PCN offsets were significantly delayed on modality-shift vs. -repetition trials (*latencies:* 266 ms vs. 260 ms, F (1, 16) = 4.76, $p = .044$, $\eta^2 = .031$; *offsets:* 309 ms vs. 301 ms; F (1, 16) = 6.12, $p = .025$, $\eta^2 = .017$). However, the delay caused by a modality shift was not larger than that of an (intra-modal) dimension-shift for visual targets (p = 0.332). For vibro-tactile targets, we also found a prominent amplitude effect in the CCN (F (1, 16) = 5.06, $p = .039$, $\eta^2 = .010$): the CCN amplitudes were less negative for shifts from visual to vibro-tactile targets (–1.05 µV vs. –1.28 µV); there was no modality-shift effect on any of the CCN timing parameters (ps > .323). Interestingly, the tactile CCN-amplitude effect was also preserved in the PCN (F (1, 16) = 7.12, $p = .017$, $\eta^2 = .058$), albeit being less negative in the latter component (modality-shifts vs. -repetitions: –0.50 µV vs. –0.76 µV). Given that the timing of the tactile CCN (*onset*: 157 ms; *latency*: 239 ms) precedes the tactile PCN (*onset*: 198 ms; *latency*: 255 ms), it would appear that the tactile component originates over central regions and propagates towards extrastriate regions, which is associated with a reduction in its negativity reduces (see Figure 4–7 & 4–8, and Supplementary Figure 4–A4).

In summary, finding the same negative reduction for modality shifts in the PCN and CCN amplitudes favors a supramodal account, indicative of reduced attentional resources for processing a given target modality (say touch) when a cross-trial target change requires switching to this modality from another modality (in the example, vision). The timing, however, appeared to be relevant only for the PCN elicited by visual targets: here, there was a temporal delay in the latency estimate, which however did not differ between dimension shifts and modality-shifts. In order to examine whether these findings in amplitude and latency are both linked to the behavioral RTs, we conducted repeated-measures correlations. Overall, we found a significant (moderately) positive correlation, with higher negativity in the PCN being associated with faster RTs (r (101) = .407, $p < .001$). Also, we found another (moderately) positive correlation for temporally sensitive visual shifts (VR, VF, VD, VM), with a shorter PCN latency being linked with faster RTs (r (50) = .421, $p = .002$).

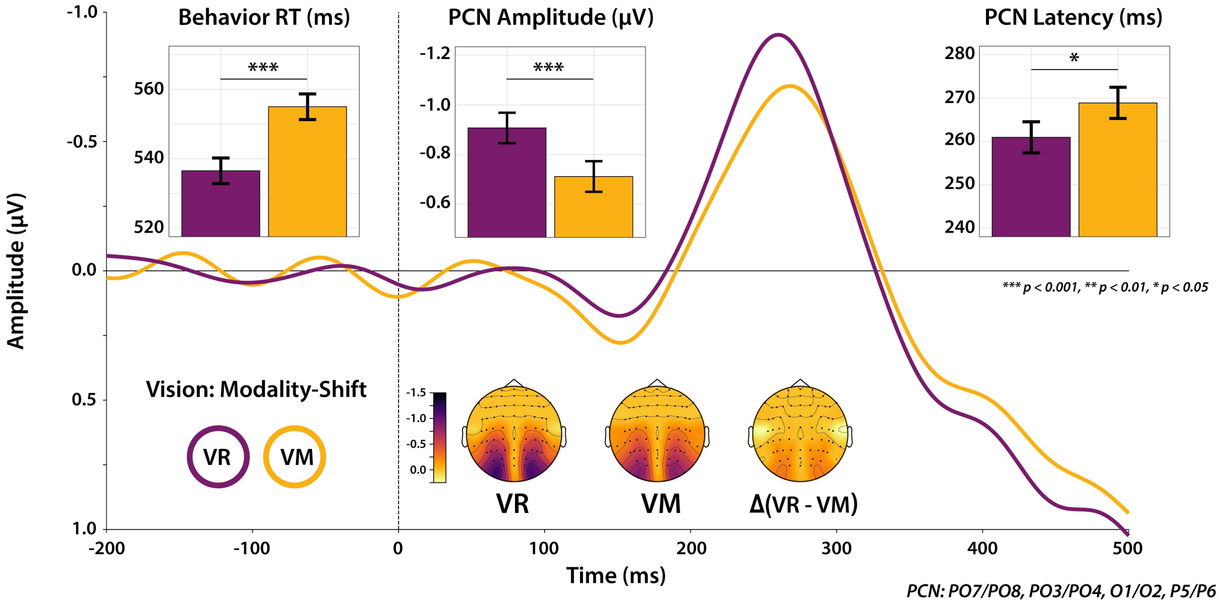

Figure 4–7. The PCN for visual targets as contrasts in the modality layer is shown. VR (purple) denotes dimension repetitions (including all intramodal repetitions) and VM (yellow) modality shifts. The PCN is presented in the center; the topomaps underneath (averaged activity in the time window 240–260 ms) depict the central and extrastriate activations. The bar plots present descriptives of the contrast in reaction times (RTs), PCN amplitude, and PCN latency, including within-subject confidence intervals. Findings are either marked as non-significant (n.s.) or significant (indicated by asterisks). Overall, visual modality-shifts produced a significant delay in reaction times and PCN *latency*, and a reduced PCN *amplitude*.

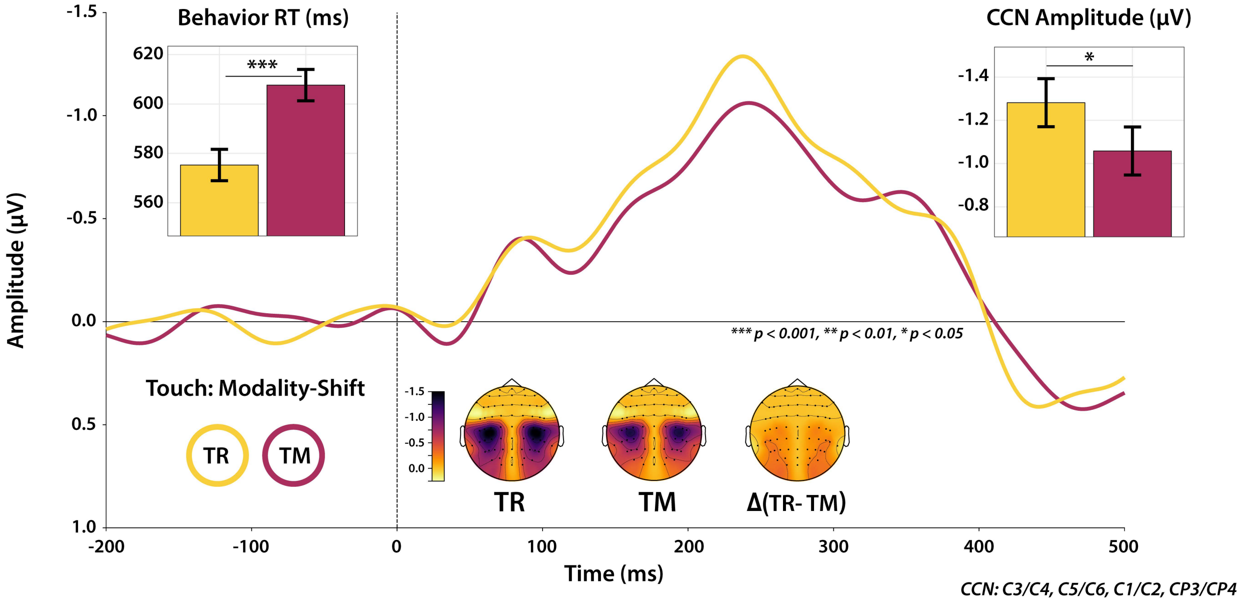

Figure 4–8. The CCN for tactile targets as contrasts in the modality layer is shown. TR (yellow) denotes dimension repetitions (including all intramodal repetitions) and TM (purple) modality shifts. The CCN is presented in the center; the topomaps underneath (average activity in the time window 240–260 ms) depict the central and extrastriate activations. The bar plots present descriptives of the contrast in reaction times (RTs) and CCN amplitude (within-subject confidence intervals), significance is indicated by asterisks. Overall, tactile modality-shifts produced a significant delay in reaction times and CCN *amplitude*.

Discussion

The present – cross-modal search – study was designed to investigate the hierarchy governing inter-trial and switch-cost effects (Found & Müller, 1996; Spence, Shore, et al., 2001; Theeuwes, 2010; Töllner et al., 2009). To do so, we combined evidence from the MSE (Miles et al., 2011; e.g., Spence, Nicholls, et al., 2001; Töllner et al., 2009) and related effects in (uni-modal) visual search (Krummenacher et al., 2009; e.g., H. J. Müller et al., 1995; Pollmann et al., 2000) guided by the hierarchical architecture of attentional-priority computation envisaged by the MWA proposed by Töllner et al. (2009), which forms the theoretical backbone of our study. Within its hierarchy of layers (see Figure 4–1), the MWA predicts that switch costs are largest for modality shifting (MSE), intermediate for (intra-modal) dimension shifting, and negligible for (intra-dimensional) feature shifting.

In terms of behavioral – RT – performance, the pattern of inter-trial effects we found is consistent with this assumed hierarchy, for both vision and touch: there were no reliable RT costs for switching (as compared to repeated) the target-defining feature within a given dimension; significant shift costs emerged when switching from one to another dimension (as compared to staying within the same dimension) within the same modality (here only tested for vision); and the shift costs were significantly higher when switching from one to another modality (as compared to staying within the same modality) (s. Figure 4–2). This pattern is in accordance with previous studies (Found & Müller, 1996; Krummenacher et al., 2009; Mortier et al., 2005; H. J. Müller et al., 1995). In other words, while dimension and modality shifts produce costs, feature shifts are handled as efficiently as repetitions, and with respect to switch costs can also be considered to produce benefits (Jonides & Mack, 1984; Spence, Nicholls, et al., 2001). In general, this pattern argues in favor of the MWA as a valid extension of the DWA to multi-modal search scenarios. In particular, the MSE reflects a true supra-modal effect that cannot be readily explained by accounts that assume just a layer of dimensions-specific saliency coding (without differentiating stimulus modalities), the outputs of which feed directly to the attention-guiding priority map; instead, the outputs of the dimension maps are first integrated per modality (modality layer) and then forwarded, in a competitively weighted fashion, to the priority map (s. Figure 4–2, Supplementary Figure 4–A1).

Before discussing our PCN and CCN findings, it is interesting to briefly consider what appear to be basic RT (and ER) differences between the three target dimensions in the present study: color, shape, and (vibro-tactile) frequency. These differences are suggestive of a processing hierarchy (RT, ER) reflecting the respective target's physical saliency (s. Figure 4–2). The advantage for colors over shapes most likely derives from the fact that both are processed differently in the visual cortex (Brincat & Connor, 2004; Gegenfurtner, 2003; Stettler et al., 2002), insofar the color encoding precedes that of shape in the ventral stream (Perry & Fallah, 2014). This would also explain why switch costs for dimension-shifts were less prominent in shapes, as the allocation of attentional resources follows the visual stream. In contrast, switching from shapes to colors becomes more costly, since attention must be retracted against the direction of its visual hierarchy.

Compared to visual targets, responding to vibro-tactile targets was delayed (and more error-prone) – indicating that they took longer to become salient than visual targets. While this would be in line with the (presumed) dominant role of vision (Spence, Nicholls, et al., 2001; Spence, Shore, et al., 2001), there are alternative explanations of the visual bias in the present study. First, participants were a-priori less familiar with tactile search (a scenario only introduced in the experiment, whereas looking for an odd-one-out color or shape corresponds more with participant's everyday life experiences); and second, the visual/tactile target ratio was by default unbalanced, with twice as many visual- than tactile-target trials. Yet, the modality-shift effects appear to be comparable between vision and touch. Accordingly, a clear conclusion about visual dominance is beyond the scope of the present study; and, more generally, how the dominance of one or another task-relevant modality (in a multi-modal scenario) is modulated by probabilistic weighting of sensory modalities remains to be investigated in future work.

Another interesting fact, which captures the gap between saliency computation and intertrial history, even though not too obvious, may be gleaned from the effect sizes (generalized η^2) in the performance (RT) analyses. While the effect-size for *dimensions* was quite large (0.216), effect-sizes for a main-effect in *shifts* were quite small in both modalities, vision (0.031) and touch (0.043) respectively, which underscores the subordinate and more variable role of inter-trial effects. That is, while bottom-up saliency computations can be considered as a first-order process in which the various feature analyzer units are activated (scaled according to the feature contrast to the units in their surround) by the current search array, the inter-trial effects reflect a second-order modulation of these signals according to the weight settings (i.e., the residual neural traces) carried over to the current from the previous trial(s).

Further evidence revealing the persistence of modality-shift across sensory modalities comes from the neural decay analysis. Here, it was shown that the residual activations from visual or tactile 'neural traces' produce the most-persisting switch costs across multiple trials (Mannuzza, 1980; Zubin, 1975; Zylberberg et al., 2009). This finding indicates that attentional resources are bound to the preceding sensory modality (vision or touch) and take time to detach completely. For future studies it would be interesting to investigate whether these neural traces can be described as an exponential decay. Another interesting shift pattern was revealed in location shifts. Whereas visual targets was generally improved for location changes, switching between left and right visual hemifields, performances for tactile targets was improved for hand repetitions. This finding indicates a spatial dissociation between vision and touch, which can be explained by inhibition-of-return' (Klein, 2000; Posner et al., 1985; Taylor & Klein, 2000) or by somatosensory pre-activation (Assumpção et al., 2018; Badde et al., 2015). Interestingly, switching from touch to vision or vice versa was always better for the opposite hemifields. This indicates that spatial orienting might be primarily guided by visual processes and thus can be regarded as evidence for visual dominance (Klein, 1977; Posner et al., 1976; Spence, Nicholls, et al., 2001).

Beyond the behavioral inter-trial effects, the present study focused on the neural dynamics governing the switching between targets within and across the modalities of vision and touch. While the pattern of behavioral (RT) results clearly favor the MWA (with the MSE reflecting a supra-modal cost function), the critical question was whether the hierarchy of inter-trial shift effects would also be manifest in two components linked to attention selection: the PCN and CCN. Given that the interplay of neuronal activations (present vs. residual) reflects distribution of dimension- and modality-specific processing resources, we expected dimensional (intra-modal) and cross-modal shifts of the target from the previous to the current trial to impact the attention-related components, due to the carry-over of the neural traces (i.e., relative balance of the respective pathway activations established as result of inter-trial history).

Importantly, there were no significant RT switch costs for intra-dimensional feature shifts vs. feature repetitions – in line with previous findings in visual search (e.g., Found & Müller, 1996; Töllner et al., 2008) and the MWA (Töllner et al., 2009); likewise, feature shifts (vs. repetitions) did not impact the dynamics seen in PCN and CCN for either visual or vibro-tactile targets. In contrast, for dimension shifts vs. dimension repetitions in the visual modality (in our design the only modality in which we could examine for such an effect), the PCN was delayed – evidenced by significant differences in both the PCN *onset* (+10 ms) and *latency* (+7 ms) parameters. Again, this finding is in line with (Töllner et al., 2008), who found a delay of 8 ms. Of note, the delay in PCN latency and the RT switch costs (+9 ms) are near-equivalent.

Finally, and theoretically most important in the present context, there was also a delay in the PCN latency for modality shifts (from vibro-tactile) to visual targets, as compared to visual-target repetitions; although this delay was robust only in the PCN *offsets* (+8 ms), the effect was significant for our aggregate *latency* estimate (+6 ms). Of note, the *latency* estimate was consistently (and significantly) delayed for both types of visual shift (i.e., for shifts from vibro-tactile targets both to color targets and to shape targets), strengthening the validity of this finding with regard to the temporal dynamics of the (visual) PCN. That this delay effect was not obtained consistently in both the PCN *onset* and *offset* measures (even though both derive from the same, 50% criterion) is likely owing to the relativeness of the 50% criterion which derives from different amplitudes in modality-shift and modality-repeat conditions. Given this fragility of the PCN in combination with its rather low signal-to-noise ratio, a valid PCN time criterion is essential. We propose that our aggregate *latency* estimate provides a better solution than any single timing measure: it reduces random variability, thus becoming more consistent.

While no temporal effects were found in tactile targets and the CCN, the timing effects in PCN were statistically prominent and modality-specific, here, the (visual) PCN latency did not differ between dimension shifts within the visual modality and modality shifts from the tactile to be the visual modality. Even though our findings of systematic PCN delays are in line with the hierarchy of the MWA and linked to the behavioral switch costs, as indicated by the significant correlations (i.e., revealing slower RTs for lower amplitudes and temporal delays in PCN), the PCN delay effects do not reflect a genuine supra-modal mechanism due to the absence of analogous temporal effects in the tactile CCN.

As for the amplitudes of the PCN and CCN, we found a consistent amplitude reduction for modality shifts in both vision and touch, with the PCN and CCN, respectively, becoming less negative. For tactile targets, the (central) CCN appeared to propagate towards the (posterior) PCN, while preserving its amplitude effect (i.e., the reduction) in extrastriate regions. Of note, there was also a significant correlation between amplitudes and RTs, that is, an association of the amplitude and switch costs in intertrial effects (as indicated by slower RTs). Thus, overall, the reduced component amplitudes can be explained in terms of the attentional resources, or ‘weight’, being reduced for a given modality following a shift (to it) from another modality – indicative of a truly supra-modal mechanism that generates costs for modality switching or, conversely, benefits from modality repetitions. In terms of the MWA, this mechanism reflects the competitive weighting of the two alternative target modalities driven by inter-trial history: the respective target-defining modality is upweighted on a given trial *n–1* and the non-defining modality down-weighted; accordingly, when the target on the subsequent trial *n* is defined in the previously non-defining modality, its weight in the priority computation is reduced and consequently it is processed with fewer attentional resources compared to a modality repetition. In other words, the amplitude results support the MWA as a suitable extension of the DWA (from intra-modal) to cross-modal search.

For visual dimension-shifts, there was no significant difference in PCN amplitudes. There may be two reasons for this. One is that the task itself may play a modulatory role: in a crossmodal-search scenario, for the participants, the major demands arise from having to handle the processing of targets defined in separate modalities, as a result of which intra-modal processes may become attenuated and so less prominent. The second is that, in visual search, clear evidence for amplitude modulations due to dimension shifts is still not fully resolved as differences in amplitude might simply derive from temporal dissociations of early vs. late PCN effects (cf. Eimer et al., 2010) (in contrast to higher amplitudes for redundancy gains, which have been shown several times, e.g., Töllner et al., 2011).

In summary, our findings suggest that the hierarchy of the MWA provides a good fit to explain the cross-trial dynamics of selective attention. As temporal effects in the visual PCN latency appear to be consistent by showing a delay for shifts across dimensions and modalities, but no effects in CCN latency, thus being modality-specific, we cannot declare a favorite hierarchy as both the DWA and MWA seems to apply. However, we found a supra-modal mechanism in amplitude modulations, favoring the MWA. Interestingly, the tactile CCN propagates from central towards extrastriate regions, while the visual PCN does not propagate in the reverse direction. For parietal regions, these findings may suggest not only a visual but also a supra-modal role in attentional shifts. This reflection on a common pool of attentional resources relates to previous findings and the debate of whether inferior parietal cortex and temporo-parietal junction are linked to crossmodal attention (Chambers et al., 2007; Chambers, Stokes, et al., 2004; Hopf et al., 2000; Quinn et al., 2014).

As the evidence to favor a MWA suggests, it should be the cause for deeper analyses. Therefore, it is worthwhile to consider some thoughts for future investigations. First, the concept of neural traces, as a gradient or second-order process, might be interesting to investigate, as it seems to reflect a function of contrast but also distance between perceptual analyzer units and sensory modalities, which perfectly captures the purpose of the MWA. As crossmodal attention seems to be linked to parietal regions, being very close to visual areas might explain the impact and presence of visual dominance. Also, because only vision produces temporal effects in PCN, which genuinely describes intertrial effects, producing benefits or costs in time.

Appendix

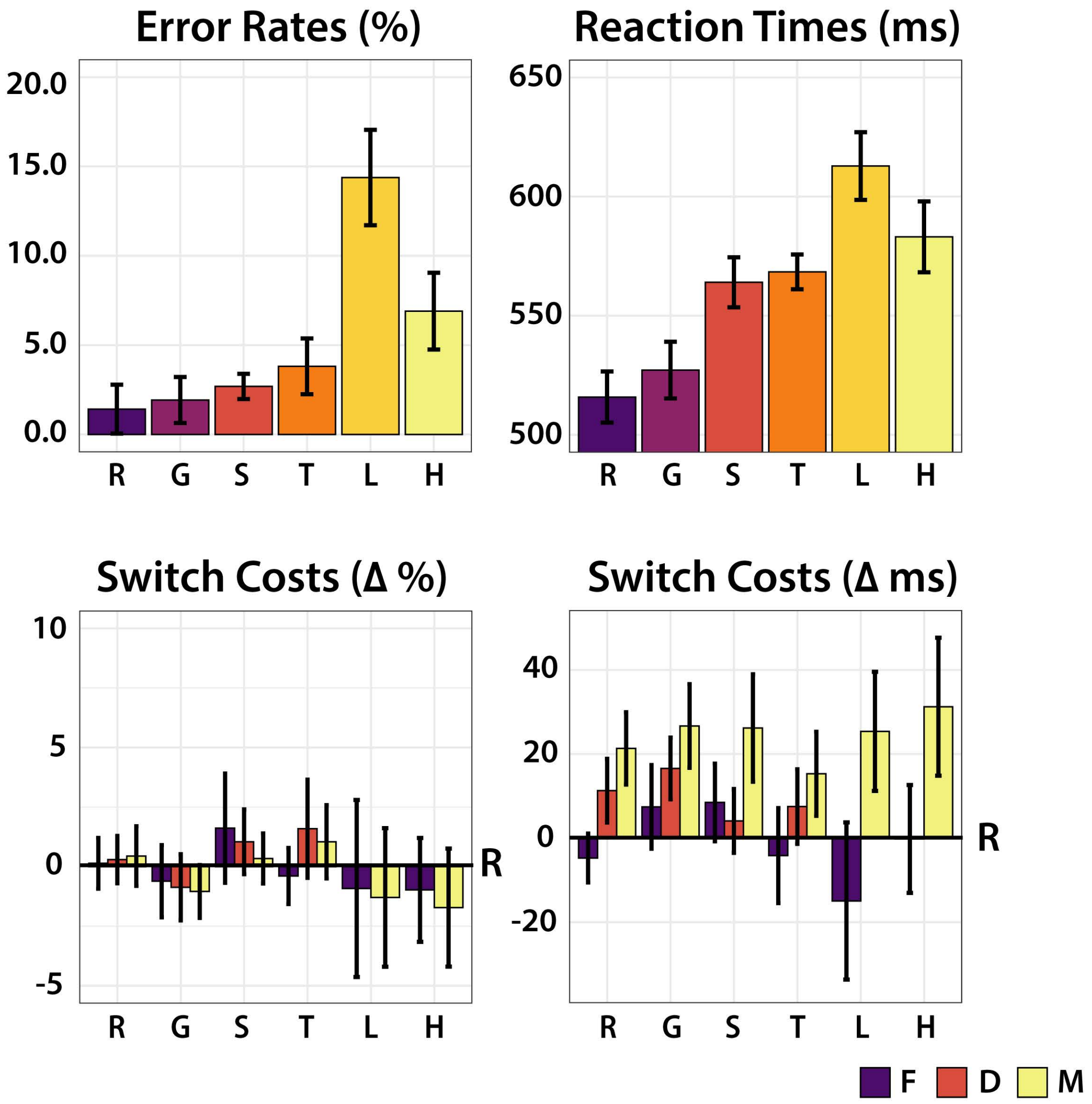

Figure 4–A1. Error rates (in percentage) and reaction times (in milliseconds) are shown for each of the six features (color: R: red, G: green; shape: S: square, T: triangle; vibration: L: low, H: high) and their respective switch costs (delta to repetition baseline) of feature-shift (F), dimension-shift (D), and modality-shift (M). Notably, tactile targets were least accurate and slowest and switch costs for modality-shifts consistently high.

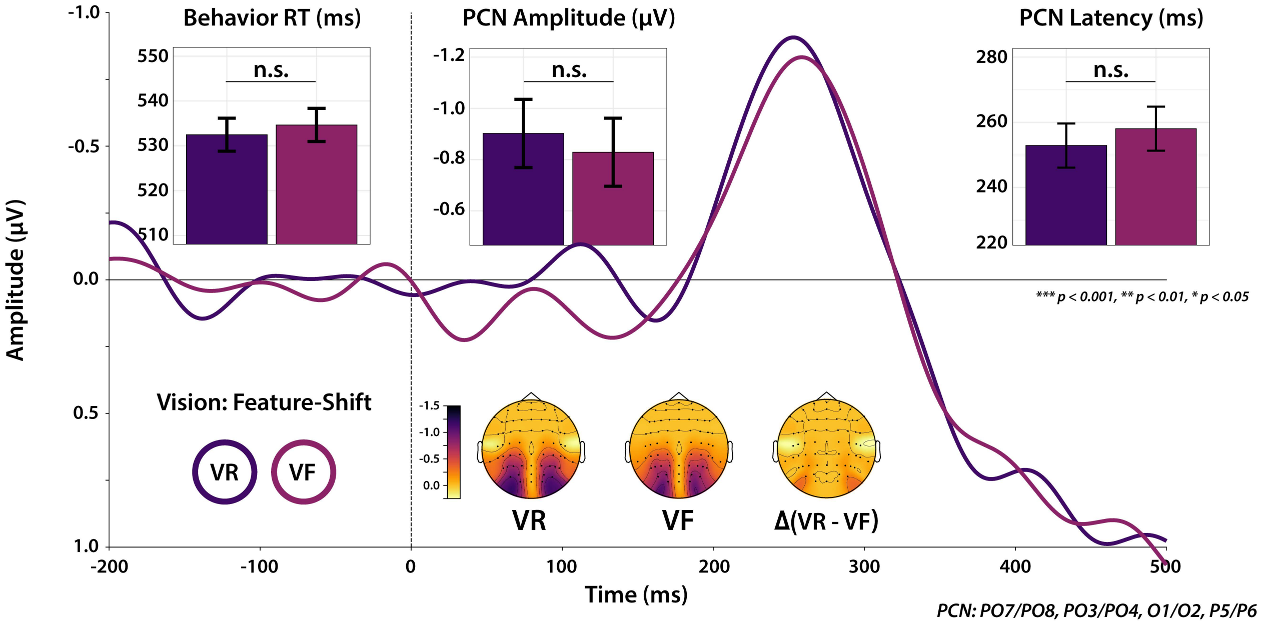

Figure 4–A2. The PCN for visual targets as contrasts in the feature layer is shown. VR (purple) denotes exact repetitions and VF (fuchsia) feature shifts (e.g., red → green). The PCN is presented in the center; the topomaps underneath (averaged activity in the time window 240–260 ms) depict the central and extrastriate activations. The bar plots present descriptives of the contrast in reaction times (RTs), PCN amplitude, and PCN latency, including within-subjects confidence intervals. Findings are either marked with asterisks or non-significant (n.s.). Overall, visual feature-shifts were comparable in reaction times an, *PCN latency*, and PCN *amplitudes*.

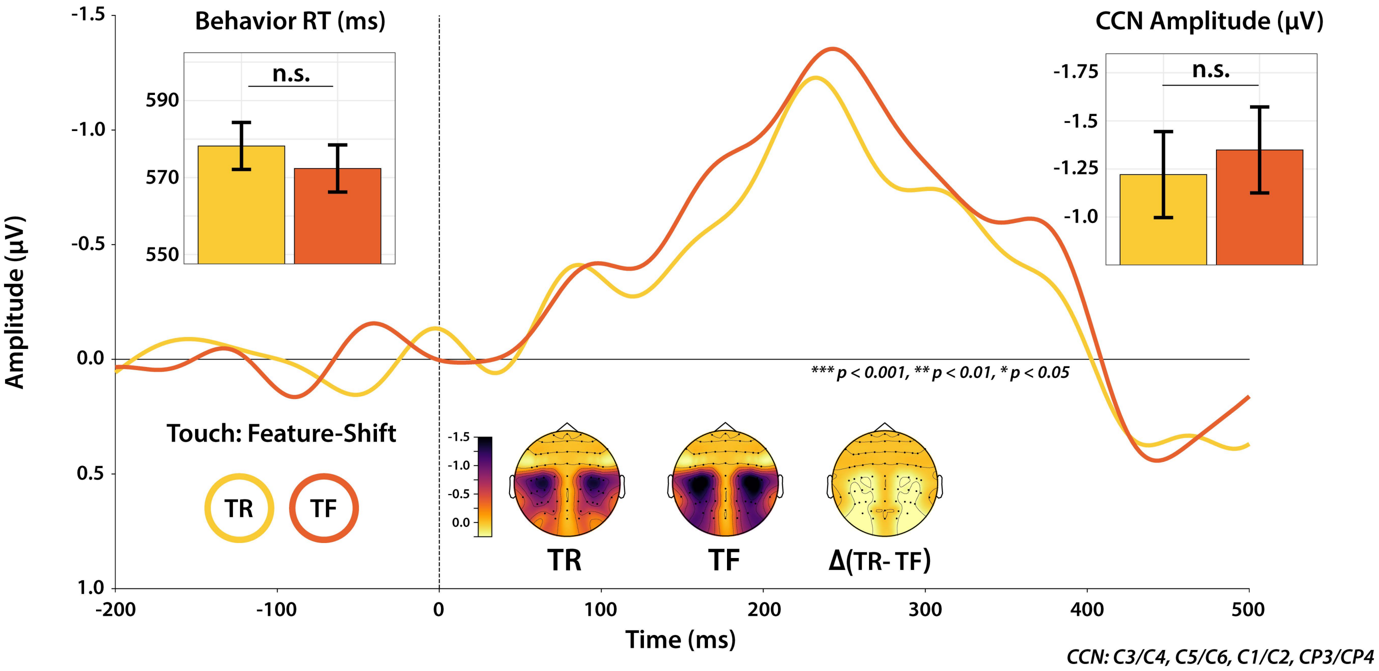

Figure 4–A3. The CCN for tactile targets as contrasts in the feature layer is shown. TR (yellow) denotes exact repetitions (including all intramodal repetitions) and TF (orange) feature shifts (e.g., low → high). The CCN is presented in the center; the topomaps underneath (average activity in the time window 240–260 ms) depict the central and extrastriate activations. The bar plots present descriptives of the contrast in reaction times (RTs) and CCN amplitude (within-subject confidence intervals), significance is indicated by asterisks or by n.s. (non-significant). Overall, tactile feature-shifts were comparable in reaction times and CCN *amplitude*.

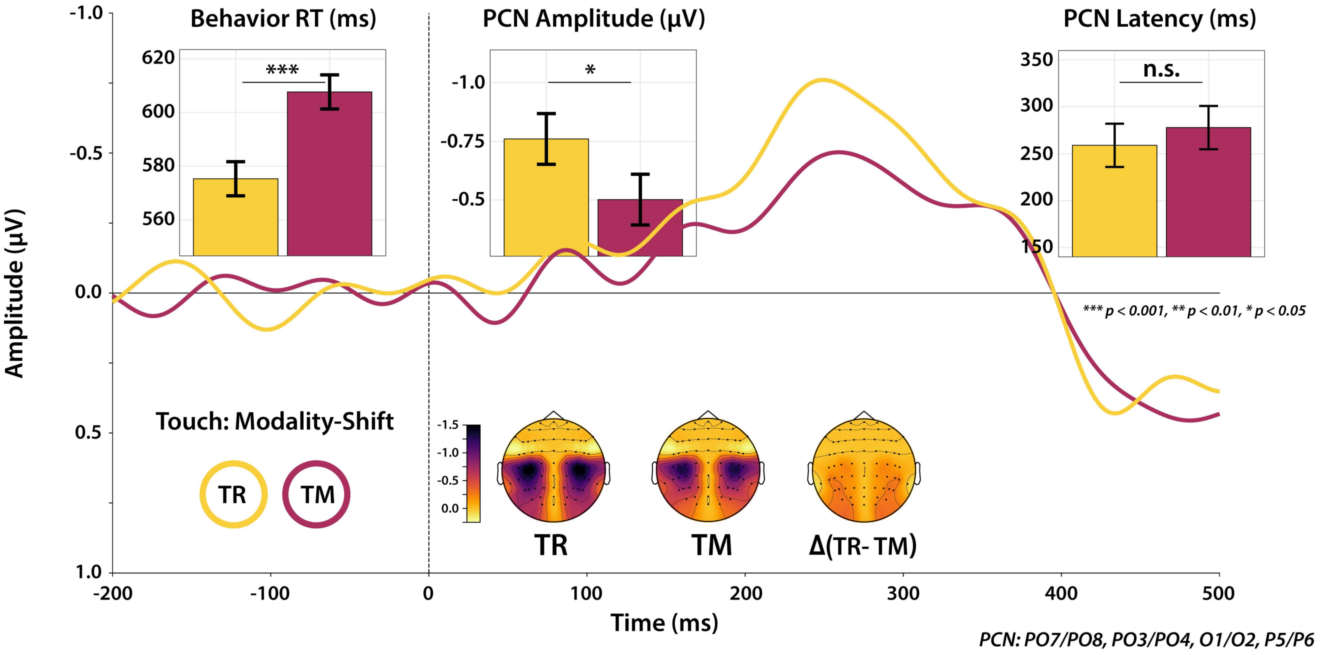

Figure 4–A4. The CCN for tactile targets as contrasts in the modality layer is shown in PCN electrode locations (PO7/PO8, PO3/PO4, O1/O2, and P5/P6). TR (yellow) denotes exact repetitions (including all intramodal repetitions) and TM (purple) modality shifts (e.g., red → high). The CCN is presented in the center; the topomaps underneath (average activity in the time window 240–260 ms) depict the central and extrastriate activations. The bar plots present descriptives of the contrast in reaction times (RTs) and CCN amplitude (within-subject confidence intervals), significance is indicated by asterisks or by n.s. (non-significant). Overall, tactile modality shifts were significantly delayed in reaction times and decreased in tactile PCN (or extrastriate CCN) amplitudes.

5 General Discussion

Marbles click and climb
While a journey's end
Shines lost in gold.

After presenting the main empirical studies, the following sections will summarize and discuss their findings based on the proposed preattentive framework: the modality-weighting account (MWA) and further link them to the superordinate research questions from the General Introduction (Chapter 1). The 'General Discussion' outline is as follows: The presentation of empirical evidence from the visuo-tactile search paradigms is based on how much the MWA is favored over an extended dimension-weighting account (DWA). These considerations will focus primarily on the findings in PCN and CCN components, the markers of choice for visual and tactile attention. Here, obtaining a clear picture of the modality-specific and supramodal functions of PCN and CCN will be essential. Based on those facts, the concept of a "supramodal priority network" (in line with the MWA) will be discussed in detail, which implies preattentive coactivations. Further, how the evidence contributes to a (general) computational attention model (CMA) will be discussed. Also, the future challenges of attention research are (roughly) outlined, and the "Dynamic Search" paradigm is proposed. Finally, with a fair degree of certainty, an answer to the central (decade-long) question of whether crossmodal attention derives from "common or separate pools of selective attention" is given.

Modality-Weighting in Crossmodal Search

First, it is vital to get an overview of the central findings and how they relate to the preattentive architecture of modality-weighting. Remember, the MWA proposes an extra layer of modality maps with modality-specific saliency computation that precedes the attentional guidance of a (supramodal) priority map during crossmodal search. In contrast to an extended (multisensory) DWA, which suggests that independent analyzers are comparable within and across sensory modalities, the MWA assumes a dissociation between intramodal (dimension-wise) and crossmodal (modality-wise) processes (Töllner et al., 2009). Hence, the main objective was to provide evidence for the theoretical proposition by conducting crossmodal studies that include visuo-tactile search paradigms.

Based on the scientific background, the redundant-signals effect (RSE) (Diederich & Colonius, 2004; Forster et al., 2002; Krummenacher et al., 2002; Miller, 1982; Töllner et al., 2011) and the modality-shift effect (MSE) or intertrial switch costs (Found & Müller, 1996; Miles et al., 2011; Spence, Nicholls, et al., 2001; Töllner et al., 2009) proved to be particularly suitable for empirically investigating the (expected) difference between MWA and DWA. The MWA predicts that crossmodal redundancies and switch costs exceed intramodal ones, producing substantially higher benefits but also costs in response behavior and during attentional selection. The experimental manipulations in the studies on redundancy gains (Chapter 2) and intertrial effects (Chapter 4) examined the preattentive layers of (visual) dimensions and modalities across distinct task conditions. The distinction between the 'modality condition' and the (visual) 'dimension condition' was central to both studies.

The study on redundancy gains (RGs) in Chapter 2 gave strong empirical support in favor of the MWA by showing substantially faster RGs in crossmodal targets (CF, SF, or CSF) than in the intramodal visual redundant target (CS), providing a clear answer to the first superordinate question (1). This outcome is strengthened by the common baseline between nonredundant targets (C, S, & F), ensuring equal saliency, and the extended feature analysis (s. Figure 2–2, & 2–A1). Further support for unveiling an RG hierarchy stems from studying spatial attention (Chapter 3), where crossmodal benefits exceed intramodal limitations. This finding closes the scientific gap between (the co-existing) visual and crossmodal search by showing a systematic dissociation between intramodal dimensions and sensory modalities.

Further, the analyses of neural dynamics in PCN and CCN provided evidence that this systematic dissociation (already) arises during the preattentive stage. Again, the PCN and CCN reflect neural markers for visual and tactile "spatial filtering" and attentional orienting in parieto-occipital and central regions, respectively. Previous studies showed that the PCN modulates sensitively in timing and magnitude based on the outcome of preattentive computations, indicative of the processing speed and the amount of allocated attentional resources (cf. Töllner et al., 2011). The CCN timing and amplitude were consistent across tactile and visuo-tactile targets, reflecting a stable somatosensory component for tactile attention and is, therefore, insensitive to depict the preattentive hierarchy of the MWA. By showing faster onsets and higher amplitudes for crossmodal redundant targets and earlier offsets in visual redundant targets, the PCN, on the contrary, revealed itself to be a suitable candidate for detecting modality-specific (visual) but also supramodal (visuo-tactile) activity. This finding is in line with visual search studies and the evidence of crossmodal links within the posterior parietal cortex (PPC), e.g., in IPS or TPJ (Downar et al., 2000; Eimer et al., 2010; Grubert et al., 2011; Macaluso et al., 2000, 2002a, 2002b; Man et al., 2015; Quinn et al., 2014; Wolber & Wascher, 2005). Although the outcome of substantial crossmodal benefits in the PCN answers the second superordinate question (2), verifying its functional relevance by excluding the option of simple linear superposition was crucial.

Another (central) claim of the MWA is the presence of coactivation during the preattentive stage, favoring an early instead of a late locus of coactivation (Feintuch & Cohen, 2002; Töllner et al., 2011; Zehetleitner et al., 2009). Based on the assumption of a "supramodal priority map or network", the unimodal saliency computations from independent sensory systems (vision and touch) need to converge in advance to enable crossmodal prioritization, which (arguably) occurs during the stage of multisensory integration (Talsma & Woldorff, 2005). Providing evidence for the presence of coactivation in response rates, e.g., using the 'maximum negative dependency' test (Colonius & Diederich, 2017), was essential to examine neural racers in the PCN component. Again, the 'maximum negative dependency' as well as the 'race-model inequality' (RMI) (Miller, 1982) measure violations of the independent racer assumption, yet, as stated before, the RMI is very conservative in that the absence of violations does not preclude the presence of coactivation (Liesefeld et al., 2017). Exactly this was observed for redundant signals. While no violations of the RMI were present, the 'maximum negative dependency' test obtained several significant enhancements in SF, CSF, CR (and TR) but not in CS or CF.

Reasonably, the approaches to measure those violations (have to) differ across behavior and PCN, yet they rely on the same assumption of independent racers. By considering the 'independence' of distinct feature activations (C, S, & F), the neural activity in PCN onsets for redundant signals should be either comparable (SF = S + F) or exceed (SF < S + F) their respective feature sums temporarily. Overall, the study on redundancy gains (Chapter 2) found a consistent pattern of (neural) race-model violations for crossmodal redundancies (except for CF) that is precisely the same as in the behavior (s. Figure 2–2, Figure 2–6, & Table 2–A2). Due to previous knowledge, the coactivation evidence can be associated with crossmodal links in the PPC within the timing of multisensory integration (e.g., Eimer & Driver, 2000; Macaluso et al., 2002a, 2002b; Talsma & Woldorff, 2005). Generally, the finding of preattentive coactivation reflects a decisive contribution to the (still) ongoing "locus of coactivation" debate (Feintuch & Cohen, 2002; Zehetleitner et al., 2009). Essentially, it favors the MWA by fulfilling one of its central claims: the assumption of combined (or co-active) loads across modality maps that precede the supramodal selective attention processes, culminating in the PCN dynamics.

Although the presence of preattentive coactivations answers the third and last superordinate question (3) for the factor of object features, several aspects must be considered.

Of course, more evidence from future studies is needed to corroborate the neural-race proposition. For example, it would be interesting to analyze the reason for the functional dissociation between shape-frequency (SF) and color-frequency (CF), being either with or without coactivation. Based on the topographic maps (s. Figure 2–8), it might be the case that the structural distance between feature-coding areas matters. For example, color-coding in V1 of the primary visual cortex (Li, 2002) is more distant to the tactile-coding somatosensory cortex than shape-coding areas in V4 or IT of the parietal cortex (Goolsby et al., 2005). Therefore, the lack of coactivation evidence might derive from less (direct) connectivity between color and frequency areas, in contrast to shape and frequency. This assumption is consistent with previous studies (Wolber & Wascher, 2005), and the main results and topographic pattern showing most negative deflections of a shape PCN in parietal electrodes (e.g., P5/P6) and of a color PCN in parieto-occipital electrodes (e.g., PO7/PO8). Furthermore, the feedforward direction of the visual cortex likely explains the consistent PCN delay between 'basic' color and 'more complex' shape targets, i.e., PCN latency in color precedes those of shapes (Δ 25 ms). Examining the PCN projections with structural information on the various visual and crossmodal feature activations would be interesting in a future study.

Furthermore, the substantial predictive impact of the PCN parameters on behavior, as shown by linear mixed-effect models, explaining up to 25 percent of the variance (s. Table 2–A3), draws a sharp distinction between PCN onset and offset. Crucially, onsets were faster for crossmodal redundant embeddings (CF, SF, & CSF, or CR & TR), and offsets were faster for visual redundant embeddings (CS, CSF, or VR & TR). This outcome indicates two separate and distinct functional processes within the PCN component: modality-specific for visual information and supramodal for visuo-tactile information. The fact that two aspects support the PCN shows 'early supramodal engagement' during its onset timing.

The first aspect is that the CCN component generally precedes the PCN and propagates toward parietal regions to co-register the somatotopic signals with external visual information. If no visual target is present, i.e., solely homogenous nontargets, visuo-tactile co-registration is not possible, leading to the purely tactile PCN. On the contrary, supramodal engagement occurs for visuo-tactile targets (e.g., SF & CSF). Here, the somatotopic tactile information aligns with the firing of relevant receptive fields due to the visual 'pop-out' feature contrast (promoted by location coherence) (cf. Macaluso et al., 2000). Second, this early supramodal engagement is again supported by the evidence from parietal crossmodal links (Man et al., 2015; Quinn et al., 2014) and coactivation in PCN onsets exceeding the linear superposition solutions.

Another issue is the PCN deflection itself during the visuo-tactile detection task, which arguably was a bit low (< 1μV), especially in the three visual targets (C, S, CS). Although the large visual angle of the display size (around 40°) might play a role, it is more plausible that the included number of seven targets is the reason for that, as (visuo-tactile) attention resources (or weights) were likely distributed among them. This conclusion is similar to Grubert et al. (2016), who found a PCN attenuation for the included number of color targets: single (- 3μV), two (- 2μV), or three (-1.5 μV) colors, which became even lower during variable color blocks. Although the detection task was primarily bottom-up driven and in singleton search mode, i.e., finding the "odd-one-out", the observed visual PCN amplitudes align with the expected decay of resources. This finding suggests that participants obtained implicit knowledge during the mixed-block design about the relevant and pseudo-random targets (Aguirre et al., 2011). It can also be that the amplification of crossmodal redundancies alters the firing modes of receptive fields, which adaptively reduces unimodal signal processing.

Finally, the relevant role of the PCN offset (for visual redundancies) gives some room for speculation. While the onset (and peak latency) indicate the waxing of selective attention processes, reflecting the outcome of preattentive computations, the offset might indicate the waning after attentional deployment, or the transition efficiency based on visual discrimination. This PCN concept relates to 'feature-selection' and 'feature-inhibition' processes (Treisman & Sato, 1990). Initially, the PCN selects (or guides attention to) the most salient information (via up-weighting or prioritization); only then irrelevant display (or scene) aspects like nontargets or distractors are suppressed (via down-weighting or inhibition). Based on this reasoning, inhibition is (or must be) a consequence of selection, which is in line with the Pd findings of active suppression that occurs after PCN timing within 290–340 ms (Feldmann-Wüstefeld et al., 2021; Hickey et al., 2009; Hilimire et al., 2012; Sawaki & Luck, 2010; Wyble et al., 2015). Since no distractors were shown in this paradigm, those distinct dynamics in the PCN indicate the optimization of the signal-to-noise ratio (SNR), similar to a broad "spatial tuning function". Again, providing empirical evidence for these speculative thoughts (e.g., by analyzing additional Gaussian properties such as the width – the temporal difference between onset and offset) would be exciting.

The following study on spatial attention (Chapter 3) complements the results on crossmodal redundancy gains (Chapter 2) by adding an in-depth spatial decomposition for each of the seven targets. Given the fact that visual sensitivity decays from the fovea toward the periphery, which is known as the eccentricity effect (Carrasco & Frieder, 1997; Staugaard et al., 2016), mean reaction times can actually be decomposed into foveal and peripheral slopes via quadratic fitting across spatial locations. On the other side is somatosensory or tactile integration, which is most sensitive in the hands, the "fovea of touch" (Forster et al., 2016), and presumably flat (or constant) across fingers. Although vision and touch rely on two distinct spatial functions, external and somatotopic, respectively, based on evidence from crossmodal cue-target pairs (Föcker et al., 2010; Gray et al., 2009), an interaction of the spatial functions can be assumed.

The initial behavioral analysis confirmed the quadratic and constant functions of vision and touch and their interaction within crossmodal targets (CF, SF, CSF). The spatial decomposition made redundancy gains for those crossmodal redundant targets much more explicit, revealing a genuine and persistent pattern between feature contrasts of color and shape. Whereas color discrimination was (as expected) optimal in the fovea (or central) slope, shape discrimination was substantially better in the peripheral slope. Fascinatingly, the visual redundant target (CS) adopted the optimal combination of both by combining the focal color slope and the peripheral shape slope yet remaining within these visual boundaries. Also, the spatial function in the purely tactile target (F) was non-quadratic and thus constant within each hand. Overall, the finding of two modality-specific spatial functions in vision and touch, quadratic and constant, answers another superordinate question (4).

As described, intramodal (visual) redundancies remained within their (spatial) limits; on the contrary, crossmodal redundancies exceeded those same limits by amplifying the focal and bending the peripheral slope, causing redundancy gains with accelerated reaction times. Intriguingly, the same pattern of redundancy benefits due to focal color and peripheral shape was repeated in crossmodal targets. This intricate finding of self-similar feature mappings led to the proposition of a vector space (s. Figure 3–6). In this vector space, distinct features can be described as eigenvectors based on their spatial coordinates (focal and peripheral slope). This vector space is likely suitable to predict the physical limits of each (visual) feature, e.g., due to vector intersections. Also, this vector space might be feasible to detect and describe coactivation areas, that is, a specific spatial configuration or a (pre-defined) range of spatial slopes. Of course, this novel eigenvector proposition needs empirical validation from future studies. Yet its utility and impact for visual and crossmodal search may be considerable, which will be made more explicit with the following example.

By calibrating the visual or crossmodal scene into a two-dimensional spatial representation similar to the priority map, which also considers the fixation center, feature activations can be localized based on their coordinates. According to the findings in Chapter 3, every feature activates a linear combination of spatial slopes (creating a vector), which, similar to Bayesian priors, can be estimated by a sufficient number of occurrences. Those vectors represent the unique identifier of each feature, e.g., a blue color or a left-tilted orientation, whose similarities or differences can be compared statistically. This “feature vector space” enables in-depth examination and accurate prediction of complex visual displays containing various feature sets, including discriminative states of coactivation and statistical facilitation. It allows comparison across samples and participants and thus promotes sparse sampling. Finally, it would be interesting to know whether these spatial priors are stable or dynamic, opening a new perspective on feature integration dynamics.

Besides the convincing evidence in behavior, the findings in PCN and CCN were congruent. Here, PCN amplitudes in visual targets showed a quadratic decay toward the periphery; the CCN for tactile targets was constant. Overall, this finding has particular value as it recognizes the two modality-specific spatial functions of external and egocentric mapping during the timing of attentional selection. While the CCN accumulates information from the “fovea of touch” across fingers, the PCN amplitude vividly represents the firing sum of relevant receptive fields, decreasing toward the periphery. Further, this outcome provides strong evidence for the presumed spatio-topographic processes that combine relevant object features with spatial locations (Chambers et al., 2004; Koch & Ullman, 1985; Treisman & Gelade, 1980; Wolfe, 1994). The finding of spatial functions in PCN and CCN answers the fifth question (5).

Although the answer to the sixth research question is already half given: the self-similar pattern of crossmodal redundancy gains due to focal amplification and peripheral bending. However, the evidence from PCN and CCN still needs to be included. While the quadratic decay in PCN amplitudes was constant across all six visual targets (C, S, CS, CF, SF, & CSF), the PCN amplitudes were generally amplified for crossmodal targets (CF, SF, & CSF). This result reveals a decisive difference between behavior and the PCN, which showed no signs of peripheral bending within crossmodal targets. Therefore, the firing rates of receptive field clusters appear unaffected for crossmodal redundancies, suggesting that the PCN's peripheral slope reflects a modality-specific process based on external visual mapping. However, the amplification, adding a constant tactile term, reveals a supramodal effect in PCN amplitudes. This presumably preattentive effect is underscored by the evidence from coactivations and neural racers from Chapter 2. Of note, the coactivation between vision and touch (in SF and CSF) becomes visible in the significant quadratic slopes of SF and CSF in the CCN (s. Figure 3–5). Although it is a minor finding, some considerations can be made from it. Arguably, crossmodal binding occurs in the PPC and can be observed in PCN amplification. Yet, coactivation from active crossmodal links might be detectable in the CCN, similar to a feedback function. Based on this consideration, which again needs further empirical evaluation, another claim can be drawn from visual inspection of focal or central activations in CF and CSF: they seem to exceed the other locations, which suggests that coactivation in color can only be detected in focal regions but vanishes in peripheral locations.

This claim also calls for a ‘spatial update’ of coactivation models, which has not yet been considered in too much depth (Colonius & Diederich, 2017; Miller, 1982; Schwarz, 1989). Again, this race-model update relates to the previously introduced feature vectors. Based on the evidence in Chapter 3, intramodal redundancies remain within the spatial limits of their single features. Arguably, this finding should be generalizable across all visually redundant combinations. Further, it opens a new perspective on how to describe coactivations: If coactivation is simply the violation of the race-model inequality or the enhancement of the maximum negative dependency test without considering the spatial properties, it may not live up to its own standards.

Although no evidence for coactivation was found for the visually redundant color-shape target, it shows a precedent based on the observed limits in spatial slopes. To make this argument more apparent, both racers show different speeds: the color racer is faster in focal locations, the shape racer is faster in peripheral locations, thus, the color-shape target merely reflects the interaction of both racers, which adds up to significant redundancy gains. There is actually no extra resource or activation beyond those racers that go along with the definition of coactivation. Generally, the 'interactive race model' seems more suitable to explain the underlying pattern (Mordkoff & Yantis, 1991, 1993) as it does not imply extra coactive resources but merely the interaction between racers or targets and nontargets. In contrast to this study, Mordkoff and Yantis (1993) found race-model violations for redundant color-shape targets above the additive sum of single color and shape.

It will be crucial for future studies to clarify coactivation and its spatial properties in visual search. Overall, there might be three possible explanations. First, coactivations occur between visual dimensions that clearly exceed their feature limits. Second, the 'interactive racers' might produce spurious violations of the race-model inequality in visually redundant targets due to increased internal noise (Otto & Mamassian, 2012). Third, the "coactivation process" itself might be more versatile. It can reflect co-activation as well as enhanced co-inhibition of irrelevant nontargets. This assumption would align with the timing effects in the PCN that favor preattentive co-activation for crossmodal redundant targets in PCN onsets and co-inhibition in visual redundant targets during the (post-)selective stage in PCN offsets.

The third study shifted the focus from the bottom-up factors of object features and spatial locations to intertrial effects or prior history (Awh et al., 2012; Found & Müller, 1996; Wolfe, 2021). Despite analyzing co-activation effects, the benefits and switch costs for various shift conditions were examined to evaluate another core assumption of the MWA that shifts between sensory modalities (e.g., vision → touch) are more costly than between (intramodal) dimensions (e.g., color → shape). Overall, the behavioral pattern supports the MWA by showing significantly higher switch costs, i.e., prolonged reaction times for modality shifts compared to a visual dimension-shift. Additionally, there was also a (predicted) null finding as exact feature repetitions (e.g., red → red) and an intradimensional feature-shift (e.g., triangle → square) were comparable. This null finding aligns with DWA and MWA, which assume no switch costs for intradimensional changes (Found & Müller, 1996). Of course, this outcome of a switch cost hierarchy answers the next superordinate research question (7).

Interestingly, the extended behavioral analyses revealed modality-specific location-shift patterns. Although visual and tactile performances were generally better for the right hand and visual hemifield (in line with previous studies by Spence, Shore, et al., 2001; Whitehead, 1991), the location-shift preferences differed for visual and tactile targets. Whereas visual target localization became faster for location changes (e.g., right → left) across all four visual conditions (VR, VF, VD, & VM), tactile target localization was improved for hand repetitions (e.g., right → right). However, the fact that a modality-shift for visual and tactile targets favored location changes (over repetitions) highlights a functional difference in spatial orienting in vision and touch. Arguably, visual search operates with the "inhibition-of-return" mechanism, suppressing (or down-weighting) the activity of the current location and promoting (up-weighting) the opposite (or unknown) location (Klein, 2000; Posner et al., 1985; Taylor & Klein, 2000). However, tactile targets seem to enhance the sensitivity of (or up-weight) the current hand, while a hand shift needs to relocate these attentional resources. Finally, the evidence that location changes are always faster regardless of the preceding targets indicates a (persisting) visual bias or dominance for spatial decoding (Posner et al., 1976; Spence, Nicholls, et al., 2001).

Furthermore, a modality-shift, i.e., a shifting of attentional resources across sensory modalities, had the most prolonged "neural trace", which remained traceable for three successive trials. Generally, it would be interesting to model a neural decay function, e.g., via exponential fitting, to investigate the temporal dynamics of residual activations (e.g., Zylberberg et al., 2009). For example, it would be interesting to study whether those "decay functions" reflect a stable, functional marker or are rather dynamic over the experiment time course, within and across participants. The absence of PCN and CCN results indicates a limitation. Therefore, the ninth research question (9) is only partially answered. However, it will be necessary to replicate those findings of "location-shift patterns" and "neural decay functions" in future studies. Investigating crossmodal switch costs by their spatio-temporal dynamics as a temporal derivative (n–1 vs. n) would be fascinating. An intuition of such an approach can be seen in the delta topomaps in Figures 4–6, 4–7, & 4–8. This approach shows attentional resources as a time-dependent variable distributed across modalities (or features). Again, those time-regressive dynamics can be made available via enhanced jackknife procedures. Finally, this time-dependent approach is not limited to the PCN or CCN but applies to all other perceptual and preattentive components, e.g., the anterior N1 (Töllner et al., 2009).

Besides the response evidence that favors the MWA due to a hierarchy of switch costs, a congruent pattern was observed in PCN and CCN components by showing substantial amplitude decreases during a modality-shift. Overall, this outcome reflects evidence for a supramodal mechanism of shared attentional resources already active within the preattentive stage – promoting an MWA. Whenever a "pop-out" target occurs, e.g., a red circle, it attracts attention, i.e., establishing a "strong" or prioritized connection between feature activations and the priority map. In case of a feature repetition (another red circle), this link is still (or already) active, and thus processing takes less time. On the contrary, a shift between (visual) dimensions or sensory modalities needs to (re-)establish a direct link. Since our attentional resources are limited, it requires to free these occupied resources from previous events (or connections). This common and supramodal mechanism, which becomes most apparent for a modality-shift, can be seen as strong evidence favoring the MWA as it implies the attentional orienting of a supramodal priority map. Also, it answers the eighth research from the General Introduction.

Furthermore, the PCN also revealed a modality-specific effect by showing a consistent delay in their PCN latency estimate, i.e., the average from the onset, latency, and offset for shifts between dimensions and modalities. The fact that both changes were the same in the PCN delay refers to a common delay mechanism, which would be interesting to study in greater detail. Again, no temporal differences were found in the CCN, similar to the previous studies. The absence of CCN timing effects indicates two things. First, the CCN might not be susceptible to temporal differences across finger locations, i.e., the "fovea of touch" (Forster et al., 2016). Second, by only including vibro-tactile frequencies, these studies may be limited to retrieving the whole functional picture of the CCN. It will be necessary for future studies to include various targets across tactile dimensions (e.g., pressure, frequency, or distinct patterns) to obtain a complete picture of tactile attention.

All three studies supported the MWA as the underlying preattentive architecture. They revealed several supramodal mechanisms that promote the existence of an attentional priority map across sensory modalities. First, PCN onsets and amplitudes are accelerated and amplified for crossmodal redundant targets, exceeding the intramodal limits of vision, which goes along with coactivations during multisensory integration. Second, switching between sensory modalities consistently decreases the PCN and CCN amplitudes, indicating that attentional resources are shared across vision and touch. While the CCN reflected a stable somatosensory component, the PCN modulated sensitively to crossmodal interactions and interferences, which is in line with previous evidence on crossmodal links in the PPC (Chambers et al., 2007; Macaluso et al., 2002b; Man et al., 2015; Quinn et al., 2014).

In summary, the main findings in performances and neural dynamics of PCN and CCN components strongly support "modality-weighting" as the underlying preattentive architecture in crossmodal search, thus extending the visual "dimension-weighting" by introducing another intermediate layer of modality maps. Essentially, crossmodal attention combines and connects their resources during attentional orienting, revealing exceeding redundancy benefits and intersensory switch costs. Although the main research questions of this thesis are answered with this, some additional considerations are made in the following sections.

A Supramodal Priority Network

As described in the General Introduction, the quest to find the priority map of visual (or crossmodal) spatio-topographic computations in saliency and relevance initiated an ongoing debate. While some favor early visual areas, such as V1 or the superior colliculi (SC) (Koch & Ullman, 1985; Koene & Zhaoping, 2007; Li, 2002; Meredith & Stein, 1986; Shim et al., 2013), others promote saliency summation during a later stage (e.g., V4, MT) and beyond the visual striate cortex in the PPC (Chambers et al., 2004; Fecteau & Munoz, 2006; Koch & Ullman, 1985; Somers & Sheremata, 2013). However, the search for a single map is limited. Based on the evidence of Chambers, Payne, Stokes et al. (2004), spatio-topographic saliency computation derives through a fast and a slow pathway, activating the spatial reference frame and subsequently providing the specific feature content during the timing of attentional selection (210–240 ms), respectively.

Interestingly, this explanation is in line with evidence from firing rates in receptive fields (Chelazzi et al., 1993, 1998) that show an initial firing of all receptive fields during the early P1 time window (90–120 ms), i.e., initiation of the spatial frame, while only a subset of relevant receptive fields continue firing until the PCN timing of attentional selection (around 250 ms) (Luck et al., 1997). The continuation of firing within a subset of receptive fields reflects the (hidden) dynamic weighting processes to prioritize specific feature activations based on their bottom-up saliency or top-down relevance (e.g., matching a target template). Essentially, the deployment of selective attention that provides a central representation for higher cognitions can only occur by establishing a strong and coherent linkage between those relevant network hubs, similar to optimizing the signal-to-noise ratio (SNR).

The proposition of a "priority network" instead of a "priority map" seems more plausible as it is embedded into the fronto-parietal attention network that combines distinct hubs (e.g., FEF, PPC, IPS, or TPJ) to enable attentional orienting (Corbetta & Shulman, 2002; Posner & Dehaene, 1994). Therefore, attentional selection describes a "coherent state of deployed resources across network hubs". Although the PCN seems suitable for depicting the supramodal priority computations from visual and tactile activations, arguably, it 'only' accounts for the parietal activations. For example, it combines crossmodal amplification and coactivation while maintaining the external visual quadratic mapping in receptive fields. The tactile somatotopic mapping might be included in the PCN amplification, yet the peripheral slopes in CCN seem to modulate sensitively for coactive, i.e., revealing significant quadratic slopes. This finding suggests reciprocal connections between parietal and central regions, which need to be considered in greater detail in future studies. Furthermore, the impact of top-down control from presumably frontal areas is not fully resolved yet. Although top-down control enhances the PCN dynamics (Eimer & Kiss, 2010; Grubert et al., 2011), evidence from attentional orienting to crossmodal templates is still missing. A "priority network" combines multiple components across fronto-parietal regions, and the PCN is one of them. Other components might be the CCN or the anterior N1.

To extend the scope of the "attention priority network", preliminary findings from the attentional templates study (short presentation at the TeaP in Ulm, 2021) showed that the PCN delay in a constant target template (square target) could be entirely attributed to the template difference of the preceding trial, being intradimensional (triangle target), intramodal (color target), or crossmodal (frequency target), respectively. This (preliminary) outcome indicates that the PCN is indeed sensitive to the prior history (see also Chapter 4). At the same time, the PCN activity of target templates (square target) is independent of their contextual embeddings. However, since performances revealed a contextual dependence of target template combinations (e.g., square vs. frequency target), the PCN evidence might not explain the behavior pattern thoroughly. Other factors need to be considered, e.g., frontal activity.

Another preliminary finding supports the relevant role of brain oscillations for sustained attention (Poster presentation at the SAMBA Salzburg 2021). Here, it was shown that performance fluctuations over the experimental time course (being either "in-the-zone" or "out-of-the-zone") (Esterman et al., 2013; Rosenberg et al., 2015) are correlated to the intertrial-coherence (ITC) of the frontal and parietal theta and alpha oscillations and also the PCN amplitude, becoming increased and more negative, respectively during good "in-the-zone" performances. This preliminary outcome contributes to other studies that propose a genuine role of brain oscillations within the attention network (Fiebelkorn et al., 2018; Helfrich et al., 2018; Michel et al., 2021; Sauseng et al., 2007). It will be essential in future studies to examine the (correlative) links between the PCN and theta and alpha oscillations.

Overall, the “supramodal priority network” reflects the adjusted weighting (SNR optimization) of relevant brain areas, which derives from “communication through coherence” that includes a hierarchy of feedforward and feedback connections (Bastos, Litvak, et al., 2015; Bastos, Vezoli, et al., 2015; Fries, 2005), which in the case of crossmodal search can be described by the MWA. Further, it can be stated that the graphical structure of the MWA architecture and the “supramodal priority network” are identical: attentional resources are deployed (and distributed) across branches. Therefore, the amount of attentional resources is extended by each network (children) node (e.g., dimension or modality map) yet limited by the capacity of the superior (parent) node. Finally, the “supramodal priority network” combines the attentional resources from vision and touch, which fastens their parallel (co-)activation but delays intersensory switching due to the more costly serial allocation of resources during attention shifts.

A Computational Model of Attention

To understand crossmodal selective attention sufficiently, it must integrate the empirical evidence into a computational model of attention (CMA). As already stated in the first chapter, the five-factor model reflects the skeleton of such a CMA, as it depicts the feedforward (object features, spatial locations) and feedback (attentional templates) directions of a “priority network”, as well as time-regressive fluctuations of the prior history (intertrial effects) and sustained attention. Similar to previous studies, a CMA entails distinct processing levels or stages, e.g., a sequence of perceptual, preattentive, post-selective, and response levels (McClelland, 1979; Miller, 1982; Wolfe, 1994). By considering a closed-loop system (deBettencourt et al., 2015) within the stimulus-to-response interval, performances are essentially the product of the preceding dynamics within those levels (similar to a neural network).

The contribution of this thesis for such a CMA is the in-depth analysis of the PCN and CCN components as neural markers for visual and tactile spatial attention. Overall, it has been shown that the PCN is a good performance predictor by explaining up to 25 % of reaction time’s variance (s. Chapter 2). However, a detailed look into the PCN becomes more complex, revealing distinct temporal stages for crossmodal and visual selection. Also, the PCN model is limited as it does not include the purely tactile target (which likely also applies to auditory targets). Again, the PCN is sensitive to crossmodal information, but only in combination with visual information. Indeed, other components or brain oscillations need to be included to complete a CMA, such as the N1(VEP, SEP, AEP), theta & alpha oscillations, the PD, CDA, or sLRP (Fiebelkorn et al., 2018; Hickey et al., 2009; Luria et al., 2016; Miller et al., 1998).

Although plenty of empirical evidence exists for each component, attention research needs an integrative CMA that combines them. Crucially, a parameterized CMA would estimate the influence of each attention factor (e.g., bottom-up saliency or switch costs) on each target to obtain a set of 'priors', e.g., based on their distinct effect sizes. These 'priors' are linked to spatio-temporal patterns, e.g., PCN onset or N1 amplitude, which must be carefully selected. A CMA can be described by a hierarchy of activation functions for each level. Like the cascade model of McClelland (1979), activation is projected from level to level until response timing, i.e., the end of the closed-loop system. Arguably, the choice of parameters and the time-dependent integration falls into the field of dynamic systems, combining stable and unstable moments or convergent and divergent neural states.

Challenges of Attention Research

Even though attention research spans more than a century, its challenges have not become any less, quite the opposite. From "every one knows what attention is" (James, 1890) to "no one knows what attention is" (Hommel et al., 2019), we are somewhere on the center line. Of course, an incredible corpus of scientific insights or a pile of evidence over many decades is available, but it is hard to get an overview. Overall, the thesis extended the theory on crossmodal attention by providing evidence for "modality-weighting" and preattentive coactivations and proposing a "supramodal priority network". Closing the scientific gap(s) between visual and crossmodal search with the MWA concept establishes a bridge between these (decade-long) separated research paradigms. Yet, it also raised many follow-up questions.

Generally, attention research reflects one topic out of many in experimental psychology. Of course, it has to evolve to keep its relevant role and expertise within cognitive research. Within the next years, attention research has to overcome several challenges. First, the issue of "setup limitations" (especially in crossmodal search) must be resolved by taking advantage of modern applications (e.g., augmented or virtual reality) (Climent et al., 2021; Drigas et al., 2022; Mitsea et al., 2022) and the developments in co-registration (Himmelstoss et al., 2020; Nikolaev et al., 2016; Nunez & Silberstein, 2000) that make natural settings more feasible (Coco et al., 2020; Leszczynski & Schroeder, 2019; Võ & Wolfe, 2015). Second, the "methodological scope" of data processing and analysis must be expanded, linked, and standardized. For example, more and more efforts are being made to provide common preprocessing pipelines in various neuroimaging techniques, such as EEG or fMRI (Bigdely-Shamlo et al., 2015; Esteban et al., 2019). Of course, machine learning will play a crucial role in the future of attention research (Lindsay, 2020). Here, it will be essential that knowledge does not become implicit but remains explicit (blackbox vs. science). Third, attention "paradigms" need to evolve by taking complexity into account. For example, the aspect of stimulus heterogeneity in targets, distractors, and nontargets across visual dimensions and feature activations has not been fully considered. It would be interesting to obtain complete saliency solutions (or estimates) for the whole color or orientation spectrum in its many combinations. Similar to sensory uncertainty, those saliency solutions can be utilized to predict attention processes (van Bergen et al., 2015).

Again, attention must remain within the scientific focus to not become a 'transient phenomenon' itself. The following section briefly introduces a new paradigm that summarizes the results and challenges presented.

Dynamic Search

Based on the assumption that attention reflects a unique cognitive resource and a transient phenomenon that links (multisensory) perception, cognitions, and actions via filtering, the focus goes inevitably to its third attribute – attention is "dynamic". This section proposes a 'new' paradigm of "dynamic search". As shown in the preceding sections, attention research must evolve to address the open issues and new challenges. Dynamic search combines visual and crossmodal search evidence, assuming an MWA. As shown by the evidence of this thesis, selective attention derives from dynamic weighting (or prioritization) processes during the preattentive stage.

So far, the evidence is based on findings in detection and localization tasks, which involve a low level of complexity, as simply detecting a "pop-out singleton" is sufficient to respond to its presence or location. Hence, examining additional task designs in the same visuo-tactile paradigm would be valuable. The implementation of an identification task should provide further insides. Whereas detection and localization tasks show implicit knowledge ("that and "where" response) of the "odd-one-out", identification requires explicit feature knowledge ("what" response) (Found & Müller, 1996; Nothdurft, 1992; Sagi & Julesz, 1985). For example, a crossmodal identification task would look like this: Participants are instructed to focus on one or multiple pop-out targets within a specific modality, either vision (e.g., a blue square or fuchsia circle among blue circles) or touch (fast or slow vibrating frequencies among a baseline frequency). Each trial has one pop-out target (e.g., blue square) that guides attention. Next, participants identify the feature of the other modality (high vs. low frequency) and give their response (left or right pedal press). This procedure can be done blockwise or mixed. Essentially, this task tests crossmodal identification processes: the switching from one modality (selection) to another (identification).

Furthermore, the “stimulus-onset asynchrony” (SOA) reflects an interesting temporal factor, which examines the optimal alignment between sensory modalities. Crucially, SOA paradigms can be linked to working memory and retrospective judgments. Here, participants are presented with a tactile display (e.g., left hand: high vibration & right hand: low vibration), followed by a visual display (with a distinct SOA, e.g., 200 ms), including a pop-out singleton, which participants have to localize and then respond to the preceding tactile target (high vs. low response), which has kept in working memory as a template.

Overall, both tasks try to gain further insights into the underlying flexibility of attentional orienting to multisensory events. Whereas identification activates complete branches of the supramodal priority network at the feature level, the SOA conditions should reveal a distinct timeline of crossmodal integration.

Besides these canonical paradigms, “dynamic search” is also suited to predict selective attention spatially. By combining visual or crossmodal displays with a fast sequence of pop-out singletons and EEG recordings, including some response-relevant oddball targets, it should be possible to track covert attention shifts and to obtain a functional map of attentional orienting, e.g., via PCN or brain oscillations. For example, in a visual display with free-moving stimuli, participants are instructed to (covertly) follow a moving target. This free-moving target can be combined with moving (heterogenous) nontargets or distractors with varying distances. Furthermore, targets can dynamically evolve from a simple feature-contrast into complex objects, e.g., a color target can transform into a shape target (e.g., via interpolation). Again, this gradual change can be measured by the neural dynamics of preattentive and attentive modes. Finally, the proposition of an eigenvector space for distinct features (dimensions and modalities), which still must be empirically substantiated, can be regarded as a requirement that enables complex dynamic search. This “live-tracking” of selective attention should enable individual “attentional calibration” within the whole attention network (feedforward and feedback connections) and allow for attention-related learning effects, e.g., in neurofeedback. Overall, this brief and speculative introduction of “dynamic search” that measures the spatiotemporal flow of neural processes is similar to recent efforts in visual search (Boettcher et al., 2022).

Generally, the future of attention research looks promising and exciting at the same time. Of course, many major challenges await, but that has always been the case. Although it will be crucial to verify the “supramodal priority network” for crossmodal selective attention, the versatility of “dynamic search” enables to integrate the many insights into a coherent computational model of attention. By actively using the many interfaces to other developing fields (e.g., consciousness, neurofeedback, or virtual reality), attention research remains state-of-the-art and assert its relevance. The evidence on preattentive “modality-weighting” may have contributed to this.

References

The art of a puzzle
Builds up in pieces
a serene harmony.

Adam, K. C. S., Robison, M. K., & Vogel, E. K. (2018). Contralateral Delay Activity Tracks Fluctuations in Working Memory Performance. *Journal of Cognitive Neuroscience*, *30*(9), 1229–1240.

Aguirre, G. K., Mattar, M. G., & Magis-Weinberg, L. (2011). de Bruijn cycles for neural decoding. *NeuroImage*, *56*(3), 1293–1300.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Aitchison, L., & Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227.

Alamia, A., & VanRullen, R. (2019). Alpha oscillations and traveling waves: Signatures of predictive coding? *PLoS Biology*, *17*(10), e3000487.

Alho, K., Lavikainen, J., Reinikainen, K., Sams, M., & Näätänen, R. (1990). Event-related brain potentials in selective listening to frequent and rare stimuli. *Psychophysiology*, *27*(1), 73–86.

Allenmark, F., Zhang, B., Liesefeld, H. R., Shi, Z., & Müller, H. J. (2019). Probability cueing of singleton-distractor regions in visual search: the locus of spatial distractor suppression is determined by colour swapping. *Visual Cognition*, *27*(5-8), 576–594.

Allport, A. D. (1971). Parallel encoding within and between elementary stimulus dimensions. *Perception & Psychophysics*, *10*(2), 104–108.

Allport, A. D. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. *Perspectives on Perception and Action*, *15*, 395–419.

Ambron, E., Mas-Casadesús, A., & Gherri, E. (2018). Hand distance modulates the electrophysiological correlates of target selection during a tactile search task. *Psychophysiology*, *55*(8), e13080.

Anderson, B. A., Kim, H., Kim, A. J., Liao, M.-R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience and Biobehavioral Reviews*, *130*, 326–350.

Ansorge, U., & Heumann, M. (2006). Shifts of visuospatial attention to invisible (metacontrast-masked) singletons: Clues from reaction times and event-related potential. *Advances in Cognitive Psychology* , *2*(1), 61–76.

Assumpção, L., Shi, Z., Zang, X., Müller, H. J., & Geyer, T. (2018). Contextual cueing of tactile search is coded in an anatomical reference frame. *Journal of Experimental Psychology. Human Perception and Performance*, *44*(4), 566–577.

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

Baars, B. J. (1997). In the theatre of consciousness. Global Workspace Theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4*(4), 292–309.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839.

Badde, S., Röder, B., & Heed, T. (2015). Flexibly weighted integration of tactile reference frames. *Neuropsychologia*, *70*, 367–374.

Bagherzadeh, Y., Baldauf, D., Pantazis, D., & Desimone, R. (2020). Alpha Synchrony and the Neurofeedback Control of Spatial Attention. *Neuron*, *105*(3), 577–587.e5.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*, 456.

Balan, P. F., & Gottlieb, J. (2006). Integration of exogenous input into a dynamic salience map revealed by perturbing attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(36), 9239–9249.

Barczak, A., Haegens, S., Ross, D. A., McGinnis, T., Lakatos, P., & Schroeder, C. E. (2019). Dynamic Modulation of Cortical Excitability during Visual Active Sensing. *Cell Reports*, *27*(12), 3447–3459.e3.

Bastos, A. M., Litvak, V., Moran, R., Bosman, C. A., Fries, P., & Friston, K. J. (2015). A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. *NeuroImage*, *108*, 460–475.

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, *85*(2), 390–401.

Bastos, A. M., Vezoli, J., & Fries, P. (2015). Communication through coherence with inter-areal delays. *Current Opinion in Neurobiology*, *31*, 173–180.

Becker, S. I., Ansorge, U., & Horstmann, G. (2009). Can intertrial priming account for the similarity effect in visual search? *Vision Research*, *49*(14), 1738–1756.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *57*(1), 289–300.

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*(5919), 696–698.

Berggren, N., & Eimer, M. (2018). Electrophysiological correlates of active suppression and attentional selection in preview visual search. *Neuropsychologia*, *120*, 75–85.

Berlyne, D. E. (1951). Attention, perception and behavior theory. *Psychological Review*, *58*(2), 137–146.

Biederman, I., & Checkosky, S. F. (1970). Processing redundant information. *Journal of Experimental Psychology*, *83*(3p1), 486–490.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*, 16.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*, 1–21.

Blurton, S. P., Greenlee, M. W., & Gondan, M. (2014). Multisensory processing of redundant information in go/no-go and choice responses. *Attention, Perception & Psychophysics*, *76*(4), 1212–1233.

Boettcher, S. E. P., Shalev, N., Wolfe, J. M., & Nobre, A. C. (2022). Right place, right time: Spatiotemporal predictions guide attention in dynamic visual search. *Journal of Experimental Psychology. General*, 151(*2*), 348–362.

Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nature Neuroscience*, *2*(11), 947–957.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the "spotlight" of visual attention. *Nature Neuroscience*, *2*(4), 370–374.

Bressler, S. L., & Richter, C. G. (2015). Interareal oscillatory synchronization in top-down neocortical processing. *Current Opinion in Neurobiology*, *31*, 62–66.

Bressler, S. L., & Seth, A. K. (2011). Wiener–Granger Causality: A well established methodology. *NeuroImage*, *58*(2), 323–329.

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, *7*(8), 880–886.

Broadbent, D. E. (1956). Listening between and during practised auditory distractions. *British Journal of Psychology* , *47*(1), 51–60.

Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press.

Bruyns-Haylett, M., Luo, J., Kennerley, A. J., Harris, S., Boorman, L., Milne, E., Vautrelle, N., Hayashi, Y., Whalley, B. J., Jones, M., Berwick, J., Riera, J., & Zheng, Y. (2017). The neurogenesis of P1 and N1: A concurrent EEG/LFP study. *NeuroImage*, *146*, 575–588.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547.

Bundesen, C. (1998). A computational theory of visual attention. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1373), 1271–1281.

Burra, N., & Kerzel, D. (2013). Attentional capture during visual search is attenuated by target predictability: evidence from the N2pc, Pd, and topographic segmentation. *Psychophysiology*, *50*(5), 422–430.

Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, *88*(1), 127–144.

Buschman, T. J., & Miller, E. K. (2009). Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron*, *63*(3), 386–396.

Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11252–11255.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(24), 7869–7876.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18751–18756.

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, *18*(6), 300–309.

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(25), 9315–9322.

Carlisle, N. B., & Nitka, A. W. (2019). Location-based explanations do not account for active attentional suppression. *Visual Cognition*, *27*(3-4), 305–316.

Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57*(8), 1241–1261.

Carrasco, M., & Frieder, K. S. (1997). Cortical magnification neutralizes the eccentricity effect in visual search. *Vision Research*, *37*(1), 63–82.

Carrasco, M., McElree, B., Denisova, K., & Giordano, A. M. (2003). Speed of visual processing increases with eccentricity. *Nature Neuroscience*, *6*(7), 699–700.

Chambers, C. D., Payne, J. M., & Mattingley, J. B. (2007). Parietal disruption impairs reflexive spatial attention within and between sensory modalities. *Neuropsychologia*, *45*(8), 1715–1724.

Chambers, C. D., Payne, J. M., Stokes, M. G., & Mattingley, J. B. (2004). Fast and slow parietal pathways mediate spatial attention. *Nature Neuroscience*, *7*(3), 217–218.

Chambers, C. D., Stokes, M. G., & Mattingley, J. B. (2004). Modality-specific control of strategic spatial attention in parietal cortex. *Neuron*, *44*(6), 925–930.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345–347.

Chen, S., Shi, Z., Zang, X., Zhu, X., Assumpção, L., Müller, H. J., & Geyer, T. (2020). Crossmodal learning of target-context associations: When would tactile context predict visual search? *Attention, Perception & Psychophysics*, *82*(4), 1682–1694.

Chen, Y., Anand, S., Martinez-Conde, S., Macknik, S. L., Bereshpolova, Y., Swadlow, H. A., & Alonso, J.-M. (2009). The linearity and selectivity of neuronal responses in awake visual cortex. *Journal of Vision*, *9*(9), 12.1–17.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.

Clark, K., Appelbaum, L. G., van den Berg, B., Mitroff, S. R., & Woldorff, M. G. (2015). Improvement in visual search with practice: mapping learning-related changes in neurocognitive stages of processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(13), 5351–5359.

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*(3), 170–187.

Clayton, M. S., Yeung, N., & Cohen Kadosh, R. (2015). The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences*, *19*(4), 188–195.

Climent, G., Rodríguez, C., García, T., Areces, D., Mejías, M., Aierbe, A., Moreno, M., Cueto, E., Castellá, J., & Feli González, M. (2021). New virtual reality tool (Nesplora Aquarium) for assessing attention and working memory in adults: A normative study. *Applied Neuropsychology. Adult*, 28(*4*), 403–415.

Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during semantic integration of object-scene information. *Journal of Cognitive Neuroscience*, *32*(4), 571–589.

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, *16*(8), 411–417.

Cohen, R., & Rist, F. (1992). The modality shift effect. Further explorations at the crossroads. *Annals of the New York Academy of Sciences*, *658*, 163–181.

Colonius, H. (1986). Measuring channel dependence in separate activation models. *Perception & Psychophysics*, *40*(4), 251–255.

Colonius, H., & Diederich, A. (2006). The race model inequality: interpreting a geometric measure of the amount of violation. *Psychological Review*, *113*(1), 148–154.

Colonius, H., & Diederich, A. (2017). Measuring multisensory integration: from reaction times to spike counts. *Scientific Reports*, *7*(1), 3023.

Colonius, H., Wolff, F. H., & Diederich, A. (2017). Trimodal Race Model Inequalities in Multisensory Integration: I. Basics. *Frontiers in Psychology*, *8*, 1141.

Corbetta, M., Miezin, F. M., Shulman, G. L., & Petersen, S. E. (1993). A PET study of visuospatial attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *13*(3), 1202–1226.

Corbetta, M., & Shulman, G. L. (1998). Human cortical mechanisms of visual attention during orienting and search. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1373), 1353–1362.

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

Cruccu, G., Aminoff, M. J., Curio, G., Guerit, J. M., Kakigi, R., Mauguiere, F., Rossini, P. M., Treede, R.-D., & Garcia-Larrea, L. (2008). Recommendations for the clinical use of somatosensory-evoked potentials. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *119*(8), 1705–1719.

Dalvit, S., & Eimer, M. (2011). Mechanisms of percept-percept and image-percept integration in vision: behavioral and electrophysiological evidence. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(1), 1–11.

Damasio, A. R. (1989). The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation*, *1*(1), 123–132.

deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, *18*(3), 470–475.

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Penguin.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14529–14534.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, *79*(1-2), 1–37.

De Meo, R., Murray, M. M., Clarke, S., & Matusz, P. J. (2015). Top-down control and early multisensory processes: chicken vs. egg. *Frontiers in Integrative Neuroscience*, *9*, 17.

Deroy, O., Crisinel, A.-S., & Spence, C. (2013). Crossmodal correspondences between odors and contingent features: odors, musical notes, and geometrical shapes. *Psychonomic Bulletin & Review*, *20*(5), 878–896.

Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review*, *20*(4), 643–664.

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

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

Deutsch, J. A., & Deutsch, D. (1963). Some theoretical considerations. *Psychological Review*, *70*(1), 80–90.

DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, *11*(5), 219–226.

Diederich, A., & Colonius, H. (1991). A further test of the superposition model for the redundant-signals effect in bimodal detection [Review of *A further test of the superposition model for the redundant-signals effect in bimodal detection*]. *Perception & Psychophysics*, *50*(1), 83–86. Springer.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*(8), 1388–1404.

Diederich, A., Colonius, H., Bockhorst, D., & Tabeling, S. (2003). Visual-tactile spatial interaction in saccade generation. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *148*(3), 328–337.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *Journal of Experimental Psychology. General*, *140*(4), 552–572.

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, *3*(3), 277–283.

Drigas, A., Mitsea, E., & Skianis, C. (2022). Virtual Reality and Metacognition Training Techniques for Learning Disabilities. *Sustainability: Science Practice and Policy*, 14(*16*), 10170.

Drisdelle, B. L., & Eimer, M. (2021). PD components and distractor inhibition in visual search: New evidence for the signal suppression hypothesis. *Psychophysiology*, *58*(9), e13878.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, *2*(7), 254–262.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*(3), 272–300.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General*, *113*(4), 501–517.

Duncan, J., & Humphreys, G. (1992). Beyond the search surface: visual search and attentional engagement [Review of *Beyond the search surface: visual search and attentional engagement*]. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(2), 578–588; discussion 589–593.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.

Eason, R. G. (1981). Visual evoked potential correlates of early neural filtering during selective attention. *Bulletin of the Psychonomic Society*, *18*(4), 203–206.

Eckstein, M. P. (2011). Visual search: a retrospective. *Journal of Vision*, *11*(5). https://doi.org/10.1167/11.5.14

Edelman, G., & Tononi, G. (2001). *A Universe Of Consciousness How Matter Becomes Imagination: How Matter Becomes Imagination*. Basic Books.

Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, *1*(4), 245–252.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234.

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in Cognitive Sciences*, *18*(10), 526–535.

Eimer, M., & Driver, J. (2000). An event-related brain potential study of cross-modal links in spatial attention between vision and touch. *Psychophysiology*, *37*(5), 697–705.

Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: evidence from event-related brain potential studies. *Neuroscience and Biobehavioral Reviews*, *25*(6), 497–511.

Eimer, M., Forster, B., Fieger, A., & Harbich, S. (2004). Effects of hand posture on preparatory control processes and sensory modulations in tactile-spatial attention. *Clinical Neurophysiology*, *115*(3), 596–608.

Eimer, M., Forster, B., & Van Velzen, J. (2003). Anterior and posterior attentional control systems use different spatial reference frames: ERP evidence from covert tactile-spatial orienting. *Psychophysiology*, *40*(6), 924–933.

Eimer, M., & Grubert, A. (2015). A dissociation between selective attention and conscious awareness in the representation of temporal order information. *Consciousness and Cognition*, *35*, 274–281.

Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*(8), 1423–1433.

Eimer, M., & Kiss, M. (2010). Top-down search strategies determine attentional capture in visual search: behavioral and electrophysiological evidence. *Attention, Perception & Psychophysics*, *72*(4), 951–962.

Eimer, M., Kiss, M., & Cheung, T. (2010). Priming of pop-out modulates attentional target selection in visual search: behavioural and electrophysiological evidence. *Vision Research*, *50*(14), 1353–1361.

Eimer, M., & Schröger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology*, *35*(3), 313–327.

Eimer, M., & Van Velzen, J. (2002). Crossmodal links in spatial attention are mediated by supramodal control processes: evidence from event-related potentials. *Psychophysiology*, *39*(4), 437–449.

Eimer, M., & van Velzen, J. (2005). Spatial tuning of tactile attention modulates visual processing within hemifields: an ERP investigation of crossmodal attention. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *166*(3-4), 402–410.

Eimer, M., van Velzen, J., & Driver, J. (2002). Cross-modal interactions between audition, touch, and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations. *Journal of Cognitive Neuroscience*, *14*(2), 254–271.

Eimer, M., van Velzen, J., Forster, B., & Driver, J. (2003). Shifts of attention in light and in darkness: an ERP study of supramodal attentional control and crossmodal links in spatial attention. *Brain Research. Cognitive Brain Research*, *15*(3), 308–323.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*(2), 201–204.

Eriksen, C. W., & St James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(*1*), 111–116.

Esterman, M., Noonan, S. K., Rosenberg, M., & Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex* , *23*(11), 2712–2723.

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180.

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: modality-specific or supramodal? *Neuropsychologia*, *27*(4), 461–470.

Fecteau, J. H. (2007). Priming of pop-out depends upon the current goals of observers. *Journal of Vision*, *7*(6), 1–1.

Fecteau, J. H., & Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nature Reviews. Neuroscience*, *4*(6), 435–443.

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.

Feintuch, U., & Cohen, A. (2002). Visual Attention and Coactivation of Response Decisions for Features From Different Dimensions. *Psychological Science*, *13*(4), 361–369.

Feldmann-Wüstefeld, T., Uengoer, M., & Schubö, A. (2015). You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. *Psychophysiology*, *52*(11), 1483–1497.

Feldmann-Wüstefeld, T., Weinberger, M., & Awh, E. (2021). Spatially Guided Distractor Suppression during Visual Search. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *41*(14), 3180–3191.

Ferrante, O., Patacca, A., Di Caro, V., Della Libera, C., Santandrea, E., & Chelazzi, L. (2018). Altering spatial priority maps via statistical learning of target selection and distractor filtering. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *102*, 67–95.

Ferstl, R., Hanewinkel, R., & Krag, P. (1994). Is the modality-shift effect specific for schizophrenia patients? *Schizophrenia Bulletin*, *20*(2), 367–373.

Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2018). A Dynamic Interplay within the Frontoparietal Network Underlies Rhythmic Spatial Attention. *Neuron*, *99*(4), 842–853.e8.

Fink, G. R., Dolan, R. J., Halligan, P. W., Marshall, J. C., & Frith, C. D. (1997). Space-based and object-based visual attention: shared and specific neural domains. *Brain: A Journal of Neurology*, *120 (Pt 11)*, 2013–2028.

Föcker, J., Hötting, K., Gondan, M., & Röder, B. (2010). Unimodal and crossmodal gradients of spatial attention: Evidence from event-related potentials. *Brain Topography*, *23*(1), 1–13.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 847–858.

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), 1030–1044.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, *45*(1), 152–170.

Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *143*(4), 480–487.

Forster, B., Tziraki, M., & Jones, A. (2016). The attentive homunculus: ERP evidence for somatotopic allocation of attention in tactile search. *Neuropsychologia*, *84*, 158–166.

Foster, J., Bsales, E., & Awh, E. (2018). The N2pc does not reflect a shift of covert spatial attention. *Journal of Vision*, *18*(10), 1220–1220.

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2017). Alpha-Band Oscillations Enable Spatially and Temporally Resolved Tracking of Covert Spatial Attention. *Psychological Science*, *28*(7), 929–941.

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: investigating a "dimension-weighting" account. *Perception & Psychophysics*, *58*(1), 88–101.

Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, *57*(6), 802–816.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138.

Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, *62*(2), 1230–1233.

Friston, K., Bastos, A. M., Pinotsis, D., & Litvak, V. (2015). LFP and oscillations-what do they tell us? *Current Opinion in Neurobiology*, *31*, 1–6.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125.

Gaillard, A. W. (1976). Effects of warning-signal modality on the contingent negative variation (CNV). *Biological Psychology*, *4*(2), 139–154.

Gamble, M. L., & Luck, S. J. (2011). N2ac: An ERP component associated with the focusing of attention within an auditory scene. *Psychophysiology*, *48*(8), 1057–1068.

García-Larrea, L., Lukaszewicz, A. C., & Mauguière, F. (1992). Revisiting the oddball paradigm. Non-target vs neutral stimuli and the evaluation of ERP attentional effects. *Neuropsychologia*, *30*(8), 723–741.

García-Larrea, L., Lukaszewicz, A. C., & Mauguière, F. (1995). Somatosensory responses during selective spatial attention: The N120-to-N140 transition. *Psychophysiology*, *32*(6), 526–537.

Gaspar, J. M., & McDonald, J. J. (2014). Suppression of salient objects prevents distraction in visual search. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(16), 5658–5666.

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2015). Direct Evidence for Active Suppression of Salient-but-Irrelevant Sensory Inputs. *Psychological Science*, *26*(11), 1740–1750.

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2017). Suppression of overt attentional capture by salient-but-irrelevant color singletons. *Attention, Perception & Psychophysics*, *79*(1), 45–62.

Gaspelin, N., & Luck, S. J. (2018a). Distinguishing among potential mechanisms of singleton suppression. *Journal of Experimental Psychology. Human Perception and Performance*, *44*(4), 626–644.

Gaspelin, N., & Luck, S. J. (2018b). The Role of Inhibition in Avoiding Distraction by Salient Stimuli. *Trends in Cognitive Sciences*, *22*(1), 79–92.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135.

Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews. Neuroscience*, *4*(7), 563–572.

Gherri, E., White, F., & Ambron, E. (2022). Searching on the Back: Attentional Selectivity in the Periphery of the Tactile Field. *Frontiers in Psychology*, *13*, 934573.

Gherri, E., Zhao, B., & Ambron, E. (2021). Behavioural and electrophysiological evidence for the effect of target-distractor separation in a tactile search task. *Biological Psychology*, *162*, 108098.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–490.

Girard, S., Pelland, M., Lepore, F., & Collignon, O. (2013). Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *224*(2), 275–285.

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*(6), 1161–1169.

Goldberg, M. E., & Wurtz, R. H. (1972). Activity of superior colliculus in behaving monkey. II. Effect of attention on neuronal responses. *Journal of Neurophysiology*, *35*(4), 560–574.

Gondan, M., Götze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception & Psychophysics*, *72*(6), 1692–1709.

Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review*, *11*(2), 307–313.

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception & Psychophysics*, *78*(3), 723–735.

Gondan, M., Vorberg, D., & Greenlee, M. W. (2007). Modality shift effects mimic multisensory interactions: an event-related potential study. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *182*(2), 199–214.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Goolsby, B. A., Grabowecky, M., & Suzuki, S. (2005). Adaptive modulation of color salience contingent upon global form coding and task relevance. *Vision Research*, *45*(7), 901–930.

Goschy, H., Bakos, S., Müller, H. J., & Zehetleitner, M. (2014). Probability cueing of distractor locations: both intertrial facilitation and statistical learning mediate interference reduction. *Frontiers in Psychology*, *5*, 1195.

Gramann, K., Töllner, T., Krummenacher, J., Eimer, M., & Müller, H. J. (2007). Brain electrical correlates of dimensional weighting: an ERP study. *Psychophysiology*, *44*(2), 277–292.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron*, *24*(1), 31–47, 111–125.

Gray, R., Mohebbi, R., & Tan, H. Z. (2009). The spatial resolution of crossmodal attention: Implications for the design of multimodal interfaces. *ACM Transactions on Applied Perception*, *6*(1), 1–14.

Graziano, M. S. A., & Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, *44*(13), 2621–2635.

Green, D. M., Swets, J. A., & Others. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677.

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., & Schnitzler, A. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 13050–13055.

Grubert, A., Carlisle, N. B., & Eimer, M. (2016). The Control of Single-color and Multiple-color Visual Search by Attentional Templates in Working Memory and in Long-term Memory. *Journal of Cognitive Neuroscience*, *28*(12), 1947–1963.

Grubert, A., & Eimer, M. (2018). The Time Course of Target Template Activation Processes during Preparation for Visual Search. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(44), 9527–9538.

Grubert, A., Fahrenfort, J., Olivers, C. N. L., & Eimer, M. (2017). Rapid top-down control over template-guided attention shifts to multiple objects. *NeuroImage*, *146*, 843–858.

Grubert, A., Krummenacher, J., & Eimer, M. (2011). Redundancy gains in pop-out visual search are determined by top-down task set: behavioral and electrophysiological evidence. *Journal of Vision*, *11*(14).

Gruber, W. R., Klimesch, W., Sauseng, P., & Doppelmayr, M. (2005). Alpha phase synchronization predicts P1 and N1 latency and amplitude size. *Cerebral Cortex* , *15*(4), 371–377.

Hagmann, C. E., & Russo, N. (2016). Multisensory integration of redundant trisensory stimulation. *Attention, Perception & Psychophysics*, *78*(8), 2558–2568.

Hanning, N. M., Jonikaitis, D., Deubel, H., & Szinte, M. (2016). Oculomotor selection underlies feature retention in visual working memory. *Journal of Neurophysiology*, *115*(2), 1071–1076.

Hansen, T., Pracejus, L., & Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of Vision*, *9*(4), 26.1–12.

Hanslmayr, S., Gross, J., Klimesch, W., & Shapiro, K. L. (2011). The role of alpha oscillations in temporal attention. *Brain Research Reviews*, *67*(1-2), 331–343.

Hanslmayr, S., Klimesch, W., Sauseng, P., Gruber, W., Doppelmayr, M., Freunberger, R., Pecherstorfer, T., & Birbaumer, N. (2007). Alpha phase reset contributes to the generation of ERPs. *Cerebral Cortex* , *17*(1), 1–8.

Hardstone, R., Zhu, M., Flinker, A., Melloni, L., Devore, S., Friedman, D., Dugan, P., Doyle, W. K., Devinsky, O., & He, B. J. (2021). Long-term priors influence visual perception through recruitment of long-range feedback. *Nature Communications*, *12*(1), 6288.

Harvey, B. M., & Dumoulin, S. O. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(38), 13604–13612.

Hayden, B. Y., & Gallant, J. L. (2005). Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. *Neuron*, *47*(5), 637–643.

Heed, T., & Röder, B. (2010). Common anatomical and external coding for hands and feet in tactile attention: evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *22*(1), 184–202.

Helfrich, R. F., Fiebelkorn, I. C., Szczepanski, S. M., Lin, J. J., Parvizi, J., Knight, R. T., & Kastner, S. (2018). Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron*, *99*(4), 854–865.e5.

Heuer, A., & Schubö, A. (2016). The Focus of Attention in Visual Working Memory: Protection of Focused Representations and Its Individual Variation. *PloS One*, *11*(4), e0154228.

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward changes salience in human vision via the anterior cingulate. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(33), 11096–11103.

Hickey, C., Chelazzi, L., & Theeuwes, J. (2014). Reward-Priming of Location in Visual Search. *PloS One*, *9*(7), e103372.

Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience*, *21*(4), 760–775.

Hilimire, M. R., Hickey, C., & Corballis, P. M. (2012). Target resolution in visual search involves the direct suppression of distractors: evidence from electrophysiology. *Psychophysiology*, *49*(4), 504–509.

Hilimire, M. R., Mounts, J. R. W., Parks, N. A., & Corballis, P. M. (2009). Competitive interaction degrades target selection: an ERP study. *Psychophysiology*, *46*(5), 1080–1089.

Hilimire, M. R., Mounts, J. R. W., Parks, N. A., & Corballis, P. M. (2010). Event-related potentials dissociate effects of salience and space in biased competition for visual representation. *PloS One*, *5*(9).

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180.

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: an analysis with event-related brain potentials. *Perception & Psychophysics*, *36*(2), 185–198.

Himmelstoss, N. A., Schuster, S., Hutzler, F., Moran, R., & Hawelka, S. (2020). Co-registration of eye movements and neuroimaging for studying contextual predictions in natural reading. *Language, Cognition and Neuroscience*, 35(*5*), 595–612.

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

Holmes, N. P., & Spence, C. (2004). The body schema and the multisensory representation(s) of peripersonal space. *Cognitive Processing*, *5*(2), 94–105.

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception & Psychophysics*, 81(*7*), 2288–2303.

Hopfield, J. J. (1991). Olfactory computation and object perception. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(15), 6462–6466.

Hopf, J. M., Boelmans, K., Schoenfeld, M. A., Luck, S. J., & Heinze, H.-J. (2004). Attention to features precedes attention to locations in visual search: evidence from electromagnetic brain responses in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(8), 1822–1832.

Hopf, J. M., Luck, S. J., Girelli, M., Hagner, T., Mangun, G. R., Scheich, H., & Heinze, H. J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex* , *10*(12), 1233–1241.

Hopf, J. M., & Mangun, G. R. (2000). Shifting visual attention in space: an electrophysiological analysis using high spatial resolution mapping. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *111*(7), 1241–1257.

Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(2), 423–442.

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*(3), 599–631.

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A Connectionist Model of Visual Search. *Cognitive Psychology*, *25*(1), 43–110.

Iacoboni, M., & Zaidel, E. (2003). Interhemispheric visuo-motor integration in humans: the effect of redundant targets. *The European Journal of Neuroscience*, *17*(9), 1981–1986.

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology*, *103*(4), 1963–1968.

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

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, *2*(3), 194–203.

James, W. (1890). Attention. *The Principles of Psychology*, *1*, 402–458.

Jaśkowski, P., van der Lubbe, R. H. J., Schlotterbeck, E., & Verleger, R. (2002). Traces left on visual selective attention by stimuli that are not consciously identified. *Psychological Science*, *13*(1), 48–54.

Jenkins, M., Grubert, A., & Eimer, M. (2016). Rapid Parallel Attentional Selection Can Be Controlled by Shape and Alphanumerical Category. *Journal of Cognitive Neuroscience*, *28*(11), 1672–1687.

Jenkins, M., Grubert, A., & Eimer, M. (2018). The speed of voluntary and priority-driven shifts of visual attention. *Journal of Experimental Psychology. Human Perception and Performance*, *44*(1), 27–37.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *The European Journal of Neuroscience*, *15*(8), 1395–1399.

Jolicoeur, P., Brisson, B., & Robitaille, N. (2008). Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. *Brain Research*, *1215*, 160–172.

Jonas, J. B., Schneider, U., & Naumann, G. O. (1992). Count and density of human retinal photoreceptors. *Graefe's Archive for Clinical and Experimental Ophthalmology = Albrecht von Graefes Archiv Fur Klinische Und Experimentelle Ophthalmologie*, *230*(6), 505–510.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, *96*(1), 29.

Kahneman, D. (1973). *Attention and Effort*. New York: Prentice-Hall.

Kanayama, N., Tamè, L., Ohira, H., & Pavani, F. (2012). Top down influence on visuo-tactile interaction modulates neural oscillatory responses. *NeuroImage*, *59*(4), 3406–3417.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews. Neuroscience*, *1*(2), 91–100.

Katus, T., & Eimer, M. (2019). The N2cc component as an electrophysiological marker of space-based and feature-based attentional target selection processes in touch. *Psychophysiology*, *56*(9), e13391.

Katus, T., Grubert, A., & Eimer, M. (2015). Electrophysiological Evidence for a Sensory Recruitment Model of Somatosensory Working Memory. *Cerebral Cortex* , *25*(12), 4697–4703.

Kauer, J. S., & White, J. (2001). Imaging and coding in the olfactory system. *Annual Review of Neuroscience*, *24*, 963–979.

Keller, A. S., Payne, L., & Sekuler, R. (2017). Characterizing the roles of alpha and theta oscillations in multisensory attention. *Neuropsychologia*, *99*, 48–63.

Keren, A. S., Yuval-Greenberg, S., & Deouell, L. Y. (2010). Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. *NeuroImage*, 49(*3*), 2248–2263.

Kerzel, D., & Burra, N. (2020). Capture by Context Elements, Not Attentional Suppression of Distractors, Explains the PD with Small Search Displays. *Journal of Cognitive Neuroscience*, *32*(6), 1170–1183.

Kiesel, A., Miller, J., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: a comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*(2), 250–274.

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception & Psychophysics*, *15*(1), 149–158.

Kiss, M., Driver, J., & Eimer, M. (2009). Reward priority of visual target singletons modulates event-related potential signatures of attentional selection. *Psychological Science*, *20*(2), 245–251.

Kiss, M., Grubert, A., Petersen, A., & Eimer, M. (2012). Attentional capture by salient distractors during visual search is determined by temporal task demands. *Journal of Cognitive Neuroscience*, *24*(3), 749–759.

Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology*, *45*(2), 240–249.

Klein, R. M. (1977). Attention and visual dominance: A chronometric analysis. *Journal of Experimental Psychology. Human Perception and Performance*, *3*(3), 365–378.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88.

Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., & Sauseng, P. (2004). Phase-locked alpha and theta oscillations generate the P1–N1 complex and are related to memory performance. *Cognitive Brain Research*, *19*(3), 302–316.

Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*(4), 961–980.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22.

Koch, C., & Tsuchiya, N. (2012). Attention and consciousness: related yet different. *Trends in Cognitive Sciences*, *16*(2), 103–105.

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

Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: evidence for a bottom-up saliency map in V1. *Journal of Vision*, *7*(7), 6.1–14.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, *11*(2), 224–231.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kristjánsson, A., Vuilleumier, P., Schwartz, S., Macaluso, E., & Driver, J. (2007). Neural basis for priming of pop-out during visual search revealed with fMRI. *Cerebral Cortex* , *17*(7), 1612–1624.

Krummenacher, J., Grubert, A., & Müller, H. J. (2010). Inter-trial and redundant-signals effects in visual search and discrimination tasks: separable pre-attentive and post-selective effects. *Vision Research*, *50*(14), 1382–1395.

Krummenacher, J., Grubert, A., Töllner, T., & Müller, H. J. (2014). Salience-based integration of redundant signals in visual pop-out search: evidence from behavioral and electrophysiological measures. *Journal of Vision*, *14*(3), 26.

Krummenacher, J., & Müller, H. J. (2014). Visual search for singleton targets redundantly defined in two feature dimensions: Coactive processing of color-motion targets? *Journal of Experimental Psychology. Human Perception and Performance*, *40*(5), 1926–1939.

Krummenacher, J., Müller, H. J., & Heller, D. (2001a). Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Perception & Psychophysics*.

Krummenacher, J., Müller, H. J., & Heller, D. (2001b). Visual search for dimensionally redundant pop-out targets: evidence for parallel-coactive processing of dimensions. *Perception & Psychophysics*, *63*(5), 901–917.

Krummenacher, J., Müller, H. J., & Heller, D. (2002a). Visual search for dimensionally redundant pop-out targets: parallel-coactive processing of dimensions is location specific. *Journal of Experimental Psychology. Human Perception and Performance*, *28*(6), 1303–1322.

Krummenacher, J., Müller, H. J., & Heller, D. (2002b). Visual search for dimensionally redundant pop-out targets: Redundancy gains in compound tasks. *Visual Cognition*, *9*(7), 801–837.

Krummenacher, J., Müller, H. J., Zehetleitner, M., & Geyer, T. (2009). Dimension- and space-based intertrial effects in visual pop-out search: modulation by task demands for focal-attentional processing. *Psychological Research*, *73*(2), 186–197.

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2017). Dynamic Brain Network Correlates of Spontaneous Fluctuations in Attention. *Cerebral Cortex* , *27*(3), 1831–1840.

LaBerge, D. (1973a). Attention and the measurement of perceptual learning. *Memory & Cognition*, *1*(3), 268–276.

LaBerge, D. (1973b). Identification of two components of the time to switch attention: A test of a serial and a parallel model of attention. *Attention and Performance IV*, 71–85.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology. Human Perception and Performance*, *9*(3), 371–379.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, *96*(1), 101–124.

Laing, D. G., & Glemarec, A. (1992). Selective attention and the perceptual analysis of odor mixtures. *Physiology & Behavior*, *52*(6), 1047–1053.

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron*, *64*(3), 419–430.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Lamy, D. F., & Kristjánsson, A. (2013). Is goal-directed attentional guidance just intertrial priming? A review. *Journal of Vision*, *13*(3), 14.

Lappin, J. S. (1967). Attention in the identification of stimuli in complex visual displays. *Journal of Experimental Psychology*, *75*(3), 321–328.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10), 6.1–16.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(3), 451–468.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197.

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, *2*(4), 375–381.

Leonards, Sunaert, & Van Hecke. (2000). Attention mechanisms in visual search—an fMRI study. *Journal of Cognitive and Behavioral Psychotherapies: The Official Journal of the International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health*.

Leszczynski, M., & Schroeder, C. E. (2019). The Role of Neuronal Oscillations in Visual Active Sensing. *Frontiers in Integrative Neuroscience*, *13*, 32.

Levit, R. A., Sutton, S., & Zubin, J. (1973). Evoked potential correlates of information processing in psychiatric patients. *Psychological Medicine*, *3*(4), 487–494.

Liesefeld, H. R., Liesefeld, A. M., & Müller, H. J. (2021). Preparatory Control Against Distraction Is Not Feature-Based. *Cerebral Cortex*

Liesefeld, H. R., Liesefeld, A. M., Müller, H. J., & Rangelov,

D. (2017). Saliency maps for finding changes in visual scenes? *Attention, Perception & Psychophysics*, *79*(7), 2190–2201.

Liesefeld, H. R., Liesefeld, A. M., Sauseng, P., Jacob, S. N., & Müller, H. J. (2020). How visual working memory handles distraction: cognitive mechanisms and electrophysiological correlates. *Visual Cognition*, *28*(5-8), 372–387.

Liesefeld, H. R., & Müller, H. J. (2019). Distractor handling via dimension weighting. *Current Opinion in Psychology*, *29*, 160–167.

Liesefeld, H. R., & Müller, H. J. (2021). Modulations of saliency signals at two hierarchical levels of priority computation revealed by spatial statistical distractor learning. *Journal of Experimental Psychology. General*, *150*(4), 710–728.

Lindsay, G. W. (2020). Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*, 14, 29.

Lisman, J. E., & Jensen, O. (2013). The θ-γ neural code. *Neuron*, *77*(6), 1002–1016.

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(18), 10530–10535.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16.

Lleras, A., Kawahara, J.-I., Wan, X. I., & Ariga, A. (2008). Intertrial inhibition of focused attention in pop-out search. *Perception & Psychophysics*, *70*(1), 114–131.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490.

Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychological Review*, *103*(4), 603–649.

Luck, S. J. (2011). *Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components*. Oxford University Press.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luck, S. J., Fan, S., & Hillyard, S. A. (1993). Attention-related modulation of sensory-evoked brain activity in a visual search task. *Journal of Cognitive Neuroscience*, *5*(2), 188–195.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the Gap between Monkey Neurophysiology and Human Perception: An Ambiguity Resolution Theory of Visual Selective Attention. *Cognitive Psychology*, *33*(1), 64–87.

Luck, S. J., & Hillyard, S. A. (1994a). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*(3), 291–308.

Luck, S. J., & Hillyard, S. A. (1994b). Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(5), 1000–1014.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440.

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience and Biobehavioral Reviews*, *62*, 100–108.

Lu, Z. L., & Dosher, B. A. (2000). Spatial attention: different mechanisms for central and peripheral temporal precues? *Journal of Experimental Psychology. Human Perception and Performance*, *26*(5), 1534–1548.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention [Review of *Modulation of human visual cortex by crossmodal spatial attention*]. *Science*, *289*(5482), 1206–1208.

Macaluso, E., Frith, C. D., & Driver, J. (2002a). Directing attention to locations and to sensory modalities: multiple levels of selective processing revealed with PET. *Cerebral Cortex* , *12*(4), 357–368.

Macaluso, E., Frith, C. D., & Driver, J. (2002b). Supramodal effects of covert spatial orienting triggered by visual or tactile events. *Journal of Cognitive Neuroscience*, *14*(3), 389–401.

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The Curious Incident of Attention in Multisensory Integration: Bottom-up vs. Top-down. *Multisensory Research*, *29*(6-7), 557–583.

Mackworth, J. F. (1968). Vigilance, arousal, and habituation. *Psychological Review*, *75*(4), 308–322.

MacLean, K. A., Aichele, S. R., Bridwell, D. A., Mangun, G. R., Wojciulik, E., & Saron, C. D. (2009). Interactions between endogenous and exogenous attention during vigilance. *Attention, Perception & Psychophysics*, *71*(5), 1042–1058.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657–672.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics*, *58*(7), 977–991.

Maljkovic, V., & Nakayama, K. (2000). Priming of popout: III. A short-term implicit memory system beneficial for rapid target selection. *Visual Cognition*, *7*(5), 571–595.

Mangun, G. R. (1994). Orienting Attention in the Visual Fields: An Electrophysiological Analysis. In H.-J. Heinze, T. F. Münte, & G. R. Mangun (Eds.), *Cognitive Electrophysiology* (pp. 81–101). Birkhäuser Boston.

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18.

Mangun, G. R., Hansen, J. C., & Hillyard, S. A. (1987). The spatial orienting of attention: Sensory facilitation or response bias. *Current Trends in Event-Related Potential Research*, 118–124.

Mangun, G. R., & Hillyard, S. A. (1987). The spatial allocation of visual attention as indexed by event-related brain potentials. *Human Factors*, *29*(2), 195–211.

Mangun, G. R., & Hillyard, S. A. (1988). Spatial gradients of visual attention: behavioral and electrophysiological evidence. *Electroencephalography and Clinical Neurophysiology*, *70*(5), 417–428.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(4), 1057–1074.

Man, K., Damasio, A., Meyer, K., & Kaplan, J. T. (2015). Convergent and invariant object representations for sight, sound, and touch. *Human Brain Mapping*, *36*(9), 3629–3640.

Mannuzza, S. (1980). Cross-Modal Reaction Time and Schizophrenic Attentional Deficit: A Critical Review. In *Schizophrenia Bulletin* (Vol. 6, Issue 4, pp. 654–675)

Mathewson, K. E., Prudhomme, C., Fabiani, M., Beck, D. M., Lleras, A., & Gratton, G. (2012). Making waves in the stream of consciousness: entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. *Journal of Cognitive Neuroscience*, *24*(12), 2321–2333.

Matusz, P. J., & Eimer, M. (2013). Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*, *50*(10), 996–1009.

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *181*(3), 531–536.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287–330.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906–908.

McDonald, J. J., Teder-Sälejärvi, W. A., Russo, F. D., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*, *15*(1), 10–19.

McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., & Horwitz, B. (1994). Network analysis of cortical visual pathways mapped with PET. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *14*(2), 655–666.

McLelland, D., & VanRullen, R. (2016). Theta-Gamma Coding Meets Communication-through-Coherence: Neuronal Oscillatory Multiplexing Theories Reconciled. *PLoS Computational Biology*, *12*(10), e1005162.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, *332*(6160), 154–155.

Medina, J., & Coslett, H. B. (2010). From maps to form to space: touch and the body schema. *Neuropsychologia*, *48*(3), 645–654.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662.

Mesulam, M. M. (1998). From sensation to cognition. *Brain: A Journal of Neurology*, *121 (Pt 6)*, 1013–1052.

Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, *32*(7), 376–382.

Michel, R., Dugué, L., & Busch, N. A. (2021). Distinct contributions of alpha and theta rhythms to perceptual and attentional sampling. *The European Journal of Neuroscience*

Miles, E., Brown, R., & Poliakoff, E. (2011). Investigating the nature and time-course of the modality shift effect between vision and touch. *Quarterly Journal of Experimental Psychology* , *64*(5), 871–888.

Miller, J. (1978). Multidimensional same--different judgments: Evidence against independent comparisons of dimensions. *Journal of Experimental Psychology. Human Perception and Performance*, *4*(3), 411.

Miller, J. (1981). Global precedence in attention and decision. *Journal of Experimental Psychology. Human Perception and Performance*. https://psycnet.apa.org/record/1982-22665-001

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331–343.

Miller, J. (2016). Statistical facilitation and the redundant signals effect: What are race and coactivation models? *Attention, Perception & Psychophysics*, *78*(2), 516–519.

Miller, J., Beutinger, D., & Ulrich, R. (2009). Visuospatial attention and redundancy gain. *Psychological Research*, *73*(2), 254–262.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*(1), 99–115.

Miller, J., & Reynolds, A. (2003). The locus of redundant-targets and nontargets effects: evidence from the psychological refractory period paradigm. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(6), 1126–1142.

Miller, J., Ulrich, R., & Lamarre, Y. (2001). Locus of the redundant-signals effect in bimodal divided attention: a neurophysiological analysis. *Perception & Psychophysics*, *63*(3), 555–562.

Miller, J., Ulrich, R., & Schwarz, W. (2009). Why jackknifing yields good latency estimates. *Psychophysiology*, *46*(2), 300–312.

Milner, A. D. (2017). How do the two visual streams interact with each other? *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *235*(5), 1297–1308.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417.

Mitsea, E., Drigas, A., & Skianis, C. (2022). Breathing, attention & consciousness in sync: The role of breathing training, metacognition & virtual reality. *Technium Soc. Sci. J.*, 29, 79.

Mombaerts, P. (1999). Molecular biology of odorant receptors in vertebrates. *Annual Review of Neuroscience*, *22*, 487–509.

Mordkoff, J. T., & Miller, J. (1993). Redundancy gains and coactivation with two different targets: The problem of target preferences and the effects of display frequency. *Perception & Psychophysics*, *53*(5), 527–535.

Mordkoff, J. T., Miller, J., & Roch, A. C. (1996). Absence of coactivation in the motor component: evidence from psychophysiological measures of target detection. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(1), 25–41.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(2), 520–538.

Mordkoff, J. T., & Yantis, S. (1993). Dividing attention between color and shape: evidence of coactivation. *Perception & Psychophysics*, *53*(4), 357–366.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, *4*(2), 61–64.

Mortier, K., Theeuwes, J., & Starreveld, P. (2005). Response selection modulates visual search within and across dimensions. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(3), 542–557.

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

Müller, H. J., & Krummenacher, J. (2006). Locus of dimension weighting: Preattentive or postselective? *Visual Cognition*, *14*(4-8), 490–513.

Müller, H. J., 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.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(2), 315–330.

Müller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(5), 1021–1035.

Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A Physiological Correlate of the "Zoom Lens" of Visual Attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(9), 3561–3565.

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection. *Psychological Bulletin*, *92*(3), 605–640.

Näätänen, R. (1988). Implications of ERP data for psychological theories of attention. *Biological Psychology*, *26*(1-3), 117–163.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *The Behavioral and Brain Sciences*, *13*(2), 201–233.

Näätänen, R., & Gaillard, A. W. K. (1983). 5 The orienting reflex and the N2 deflection of the event-related potential (ERP). In *Advances in psychology* (Vol. 10, pp. 119–141). Elsevier.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*(6059), 264–265.

Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*(4), 605–617.

Neisser, U. (1967). *Cognitive Psychology: Classic Edition*. Psychology Press.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: energy summation or preparation enhancement? *Psychological Review*, *80*(6), 489–509.

Nikolaev, A. R., Meghanathan, R. N., & van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, 107, 55–83.

Nothdurft, H. (2000). Salience from feature contrast: additivity across dimensions. *Vision Research*, *40*(10-12), 1183–1201.

Nothdurft, H. C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception &*

Psychophysics, *52*(4), 355–375.

Nothdurft, H. C. (1993). Saliency effects across dimensions in visual search. *Vision Research*, *33*(5-6), 839–844.

Novitskiy, N., Ramautar, J. R., Vanderperren, K., De Vos, M., Mennes, M., Mijovic, B., Vanrumste, B., Stiers, P., Van den Bergh, B., Lagae, L., Sunaert, S., Van Huffel, S., & Wagemans, J. (2011). The BOLD correlates of the visual P1 and N1 in single-trial analysis of simultaneous EEG-fMRI recordings during a spatial detection task. *NeuroImage*, *54*(2), 824–835.

Nunez, P. L., & Silberstein, R. B. (2000). On the Relationship of Synaptic Activity to Macroscopic Measurements: Does Co-Registration of EEG with fMRI Make Sense? *Brain Topography*, 13(*2*), 79–96.

Nunez, P. L., Wingeier, B. M., & Silberstein, R. B. (2001). Spatial-temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale measurements, and global binding of local networks. *Human Brain Mapping*, *13*(3), 125–164.

Oberauer, K. (2019). Working memory and attention—A conceptual analysis and review. *Journal of Cognition*, *2*(1), 36.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: when it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334.

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage*, *27*(2), 341–356.

Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology: CB*, *22*(15), 1391–1396.

Palmer, E. M., Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2019). Measuring the time course of selection during visual search. *Attention, Perception & Psychophysics*, *81*(1), 47–60.

Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(16), 7580–7585.

Papaioannou, O., & Luck, S. J. (2020). Effects of eccentricity on the attention-related N2pc component of the event-related potential waveform. *Psychophysiology*, *57*(5), e13532.

Pashler, H. (1987). Detecting conjunctions of color and form: reassessing the serial search hypothesis. *Perception & Psychophysics*, *41*(3), 191–201.

Pashler, H. (1997). *The psychology of attention*. The MIT Press.

Perry, C. J., & Fallah, M. (2014). Feature integration and object representations along the dorsal stream visual hierarchy. *Frontiers in Computational Neuroscience*, *8*, 84.

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*(1), 73–89.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalography and Clinical Neurophysiology*, *36*(2), 191–199.

Polich, J. (1986). Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, *63*(3), 251–259.

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.

Posner, M. I. (1978). *Chronometric explorations of mind. 271*. https://psycnet.apa.org/fulltext/1980-02340-000.pdf

Posner, M. I. (1980). Orienting of Attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(16), 7398–7403.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neurosciences*, *17*(2), 75–79.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance. *Psychological Review*, *83*(2), 157–171.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, *2*(3), 211–228.

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

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38.

Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: an examination of the task constraints on feature and conjunction searches. *Perception & Psychophysics*, *41*(5), 455–472.

Quinn, B. T., Carlson, C., Doyle, W., Cash, S. S., Devinsky, O., Spence, C., Halgren, E., & Thesen, T. (2014). Intracranial cortical responses during visual-tactile integration in humans. *Journal of Neuroscience*, *34*(1), 171–181.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.

Rangelov, D., Müller, H. J., & Zehetleitner, M. (2013). Visual search for feature singletons: multiple mechanisms produce sequence effects in visual search. *Journal of Vision*, *13*(3), 22:1–16.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*(3), 618–660.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, *7*, 361.

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, *24*(1), 19–29, 111–125.

Riccio, C. A., Reynolds, C. R., Lowe, P., & Moore, J. J. (2002). The continuous performance test: a window on the neural substrates for attention? *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *17*(3), 235–272.

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature*, *400*(6740), 162–166.

Rodway, P. (2005). The modality shift effect and the effectiveness of warning signals in different modalities. *Acta Psychologica*, *120*(2), 199–226.

Roland, P. E. (1982). Cortical regulation of selective attention in man. A regional cerebral blood flow study. *Journal of Neurophysiology*, *48*(5), 1059–1078.

Rosenberg, M. D., Finn, E. S., Constable, R. T., & Chun, M. M. (2015). Predicting moment-to-moment attentional state. *NeuroImage*, *114*, 249–256.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171.

Rosenberg, M. D., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. *Attention, Perception & Psychophysics*, *75*(3), 426–439.

Rosli, R. M., Tan, H. Z., Proctor, R. W., & Gray, R. (2008). Attentional gradient for crossmodal proximal-distal tactile cueing of visual spatial attention. *ACM Transactions on Applied Perception*, *8*(4), 1–12.

Ruge, H., & Naumann, E. (2006). Brain-electrical correlates of negative location priming under sustained and transient attentional context conditions. *Journal of Psychophysiology*, *20*(3), 160–169.

Sagi, D., & Julesz, B. (1985). "Where" and "what" in vision. *Science*, *228*(4704), 1217–1219.

Sams, M., Alho, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology*, *17*(1), 41–58.

Sato, T., Okada, Y., Miyamoto, T., & Fujiyama, R. (1999). Distributions of sensory spots in the hand and two-point discrimination thresholds in the hand, face and mouth in dental students. *Journal of Physiology, Paris*, *93*(3), 245–250.

Sauter, M., Liesefeld, H. R., & Müller, H. J. (2019). Learning to suppress salient distractors in the target dimension: Region-based inhibition is persistent and transfers to distractors in a nontarget dimension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*(11), 2080–2097.

Sauter, M., Liesefeld, H. R., Zehetleitner, M., & Müller, H. J. (2018). Region-based shielding of visual search from salient distractors: Target detection is impaired with same- but not different-dimension distractors. *Attention, Perception & Psychophysics*, *80*(3), 622–642.

Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception & Psychophysics*, *72*(6), 1455–1470.

Sawaki, R., & Luck, S. J. (2013). Active suppression after involuntary capture of attention. *Psychonomic Bulletin & Review*, *20*(2), 296–301.

Schack, B., Weiss, S., & Rappelsberger, P. (2003). Cerebral information transfer during word processing: where and when does it occur and how fast is it? *Human Brain Mapping*, *19*(1), 18–36.

Schaffer, S., Schubö, A., & Meinecke, C. (2011). Electrophysiological correlates of target eccentricity in texture segmentation. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *80*(3), 198–209.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Schreij, D., Owens, C., & Theeuwes, J. (2008). Abrupt onsets capture attention independent of top-down control settings. *Perception & Psychophysics*, *70*(2), 208–218.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*(2), 172–176.

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755–759.

Schröter, H., Ulrich, R., & Miller, J. (2007). Effects of redundant auditory stimuli on reaction time.

Psychonomic Bulletin & Review, *14*(1), 39–44.

Schwarz, W. (1989). A new model to explain the redundant-signals effect. *Perception & Psychophysics*, *46*(5), 498–500.

Schwarz, W. (1994). Diffusion, Superposition, and the Redundant-Targets Effect. *Journal of Mathematical Psychology*, *38*(4), 504–520.

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, *117*(2), 230.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences*, *31*(8), 401–409.

Sereno, M. I., McDonald, C. T., & Allman, J. M. (1994). Analysis of retinotopic maps in extrastriate cortex. *Cerebral Cortex* , *4*(6), 601–620.

Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*(5545), 1350–1354.

Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(16), 10917–10922.

Shanahan, M., & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition*, *98*(2), 157–176.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190.

Shim, W. M., Jiang, Y. V., & Kanwisher, N. (2013). Redundancy gains in retinotopic cortex. *Journal of Neurophysiology*, *110*(9), 2227–2235.

Shulman, G. L., d'Avossa, G., Tansy, A. P., & Corbetta, M. (2002). Two attentional processes in the parietal lobe. *Cerebral Cortex* , *12*(11), 1124–1131.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, *4*(4), 147–155.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, *24*(1), 49–65, 111–125.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–586.

Singh-Curry, V., & Husain, M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia*, *47*(6), 1434–1448.

Slagter, H. A., Prinssen, S., Reteig, L. C., & Mazaheri, A. (2016). Facilitation and inhibition in attention: Functional dissociation of pre-stimulus alpha activity, P1, and N1 components. *NeuroImage*, *125*, 25–35.

Smulders, F. T. Y. (2010). Simplifying jackknifing of ERPs and getting more out of it: retrieving estimates of participants' latencies. *Psychophysiology*, *47*(2), 387–392.

Solovey, G., Graney, G. G., & Lau, H. (2015). A decisional account of subjective inflation of visual perception at the periphery. *Attention, Perception & Psychophysics*, *77*(1), 258–271.

Somers, D. C., & Sheremata, S. L. (2013). Attention maps in the brain. *Wiley Interdisciplinary Reviews. Cognitive Science*, *4*(4), 327–340.

Soto, D., Humphreys, G. W., & Heinke, D. (2006). Working memory can guide pop-out search. *Vision Research*, *46*(6-7), 1010–1018.

Soto, D., Humphreys, G. W., & Rotshtein, P. (2007). Dissociating the neural mechanisms of memory-based guidance of visual selection. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17186–17191.

Spence, C. (2002). Multisensory attention and tactile information-processing. *Behavioural Brain Research*, *135*(1-2), 57–64.

Spence, C., & Driver, J. (1997a). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*(1), 1–22.

Spence, C., & Driver, J. (1997b). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, *59*(3), 389–403.

Spence, C., & Driver, J. (2004). *Crossmodal Space and Crossmodal Attention*. OUP Oxford.

Spence, C., & Gallace, A. (2007). Recent developments in the study of tactile attention. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *61*(3), 196–207.

Spence, C., Kettenmann, B., Kobal, G., & McGlone, F. P. (2000). Selective attention to the chemosensory modality. *Perception & Psychophysics*, *62*(6), 1265–1271.

Spence, C., & McGlone, F. P. (2001). Reflexive spatial orienting of tactile attention. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *141*(3), 324–330.

Spence, C., McGlone, F. P., Kettenmann, B., & Kobal, G. (2001). Attention to olfaction. A psychophysical investigation. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *138*(4), 432–437.

Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*(2), 330–336.

Spence, C., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(4), 1298–1319.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: a review. *Hearing Research*, *258*(1-2), 134–142.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology. General*, *130*(4), 799–832.

Spence, C., Wan, X., Woods, A., Velasco, C., Deng, J., Youssef, J., & Deroy, O. (2015). On tasty colours and colourful tastes? Assessing, explaining, and utilizing crossmodal correspondences between colours and basic tastes. *Flavour*, *4*(1), 23.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*, *193*(4258), 1142–1146.

Stahl, J., & Gibbons, H. (2004). The application of jackknife-based onset detection of lateralized readiness potential in correlative approaches. *Psychophysiology*, *41*(6), 845–860.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *31*(1), 137–149.

Staugaard, C. F., Petersen, A., & Vangkilde, S. (2016). Eccentricity effects in vision and attention. *Neuropsychologia*, *92*, 69–78.

Stein, B. E., & Meredith, A. M. (1993). *The Merging of the Senses*. MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, *9*(4), 255–266.

Stein, J. F. (1989). Representation of egocentric space in the posterior parietal cortex. *Quarterly Journal of Experimental Physiology* , *74*(5), 583–606.

Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*(4), 739–750.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188.

Talcott, T. N., & Gaspelin, N. (2020). Prior target locations attract overt attention during search. *Cognition*, *201*, 104282.

Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*, 19.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098–1114.

Tay, D., McIntyre, D. L., & McDonald, J. J. (2022). Searching for Visual Singletons Without A Feature to Guide Attention. *Journal of Cognitive Neuroscience*, 1–18.

Taylor, T. L., & Klein, R. M. (2000). Visual and motor effects in inhibition of return. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(5), 1639–1656.

Teder-Sälejärvi, W. A., Hillyard, S. A., Röder, B., & Neville, H. J. (1999). Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials. *Brain Research. Cognitive Brain Research*, *8*(3), 213–227.

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Research. Cognitive Brain Research*, *14*(1), 106–114.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, *50*(2), 184–193.

Theeuwes, J. (1992). Perceptual selectivity for color and form. In *Perception & Psychophysics* (Vol. 51, Issue 6, pp. 599–606).

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2), 77–99.

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97–101.

Theeuwes, J., Bogaerts, L., & van Moorselaar, D. (2022). What to expect where and when: how statistical learning drives visual selection. *Trends in Cognitive Sciences*, *26*(10), 860–872.

Thesen, T., Vibell, J. F., Calvert, G. A., & Österbauer, R. A. (2004). Neuroimaging of multisensory processing in vision, audition, touch, and olfaction. *Cognitive Processing*, *5*(2), 84–93.

Thibault, R. T., MacPherson, A., Lifshitz, M., Roth, R. R., & Raz, A. (2018). Neurofeedback with fMRI: A critical systematic review. *NeuroImage*, *172*, 786–807.

Todd, J. W. (1912). *Reaction to multiple stimuli*. The Science Press.

Töllner, T., Gramann, K., Müller, H. J., & Eimer, M. (2009). The anterior N1 component as an index of modality shifting. *Journal of Cognitive Neuroscience*, *21*(9), 1653–1669.

Töllner, T., Gramann, K., Müller, H. J., Kiss, M., & Eimer, M. (2008). Electrophysiological markers of visual dimension changes and response changes. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(3), 531–542.

Töllner, T., Rangelov, D., & Müller, H. J. (2012). How the speed of motor-response decisions, but not focal-attentional selection, differs as a function of task set and target prevalence. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), E1990–E1999.

Töllner, T., Zehetleitner, M., Gramann, K., & Müller, H. J. (2010). Top-down weighting of visual dimensions: behavioral and electrophysiological evidence. *Vision Research*, *50*(14), 1372–1381.

Töllner, T., Zehetleitner, M., Gramann, K., & Müller, H. J. (2011). Stimulus saliency modulates pre-attentive processing speed in human visual cortex. *PloS One*, *6*(1), e16276.

Töllner, T., Zehetleitner, M., Krummenacher, J., & Müller, H. J. (2011). Perceptual basis of redundancy gains in visual pop-out search. *Journal of Cognitive Neuroscience*, *23*(1), 137–150.

Townsend, J. T., & Ashby, G. F. (1983). *Stochastic Modeling of Elementary Psychological Processes*. Cambridge University Press.

Treisman, A. M. (1960). Contextual cues in selective listening. *The Quarterly Journal of Experimental Psychology*, *12*(4), 242–248.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299.

Treisman, A. M. (1988). Features and objects: the fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *40*(2), 201–237.

Treisman, A. M. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178.

Treisman, A. M. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1373), 1295–1306.

Treisman, A. M. (1999). Solutions to the binding problem: progress through controversy and convergence. *Neuron*, *24*(1), 105–110, 111–125.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treisman, A. M., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, *33*(6), 527–532.

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(3), 459–478.

Turatto, M., Benso, F., Galfano, G., & Umiltà, C. (2002). Nonspatial attentional shifts between audition and vision. *Journal of Experimental Psychology. Human Perception and Performance*, *28*(3), 628–639.

Turatto, M., Galfano, G., Bridgeman, B., & Umiltà, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *155*(3), 301–310.

Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: testability and checking of cross-channel dependency. *Perception & Psychophysics*, *39*(4), 248–254.

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, *38*(5), 816–827.

Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: an algorithm and computer programs. *Behavior Research Methods*, *39*(2), 291–302.

van Bergen, R. S., & Jehee, J. F. M. (2019). Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *39*(41), 8164–8176.

van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*, *18*(12), 1728–1730.

van Es, M. W. J., Marshall, T. R., Spaak, E., Jensen, O., & Schoffelen, J.-M. (2022). Phasic modulation of visual representations during sustained attention. *The European Journal of Neuroscience*, *55*(11-12), 3191–3208.

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375.

van Moorselaar, D., Lampers, E., Cordesius, E., & Slagter, H. A. (2020). Neural mechanisms underlying expectation-dependent inhibition of distracting information. *eLife*, *9*.

van Moorselaar, D., & Slagter, H. A. (2019). Learning What Is Irrelevant or Relevant: Expectations Facilitate Distractor Inhibition and Target Facilitation through Distinct Neural Mechanisms. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *39*(35), 6953–6967.

VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Sciences*, *20*(10), 723–735.

Van Voorhis, S., & Hillyard, S. A. (1977). Visual evoked potentials and selective attention to points in space. *Perception & Psychophysics*, *22*(1), 54–62.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology. General*, *123*(2), 146–160.

Veldhuizen, M. G., & Small, D. M. (2011). Modality-specific neural effects of selective attention to taste and odor. *Chemical Senses*, *36*(8), 747–760.

Verleger, R., & Cohen, R. (1978). Effects of certainty, modality shift and guess outcome on evoked potentials and reaction times in chronic schizophrenics. *Psychological Medicine*, *8*(1), 81–93.

Verleger, R., Zurawska Vel Grajewska, B., & Jaśkowski, P. (2012). Time-course of hemispheric preference for processing contralateral relevant shapes: P1pc, N1pc, N2pc, N3pc. *Advances in Cognitive Psychology / University of Finance and Management in Warsaw*, *8*(1), 19–28.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751.

Võ, M. L.-H., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences*, *1339*(1), 72–81.

von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *38*(3), 301–313.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1993). Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *Journal of Neurophysiology*, *69*(6), 1797–1809.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*(4943), 380–384.

Walz, J. M., Goldman, R. I., Carapezza, M., Muraskin, J., Brown, T. R., & Sajda, P. (2014). Simultaneous EEG-fMRI reveals a temporal cascade of task-related and default-mode activations during a simple target detection task. *NeuroImage*, *102 Pt 1*, 229–239.

Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture independent of search strategy. *Attention, Perception & Psychophysics*, *80*(7), 1763–1774.

Wang, B., & Theeuwes, J. (2020). Salience determines attentional orienting in visual selection. *Journal of Experimental Psychology. Human Perception and Performance*, *46*(10), 1051–1057.

Wang, J., Zhou, T., Qiu, M., Du, A., Cai, K., Wang, Z., Zhou, C., Meng, M., Zhuo, Y., Fan, S., & Chen, L. (1999). Relationship between ventral stream for object vision and dorsal stream for spatial vision: An fMRI+ERP study. *Human Brain Mapping*, *8*(4), 170–181.

Warden, M. R., & Miller, E. K. (2010). Task-dependent changes in short-term memory in the prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(47), 15801–15810.

Wei, P., & Ji, L. (2021). Reward expectation modulates N2pc for target selection: Electrophysiological evidence. *Psychophysiology*, *58*(8), e13837.

Wei, P., Müller, H. J., Pollmann, S., & Zhou, X. (2009). Neural basis of interaction between target presence and display homogeneity in visual search: an fMRI study. *NeuroImage*, *45*(3), 993–1001.

Whitehead, R. (1991). Right hemisphere processing superiority during sustained visual attention. *Journal of Cognitive Neuroscience*, 3(*4*), 329–334.

Wolber, M., & Wascher, E. (2005). The Posterior Contralateral Negativity as a Temporal Indicator of Visuo-Spatial Processing. *Journal of Psychophysiology*, *19*(3), 182–194.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*(3), 170–191.

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). New York: Psychology Press.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, *24*(1), 11–17, 111–125.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(3), 419–433.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1–8.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*(12), 1411–1426.

Wolfe, J. M., O'Neill, P., & Bennett, S. C. (1998). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception & Psychophysics*, *60*(1), 140–156.

Wollenberg, L., Deubel, H., & Szinte, M. (2018). Visual attention is not deployed at the endpoint of averaging saccades. *PLoS Biology*, *16*(6), e2006548.

Won, B.-Y., Kosoyan, M., & Geng, J. J. (2019). Evidence for second-order singleton suppression based on probabilistic expectations. *Journal of Experimental Psychology. Human Perception and Performance*, *45*(1), 125–138.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*(6747), 867–869.

Woodman, G. F., & Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science*, *14*(6), 605–611.

Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex* , *17 Suppl 1*, i118–i124.

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, *15*(1), 223–229.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, *12*(3), 219–224.

Wyble, B., Chen, H., Stucynski, J., Callahan-Flintoft, C., & Tan, M. (2015). Understanding the PD and the N2pc: modeling the neural mechanisms underlying spatial attention shifts. *Journal of Vision*, *15*(12), 1250–1250.

Xia, J., Zhang, W., Jiang, Y., Li, Y., & Chen, Q. (2018). Neural practice effect during cross-modal selective attention: Supra-modal and modality-specific effects. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *106*, 47–64.

Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: evidence from focused attention tasks. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(1), 135–149.

Yoshor, D., Bosking, W. H., Ghose, G. M., & Maunsell, J. H. R. (2007). Receptive fields in human visual cortex mapped with surface electrodes. *Cerebral Cortex*, *17*(10), 2293–2302.

Zehetleitner, M., Koch, A. I., Goschy, H., & Müller, H. J. (2013). Salience-based selection: attentional capture by distractors less salient than the target. *PloS One*, *8*(1), e52595.

Zehetleitner, M., Krummenacher, J., & Müller, H. J. (2009). The detection of feature singletons defined in two dimensions is based on salience summation, rather than on serial exhaustive or interactive race architectures. *Attention, Perception & Psychophysics*, *71*(8), 1739–1759.

Zehetleitner, M., Muller, H. J., & Krummenacher, J. (2008). The redundant-signals paradigm and preattentive visual processing. *Frontiers in Bioscience: A Journal and Virtual Library*, *13*, 5279–5293.

Zehetleitner, M., Ratko-Dehnert, E., & Müller, H. J. (2015). Modeling violations of the race model inequality in bimodal paradigms: co-activation from decision and non-decision components. *Frontiers in Human Neuroscience*, *9*, 119.

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the New York Academy of Sciences*, *1339*(1), 154–164.

Zhang, B., Allenmark, F., Liesefeld, H. R., Shi, Z., & Müller, H. J. (2019). Probability cueing of singleton-distractor locations in visual search: Priority-map- versus dimension-based inhibition? *Journal of Experimental Psychology. Human Perception and Performance*, *45*(9), 1146–1163.

Zinchenko, A., Conci, M., Töllner, T., Müller, H. J., & Geyer, T. (2020). Automatic Guidance (and Misguidance) of Visuospatial Attention by Acquired Scene Memory: Evidence From an N1pc Polarity Reversal. *Psychological Science*, *31*(12), 1531–1543.

Zivony, A., Allon, A. S., Luria, R., & Lamy, D. (2018). Dissociating between the N2pc and attentional shifting: An attentional blink study. *Neuropsychologia*, *121*, 153–163.

Zivony, A., & Lamy, D. (2018). Contingent Attentional Engagement: Stimulus- and Goal-Driven Capture Have Qualitatively Different Consequences. *Psychological Science*, *29*(12), 1930–1941.

Zubin, J. (1975). Problem of attention in schizophrenia. In M. L. Kietzman, S. Sutton, & J. Zubin (Eds.), *Experimental Approaches to Psychopathology* (pp. 139–166). New York: Academic Press, Inc.

Zylberberg, A., Dehaene, S., Mindlin, G. B., & Sigman, M. (2009). Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model. *Frontiers in Computational Neuroscience*, *3*, 4.

List of Author Contributions

Jan Nasemann (JN), Zhuanghua Shi (ZS), Hermann J. Müller (HJM), Thomas Töllner (TT)

1. **General Introduction**
   JN conceived and wrote the manuscript and created the figures.

2. **Redundancy Gains in Crossmodal Search**
   Nasemann, J., Töllner, T., Müller, H. J., & Shi, Z. (2023). Hierarchy of Intra- and Cross-modal Redundancy Gains in Visuo-tactile Search: Evidence from the Posterior Contralateral Negativity. *Journal of Cognitive Neuroscience* 35(*4*), 1–28.
   TT, JN, and ZS conceived and designed the study, ZS and JN programmed the experimental setup and task. JN collected and analyzed the data with statistics, created the figures, and wrote the manuscript together with HJM and ZS.

3. **Spatial Attention in Crossmodal Search**
   Crossmodal Bending in Spatial Attention (in preparation).
   JN conceived the study, collected the data, performed the analyses and statistics, made the figures, and wrote the manuscript. ZS and HJM revised and commented on the manuscript.

4. **Intertrial Effects in Crossmodal Search**
   Reduced Amplitudes in PCN and CCN Indicate a Common Mechanism for Modality-Shifts in Visuo-Tactile Search (in preparation).
   TT & ZS conceived and designed the study. ZS programmed the experiment. JN analyzed the data, performed statistics, created the figures, and wrote the manuscript with support from ZS and HJM (revision and comments).

5. **General Discussion**
   JN conceived and wrote the manuscript.

……………………………… ………………………………

Jan Nasemann Prof. Dr. Zhuanghua Shi