

Brain and Cognitive Mechanisms of Top-Down Attentional Control in a Multisensory World: Benefits of Electrical Neuroimaging

Pawel J. Matusz^{1,2,3*}, Nora Turoman², Ruxandra I. Tivadar²,
Chrysa Retsa², and Micah M. Murray^{2,3,4*}

Abstract

■ In real-world environments, information is typically multisensory, and objects are a primary unit of information processing. Object recognition and action necessitate attentional selection of task-relevant from among task-irrelevant objects. However, the brain and cognitive mechanisms governing these processes remain not well understood. Here, we demonstrate that attentional selection of visual objects is controlled by integrated top-down audiovisual object representations (“attentional templates”) while revealing a new brain mechanism through which they can operate. In multistimulus (visual) arrays, attentional selection of objects in humans and animal models is traditionally quantified via “the N2pc component”: spatially selective enhancements of neural processing of objects within ventral visual cortices at approximately 150–300 msec poststimulus. In our adaptation of Folk et al.’s [Folk, C. L., Remington, R. W., & Johnston, J. C. Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human*

Perception and Performance, 18, 1030–1044, 1992] spatial cueing paradigm, visual cues elicited weaker behavioral attention capture and an attenuated N2pc during audiovisual versus visual search. To provide direct evidence for the brain, and so, cognitive, mechanisms underlying top-down control in multisensory search, we analyzed global features of the electrical field at the scalp across our N2pcs. In the N2pc time window (170–270 msec), color cues elicited brain responses differing in strength and their topography. This latter finding is indicative of changes in active brain sources. Thus, in multisensory environments, attentional selection is controlled via integrated top-down object representations, and so not only by separate sensory-specific top-down feature templates (as suggested by traditional N2pc analyses). We discuss how the electrical neuroimaging approach can aid research on top-down attentional control in naturalistic, multisensory settings and on other neurocognitive functions in the growing area of real-world neuroscience. ■

INTRODUCTION

The ultimate goals of cognitive neuroscience are to create accurate models of how information processing occurs in everyday situations and how the governing mechanisms are orchestrated by the brain. There are increasing numbers of voices in the community calling for discussions of the optimal ways of achieving these goals (e.g., Jonas & Kording, 2017; Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017; Poldrack et al., 2017; Love, 2015). Our environments are typically cluttered and complex, and only some of the objects and events registered by the sensory systems become associated with our current aims and behavioral responses. To better understand how perception is orchestrated in such settings, research on attentional control processes

has emulated one or many of these features (Kingstone, 1992; Posner, 1980; Treisman & Gelade, 1980). Rigorous tasks isolating cognitive processes of interest and identifying their behavioral measures, combined with careful experimental manipulations, were instrumental to developing theories of attentional control (and related cognitive processes).

At the same time, studies using brain imaging and mapping methods have been fundamental in refining and refuting these theories. Early electrophysiological recordings in nonhuman mammals revealed how top-down, goal-based attention controls stimulus processing through “gain control,” that is, enhancing neuronal responses to stimuli with task-relevant features (e.g., spatial location) and resolving stimulus competition in multistimulus settings (Desimone & Duncan, 1995; Moran & Desimone, 1985). Yet, it was the use of EEG that allowed wide-scale research into attentional control mechanisms. The excellent temporal resolution of EEG helped to characterize the earliest information processing stages influenced by spatial attention and how the latter interacts with feature-based top-down attention in multistimulus

¹University of Applied Sciences Western Switzerland, ²University Hospital Centre and University of Lausanne, ³Vanderbilt University, Nashville, TN, ⁴University of Lausanne and Fondation Asile des Aveugles

*Coauthors contributed equally to this work.

contexts. Theoretical progress was further facilitated by systematic focus on specific ERP components: sequences of electric field scalp topographies occurring in a set temporal order in specific stimulus/task contexts that served as temporally resolved brain correlates of perceptual and cognitive processes (Eimer, Kiss, Press, & Sauter, 2009; Eimer, 1996; Luck & Hillyard, 1994; Heinze, Luck, Mangun, & Hillyard, 1990; Mangun & Hillyard, 1990). More recently, novel mechanisms orchestrating top-down attention control have been revealed by hemodynamic imaging methods, such as fMRI and PET. Such neuroimaging studies provided direct evidence for the purported top-down “attentional templates” (Duncan & Humphreys, 1989), that is, actively maintained working memory representations of task-relevant object features whereby the brain biases sensory processing toward our goals. These biasing signals were revealed to be generated by a network of frontoparietal “sources” (e.g., Serences, Schwarzbach, Courtney, Golay, & Yantis, 2004; Le, Pardo, & Hu, 1998; Wojciulik, Kanwisher, & Driver, 1998). More recently, fMRI studies revealed how both stable and flexible patterns of connectivity between the frontoparietal network and sensory cortices are instrumental to top-down attention (Hwang, Shine, & D’Esposito, 2017; Parks & Madden, 2013).

Research into how neurocognitive functions operate in veridical, real-world environments is increasingly popular, and EEG is becoming increasingly applied in this area. The multitude of advantages of EEG—its low cost, portability, noninvasiveness—is likely behind this popularity. Therein, measures relying on the exquisite temporal resolution of EEG have been shedding light on the mechanisms governing joint action and attention and learning in everyday situations (Bevilacqua et al., 2018; Müller, Sängler, & Lindenberger, 2018; Tseng, Rajangam, Lehew, Lebedev, & Nicoletis, 2018; Dikker et al., 2017; Ko, Komarov, Hairston, Jung, & Lin, 2017; Cohen & Parra, 2016). The way to fully capitalize on the hardware- and information-level benefits of EEG likely lies in employing signal processing techniques that extract both temporal and reasonably resolved spatial information from EEG data that are also neurobiologically interpretable (Tivadar & Murray, 2018).

EEG-based techniques such as electrical neuroimaging (EN) focus on reference-independent measures of the global, rather than local, features of the brain’s electric fields at the scalp. Such measures allow scientists to readily distinguish between ERP modulations elicited by changes in the strength of response within statistically indistinguishable brain networks and those driven by alternations in the activated brain networks, respectively. Biophysical laws dictate that differences in topographies, for example, those associated with distinct experimental conditions (as indicated by DISS and related measures: see Methods), forcibly arise from changes in the configuration of the underlying sources (Fender, 1987; Helmholtz, 1853). EN measures are perhaps best suited

for real-world neuroscientific research, as they are highly robust against small montages ($\sim \leq 20$ channels) typically used in such research. In particular, the stable patterns of EEG activity and their features (e.g., duration) can be derived with high test-retest reliability with even eight-channel data (Khanna, Pascual-Leone, & Farzan, 2014). Furthermore, EN-analyzed data can more readily be compared/shared across laboratories because the measures are both global (i.e., not relying on a specific electrode montage nor on specific electrode sites) and reference independent. Different laboratories might perhaps use the same reference, but they likely will have different conventions. This raises the issue of which laboratory/reference is the “right” one. It is this issue that impedes comparison across laboratories. An EN framework thus has major statistical and interpretational advantages over traditional voltage-based ERP analyses.

Most research on attentional control, systematic investigation remains focused on single senses and so on processes gauged by visual or auditory objects alone. As such, the understanding of attentional control via “multisensory” object templates and how these multisensory templates are represented in the brain is exceptionally poor. This is problematic for several reasons. First, real-world environments are de facto multisensory, and converging evidence points to a fundamentally multisensory fashion in which our brain encodes both space and object identity (Maidenbaum, Abboud, & Amedi, 2014; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009; Assad & Maunsell, 1995). Some multisensory processes take place in early sensory cortices at stages preceding those controlled by top-down attention (reviewed in Murray, Lewkowicz, Amedi, & Wallace, 2016; De Meo, Murray, Clarke, & Matusz, 2015; van Atteveldt, Murray, Thut, & Schroeder, 2014). This offers one explanation for why multisensory stimuli are attended more strongly than unisensory stimuli, often irrespective of the observers’ intention or demands of the current task, unlike unisensory stimuli (Matusz, Thelen, et al., 2015; Scerif, Longhi, Cole, Karmiloff-Smith, & Cornish, 2012; Matusz & Eimer, 2011a, 2011b; van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011; Santangelo & Spence, 2007). Thus, even such fundamental tenets, as the predominance of top-down over bottom-up mechanisms of attentional control (cf. Folk, Remington, & Johnston, 1992), established by unisensory research, may not hold in naturalistic, multisensory environments. Second, unisensory responses have limited power in predicting multisensory responses due to nonlinear mechanisms involved in integration of multisensory information (reviewed in Murray & Wallace, 2012; Stein, 2012; Ernst & Banks, 2002). That is, if, for example, visual and auditory (brain or behavioral) responses in a given task are summed, the response to the audiovisual stimulus will be reliably larger or smaller than that sum, and currently, we cannot easily model this difference. Third and perhaps most importantly, fundamental cognitive functions

like speech, object recognition, or reading are typically improved in multisensory settings. Yet, in naturalistic, cluttered contexts, these benefits are contingent on top-down attentional control (Alsius & Soto-Faraco, 2011; Froyen, Bonte, van Atteveldt, & Blomert, 2009; Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008).

Thus, in the real world, multisensory processes and top-down attentional control likely interact in modulating object recognition and communication, but the mechanisms governing these interactions are entirely unclear. If attentional templates are a fundamental way through which the brain instantiates top-down attentional control, multisensory object templates might be an important way through which top-down control is instantiated in real-world environments. As unisensorily gauged processes may be limited in predicting those engaged by multisensory stimuli, systematic research is required to better understand how spatial, feature- or object-based top-down control mechanisms interact with multisensory processes. Notably, research on multisensory object templates is the more pertinent, as it might be directly relevant to everyday situations, such as those where a familiar sound needs to be associated with an arbitrary visual shape, as in the case of reading or early number knowledge.

How then are attentional templates represented in the brain? There are several lines of evidence to suggest that attention is indeed controlled by integrated rather than separate representations of target objects. First, it has long been shown that features are preferentially processed when they are part of the same object (e.g., Duncan, 1984). Second, working memory that arguably mediates attentional templates represents objects by integrating their features rather than keeping them separate (Luck & Vogel, 1997). However, these findings pertain to purely unisensory, visual processes. Traditional models of working memory argue for sensory-specific systems for storage and manipulation of visual and auditory information (e.g., Baddeley, 2000). Despite the mounting evidence for multisensory representations throughout the brain, it is plausible that representations of visual and auditory task-relevant features are functionally separate. The presence of visual target-defining features would, in such cases, attract visual attention and modulate activity within visual cortices in a spatially selective manner; the presence of target-defining auditory features would trigger similar modulations within relevant auditory cortices. Consequently, the ability to capture attention of objects matching just the visual feature would be unaffected by the lack of a target-defining auditory feature, as would be their spatially selective processing within visual cortices. However, representing features in a separate manner is inconsistent with the flexible, task-contingent abilities of the frontoparietal network to represent behaviorally relevant information (e.g., Miller & Cohen, 2001). In further support of integrated multi-

sensory representations of task-relevant features, neurons in the pFC have been shown to represent, in a task-dependent fashion, arbitrary but task-relevant conjunctions of color and pitch (Fuster, Bodner, & Kroger, 2000). In addition, posterior parietal cortices, known to contribute to control of attention toward task-relevant spatial locations and task-relevant object features, represent space in a multisensory fashion (Shulman, D'Avossa, Tansy, & Corbetta, 2002; Assad & Maunsell, 1995; reviewed in Stein & Stanford, 2008). Lastly, there is converging evidence to suggest that even neurons at such early stages of cortical processing as "sensory-specific" V1 are involved in multisensory integration (reviewed in Murray, Lewkowicz, et al., 2016), making the multisensory nature of object representations in the higher-level brain areas even more plausible (Matusz, Wallace, & Murray, 2017).

In line with the latter, we have previously demonstrated that the ability of visual objects to capture attention is attenuated during audiovisual search (Matusz & Eimer, 2013). We adapted the Folk et al.'s (1992) paradigm, so that participants searched for targets defined either by visual feature alone or visual-auditory feature conjunctions (e.g., red bars vs. red bars paired with a high pitch tone; "color" vs. "color-tone" task). Search arrays were preceded by a display with a visual cue that always matched the target color. The ability of visual cues to capture attention in the visual audiovisual was attenuated and/or eliminated in the audiovisual task across both brain and behavioral responses. Subsequently, behavioral studies have demonstrated that multisensory object templates generalize across different sensory pairings and nonspatially selective attention tasks (Mast et al., 2015, 2017). To better understand the neural underpinnings of multisensory templates, we have likewise recorded EEG in our study and focused EEG analyses on the traditional ERP marker of attentional selection, that is, the N2pc component, a negative-going voltage deflection over posterior electrodes collateral to the stimulus location ~200 msec poststimulus (Eimer, 1996; Luck & Hillyard, 1994). We found that within the traditional time window (170–270 msec), N2pc was reduced in amplitude in the audiovisual task versus the visual task. These N2pc reductions have been shown also for visual feature conjunction (e.g., Kiss, Grubert, & Eimer, 2013). Regarding the underlying brain mechanism, these N2pc attenuations would be traditionally taken as evidence for top-down object templates controlling spatially selective processing within nominally visual cortices like lateral occipital cortices (Hopf et al., 2000). The typical implication, which to date has not been explicitly tested, is that N2pc modulations arise via "gain control," wherein the amplitude of neural responses, but not the network configuration itself, is modulated by attention-related processes. Thus, these canonical EEG/ERP analyses would suggest that top-down object templates reduces attention-capturing abilities of the visual distractor by

decreasing the activity it elicited within the same cortices. These results, if linked to the cognitive mechanisms under study, would have potentially important mechanistic implications for the nature of object templates. However, ERP amplitude modulations can stem from alternations in both “the gain” (the strength of response of the same network) and the configuration of activated networks (see Murray, Brunet, & Michel, 2008). As such, canonical N2pc analyses are limited in the extent they can provide strong brain-level evidence for the representations orchestrating object templates.

To shed the much needed light onto the brain and so cognitive mechanisms governing top-down multisensory object templates, here we reconsidered some of the data from Matusz and Eimer (2013) within an EN framework. EN can statistically distinguish whether the observed N2pc amplitude modulations arose from strength-based (“gain”) and network-based mechanisms (Michel & Murray, 2012; Murray et al., 2008; Lehmann & Skrandies, 1980). Thus, EN can directly support accounts proposing more integrated versus separate representations of multisensory object templates. Besides these directly neurophysiologically interpretable results, EN addresses two crucial yet always ignored limitations characterizing the canonical analyses of lateralized ERP components like N2pc. For one, the latter measure the difference between two electrode channels (or channel subsets) located in opposite hemisclaps. Whether explicit or implicit, an assumption behind these analyses is that this 2-point difference in lateralized potentials reflects all of the electrical brain activity relevant to, in the case of the N2pc, attentional control. This is a major limitation of canonical lateralized ERP analyses, as they ignore the majority of the recorded brain data (91.5% in the case of ERP analysis of data from two channels from a 23-channel montage; a percentage that would further increase with higher density montages, though the degree of nonindependence across electrodes is another important consideration). As such, they likely risk missing effects occurring elsewhere across the electric field at the scalp. Furthermore, the 2-point subtraction that underlies the mean amplitude N2pc difference might have arisen due to different contexts (see Figure 3 for detailed explanations).

We predicted that top-down control via integrated multisensory object templates should be reflected by differences in the topography (and so in the engaged brain sources) of the lateralized ERP gradient elicited by color cues between the two search tasks in the N2pc time window. When the brain is set to control sensory processing toward stimuli defined by a given color, such as “all objects in red”, responses of all neurons representing red are enhanced (Desimone & Duncan, 1995). When the brain is set toward a multisensory, color-tone defined stimulus (cf. color-pitch selective prefrontal neurons found by Fuster et al., 2000), responses should now be enhanced only in those neurons that can represent both red and the sound feature. Consequently, the same color-

defined visual cue will activate a different configuration of neuronal populations when it fully matches the representation of the target versus when the target representations include another feature. If we observed global field power (GFP; i.e., strength-based) modulations alone in the cue-induced ERPs within the N2pc time window—without concomitant topographic modulations—this would be more consistent with separate representation of visual and auditory task-relevant features. That is, similar topography (and activated neuronal populations) in cue response would suggest top-down control based on the same object representations across the two search tasks. Consequently, the up-/down-modulations of the cue-induced ERPs would arise from nonvisual signals influencing more general, tonic baseline activity changes within color encoding to reflect changes in their relevance to the audiovisual search. Such changes were found previously in studies involving unisensory versus multisensory selective attention tasks (e.g., Mozolic et al., 2008; Laurienti et al., 2002).

METHODS

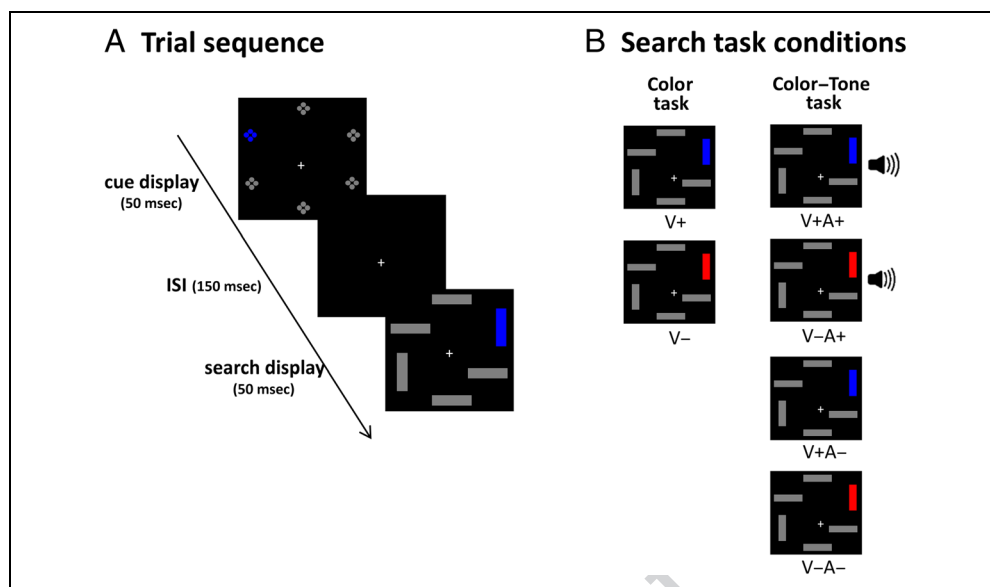
Participants

Participants were 12 right-handed paid volunteers with normal or corrected-to-normal vision (mean age = 25.8 years, age range = 21–40 years, five women). None of the participants had current or prior neurological or psychiatric illnesses. All had normal or corrected-to-normal vision and reported normal hearing. All participants provided informed consent before the start of the experiment. Some of the data were reported as part of a study where top-down control by audiovisual templates was investigated using exclusively traditional behavioral (RT spatial cueing effects) and EEG/ERP markers of attentional selection (N2pc amplitude; Matusz & Eimer, 2013).

Stimuli and Procedure

Behavioral and brain indices of attentional capture of visual cues as a function of top-down templates were assessed across a visual and an audiovisual task. As visible in Figure 1, every trial consisted of a cue display, followed by a blank screen with a fixation cross, in turn followed by a search display. Both the cue and the search array always contained a color singleton (either red or blue), with the cue color always matching the color of the target. What differed between the two tasks was how the targets and the nontarget search arrays were defined. In the color task, participants had to respond to color singleton bars when they appeared in the target color (e.g., blue) and ignore bars defined by nontarget color (e.g., red when searching for blue). Both trial types were presented equiprobably and in a random order. In the color-tone task, participants had to respond to bars of the same color, but only when they were accompanied by a tone (e.g., blue

Figure 1. Schematic presentation of the (A) trial sequence and (B) trial conditions across the color task and the color-tone task. On each trial, a cue display (50 msec duration) was followed after a 150-msec-long ISI by a search array (50-msec duration). The intertrial interval was 1450 msec.



presented together with a tone: V+A+). As in the visual task, targets appeared on half of all trials. The other half consisted of equiprobable trials with one or both features mismatching the target and required no response (e.g., red bar/tone: V-A+; blue bar/no tone: V+A-; red bar/no tone: V-A-). The two search tasks were performed by participants in a counterbalanced order together with a second audiovisual search task, which was not crucial considering the aims of this study. Each participant looked for either a blue or a red target color bar, paired with either low- or high-pitch sound (in the audiovisual search task) and assessed their orientation (vertical vs. horizontal) by pressing one of two vertically aligned keys with two index fingers (e.g., pressing a top key for vertical targets and the bottom one for horizontal). Thus, to reiterate, both target color and target pitch (as well as mapping hand key) were counterbalanced across participants. The between-subject counterbalancing of color and sound pitch target feature pairings served to prevent our results from being influenced by some participants potentially possessing any preexisting associations between colors and sound pitches, where they would regard subjectively a particular pitch level “more congruent” with a given color. The ERP data were collapsed across participants searching for different audiovisual feature combinations. Each search task was performed across four consecutive blocks, each consisting of 96 trials (48 target, 48 nontarget).

Visual stimuli were presented against a black background on a 22-in. LCD monitor (100 Hz refresh rate; 100 cm viewing distance; Samsung wide SyncMaster 2233). In the cue array, each of the six elements was composed of four closely aligned dots ($0.17^\circ \times 0.17^\circ$). The color singleton cue was blue or red (CIE x/y chromaticity coordinates 0.161/0.128 and 0.621/0.128, respectively), and the remaining items were gray (0.308/0.345). This

color singleton was presented equiprobably and randomly at one of the lateral locations, rendering it uninformative with respect to the spatial location of the upcoming target. This enabled us to measure control of visual cue-induced attentional capture as a function of top-down object templates. The lateral localization of the cues enabled us to record an N2pc in response to each cue (e.g., Hickey et al., 2006). Search arrays contained six horizontal or vertical bars ($1.1^\circ \times 0.3^\circ$) at the same positions as the preceding cue elements, with bar orientation chosen randomly for each position. Colored bars appeared with equal probability at one of the four lateral locations. All gray, blue, and red stimuli in the cue and search displays were equiluminant ($\sim 11 \text{ cd/m}^2$). In the color-tone task, the sound was a pure sine wave tone (50-msec duration; 65 dB SPL) of high or low pitch (2000 and 300 Hz, respectively, for participants searching) that was presented concurrently with search array onset from a loudspeaker located centrally behind the monitor.

EEG Acquisition and Preprocessing

EEG was DC-recorded with a BrainAmps DC Amplifier from 23 Ag-AgCl scalp electrodes in an elastic cap, positioned according to the international 10–20 system. Two additional electrodes were also located at the level of the outer canthi of the eyes. Signals from the left and the right earlobe were also recorded, and during the recording, all channels were referenced to the left earlobe. During the recording, EEG was sampled at 500 Hz, and impedances were kept below 5 k Ω . Cartool (available at www.fbmlab.com/cartool-software/; Brunet, Murray, & Michel, 2011) was used for data preprocessing and the statistical analyses. Next and before averaging, the EEG was filtered offline with a second-order Butterworth filter (-12 dB/octave roll-off, 0.1 Hz high pass, 40 Hz low

pass). The filters were computed linearly in both forward and backward directions to eliminate phase shifts. Then, the continuous EEG was segmented into peristimulus epochs, relative to the color cue onset, spanning from 100 msec prestimulus to 500 msec poststimulus onset. Subsequently, data quality was controlled with a semi-automated artifact rejection criterion of $\pm 80 \mu\text{V}$ at each channel as well as visual inspection to exclude any remaining transient noise, eye movements, and muscle artifacts.

To obtain lateralized ERPs, for each participant, the artifact-free single-trial epochs were averaged and prestimulus baseline-corrected (using the -100 to 0 msec time interval), separately for trials with color singleton cues presented in the left and right hemifield, for each of the two search tasks, resulting in four average ERPs. Then, the two weighted ERP averages (weighted according to number of accepted epochs) from conditions with cues presented in the left were relabeled, so that electrodes over the left hemiscalp now represented brain activity over the right hemiscalp, and vice versa. Following this step, the “mirror cue-on-the-right” ERP average and the veridical “cue-on-the-right” ERP average condition were collapsed, creating a single lateralized ERP file. As this was done separately for each of the two search tasks, this resulted finally in two lateralized cue-elicited ERP

averages: one for the tone task and one for the color–tone task. From this step onward, ERP data were always coded in terms of their contralaterality (contralateral vs. ipsilateral to the cue side), and we refer exclusively to contralateral and ipsilateral scalp sites with respect to the cue presentation side.

ERP Analyses

The preprocessing of the ERPs triggered by the visual cues across the color task and the color–tone task created ERP averages in which the contralateral versus ipsilateral ERP voltage gradients across the whole scalp are preserved. As the cues were physically identical in both tasks, we were able to directly contrast the insights offered by traditional N2pc analyses and the EN framework regarding the effects of top–down object templates on visual object attentional selection. An overview of our multistep analysis is detailed in Figure 2.

Step 1. Canonical N2pc Analyses

We first aimed to bridge the present EN analyses with the previous canonical N2pc analyses. Specifically, in this step, we extracted from the lateralized ERPs mean amplitude measures across the 170–270 msec postcue onset

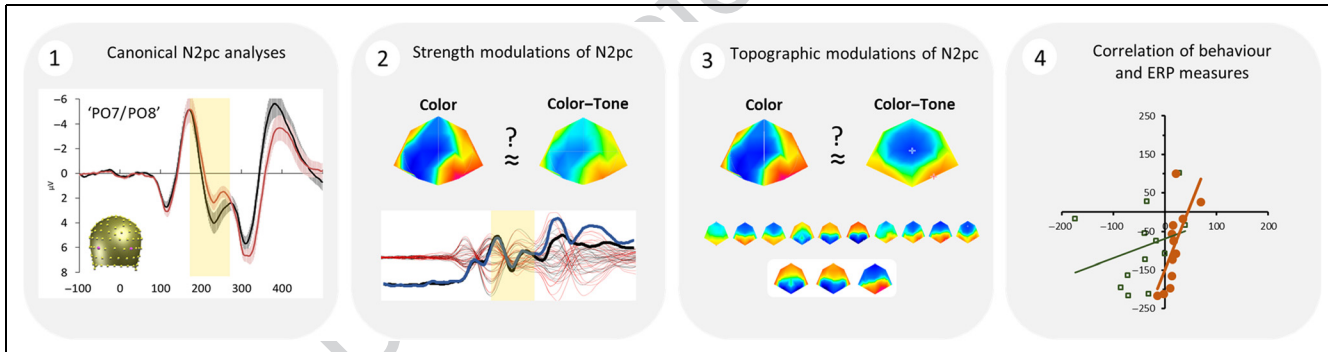


Figure 2. Flowchart of analytical steps performed in this study to identify the brain mechanisms governing the top–down control of visual attentional selection during multisensory search. In Step 1 and to bridge the present EN analyses with the canonical N2pc analyses, we compared average referenced ERPs elicited by the color cues between the color task versus the color–tone task using the traditional approach. That is, we extracted and compared between the two tasks mean amplitudes in the lateralized cue-elicited ERPs measured in the traditional N2pc (170–270 msec poststimulus) time window and over the traditional PO7/PO8 electrodes. When ERPs in the two tasks were compared, N2pc attenuations were found, which would be typically interpreted in terms of a “gain” control mechanism. However, ERP amplitude modulations can stem from alternations in both “the gain” (the strength of response of the same network) and in the configuration of activated networks. In Step 2, we directly tested whether the observed mean N2pc amplitude differences were driven by search task modulating the strength (i.e., “gain”) of cue-elicited lateralized ERP responses within statistically indistinguishable brain networks. For this purpose, we compared between the two search tasks the mean GFP values over the same 170–270 msec postcue onset time window. In Step 3, we tested whether the mean N2pc amplitude differences were driven by alternations in ERP topography and so in the configurations of brain sources that the color cues activated between the two search tasks. To this aim, we performed hierarchical clustering on the group-averaged cue-elicited ERP data across both tasks over the whole 500-msec postcue time period. The clustering revealed 13 different “template maps,” that is, periods of stable ERP activity, characterizing cue-elicited lateralized brain responses within that time window. Subsequently, the three template maps identified within the group-averaged ERPs as differentially characterizing the two search tasks over the 142–260 msec time period were submitted to a fitting procedure. During fitting, each time point of the single-subject ERP in the 142–260 msec postcue time period was “fitted” to each of the three template maps derived from group-averaged hierarchical clustering. The relative duration of each of the three template maps in the lateralized ERPs across the two search tasks was then statically assessed. Finally, in Step 4, we investigated the behavioral relevance of the canonical and EN measures of the N2pc. That is, we assessed whether the changes induced as a function of search task in the behavioral measures of attentional capture abilities of visual cues were associated with changes in the canonical and/or the EN measures of the N2pc activity. To this aim, we calculated nonparametric brain–behavioral measure correlations using the Spearman’s rho and combined it with the jackknife data resampling procedure.

time window from electrodes PO7 and PO8 and submitted these to a 2×2 within-subject repeated-measures ANOVA with factors task (color task vs. color-tone task) and contralaterality (contralateral vs. ipsilateral). As described in the section above, we used average-referenced, rather than linked earlobe referenced, as in our original study (Matusz & Eimer, 2013), ERP data. With regard to the N2pc specifically, the choice of the reference is moot, as the contralateral versus ipsilateral difference will always be the same regardless of the reference (i.e., a spatial gradient is being calculated). Specifically, the paired lateral electrode measurement captures a portion of the topography of the electric field across the scalp, and biophysical laws dictate that scalp topography is reference independent (Michel & Murray, 2012; Murray et al., 2008). To use an analogy, the altitude difference between two mountain peaks on opposite sides of a valley is independent of where the sea level is measured. Thus, for a lateralized component, such as the N2pc, we did not expect any differences between the original and the current N2pc results. This is because, to follow the analogy, the differences in altitude between two mountain peaks is equivalent independent of whether the altitude of these peaks is measured versus sea level or versus Mount Everest. In contrast, the shape of the ERP waveform recorded at one electrode (and not for a ERP waveform of a difference between two electrodes, like in the case of N2pc) within one hemiscalp will change depending on the reference electrode(s) chosen (see, e.g., Lehmann, Ozaki, & Pal, 1987; see also Figure 2 in Murray et al., 2008).

The reference independence of the canonical N2pc analyses, however, does not resolve their highly limited neurophysiologic interpretability. For one, as only a portion of the topography is considered, there is a reasonable likelihood of missing ERP effects (topographic or strength-based ERP modulations) occurring during the N2pc time period outside the two “mountain peak” points, for example, within the “mountain valleys.” Second and most importantly, as a mere subtraction of values between two opposite hemiscalp electrodes, the canonical N2pc analyses would indicate that attentional selection across two conditions is comparable in magnitude even if two very different neurophysiological situations gave rise to it. This point is well illustrated in Figure 3, which displays three hypothetical lateralized data matrices (i.e., the potential values recorded from 16 electrodes at a N2pc-like latency). Condition 2 is precisely twice that of Condition 1 at each electrode, resulting in an identical spatial distribution of values that are simply stronger in Condition 2. The values of Condition 3 are identical to those of Condition 2, though partially shuffled in their locations. The canonically measured N2pc (Figure 3A) is the difference between a contralateral electrode (black circle) and a respective ipsilateral electrode (gray circle). For Conditions 1–3, the N2pc value would be measured as -4 , -16 , and -4 μV ,

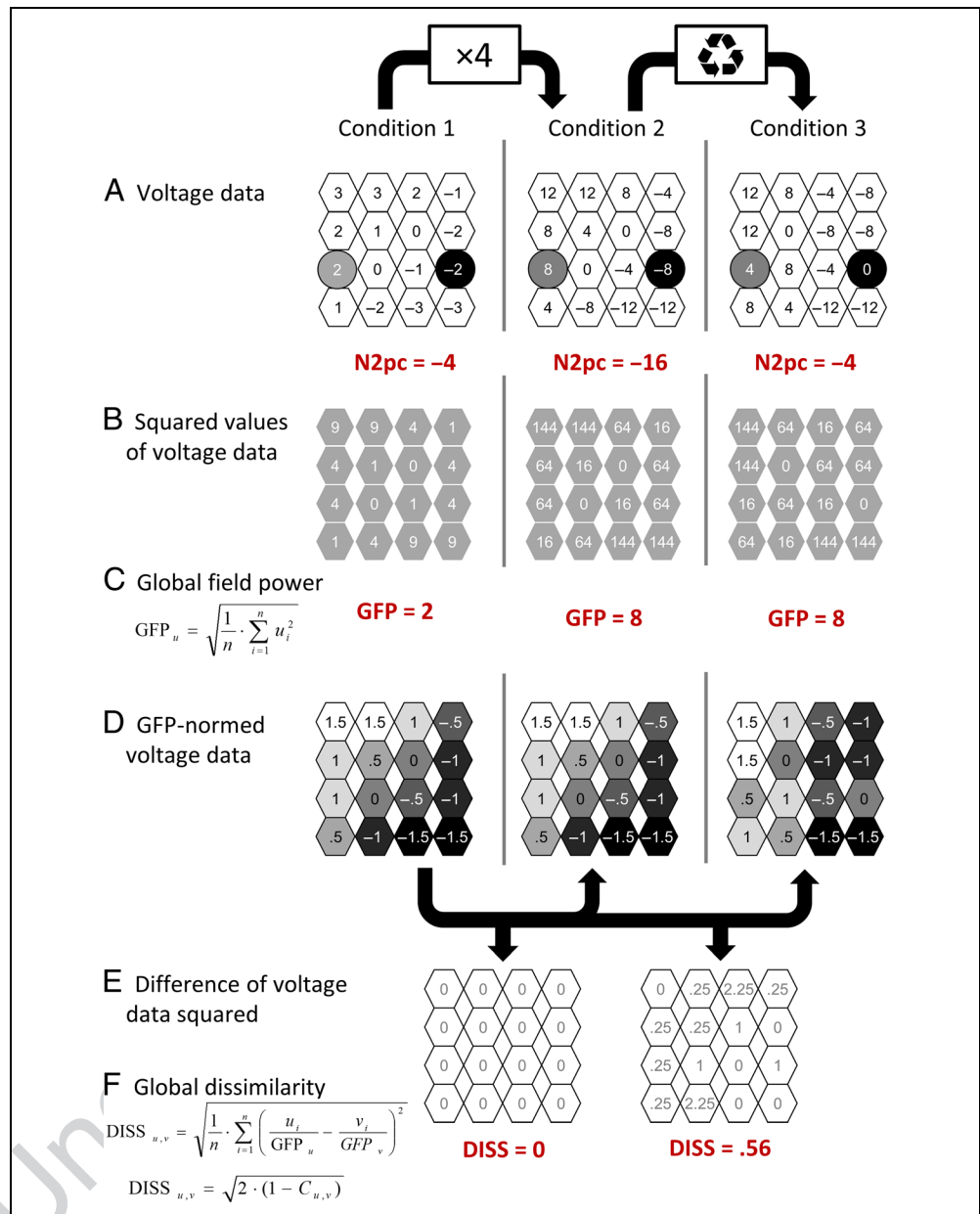
respectively. That is, canonical N2pc analyses would report no difference between Conditions 1 and 3, despite the clear, abovementioned differences in how the data were generated. In contrast, the strength difference between Conditions 1–2 and the topography difference between Conditions 2–3 are both readily captured by GFP (Figure 3B, C) and global dissimilarity (Figure 3D–F), respectively. As such, the remainder of the analyses focused on how these measures of the global attributes of the electrical field at the scalp (i.e., how the voltages behave across the whole scalp) can inform our understanding of top-down attentional control via object templates.

Step 2. Strength-based Modulations of the N2pc Component

As part of EN analyses, we first assessed whether the observed mean N2pc amplitude differences were driven by search task modulating the strength of responses within statistically indistinguishable brain networks. For this purpose, we used GFP, which equals the root mean square, or standard deviation, across the average-referenced electrode values at a given moment (as described in Lehmann & Skrandies, 1980). The GFP waveform is a moment-to-moment measure of standard deviation of potential (μV) across the whole montage. The GFP differences in Figure 3B–C directly reflect the fourfold increase in the “global” response strength between Conditions 1 and 2. What GFP does not provide insight into is how the potentials are distributed across the scalp. The most parsimonious explanation of differences in GFP between two conditions “without” concomitant statistically reliable differences in the scalp topography (as measured with global dissimilarity) is a change in the gain within statistically indistinguishable generators between two stimulus conditions. We remind the reader that GFP and DISS are reference independent. Average reference is used nevertheless in EN analyses because source estimations are typically part of the analysis pipeline (discussed in Michel & Murray, 2012). However, all source estimation methods apply a common average reference to the data as part of the biophysical principle of quasi-stationarity (i.e., that the sum of all currents at a given moment in time is zero).

The GFP waveform can be assessed statistically just like any other ERP waveform. To maintain consistency with the canonical N2pc analyses, we extracted the mean voltages of the GFP waveform over the same as before, 170–270 msec postcue time window and then compared them directly between the color task and the color-tone task using a paired t test. Global characteristics of the electric scalp field gradients, compared with the measures of local field potentials as represented, for example, by N2pc, should provide a more complete answer as to whether top-down attentional control via object

Figure 3. Illustration of the added value of an EN analysis framework. (A) Condition 1 displays exemplar voltage data across a hypothetical 16-electrode montage. The values of Condition 2 are increased precisely fourfold over those of Condition 1. The values of Condition 3 are identical to those of Condition 2, though partially shuffled in their locations. The N2pc is measured canonically as the difference between a contralateral electrode (black circle) and a respective ipsilateral electrode (gray circle). For Conditions 1–3, the N2pc value would be measured as -4 , -16 , and -4 μV , respectively. That is, there would be no N2pc difference between Conditions 1 and 3, despite the clear, abovementioned differences in how the data were generated. (B) The squared values of the voltage data in A are displayed for each of the three hypothetical conditions. (C) The GFP equals the root mean square of the voltage values across the whole scalp at a given time point (formula indicated). In the case of these hypothetical conditions, the GFP would equal 2, 8, and 8 μV , respectively. (D) The GFP-normalized voltage data are displayed and are calculated by dividing the values at each electrode in A by the GFP calculated in C. The shading illustrates the topographic distribution of these values across the electrode montage. (E) The difference of voltage data squared is displayed. Such differences allow for the quantitative measurement of global dissimilarity (DISS; equation in F) between conditions (as well as time points or groups). The DISS between Conditions 1 and 2 equals 0, meaning that the topographic distribution of the values across the hypothetical 16-electrode montage is identical. In contrast, the DISS between Conditions 1 and 3 equals 0.56; in extremum, DISS equals 2, which means that topographic distributions are perfectly inverted. Crucially for the aims of this study, we note that, in the example here, the topographic differences between the two last conditions would be completely overlooked if only traditional N2pc measures were considered.



templates can operate by altering the overall strength of the lateralized voltage potentials.

Step 3. Topographic Modulations of the N2pc Component

Next, we tested whether the mean N2pc amplitude differences were driven by alternations in ERP topography and so in the configurations of brain sources that the color cues activated between the two search tasks.

Differences between two electric fields (independent of their strength) are indexed by global dissimilarity (DISS). DISS equals the root mean square of the squared differences between the potentials measured at each electrode (vs. the average reference), each of which is first scaled to unitary strength by dividing it by the instantaneous GFP (Lehmann & Skrandies, 1980). The calculation of DISS becomes easier to understand if one considers again the data in Figure 3. As already mentioned above, Conditions 1 and 2 have the same topography but

different strengths, whereas Conditions 2 and 3 have the same strength but different topographies. Figure 3A depicts the original data from the three conditions, whereas Figure 3D shows the same data that have been GFP-normalized. Thus, after rescaling all three conditions to have the same GFP, the topographic similarities and differences between conditions become readily apparent. As visible in Figure 3E, the topographic distribution of the values across the hypothetical 16-electrode montage is identical between Conditions 1 and 2, and this is reflected by DISS equaling 0. In contrast, the DISS between Conditions 1 and 3 equals 0.56, and this reflects the relatively weak reshuffling of the values carried out between the two matrices; in extremum, DISS equals 2, which means that topographic distributions of the values (voltages) across the whole scalp are perfectly inverted at a given moment. Crucially for the aims of this study, we note that, in the example here, the topographic differences between the two last conditions would be completely overlooked if only traditional N2pc measures were considered.

DISS is directly related to the spatial Pearson's product-moment correlation coefficient between the potentials of the two compared voltage scalp maps. That is, a spatial correlation coefficient value of -1 at a given moment would indicate that two ERP topographies are perfectly inverted (i.e., DISS value of 2), and this relationship is expressed by spatial correlation being equal to $((1 - \text{DISS}^2) / 2)$. If two ERPs differ in topography independent of their strength, it directly indicates that the two maps were generated by a different configuration of sources in the brain. Display and comparison of DISS across time allow defining periods of stable patterns of ERP activity and changes therein. In fact, GFP and DISS are inversely related, that is, when GFP is high, ERP topographic activity tends to remain stable (i.e., DISS is low; see above), whereas it changes when GFP is low. Displaying DISS across time shows a highly characteristic behavior, where topographic activity remains stable for tens to hundreds of milliseconds and then changes suddenly to a new configuration, lasting again tens to hundreds of milliseconds. These highly reproducible and sequentially organized configurations have been shown to represent successive steps along the information processing pathway from perception to action (also known as "functional microstates"; Michel & Koenig, 2018; Brandeis, Lehmann, Michel, & Mingrone, 1995; Lehmann et al., 1987; Lehmann & Skrandies, 1980).

Following the above-described ideas, we focused analyses of topographic differences on hierarchical clustering (specifically, we used the modified agglomerative hierarchical clustering algorithm called Topographic Atomize and Agglomerate Hierarchical Clustering) to identify stable electric field topographies (henceforth "template maps") present in the group-averaged cue-elicited ERPs between the two tasks within the whole 500-msec post-cue time period. The aim of this step is to obtain the

minimal number of template maps that accounts for the greatest variance of the whole group-averaged data set. Within concatenated group-averaged data across all (here, two) conditions, each data point (here, map) first is defined as a single cluster. Following iterations, clusters start defining groups of data points (maps), whose mathematical mean (i.e., centroid) represents the template map for that cluster. Subsequently, the "worst" cluster is identified, that is, that contributing the least to the quality of the clustering, as indexed by lowest GEV cluster. The maps contributing to that former cluster are then "freed," that is, they cease to belong to any cluster. In iterative procedure, one map at a time is separately re-assigned to one of the remaining clusters, based on the highest spatial correlation (derived from DISS) between each "freed" map and the centroid of each remaining cluster. The clustering makes no assumption regarding the orthogonality of the derived template maps (Michel & Koenig, 2018). The end product here is effectively one final cluster (which is not informative), and so it is important in the Topographic Atomize and Agglomerate Hierarchical Clustering to be able to determine the optimal number of clusters. We achieve this by the application of a modified Krzanowski-Lai criterion (Murray et al., 2008), which identifies the optimal number of temporally stable ERP clusters, that is, the minimal number of stable maps accounting for the most amount of variance in the concatenated group-averaged data between conditions.

Subsequently, we submitted the template maps identified within the group-averaged ERPs to a fitting procedure. During this analytical step, each time point of the single-subject ERP (here measured in the 142- to 260-msec postcue time period) was "fitted" to the template maps that in the group-averaged hierarchical clustering were differentially characterizing the two search tasks over the 142- to 260-msec time period. Thus, topography at each time point in the cue-induced ERPs across the two tasks was assigned as representing that group-averaged template map, which it best correlates with spatially, in a "winner takes all" fashion (Murray et al., 2008). As an output, for each participant and each of the two conditions, we obtained the number of time samples in which single-subject data best correlated spatially with a given group-averaged ERP template map. The relative presence of each template map in the scalp topography of the two lateralized ERPs was then submitted to repeated-measures 2×3 ANOVA with within-subject factors of task (2) and map (3).

Step 4. Associations between Behavioral and ERP Measures of Attentional Selection

Next, we investigated whether the changes induced as a function of search task in the behavioral measures of attentional capture abilities of visual cues were associated with changes in both the canonical versus the EN

measures of the N2pc activity. To this aim, we calculated nonparametric brain-behavioral measure correlations using the Spearman's rho and combined it with the jackknife data resampling procedure. The jackknife method, like bootstrapping, resamples the data to estimate how big the bias is in a given statistic. The difference between the two methods lies in that jackknife resamples systematically, not at random, by computing sample statistics on n separate samples of size $n - 1$. We first computed the sample correlation of our data and then the correlations for each of the jackknife samples as to calculate their overall mean. We then used this mean to compute the estimate of the bias existing in our two rho estimates, so that our estimates could be corrected. The jackknife procedure was implemented by the Statistics and Machine Learning Toolbox "jackknife" function, in MATLAB (Version R2008a, MathWorks).

That is, mean N2pc percent amplitude changes and mean GFP percent changes in the color-tone task relative to the color task were correlated separately with the percent changes in the behavioral capture effects between the two tasks (i.e., the difference in capture effects between the color-tone and color task divided by the capture effect on the color task, which can therefore result in percentages in excess of $\pm 100\%$). Similar analyses were conducted for the percent change in the mean duration of each of the three template ERP maps measured within each participant separately during the fitting procedure, as compared in the color-tone task relative to the color task.

RESULTS

Behavioral Results

Only RTs from correct trials within 200–1000 msec and within ± 3 SDs from the mean were analyzed (leading to a loss of $< 1\%$ data). A 2×2 within-subject design was used, that is, Search Task (color vs. color-tone task) \times Cue-Target Location (same vs. different). Performance was quantified using RT* (inverse efficiency; Townsend & Ashby, 1978), which is an aggregate measure that takes into account both reaction speed and accuracy (for RT and accuracy rate analyses, see Matusz & Eimer, 2013, Experiment 1), and has been successfully used in such areas of multisensory research, as multisensory correspondence and congruence, brain plasticity following sensory deprivation as well as selective attention (Ludwig, Adachi, & Matsuzawa, 2011; Ngo & Spence, 2010; Putzar, Goerendt, Lange, Rösler, & Röder, 2007; Kitagawa, Zampini, & Spence, 2005). Overall, performance was similar across the two tasks, $F(1, 11) = 1.02, p = .34$. RT* showed that the cues captured attention behaviorally, but this ability differed between the two tasks, with the main effect of cue-target location, $F(1, 11) = 10.5, p = .008, \eta_p^2 = .49$, modulated further by search task, $F(1, 11) = 13.21, p = .004, \eta_p^2 = .55$. Follow-up planned

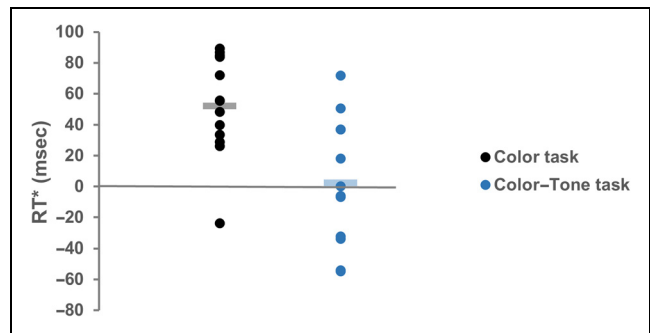


Figure 4. Behavioral spatial cueing effects, as measured by RT*, presented separately for the two tasks. The vertical scatterplots present single-subject (full circles) behavioral capture effects, together with group averages (horizontal lines), elicited by the visual cues in the color task and in the color-tone task (in black and in blue, respectively).

comparisons revealed that this interaction was driven by the fact that visual cues triggered reliable RT* attention capture effects in the color task, $t(11) = 5.4, p = .001$, but this capture was so attenuated in the color-tone task that it was no longer reliably present, $t < 1$ (Figure 4).

ERP Results

Step 1. Canonical N2pc Analyses

The 2×2 repeated-measures ANOVA on the mean N2pc amplitudes over 170–270 msec postcue recorded over PO7/PO8 electrodes replicated previous findings based on a linked ears reference (Matusz & Eimer, 2013). A main effect of contralaterality, $F(1, 11) = 20.27, p = .001, \eta_p^2 = .65$, which was modulated by task, $F(1, 11) = 5.75, p = .034, \eta_p^2 = .35$, demonstrated that the cue-induced N2pc amplitude differed between the two tasks. As before, the N2pc was attenuated yet reliably present in both the color task, $t(11) = 4.87, p = .001$, and in the color-tone task, $t(11) = 3.73, p = .003$ (see Figure 5A–C). There was no main effect of task, $F < 1$. As explained in the Methods section, this replication is unsurprising based on simple biophysical laws. Such notwithstanding, as canonical analyses cannot provide insights into the neurophysiologic mechanisms governing the ERP modulations, we focused the remainder of our analyses on ERP analyses within an EN framework.

Step 2. Strength-based Modulations of the N2pc Component

The pairwise t test carried out on the mean amplitude of the GFP waveform over the same, 170–270 msec postcue time window revealed a pattern contrasting with that found for the mean N2pc amplitudes. The mean GFP was enhanced in the color-tone task, when compared with the color task, $t(11) = 2.62, p = .012$ (paired t test; Figure 5D). That is, the gradients within the scalp-wide distribution of voltages triggered in response to the same

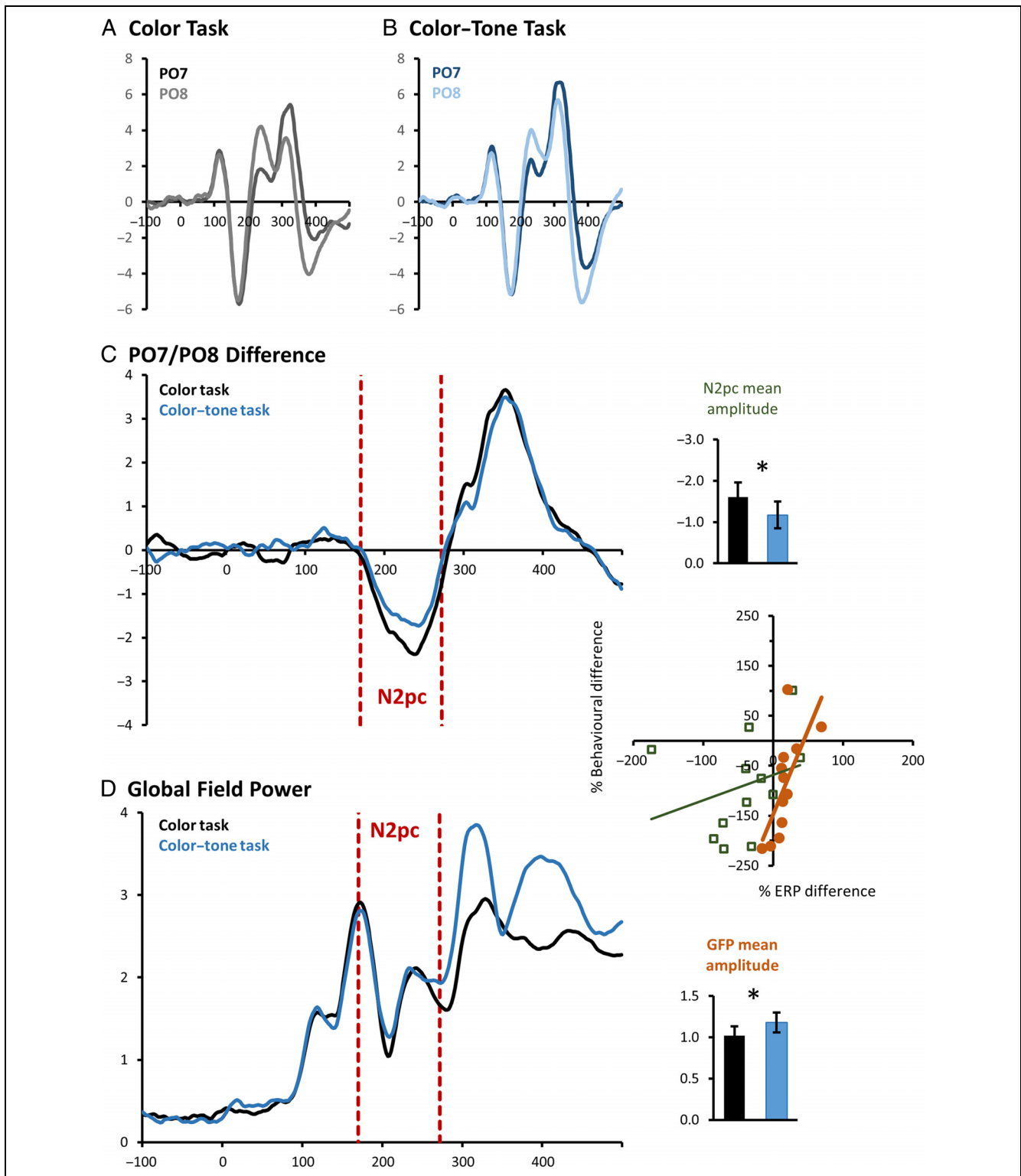
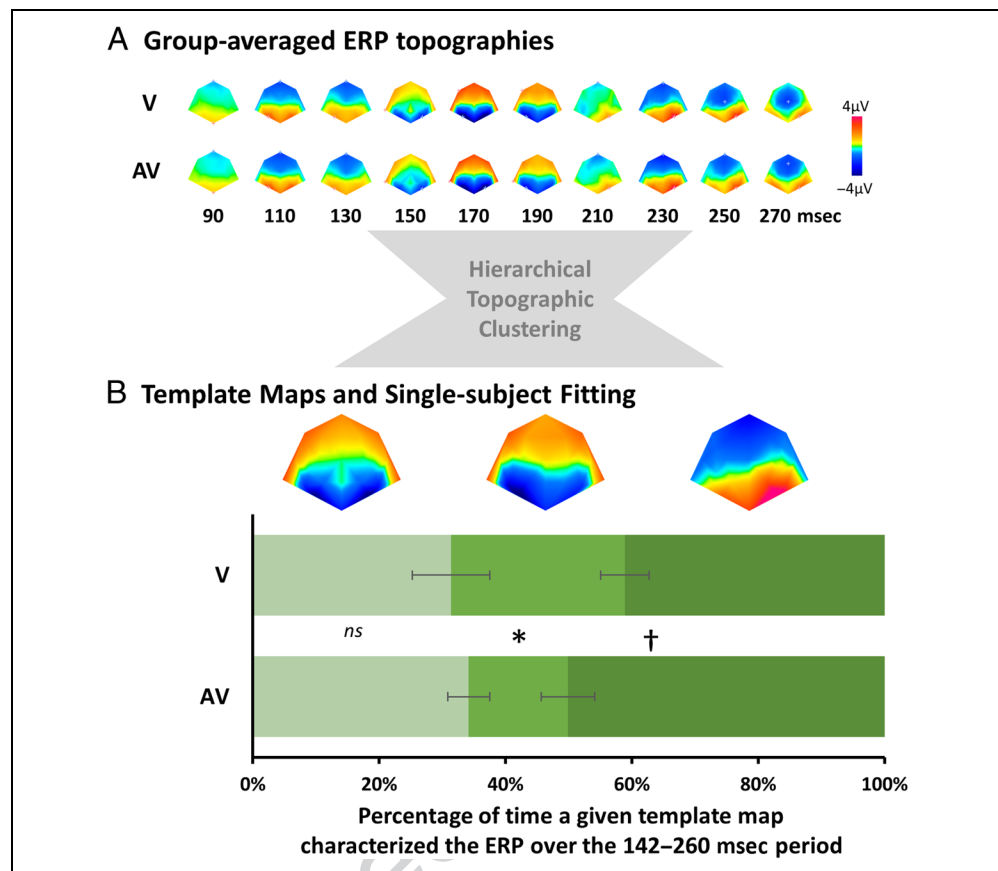


Figure 5. ERP and GFP quantification of the N2pc. (A) Group-averaged voltage waveforms are displayed from the visual task at PO7 and PO8 (dark and light gray traces, respectively). (B) Group-averaged voltage waveforms are displayed from the auditory-visual task at PO7 and PO8 (dark and light blue traces, respectively). (C) Group-averaged difference waveforms (PO7 minus PO8) are displayed for the color task and color-tone task (black and blue traces, respectively). A larger N2pc was measured over the 170–270 msec postcue interval for the color task versus color-tone task. The inset bar graph displays the N2pc mean amplitude (note y-axis is inverted for clarity; asterisk indicates $p < .05$). (D) Group-averaged GFP waveforms are displayed for the color task and color-tone task (black and blue traces, respectively). A larger GFP was measured over the N2pc, 170–270 msec postcue interval for the color task versus color-tone task. The inset bar graph displays the mean GFP amplitude in the N2pc time window (asterisk indicates $p < .05$). The scatterplot displays the relationship between differences in the two tasks in terms of behavioral measures of attentional capture (y-axis) and ERP measures (x-axis). The green points are based on the voltage waveform data (Spearman's $\rho = 0.25$), whereas the orange points are based on the GFP data (Spearman's $\rho = 0.88$).

Figure 6. Topographic analyses. (A) Topographic distributions of the ERP over the 90–270 msec postcue interval are displayed on a flattened montage on the left hemiscalp on the left; scale indicated). (B) Topographic cluster analysis of the group-averaged data identified three template maps over the 142–260 msec postcue interval that were then fitted to single-subject data. The bar graph displays the average percentage of time (SEM indicated) each of these three template maps best correlated spatially with single-subject data. The significant Map \times Task interaction indicates that different patterns of template maps best characterized responses from each task.



visual cue were, in fact, overall stronger in the color–tone task as compared with the color task.

Step 3. Topographic Modulations of the N2pc Component

Thirteen maps across 18 clusters were identified in the group-averaged ERPs, which accounted for 91.4% of the global explained variance. Until 140 msec postcue, the same template maps characterized both tasks (Figure 6A). The fitting procedure, utilizing spatial correlation between template maps identified in the group-averaged data and single-subject data from each of the two conditions, showed that over the 142–260 msec period, three different template maps appeared to differentially characterize each task (Figure 6B, top). A subsequent 2×3 repeated-measures Task \times Map ANOVA on the percentage of time each template map best correlated spatially with single-subject data over the 142–260 msec time period revealed a significant main effect of map, $F(2, 22) = 7.61, p = .032, \eta_p^2 = .27$, as well as a significant interaction, $F(2, 22) = 4.06, p = .032, \eta_p^2 = .27$ (Figure 6B, bottom). Post hoc nonparametric tests (Wilcoxon signed rank tests) were conducted for each map, comparing their relative characterization of responses to each task. The light green template map characterized responses to both tasks equally well ($p =$

.332). The middle green template map characterized responses to the color task more than the color–tone task ($p = .040$). The dark green template map characterized responses to the color–tone task more than the color task, though this exhibited a nonsignificant trend ($p = .091$).

Step 4. Brain–Behavioral Response Correlations

There was a strong correlation between the relative changes between the color task and the color–tone task in behavioral capture effects and mean GFP, $r_{ho_c}(10) = .88, p < .001$. In contrast, mean N2pc amplitude measured from the PO7/PO8 electrode pair did not show a similar correlation with behavior, $r_{ho_c}(10) = .246, p = .43$ (Figure 5C, D, right). These correlation coefficients significantly differed ($z = 2.39; p = .017$, two-tailed). There were no reliable correlations with similar measures from the fitting above.

DISCUSSION

We investigated whether reconsideration of lateralized ERPs within an EN analytical framework could provide direct and novel evidence into the brain and cognitive mechanisms governing top–down control by multisensory object templates, as compared with the canonical N2pc

analyses. We first interpret our results in the context of the existing knowledge on the brain mechanisms governing top-down attentional control and object templates. Second, we discuss the added benefits that an EN framework offers in research on the mechanisms governing attentional control and other neurocognitive functions as they occur in naturalistic and real-world environments.

The mean N2pc measured across PO7/PO8 in the typical 170–270 msec time window was reduced in the color-tone task, as compared with the color task. When we assessed the mean cue-induced GFP in the same time window, it was in fact enhanced in the color-tone task compared with the color task. Furthermore and most interestingly, there were topographic differences between the lateralized cue-induced ERPs, that is, visual cues activated distinct brain source configurations in task contexts where they matched the target template fully (color task) versus only partly (color-tone task). What do these findings suggest for top-down control by multisensory object templates? First, in light of the behavioral results, the N2pc attenuations between the two tasks would be traditionally interpreted in terms of a “gain” mechanism. Namely, these results would be interpreted as spatially selective processing of the color cues elicited within the same brain network being suppressed when these cues matched the target template only partly. However, canonical N2pc amplitudes analyses cannot provide such evidence, as ERP amplitude changes can stem from both strength- and/or network-based mechanisms (e.g., Murray et al., 2008). In contrast, the EN analytical framework readily distinguishes between these brain mechanisms. Topographic differences between the lateralized ERPs are directly interpretable in terms of top-down control via integrated multisensory object templates. In the color search task, the spatially selective brain responses to visual cues would be driven by populations of neurons that represent color (e.g., “red”). In contrast, in the color-tone task, the responses to the same visual cues would now involve neuronal populations that encode both color and tone. One could imagine this could involve a subset of the same color-coding neuronal population active in the color task, but it is more likely that these are neighboring yet distinct populations coding unisensory and multisensory stimulus features (see, e.g., Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004, for evidence for such organization in the STS). Posterior parietal cortex and lateral occipital cortex, two areas shown to give rise to the N2pc in the only existing source localization study (using magnetoencephalography [MEG]; Hopf et al., 2000), are both known multisensory hubs (e.g., Rohe & Noppeney, 2018; Reich, Maidenbaum, & Amedi, 2012). The specific sources, naturally, could change; the effects of stimulus and task contingencies on N2pc sources are yet to be systematically studied even in the visual domain. Notwithstanding, differences in the engaged neuronal populations (e.g., in their relative duration) would be readily detected

and statistically assessed with our topographic analyses. The idea of network-based mechanisms involved in top-down multisensory object template control is indirectly supported also by one of the few studies where N2pc was recorded in a multisensory search task (van der Burg et al., 2011). Therein, N2pc was sensitive to both the multisensoriness as well as to task relevance of the stimuli, with visual inspection suggesting scalp topography differences between the ERPs to audiovisual targets and distractors (albeit these were not statistically assessed; van der Burg et al., 2011, Figures 8–9).

In contrast, the engagement of a gain control mechanism purported in this study by the canonical N2pc analyses are de facto more consistent with top-down control by separate sensory-specific template mechanisms. The presence of GFP modulations that we have observed here—concomitant to topographic modulations—would suggest that indeed gain control processes within the same neuronal population also contributed to the observed ERP differences. This effect could be potentially driven by relative rigidity of (some of) the networks giving rise to the spatially selective responses captured by N2pc, where one specific network generates responses to color-defined stimuli (or even, specific colors), another for shape stimuli, and so on for other visual dimensions. In this case, nonvisual top-down signals impact merely the overall levels of this brain activity. However, the cue-induced GFP was stronger (rather than weaker, as in traditional N2pc analyses) in the audiovisual task compared with the visual task. The opposite sign of this effect may reflect top-down inhibitory brain processes, activated in the task context involving partial cue-target match, and the reliable GFP-behavior correlation (not found for N2pc) supports this possible explanation. Although this effect would first needs to be replicated, it underlines the potential utility of GFP as a direct measure of gain control mechanisms during spatially selective brain processing, which are not readily captured by traditional measures of the N2pc.

Our study has equally provided novel insights into the brain underpinnings of the N2pc. First, we reveal here that N2pc amplitude differences between different experimental conditions can also be driven by topographic modulations reflective of network-based mechanisms. This is important inasmuch as there is an assumption, albeit perhaps implicit, that N2pc amplitude differences reflect the modulation in the strength of activity of brain circuits involved in attentional selection, that is, a “gain” mechanism. Our ERPs were indeed modulated in their strength between the two tasks, but we likewise provided direct statistical evidence that N2pc differences arose from the cues engaging different brain networks viz. topographic differences in the ERPs. Second, our scalp-level analyses showed that the N2pc, even in the visual color task, is itself composed of multiple, rather than one single, stable pattern of ERP topographic activity and so configurations of brain sources. Implications of

this, although typically ignored, are that one needs to first statistically establish when a given network configuration started and stopped its activity and limit mean amplitude analyses to that time window, if these analyses are meant to be a valid measure of “gain” control within a given configuration of brain sources. Again, these important measures of ERP strength, topography, as well as sequences of topographic stability are readily offered by EN. Although we focused here predominantly on the cognitive implications of our work, we do note that, to our knowledge, we offer the first direct evidence for a “gain control” mechanism orchestrating the N2pc. Our cue-induced N2pc findings should certainly be replicated with the N2pc elicited by targets in search tasks and other indices of top-down attentional control. Yet, we think it crucial to underscore that most prior work was ill-posed to address the question of gain control or any other mechanism because of the nature of the analyses—and not the nature of search-induced versus cue-induced N2pc. Gain control has been traditionally invoked as a putative general mechanism governing top-down attentional control. Although this has been done predominantly in the context of early, stimulus-elicited ERPs (Handy & Khoe, 2005; Mangun, 1995; but see Couperus & Quirk, 2015), we expected that gain-based mechanisms equally impact N2pc. Authors of some of the seminal works on N2pc write of the component being “attenuated” (e.g., Eimer et al., 2009), which implies reduced activity of a common brain network. We are, however, the first to directly demonstrate it. More generally, as our paradigm separates cognitive and response-related processes, unlike more traditional paradigms, the more rigorous our evidence is for gain-based control of N2pc.

To summarize, we have provided direct evidence for top-down control by multisensory object templates: The ability of a visual object to capture attention is reduced when its features match only partly those defining the multisensory audiovisual target because the top-down object template will alter the configuration of brain sources activated by the visual object (compared with the full-match task context). This account is consistent with the mounting, yet typically unisensory, evidence that top-down control brain systems represent task-relevant information flexibly (Duncan, 2010; Miller & Cohen, 2001). Indeed, the spatially selective brain responses that are captured by the N2pc and believed to reflect attentional selection seem to integrate the bottom-up and top-down inputs also in multisensory contexts (van der Burg et al., 2011; see also Sarmiento, Matusz, Sanabria, & Murray, 2016). Our results extend these results by demonstrating that traditional, sensory-specific definitions of “object” and the processes that objects engender, in terms of attentional selection (e.g., Duncan, 1984), goal representation (e.g., Duncan et al., 2010), and memory-related processes (Baddeley, 2000; Luck & Vogel, 1997), extend to multisensory stimuli (ten Oever et al., 2016). Our findings also showcase how rich spatiotemporal

EEG information offered by EN analyses can provide robust and in-depth understanding of the brain and cognitive mechanisms governing multisensory object templates and top-down attentional control more generally. Although typical EN analyses “do not” identify the specific networks and/or changes in the strength of their connectivity as a function of task, they offer temporally resolved and robust means to distinguish between strength- versus network-based brain mechanisms, that is, the type of information typically not assumed to be available from EEG. Our EN-derived ERP measures also have behavioral relevance. We urge the reader to note that this information was entirely sufficient to provide the wide variety of insights into the cognitive as well as brain underpinnings of top-down object template control that we have reported here. We now discuss how the EN framework can be particularly useful in testing how top-down attentional control and other neurocognitive functions operate in naturalistic environments.

Some of the most important solutions to the current problems with reliability of experimental findings in psychology and neuroscience involve replications and large, multicenter studies (e.g., Frank et al., 2017). Regarding brain imaging more specifically, many of the practical limitations characterizing fMRI and MEG are being addressed by calling for more robust measures and information sharing (e.g., Poldrack et al., 2017). As these corrective steps are being taken at the same time as neuroscientific research is venturing outside rigorous settings controlling stimulus parameters and the surrounding environment, one should carefully consider the elements of one’s (neuro) scientific approach that will ensure its validity in creating and testing real-world models of neurocognitive functions. The wide range of insights that our study have provided into top-down attentional control was possible because it built on the scientific and methodological achievements of research in this area. Specifically, our approach combines (1) adaptations of rigorous paradigms evoking specific cognitive processes (e.g., attentional capture and its task set contingency) to more naturalistic settings, with (2) the portable and easy to administer nature of EEG as a method of measuring brain activity, and (3) signal processing techniques that provide robust, easily replicable and directly neurophysiologically interpretable mechanistic insights into neurocognitive functions. We treat each of these elements in more detail, explaining its relevance to both corrective and real-world investigations within cognitive neuroscience.

With respect to experimental paradigms, one critical point is the capacity of contemporary models of neurocognitive functions to account for the information processing demands characterizing real-world environments, and this is in fact the main motivation behind the Special Focus that this study is part of (Matusz, Dikker, Huth, & Perrodin, in press, this issue; see also Peelen & Kastner, 2014). Large-sample and cognitive modeling studies are going to bring us closer to explaining functional brain

organization and cognitive functions as they occur in real-world environments the sooner they employ paradigms and create contexts that emulate information processing demands that characterize these environments. The tasks developed within research on visual (or auditory, tactile) top-down attentional control are exemplary here, as they managed to emulate many of the attributes of real-world environments: their multistimulus, competition-inducing nature (that necessitates top-down goal-based control) or the variabilities in stimulus task relevance and task difficulty. However, these traditional paradigms have typically omitted the inherently multisensory nature of real-world environments. Indeed, most research on attentional control, as well as on learning and memory, has focused on one sense at a time. This research has been invaluable in providing important insights into such areas of everyday functioning as scholastic achievement by showing how these skills are shaped by the interplay between attentional control and learning/memory processes (Merkley & Ansari, 2016; Merkley, Thompson, & Scerif, 2016; Cragg & Gilmore, 2014; Purpura & Ganley, 2014; Astle & Scerif, 2011; Bull, Espy, & Wiebe, 2008). Many of these insights may generalize to multisensory settings, as many multisensory processes might require years of experience to reach adult levels (reviewed in, e.g., Murray, Thelen, et al., 2016; ten Oever et al., 2016). However, other insights may not generalize as easily: Multisensory integration, often involuntary and effortless, is known to enhance/alter a wide range of cognitive functions: from faster, more accurate and less variable perception through stronger distraction and interference to more robust learning and memory (Matusz et al., 2017; Sarmiento et al., 2016; Matusz, Thelen, et al., 2015; Thelen, Matusz, & Murray, 2014; Murray & Wallace, 2012; Stein, 2012; Shams & Seitz, 2008; Taylor, Moss, Stamatakis, & Tyler, 2006; von Kriegstein & Giraud, 2006; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005; Ernst & Banks, 2002; Gibson, Maunsell, & Island, 1997). The scarcity of research on efficacy and strength of multisensory processes in real-world or even lab-based settings (like created here) and the non-linearity of mechanisms governing multisensory processes render attempts at modeling “signal” or “noise” in naturalistic settings without overt unisensory versus multisensory condition manipulations limited in their validity.

The present and other studies from our group aim to bridge these traditional approaches with the demands put on information processing in environments such classrooms or busy high street. We utilize well-understood behavioral measures of cognitive processes and paradigms emulating the top-down attention-demanding contexts (e.g., Lavie & Cox, 1997; Folk et al., 1992) and combine them with particular multisensory processes of interest (e.g., Matusz, Merkley, Faure, & Scerif, 2018; Matusz, Broadbent, et al., 2015; Matusz & Eimer, 2011a, 2011b, 2013). These, behavior- and model-focused investigations can then be enriched by the advantages afforded by EEG

and EN. Similar approaches are proposed by others in the neuroscientific community as means to address problems with (neuro)science reproducibility (e.g., Krakauer et al., 2017). We have employed our approach to better understand top-down control, but our group extends this approach to such areas as education and cognitive development, brain plasticity, and sensory disorders. One can easily imagine extensions of this approach to other neurocognitive functions by manipulating, for example, stimuli (e.g., dynamic, linguistic and/or familiar) and their context (e.g., scene), task (e.g., memory-encoding and retrieval), or the social nature of the experimental context. Such extensions would certainly help facilitate more complete real-world nature of our paradigm (cf. Peelen & Kastner, 2014). Notwithstanding, the advantages of our paradigm lie in “emulating” several real-world features: (1) clutter, (2) multisensoriness of stimulation, (3) unpredictability of task relevance of upcoming stimulus, and (4) stimulus’ spatial location. By separating cognitive and motor-related processes, our paradigm has the unique added advantage of process specificity (i.e., object template-based top control), unlike other paradigms.

Regarding the analyses of brain data, in the study of neurocognitive functions, combining well-understood ERP components like the N2pc with information richness of EN offers several clear advantages: (1) It reveals the brain mechanisms underlying the components and changes therein, which in turn inform cognitive hypotheses, (2) EN indices readily address the limited interpretability of canonical ERP analyses, and (3) the range of processes potentially reflected by the EN measures is constrained when combined with well-known ERP components and/or behavioral measures. The utility of this approach extends well beyond research on attentional control or laboratory walls. In extant real-world neuroscientific research, the advantages of EEG have been harnessed by extracting processes reflecting interpersonal brain activity synchronization, and this has shed a new light on the highly dynamic processes that scaffold social interactions and cognitive functions therein, like sustained attention, and learning and memory (Bevilacqua et al., 2018; Müller et al., 2018; Tseng et al., 2018; Dikker et al., 2017; Ko et al., 2017). Focusing EEG analyses on well-understood ERP components would offer the researchers valuable correlates of cognitive processes that do not require overt behavioral responses that may be hard to obtain in natural social interactions. The EN approach would then equip these analyses with robustness and direct neurophysiological interpretability.

Finally, regarding EEG itself, recent technological advances in creating cheap and effective portable EEG headsets open a new exciting avenue to utilize again on a wide scale its hardware and information-level advantages—this time, in neurocognitive research in the real world. The low-cost and easy-to-administer nature

EEG, combined with the rich and robust EN analyses, equip it with the potential to tackle many of the practical limitations characterizing fMRI and MEG (e.g., Poldrack et al., 2017). An important advantage of EN analyses is the ability to compare results across laboratories with different setups, as EN employs reference-independent EEG measures. Currently, in traditional EEG research, scientists choose to analyze results using the same reference electrodes, the active electrodes where the effect is measured as well as the same time window as their predecessors, as “traditionally used” and/or as they have done in the past. Voltage-based analyses are fundamentally reference dependent, which limits the interpretability of the “sign” or scalp localization of the effect (e.g., Murray et al., 2008). As such, we want to reiterate, an issue will always remain as to which electrode/approach of which laboratory is the “right one.” Furthermore and as we directly demonstrated here, there is no guarantee whatsoever that changes to the stimuli, stimulus design or task instructions, will not lead to changes in the active brain networks and so scalp topography triggered by “the same” stimulus in the present compared with past studies. Taking these points together, the global and data-driven nature of measures employed by the EN approach renders it an important advancement over traditional EEG analyses. As parameters of EN measures are highly reliable even with montages as small as eight channels (Khanna et al., 2014), EN has a big potential for supporting replicable neuroscientific research, in and outside the laboratory.

To summarize, we used EN to provide direct evidence for visual attention control by integrated top-down object templates in naturalistic, multisensory laboratory settings. These insights were possible by our approach that combined adaptations of rigorous selective attention paradigms, the advantages of EEG as a brain activity measure, and the robust and information-rich measures offered by EN. We believe that this “naturalistic laboratory” approach constitutes a crucial, intermediate stage between “classic” laboratory research and “fully naturalistic” research conducted in veridical real-world situations (see Matusz et al., in press, this issue, for detailed discussion). Traditional studies offer maximal control over stimulation and the testing environment but involve settings far detached from the real world, and as such offer partial—but highly rigorous—tests of models of perception and action. The studies conducted in everyday situations involve maximally natural conditions for investigating neurocognitive functions, but this occurs at the expense of the control over the stimulation or the environment. By “emulating,” in a controlled fashion, the demands of information processing in real-world environments, our approach, as demonstrated here, allows for more careful testing, inside the laboratory, of hypotheses regarding neurocognitive functions as they occur in everyday situations and retaining elements of the same approach, doing so also in the real world.

UNCITED REFERENCE

Seltzer & Pandya, 1994

Acknowledgments

P. J. M. is funded by the Pierre Mercier Foundation and the Swiss National Science Foundation (Grant PZ00P1_174150). M. M. M. is funded by the Swiss National Science Foundation (Grants 320030-149982 and 320030-169206 and the National Centre of Competence in Research Project “SYNAPSY, The Synaptic Bases of Mental Disease,” Project 51NF40-158776) as well as by a generous grantor advised by Carigest SA. M. M. M. and P. J. M. are both funded by the Fondation Asile des Aveugles.

Reprint requests should be sent to Pawel J. Matusz, Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO Valais), Technopole 3, 3960 Sierre, Switzerland, or via e-mail: pawel.matusz@gmail.com.

REFERENCES

- Alsius, A., & Soto-Faraco, S. (2011). Searching for audiovisual correspondence in multiple speaker scenarios. *Experimental Brain Research*, *213*, 175–183.
- Assad, J. A., & Maunsell, J. H. R. (1995). Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature*, *373*, 518–521.
- Astle, D. E., & Scerif, G. (2011). Interactions between attention and visual short-term memory (VSTM): What can be learnt from individual and developmental differences? *Neuropsychologia*, *49*, 1435–1445.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*, 1190–1192.
- Bevilacqua, D., Davidesco, I., Wan, L., Oostrik, M., Chaloner, K., Rowland, J., et al. (2018). Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography study. *Journal of Cognitive Neuroscience*, *1*, 194–198.
- Brandeis, D., Lehmann, D., Michel, C. M., & Mingrone, W. (1995). Mapping event-related brain potential microstates to sentence endings. *Brain Topography*, *8*, 145–159.
- Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience*, *2011*, Article ID 813870.
- Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205–228.
- Cohen, S. S., & Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *eNeuro*, *3*, 1–11.
- Couperus, J. W., & Quirk, C. (2015). Visual search and the N2pc in children. *Attention, Perception, and Psychophysics*, *77*, 768–776.
- Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*, 63–68.
- 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, 1–6.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.
- Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., et al. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, 27, 1375–1380.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–517.
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14, 172–179.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458.
- Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, 99, 225–234.
- Eimer, M., Kiss, M., Press, C., & Sauter, D. (2009). The roles of feature-specific task set and bottom-up salience in attentional capture: An ERP study. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1316–1328.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Fender, D. H. (1987). Source localization of brain electrical activity. In A. S. Gevins & A. Remond (Eds.), *Methods of analysis of brain electrical and magnetic signals EEG handbook* (pp. 309–354). Amsterdam: Elsevier.
- 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, 1030–1044.
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., et al. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, 22, 421–435.
- Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2009). The long road to automation: Neurocognitive development of letter–speech sound processing. *Journal of Cognitive Neuroscience*, 21, 567–580.
- Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, 405, 347–351.
- Gibson, J. A. Y. R., Maunsell, J. H. R., & Island, R. (1997). Sensory modality specificity of neural activity related to memory in visual cortex. *Cortex*, 78, 1263–1275.
- Handy, T. C., & Khoe, W. (2005). Attention and sensory gain control: A peripheral visual process? *Journal of Cognitive Neuroscience*, 17, 1936–1949.
- Heinze, H.-J., Luck, S. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focus attention within bilateral arrays. I. Evidence for early selection. *Electroencephalography & Clinical Neurophysiology*, 75, 511–527.
- Hopf, J.-M., Luck, S. J., Girelli, M., Mangun, G. R., Scheich, H., & Heinze, H.-J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, 10, 1233–1241.
- Hwang, K., Shine, J. M., & D'Esposito, M. (2017). Fronto-parietal interactions with task-evoked functional connectivity during cognitive control. *BioRxiv*, 1–23. Retrieved from <http://biorxiv.org/content/early/2017/05/03/133611.abstract>.
- Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, 15, 548–554.
- Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS Computational Biology*, 13, e1005268.
- Khanna, A., Pascual-Leone, A., & Farzan, F. (2014). Reliability of resting-state microstate features in electroencephalography. *PLoS One*, 9, 1–21.
- Kingstone, A. (1992). Combining expectancies. *The Quarterly Journal of Experimental Psychology Section A*, 44, 69–104.
- Kiss, M., Grubert, A., & Eimer, M. (2013). Top-down task sets for combined features: Behavioral and electrophysiological evidence for two stages in attentional object selection. *Attention, Perception & Psychophysics*, 75, 216–228.
- Kitagawa, N., Zampini, M., & Spence, C. (2005). Audiotactile interactions in near and far space. *Experimental Brain Research*, 166, 528–537.
- Ko, L.-W., Komarov, O., Hairston, W. D., Jung, T.-P., & Lin, C.-T. (2017). Sustained attention in real classroom settings: An EEG study. *Frontiers in Human Neuroscience*, 11, 1–10.
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93, 480–490.
- Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, 14, 420–429.
- Lavie, N., & Cox, S. (1997). On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. *Psychological Science*, 8, 395–398.
- Le, T. H., Pardo, J. V., & Hu, X. (1998). 4 T-fMRI study of nonspatial shifting of selective attention: Cerebellar and parietal contributions. *Journal of Neurophysiology*, 79, 1535–1548.
- Lehmann, D., Ozaki, H., & Pal, I. (1987). EEG alpha map series: Brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and Clinical Neurophysiology*, 67, 271–288.
- Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, 48, 609–621.
- Love, B. C. (2015). The algorithmic level is the bridge between computation and brain. *Topics in Cognitive Science*, 7, 230–242.
- Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31, 291–308.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proceedings of the National Academy of Sciences, U.S.A.*, 108, 20661–20665.
- Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, 63, 397–405.
- Maidenbaum, S., Abboud, S., & Amedi, A. (2014). Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience and Biobehavioral Reviews*, 41, 3–15.
- Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, 32, 4–18.
- Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial locations: Tradeoff functions for event-

- related brain potentials and detection performance. *Perception & Psychophysics*, *47*, 532–550.
- Matusz, P. J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R., & Scerif, G. (2015). Multi-modal distraction: Insights from children's limited attention. *Cognition*, *136*, 156–165.
- Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (in press). Are we ready for real-world neuroscience? *Journal of Cognitive Neuroscience*.
- Matusz, P. J., & Eimer, M. (2011a). Multisensory enhancement of attentional capture in visual search. In *Perception 40 ECVF 2011 Abstract Supplement* (p. 184). Retrieved from www.perceptionweb.com/abstract.cgi?id=v110603.
- Matusz, P. J., & Eimer, M. (2011b). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*, *18*, 904–909.
- Matusz, P. J., & Eimer, M. (2013). Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*, *50*, 996–1009.
- Matusz, P. J., Merkley, R., Faure, M., & Scerif, G. (2018, 16 June). Expert attention: Attentional allocation depends on the differential development of multisensory number representations. *PsyArXiv*.
- Matusz, P. J., Thelen, A., Amrein, S., Geiser, E., Anken, J., & Murray, M. M. (2015). The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. *European Journal of Neuroscience*, *41*, 699–708.
- Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*, *105*, 243–252.
- Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behavior. *Current Opinion in Behavioral Sciences*, *10*, 14–20.
- Merkley, R., Thompson, J., & Scerif, G. (2016). Of huge mice and tiny elephants: Exploring the relationship between inhibitory processes and preschool math skills. *Frontiers in Psychology*, *6*, 1–14.
- Michel, C. M., & Koenig, T. (2018). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *Neuroimage*, *180*, 577–593.
- Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *Neuroimage*, *61*, 371–385.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review Neuroscience*, *24*, 167–202.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.
- Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., et al. (2008). Cross-modal deactivations during modality-specific selective attention. *BMC Neurology*, *8*, 1–11.
- Müller, V., Sängler, J., & Lindenberger, U. (2018). Hyperbrain network properties of guitarists playing in quartet. *Annals of the New York Academy of Sciences*, *1423*, 198–210.
- Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*, *20*, 249–264.
- Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory processes: A balancing act across the lifespan. *Trends in Neurosciences*, *39*, 567–579.
- Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia*, *83*, 161–169.
- Murray, M. M., & Wallace, M. T. (2012). *The neural bases of multisensory processes*. CRC Press.
- Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, *72*, 1654–1665.
- Parks, E. L., & Madden, D. J. (2013). Brain connectivity and visual attention. *Brain Connectivity*, *3*, 317–338.
- Peelen, M. V., & Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. *Trends in Cognitive Sciences*, *18*, 242–250.
- Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., et al. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*, 115–126.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology*, *122*, 104–121.
- Putzar, L., Goerendt, I., Lange, K., Rösler, F., & Röder, B. (2007). Early visual deprivation impairs multisensory interactions in humans. *Nature Neuroscience*, *10*, 1243–1245.
- Reich, L., Maidenbaum, S., & Amedi, A. (2012). The brain as a flexible task machine: Implications for visual rehabilitation using noninvasive vs. invasive approaches. *Current Opinion in Neurology*, *25*, 86–95.
- Rohe, T., & Noppeney, U. (2018). Reliability-weighted integration of audiovisual signals can be modulated by top-down control. *ENeuro*, *5*, ENEURO.0315.
- Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1311–1321.
- Sarmiento, B. R., Matusz, P. J., Sanabria, D., & Murray, M. M. (2016). Contextual factors multiplex to control multisensory processes. *Human Brain Mapping*, *37*, 273–288.
- Scerif, G., Longhi, E., Cole, V., Karmiloff-Smith, A., & Cornish, K. (2012). Attention across modalities as a longitudinal predictor of early outcomes: The case of fragile X syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*, 641–650.
- Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. *The Journal of Comparative Neurology*, *343*, 445–463.
- Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cerebral Cortex*, *14*, 1346–1357.
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*, 411–417.
- Shulman, G. L., D'Avossa, G., Tansy, A. P., & Corbetta, M. (2002). Two attentional processes in the parietal lobe. *Cerebral Cortex*, *12*, 1124–1131.
- Stein, B. (2012). *The new handbook of multisensory processing*. Cambridge, MA: MIT Press.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–267.
- Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 8239–8244.
- ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M. M., & Matusz, P. J. (2016). The COGs (context, object, and goals) in multisensory processing. *Experimental Brain Research*, *234*, 1307–1323.

- Thelen, A., Matusz, P. J., & Murray, M. M. (2014). Multisensory context portends object memory. *Current Biology*, *24*, R734–R735.
- Tivadar, R. I., & Murray, M. M. (2018). A primer on electroencephalography and event-related potentials for organizational neuroscience. *Organizational Research Methods*, *22*.
- Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. *Cognitive Theory*, *3*, 199–239.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Tseng, P. H., Rajangam, S., Lehew, G., Lebedev, M. A., & Nicolelis, M. A. L. (2018). Interbrain cortical synchronization encodes multiple aspects of social interactions in monkey pairs. *Scientific Reports*, *8*, 1–15.
- van Atteveldt, N. M., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. *Neuron*, *81*, 1240–1253.
- van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage*, *55*, 1208–1218.
- von Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche (Schluss.). *Annalen Der Physik*, *165*, 353–377.
- von Kriegstein, K., & Giraud, A. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, *4*, e326.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A.-L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, *17*, 367–376.
- Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology*, *79*, 1574–1578.

Uncorrected Proof

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

During the preparation of your manuscript, the questions listed below arose. Kindly supply the necessary information.

1. Please verify the affiliations.
2. In the abstract section, kindly check if the changes made in the statement “spatially selective enhancements of neural processing of objects within ventral visual cortices at approximately 150–300 msec poststimulus” are appropriate.
3. Please check if the use of the phrase “visual audiovisual” here is appropriate (Introduction, para. 8).
4. Mast et al., 2015, 2017, were not cited in the reference list. Please check.
5. Hickey et al., 2006, was not cited in the reference list. Please check.
6. Please spell out GEV.
7. Duncan et al., 2010, was not cited in the reference list. Please check.
8. Please insert a citation of Seltzer & Pandya, 1994, in the body.
9. Please provide the publisher location of Murray & Wallace, 2012.

END OF ALL QUERIES

Uncorrected Proof