Brain and cognitive mechanisms of top-down attentional control in a multisensory world: Benefits of electrical neuroimaging

Pawel J. Matusz[1,2,3 *†], Nora Turoman[2], Ruxandra I. Tivadar[2], Chrysa Retsa[2] & Micah M. Murray[2-5*]

[1] Information Systems Institute at the University of Applied Sciences Western Switzerland

(HES-SO Valais), 3960 Sierre, Switzerland

[2] The LINE (Laboratory for Investigative Neurophysiology), Department of Radiology, University Hospital

Centre and University of Lausanne, 1011 Lausanne, Switzerland

[3] Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN 37203-

5721, USA

[4] The EEG Brain Mapping Core, Center for Biomedical Imaging (CIBM), University Hospital

Center and University of Lausanne, 1011 Lausanne, Switzerland

[5] Department of Ophthalmology, University of Lausanne and Fondation Asile des Aveugles,

Lausanne, Switzerland

* co-authors contributed equally to this work

† Corresponding Author:

pawel.matusz@gmail.com

Information Systems Institute

University of Applied Sciences Western Switzerland (HES-SO Valais)

Technopole 3, 3960 Sierre, Switzerland

ABSTRACT

In real-world environments information is typically multisensory, and objects are a primary unit of information processing. Object recognition and action necessitates 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 multi-stimulus (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 app. 150–300ms post-stimulus. In our adaptation of Folk et al.'s (1992) spatial-cueing paradigm, visual cues elicited weaker behavioural 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 analysed global features of the electrical field at the scalp across our N2pcs. In the N2pc time-window (170–270ms), colour 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.

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 et al., 2017; Love, 2015; Poldrack et al., 2017). 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 behavioural 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 behavioural 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 non-human mammals revealed how top-down, goal-based attention controls stimulus processing through "gain control", i.e., enhancing neuronal responses to stimuli with task-relevant features (e.g. spatial location) and resolving stimulus competition in multi-stimulus 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 characterise the earliest information processing stages influenced by spatial attention and how the latter interacts with feature-based top-down attention in multi-stimulus contexts. Theoretical progress was further facilitated by systematic focus on specific event-related potential (ERPs) 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, 1996; Eimer et al., 2009; Heinze et al., 1990; Luck & Hillyard, 1994; Mangun & Hillyard, 1990). More recently, novel mechanisms orchestrating top-down attention control have been revealed by haemodynamic imaging methods, such as fMRI and PET. Such neuroimaging studies provided direct evidence for the purported top-down "attentional templates" (Duncan & Humphreys, 1989), i.e., actively maintained working memory representations of task-relevant object features whereby the brain biases sensory processing towards our goals. These biasing signals were revealed to be generated by a network of fronto-parietal "sources" (e.g. Le et al., 1998; Serences et al., 2004; Wojciulik et al. 2009). More recently, fMRI studies revealed how both stable and flexible patterns of connectivity between the fronto-parietal network and sensory cortices are instrumental to top-down attention (Hwang et al. 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, non-invasiveness - 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; Cohen & Parra, 2016; Dikker et al., 2017; Ko et al. , 2017; Müller et al., 2018; Tseng et al. 2018). The way to fully capitalise 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, e.g. 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 (~≤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 8-channel data (Khanna et al. 2014). Further, EN-analysed 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 labs 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 (Assad & Maunsell, 1995a; Mahon et al., 2009; Maidenbaum et al., 2014). Some multisensory processes take place in early sensory cortices, at stages preceding those controlled by top-down attention (reviewed in De Meo et al., 2015; Murray et al., 2016a; van Atteveldt et al. , 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 & Eimer, 2011; Matusz et al., 2015a; Santangelo & Spence, 2007; Scerif et al. 2012; van der

Burg et al. 2011). Thus, even such fundamental tenets, as the predominance of top-down over bottom-up mechanisms of attentional control (cf. Folk et al. 1992), established by unisensory research, may not hold in naturalistic, multisensory environments. Second, unisensory responses have limited power in predicting multisensory responses due to non-linear mechanisms involved in integration of multisensory information (reviewed in Ernst & Banks, 2002; Murray & Wallace, 2012; Stein, 2012). That is, if, e.g., visual and auditory (brain or behavioural) 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 et al. 2009; Iordanescu et al., 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, represent 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 fronto-parietal network to represent behaviourally relevant information (e.g. Miller & Cohen, 2001). In further support of integrated multisensory representations of task-relevant features, neurons in the prefrontal cortex have been shown to represent, in a task-dependent fashion, arbitrary but task-relevant conjunctions of colour and pitch (Fuster et al. 2000). Additionally, posterior parietal cortices, known to contribute to control of attention towards task-relevant spatial locations and task-relevant object features, represent space in a multisensory fashion (Assad & Maunsell, 1995; Shulman et al., 2002; reviewed in Stein & Stanford, 2008). Lastly, there is converging evidence to suggests that even neurons at such early stages of cortical processing as "sensory-specific" V1 are involved in multisensory integration (reviewed in Murray et al. 2016a), making the multisensory nature of object representations in the higher-level brain areas even more plausible (Matusz et al., 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; Colour vs. Colour-Tone task). Search arrays were preceded by a display with a visual cue that always matched the target colour. The ability of visual cues to capture attention in the visual task was attenuated and/or eliminated in the audiovisual task, across both brain and behavioural responses. Subsequently, behavioural studies have demonstrated that multisensory object templates generalise across different sensory pairings and non-spatial 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, i.e., the N2pc component, a negative-going voltage deflection over posterior electrodes collateral to the stimulus location ~200ms post-stimulus (Eimer, 1996; Luck & Hillyard, 1994). We found that within the traditional time-window (170-270ms), 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) as well as in the configuration of activated networks (see Murray et al. 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 re-considered 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- ("gain") vs. network-based mechanisms (Lehmann & Skrandies, 1980; Michel & Murray, 2012; Murray et al., 2008). 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 characterising the canonical analyses of lateralised ERP components like N2pc. For one, the latter measure the difference between two electrode channels (or channel subsets) located in opposite hemiscalps. Whether explicit or implicit, an assumption behind these analyses is that this two-point difference in lateralised 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 lateralised ERP analyses, as they ignore the majority of the recorded brain data (91.5% in the case of ERP analysis of data from 2 channels from a 23-channel montage; a percentage that would further increase with higher-density montages, though the degree of non-independence across electrodes is another important consideration). As such, they likely risk missing effects occurring elsewhere across the electric field at the scalp. Furthermore, the two-point subtraction that underlies the mean amplitude N2pc difference might have arisen due to different contexts (see **Fig.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 lateralised ERP gradient elicited by colour cues between the two search tasks in the N2pc time-window. When the brain is set to control sensory processing towards stimuli defined by a given colour, such as "all objects in red", responses of all neurons representing red are enhanced (Desimone & Duncan, 1995). When the brain is set towards a multisensory, colour-tone defined stimulus (cf. colour-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 colour-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 includes another feature. If we observed 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 non-visual signals influencing more general, tonic baseline activity changes within colour-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., Laurienti et al., 2000; Mozolic et al., 2008).

METHODS

Participants

Participants were twelve right-handed paid volunteers with normal or corrected-to-normal vision (mean age 25.8 years, age range 21–40 years, 5 females). None of the subjects 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 prior to 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 behavioural (RT spatial cueing effects) and EEG/ERP markers of attentional selection (N2pc amplitude; Matusz & Eimer 2013).

Stimuli and procedure

Behavioural 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 colour singleton (either red or blue), with the cue colour always matching the colour of the target. What differed between the two tasks was how the targets and the non-target search arrays were defined. In the *Colour* task, participants had to respond to colour-singleton bars when they appeared in the target colour (e.g., blue), and ignore bars defined by non-target colour (e.g., red when searching for blue). Both trial types were presented equiprobably and in a random order. In the *Colour-Tone* task, participants had to respond to bars of the same colour, but only when they were accompanied by a tone (e.g., blue 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 the present study. Each participant looked for either a blue or a red target-colour 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 colour and target pitch (as well as mapping hand-key) were counterbalanced across participants. The between-subject counterbalancing of colour and sound pitch target feature pairings served to prevent our results from being influenced by some participants potentially possessing any pre-existing associations between colours and sound pitches, where they would regard subjectively a particular pitch level "more congruent" with a given colour. 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 non-target).

Visual stimuli were presented against a black background on a 22" LCD monitor (100 Hz refresh rate; 100cm viewing distance; Samsung wide SyncMaster 2233). In the cue array, each of the six elements was composed of four closely aligned dots (0.17° x 0.17°). The colour singleton cue was blue or red (CIE x/y chromaticity coordinates .161/.128 and .621/.128, respectively), and the remaining items were grey (.308/.345). This colour 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 localisation 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° x 0.3°) at the same positions as the preceding cue elements, with bar orientation chosen randomly for each position. Coloured bars appeared with equal probability at one of the four lateral locations. All grey, blue, and red stimuli in the cue and search displays were equiluminant (~11 cd/m2). In the Colour-Tone task, the sound was a pure sine-wave tone (50 ms duration; 65 dB SPL) of high- or low-pitch (2000Hz and 300Hz, respectively for participants searching) that was presented concurrently with search array onset from a loudspeaker located centrally behind the monitor.

EEG acquisition and pre-processing

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 500Hz and impedances were kept below 5kΩ. Cartool (available at http://www.fbmlab.com/cartool-software/) (Brunet et al., 2011) was used for data pre-processing and the statistical analyses. Next and prior to averaging, the EEG was filtered offline with a 2nd order Butterworth filter (-12dB/octave roll-off; 0.1Hz high-pass; 40Hz low-pass). The filters were computed linearly in both

forward and backward directions to eliminate phase shifts. Then, the continuous EEG was segmented into peri-stimulus epochs, relative to the colour cue onset, spanning from 100ms pre-stimulus to 500ms post-stimulus onset. Subsequently, data quality was controlled with a semi-automated artefact rejection criterion of ±80μV at each channel as well as visual inspection to exclude any remaining transient noise, eye movements and muscle artefacts.

To obtain lateralised ERPs, for each subject, the artefact-free single-trial epochs were averaged and pre-stimulus baseline corrected (using the -100ms to 0ms time-interval), separately for trials with colour singleton cues presented in the left and right hemifield, for each of the two search tasks, resulting in 4 average ERPs. Then, the 2 weighted ERP averages (weighted according to number of accepted epochs) from conditions with cues presented in the left were re-labelled, 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 lateralised ERP file. As this was done separately for each of the two search tasks, this resulted finally in two lateralised cue-elicited ERP averages: 1 for the Tone task, and 1 for the Colour-Tone task. From this step onwards, 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 pre-processing of the ERPs triggered by the visual cues across the Colour task and the Colour-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 multi-step 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 lateralised ERPs mean amplitude measures across the 170-270ms post-cue onset time-window from electrodes PO7 and PO8 and submitted these to a 2x2 within-subject repeated-measures ANOVA with factors Task (Colour task vs. Colour-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 vs. 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 (Murray et al., 2008; Michel & Murray, 2012). 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 lateralised 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 vs. sea level or vs. 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, 1987; see also Fig.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 of the two "mountain peak" points, e.g., 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 3 hypothetical lateralised-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 (**Fig.3A**) is the difference between a contralateral electrode (black circle) and a respective ipsilateral electrode (grey circle). For Conditions 1–3, the N2pc value would be measured as -4, -16, and -4µV, 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 Global Field Power (**Fig.3BC**) and Global Dissimilarity (**Fig.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 (RMS), 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 4-fold 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, 170-270ms post-cue time-window and then compared them directly between the Colour and the Colour-Tone task using a paired t-test. Global characteristics of the electric scalp field gradients, compared to the measures of local field potentials as represented, e.g., by N2pc, should provide a more complete answer as to whether top-down attentional control via object templates can operate by altering the overall strength of the lateralised 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 colour 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 (versus 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, while Conditions 2 and 3 have the same strength but different topographies. **Figure 3A** depicts the original data from the 3 conditions, while panel D shows the same data that have been GFP-normalized. Thus, after re-scaling all 3 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 equalling 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 2 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 the present 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 $[(1-DISS^2)/2]$. If two ERPs differ in topography independently 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 allows defining periods of stable patterns of ERP activity and changes therein. In fact, GFP and DISS are inversely related, i.e., when GFP is high, ERP topographic activity tends to remain stable (i.e., DISS is low; see above), whereas its changes when GFP is low. Displaying DISS across time shows a highly characteristic behaviour, 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 organised configurations have been shown to represent successive steps along the information processing pathway from perception to action (also known as "functional microstates"; Brandeis et al., 1995; Lehmann et al., 1987; Lehmann & Skrandies, 1980; Michel & Koenig, 2017).

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, or TAAHC) to identify stable electric field topographies (henceforth "template maps") present in the *group*-averaged cue-elicited ERPs between the two tasks within the whole 500ms 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 dataset. 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, i.e., 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", i.e., 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, 2017). The end-product here is effectively 1 final cluster (which is not informative), and so it is important in the TAAHC 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, i.e. 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 142–260ms post-cue time-period) was "fitted" to the template maps that in the group-averaged hierarchical clustering were differentially characterising the two search tasks over the 142–260ms 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 subject and each of the 2 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 2 lateralised ERPs was then submitted to repeated-measures 2x3 ANOVA with within-subject factors of Task (2) and Map (3).

Step 4. Associations between behavioural and ERP measures of attentional selection

Next, we investigated whether the changes induced as a function of search task in the *behavioural* 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 non-parametric brain–behavioural 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 jacknife 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 %-amplitude changes and mean GFP %-changes in the Colour-Tone task relative to the Colour task were correlated separately with the % changes in the behavioural capture effects between the two tasks (i.e. the difference in capture effects between the Colour-Tone and Colour task divided by the capture effect on the colour task, which can therefore result in percentages in excess of ±100%). Similar analyses were conducted for the %-change in the mean duration of each of the 3 template ERP maps

measured within each participant separately during the fitting procedure, as compared in the Colour-Tone task relative to the Colour task.

<div style="text-align:center">

RESULTS

</div>

Behavioural results

Only RTs from correct trials within 200 to 1000ms, and within ±3 SDs from the mean were analysed (leading to a loss of <1% data). A 2 x 2 within-subjects design was used; i.e. Search task (Colour vs. Colour-Tone task) x Cue-Target Location (same vs. different). Performance was quantified using RT* (inverse efficiency; Townsend & Ashby, 1978), which is an aggregate measure that takes into the account both reaction speed and accuracy (for RT and accuracy rate analyses, see Matusz & Eimer 2013, Exp.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 (Kitagawaet al., 2005; Ludwig et al., 2011; Ngo & Spence, 2010; Putzar et al., 2007). Overall, performance was similar across the two tasks, $F(1,11)=1.02$, $p=.34$. RT* showed that the cues captured attention behaviourally, 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 comparisons revealed that this interaction was driven by the fact that visual cues triggered reliable RT* attention capture effects in the Colour task, $t(11)=5.4$, $p=.001$, but this capture was so attenuated in the Colour-Tone task that it was no longer reliably present, $t<1$ (**Fig.4**).

ERP results

Step 1. Canonical N2pc analyses

The 2x2 repeated-measures ANOVA on the mean N2pc amplitudes over 170-270ms post-cue recorded over PO7/8 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 Colour task ($t(11)=4.87$, $p=.001$) and in the Colour-Tone task ($t(11)=3.73$, $p=.003$) (see **Fig.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

<div style="text-align:center">

15

</div>

The pairwise t-test carried out on the mean amplitude of the GFP waveform over the same, 170-270ms post-cue time-window revealed a pattern contrasting with that found for the mean N2pc amplitudes. The mean GFP was *enhanced* in the Colour-Tone task, when compared to the Colour task, $t(11)=2.62$, $p=0.012$ (paired t-test; Fig.5D). That is, the gradients within the scalp-wide distribution of voltages triggered in response to the same visual cue were in fact overall stronger in the Colour-Tone task as compared to the Colour 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 140ms post-cue, the same template maps characterised both tasks **(Fig.6A)**. The fitting procedure, utilising spatial correlation between template maps identified in the group-averaged data and single-subject data from each of the 2 conditions, showed that over the 142–260ms period, three different template maps appeared to differentially characterise each task (**Fig.6B, upper panel**). A subsequent 2x3 repeated-measures Task x Map ANOVA on the percentage of time each template map best correlated spatially with single-subject data over the 142-260ms time-period revealed a significant main effect of Map ($F(2,22)=7.61$; $p=0.032$; $\eta_p^2=0.27$), as well as a significant interaction ($F(2,22)=4.06$; $p=0.032$; $\eta_p^2=0.27$) (**Fig.6B, bottom panel**). Post-hoc non-parametric tests (Wilcoxon signed rank tests) were conducted for each map, comparing their relative characterisation of responses to each task. The light green template map characterised responses to both tasks equally well ($p=0.332$). The middle green template map characterised responses to the Colour task more than the Colour-Tone task ($p=0.040$). The dark green template map characterised responses to the Colour-Tone task more than the Colour task, though this exhibited a non-significant trend ($p=0.091$).

Step 4. Brain-behavioural response correlations

There was a strong correlation between the relative changes between the Colour and the Colour-Tone task in behavioural capture effects and mean GFP, $rho_c(10)=.88$, $p<.001$. In contrast, mean N2pc amplitude measured from the PO7/8 electrode pair did not show a similar correlation with behaviour, $rho_c(10) = .246$, $p=.43$ (**Fig.5CD, right panel**). These correlation coefficients significantly differed ($z=2.39$; $p=0.017$, 2-tailed). There were no reliable correlations with similar measures from the fitting above.

DISCUSSION

We investigated whether re-consideration of lateralised ERPs within an EN analytical framework could provide direct and novel evidence into the brain and cognitive mechanisms governing top-down attentional

control by multisensory object templates, as compared to 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 PO7/8 in the typical 170–270ms time-window was reduced in the Colour-Tone task, as compared to the Colour task. When we assessed the mean cue-induced GFP in the same time-window, it was in fact enhanced in the Colour-Tone task compared to the Colour task. Furthermore and most interestingly, there were topographic differences between the lateralised cue-induced ERPs, i.e., visual cues activated distinct brain source configurations in task contexts where they matched the target template fully (Colour task) vs. only partly (Colour-Tone task). What do these findings suggest for top-down control by multisensory object templates? First, in light of the behavioural 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 colour 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 lateralised ERPs are directly interpretable in terms of top-down control via integrated multisensory object templates. In the Colour search task, the spatially-selective brain responses to visual cues would be driven by populations of neurons that represent colour (e.g., "red"). In contrast, in the Colour-Tone task, the responses to the same visual cues would now involve neuronal populations that encode both colour *and* tone. One could imagine this could involve a subset of the same, colour-coding neuronal population active in the Colour task, but it is more likely that these are neighbouring yet distinct populations coding unisensory and multisensory stimulus features (see, e.g., Beauchamp et al., 2004, for evidence for such organisation in the superior temporal sulcus). Posterior parietal cortex (PPC) and lateral occipital cortex (LOC), two areas shown to give rise to the N2pc in the only existing source localisation study (using MEG; Hopf et al. 2000), are both known multisensory hubs (e.g. Reich et al., 2012; Rohe & Noppeney, 2018). 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 multisensory-

ness 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. Figs.8–9).

In contrast, the engagement of a gain control mechanism purported in the present 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 colour-defined stimuli (or even, specific colours), another for shape stimuli, and so on for other visual dimensions. In this case, non-visual 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 to 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–behaviour correlation (not found for N2pc) supports this possible explanation. While 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 important inasmuch as there is a an assumption, albeit perhaps implicit, that N2pc amplitude differences reflect the modulation in the strength of activity of brain circuits involved in attentional selection; i.e. 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 colour 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, while 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. While 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 vs. cue-induced N2pc. Gain control has been traditionally invoked as a putative general mechanism governing top-down attentional control. While 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 summarise, 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 to 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 et al., 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 spatio-temporal 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. While 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, i.e., the type of information typically not assumed to be available from EEG. Our EN-derived ERP measures also have behavioural 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 as well as 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, multi-centre studies (e.g., Frank et al., 2017). Regarding brain imaging more specifically, many of the practical limitations characterising 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 of 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 contingence) 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 characterising real-world environments, and this is in fact the main motivation behind the Special Focus that this study is part of (Matusz et al. 2018 *this issue;* see also Peelen & Kastner, 2014). Large-sample and cognitive modelling studies are going to bring us closer to explaining functional brain organisation and cognitive functions as they occur in real-world environments the sooner they employ paradigms and create contexts that emulate information processing demands that characterise 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 multi-stimulus, 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 (Astle & Scerif, 2011; Bull et al., 2008; Cragg & Gilmore, 2014; Merkley & Ansari, 2016; Merkley et al., 2016; Purpura & Ganley, 2014). Many of these insights may generalise to multisensory settings, as many multisensory processes might require years of experience to reach adult levels (reviewed in e.g. Murray et al., 2016; ten Oever et al. 2016). However,

other insights may not generalise 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 (Ernst & Banks, 2002; Gibson et al., 1997; Kriegstein & Giraud, 2006; Kriegstein et al., 2005; Matusz et al., 2015a, 2017; Murray & Wallace, 2012; Sarmiento et al., 2016; Shams & Seitz, 2008; Stein, 2012; Taylor et al., 2006; Thelen et al., 2014). 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 modelling "signal" or "noise" in naturalistic settings without overt unisensory vs. 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 utilise well-understood behavioural measures of cognitive processes and paradigms emulating the top-down attention-demanding contexts (e.g. Folk et al., 1992; Lavie & Cox, 1997), and combine them with particular multisensory processes of interest (e.g., Matusz & Eimer, 2011, 2013; Matusz, 2015b, 2018). These, behaviour- and model-focused, investigations can then 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, e.g., 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 (c.f. Peelen & Kastner 2014). Notwithstanding, the advantage(s) 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), unike 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 behavioural 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 inter-personal brain activity synchronisation, and this has shed 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; Dikker et al., 2017; Ko et al., 2017; Müller et al., 2018; Tseng et al., 2018). Focusing EEG analyses on well-understood ERP components would offer the researchers valuable correlates of cognitive processes that do not require overt behavioural 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 utilise 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 characterising fMRI and MEG (e.g., Poldrack et al., 2017). An important advantage of EN analyses is the ability to compare results across labs with different setups, as EN employs reference-independent EEG measures. Currently, in traditional EEG research, scientists choose to analyse 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 localisation 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 lab 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 to 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 8-channel (Khanna et al. 2014), EN has big potential for supporting replicable neuroscientific research, in and outside of the laboratory.

To summarise, we used EN to provide direct evidence for visual attention control by integrated top-down object templates in naturalistic, multisensory lab 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. 2018, *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 lab, 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.

Figure legends

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

Figure 2. Flowchart of analytical steps performed in the present 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 colour cues between the Colour vs. the Colour-Tone task using the traditional approach. That is, we extracted and compared between the two tasks mean amplitudes in the lateralised cue-elicited ERPs measured in the traditional N2pc (170–270ms post-stimulus) time-window and over the traditional PO7/8 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) as well as 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 lateralised ERP responses within statistically indistinguishable brain networks. For this purpose, we compared between the two search tasks the mean GFP values over the same, 170–270ms post-cue 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 colour 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 500ms post-cue time-period. The clustering revealed 13 different "template maps", i.e., periods of stable ERP activity, characterising cue-elicited lateralised brain responses within that time-window. Subsequently, the 3 template maps identified within the group-averaged ERPs as differentially characterising the two search tasks over the 142–260ms time-period were submitted to a fitting procedure. During fitting, each time point of the single-subject ERP in the 142–260ms post-cue time-period was "fitted" to each of the 3 template maps derived from group-averaged hierarchical clustering. The relative duration of each of the 3 template maps in the lateralised ERPs across the two search tasks was then statically assessed. Finally, in Step 4, we investigated the behavioural 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 behavioural 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 non-parametric brain–behavioural measure correlations using the Spearman's rho and combined it with the jackknife data resampling procedure.

Figure 3. Illustration of the added value of an electrical neuroimaging analysis framework. A. Condition 1 displays exemplar voltage data across a hypothetical 16-electrode montage. The values of Condition 2 are increased precisely 4-fold 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 (grey circle). For Conditions 1–3, the N2pc value would be measured as -4, -16, and -4μV, 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 panel A are displayed for each of the 3 hypothetical conditions. C. The Global Field Power (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 8μV, respectively. D. The GFP-normalised voltage data are displayed and are calculated by dividing the values at each electrode in panel A by the GFP calculated in panel 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 panel 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 the present 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.

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

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 grey 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 Colour task and Colour-Tone task (black and blue traces, respectively). A larger N2pc was measured over the 170-270ms post-cue interval for the Colour task vs. Colour-Tone task. The inset bar graph displays the N2pc mean amplitude (note y-axis is inverted for clarity; asterisk indicates $p<0.05$). D. Group-averaged GFP waveforms are displayed for the Colour task and Colour-Tone task (black and blue traces, respectively). A larger GFP was measured over the N2pc, 170-270ms post-cue interval for the Colour task vs. Colour-Tone task. The inset bar graph displays the mean GFP amplitude in the N2pc time-window (asterisk indicates $p<0.05$). The scatterplot displays the relationship between differences in the two tasks in terms of behavioural 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), while 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 90-270ms post-cue interval are displayed on a flattened montage for both Colour task vs. Colour-Tone task (anterior is upward and left hemiscalp on the left; scale indicated). B. Topographic cluster analysis of the group-averaged data identified three template maps over the 142-260ms post-cue interval that were then fitted to single-subject data. The bar graph displays the average percentage of time (s.e.m. indicated) each of these three template maps best correlated spatially with single-subject data. The significant Map × Task interaction indicates that different patterns of template maps best characterised responses from each task.

Acknowledgements

REFERENCES

Alsius, A., & Soto-Faraco, S. (2011). Searching for audiovisual correspondence in multiple speaker scenarios. *Experimental Brain Research*, *213*, 175–183. https://doi.org/10.1007/s00221-011-2624-0

Assad, J. A., & Maunsell, J. H. R. (1995a). Neural correlates of inferred motion in primate posterior parietal cortex. *Nature*, *373*, 518–521. https://doi.org/10.1038/373518a0

Assad, J. A., & Maunsell, J. H. R. (1995b). 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*(6), 1435–1445. https://doi.org/10.1016/j.neuropsychologia.2010.12.001

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 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*(11), 1190–1192. https://doi.org/10.1038/nn1333

Bevilacqua, D., Davidesco, I., Wan, L., Oostrik, M., Chaloner, K., Rowland, J., … Dikker, S. (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.

https://doi.org/10.1162/jocn_a_01274

Brandeis, D., Lehmann, D., Michel, C. M., & Mingrone, W. (1995). Mapping event-related brain potential microstates to sentence endings. *Brain Topography*, *8*(2), 145–159. https://doi.org/10.1007/BF01199778

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2011/813870

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*(3), 205–228. https://doi.org/10.1080/87565640801982312.Short-Term

Cohen, S. S., & Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *ENeuro*, *3*(December), 1–11. https://doi.org/10.1523/ENEURO.0203-16.2016

Couperus, J. W., & Quirk, C. (2015). Visual search and the N2pc in children. *Attention, Perception, and Psychophysics*, *77*(3), 768–776. https://doi.org/10.3758/s13414-015-0833-5

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*(2), 63–68. https://doi.org/10.1016/j.tine.2013.12.001

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. https://doi.org/10.3389/fnint.2015.00017

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222. Retrieved from http://dx.doi.org/10.1146/ANNUREV.NEURO.18.1.193

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., … Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, *27*(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517. https://doi.org/10.1037/0096-3445.113.4.501

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234. https://doi.org/10.1016/0013-4694(96)95711-9

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*(5), 1316–1328. https://doi.org/10.1037/a0015872

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 Science Publishers,.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., … Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *in press*, 1–26.

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*(3), 567–580. https://doi.org/10.1162/jocn.2009.21061

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*, 347–351. https://doi.org/10.1038/35012613

Gibson, J. A. Y. R., Maunsell, J. H. R., & Island, R. (1997). Sensory Modality Speci city of Neural Activity Related to Memory in Visual Cortex. *Cortex*, 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.

Helmholtz, H. von. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche (Schluss.). *Annalen Der Physik*, *165*(7), 353–377.

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*(12), 1233–1241. https://doi.org/10.1093/cercor/10.12.1233

Hwang, K., Shine, J. M., D'Esposito, M., & Esposito, M. D. (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*(3), 548–554. https://doi.org/10.3758/PBR.15.3.548

Jonas, E., & Kording, K. P. (2017). Could a Neuroscientist Understand a Microprocessor ? *PLoS Computational Biology*, *13*, e1005268. https://doi.org/10.1371/journal.pcbi.1005268

Khanna, A., Pascual-Leone, A., & Farzan, F. (2014). Reliability of resting-state microstate features in electroencephalography. *PLoS ONE*, *9*(12), 1–21. https://doi.org/10.1371/journal.pone.0114163

Kingstone, A. (1992). Combining expectancies. *The Quarterly Journal of Experimental Psychology Section A*, *44*(1), 69–104. https://doi.org/10.1080/14640749208401284

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*(2), 216–228. https://doi.org/10.3758/s13414-012-0391-z

Kitagawa, N., Zampini, M., & Spence, C. (2005). Audiotactile interactions in near and far space. *Experimental Brain Research*, *166*(3–4), 528–537. https://doi.org/10.1007/s00221-005-2393-8

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*(July), 1–10. https://doi.org/10.3389/fnhum.2017.00388

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Kriegstein, K. Von, & Giraud, A. (2006). Implicit Multisensory Associations Influence Voice Recognition. *PLoS Biology*, *4*, e326. https://doi.org/10.1371/journal.pbio.0040326

Kriegstein, K. von, Kleinschmidt, A., Sterzer, P., & Giraud, A.-L. (2005). Interaction of Face and Voice Areas during Speaker Recognition. *Journal of Cognitive Neuroscience*, *17*(3), 367–376. https://doi.org/10.1162/0898929053279577

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y., Field, A. S., & Stein, B. E. (2000). Deactivation of Sensory-Specific Cortex by Cross-Modal Stimuli, 420–429.

Lavie, N., & Cox, S. (1997). On the efficiency of visual selectve attention: Efficient visual search leads to inneficent distractor rejection. *Psychological Science*, *8*(5), 395–398. https://doi.org/10.1111/j.1467-9280.1997.tb00432.x

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*(3), 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. https://doi.org/10.1111/tops.12131

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*(6657), 279–281. https://doi.org/10.1038/36846

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*, *108*(51), 20661–20665. https://doi.org/10.1073/pnas.1112605108

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*(3), 397–405. https://doi.org/10.1016/j.neuron.2009.07.012

Maidenbaum, S., Abboud, S., & Amedi, A. (2014). Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2013.11.007

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*, 4–18. https://doi.org/10.1111/j.1469-8986.1995.tb03400.x

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*(6), 532–550. https://doi.org/10.3758/BF03203106

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. https://doi.org/10.1111/ejn.12804

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. https://doi.org/10.1016/j.cognition.2014.11.031

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2018). Are we ready for real-world neuroscience? *Journal of Cognitive Neuroscience*, (in press).

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. In *Perception 40 ECVP 2011 Abstract Supplement* (p. 184). Retrieved from http://www.perceptionweb.com/abstract.cgi?id=v110603

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search.

Psychonomic Bulletin & Review, *18*(5), 904–909. https://doi.org/10.3758/s13423-011-0131-8

Matusz, P. J., & Eimer, M. (2013). Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*, *50*, 996–1009. https://doi.org/https://doi.org/10.1111/psyp.12086

Matusz, P. J., Merkley, R., Faure, M., & Scerif, G. (2018). Expert Attention: Attentional allocation depends on the differential development of multisensory number representations. *PsyArXiv*, *16 June*.

Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*, *105*, 243–252. https://doi.org/10.1016/j.neuropsychologia.2017.04.008

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*(April), 14–20. https://doi.org/10.1016/j.cobeha.2016.04.006

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*(JAN), 1–14. https://doi.org/10.3389/fpsyg.2015.01903

Michel, C. M., & Koenig, T. (2017). NeuroImage EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks : A review. *NeuroImage*, *in press*, 1–17. https://doi.org/10.1016/j.neuroimage.2017.11.062

Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *NeuroImage*, *61*(2), 371–385. https://doi.org/10.1016/j.neuroimage.2011.12.039

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*(4715), 782–784. https://doi.org/10.1126/science.4023713

Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., & Laurienti, P. J. (2008). Cross-modal deactivations during modality-specific selective attention. *BMC Neurology*, *8*, 1–11. https://doi.org/10.1186/1471-2377-8-35

Müller, V., Sänger, J., & Lindenberger, U. (2018). Hyperbrain network properties of guitarists playing in quartet. *Annals of the New York Academy of Sciences*, *1423*, 198–210. https://doi.org/10.1111/nyas.13656

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*. https://doi.org/10.1007/s10548-008-0054-5

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory Processes: A Balancing Act across the Lifespan. *Trends in Neurosciences*, *39*(8), 567–579. https://doi.org/10.1016/j.tins.2016.05.003

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.

https://doi.org/10.1016/j.neuropsychologia.2015.08.011

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*(6), 1654–1665. https://doi.org/10.3758/APP

Parks, E. L., & Madden, D. J. (2013). Brain Connectivity and Visual Attention. *Brain Connectivity*, *3*(4), 317–338. https://doi.org/10.1089/brain.2012.0139

Peelen, M. V, & Kastner, S. (2014). Attention in the real worls: Toward understanding its neural basis. *Trends in Cognitive Sciences*, *18*(5), 242–250. https://doi.org/10.1016/j.tics.2014.02.004

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

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*(1), 104–121. https://doi.org/10.1016/j.jecp.2013.12.009

Putzar, L., Goerendt, I., Lange, K., Rösler, F., & Röder, B. (2007). Early visual deprivation impairs multisensory interactions in humans. *Nature Neuroscience*, *10*(10), 1243–1245. https://doi.org/10.1038/nn1978

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*(1), 86–95. https://doi.org/10.1097/WCO.0b013e32834ed723

Rohe, T., & Noppeney, U. (2018). Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Control. *ENeuro*, *5*, ENEURO.0315. https://doi.org/10.1523/ENEURO.0315-17.2018

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1311–1321. https://doi.org/10.1037/0096-1523.33.6.1311

Sarmiento, B. R., Matusz, P. J., Sanabria, D., & Murray, M. M. (2016). Contextual factors multiplex to control multisensory processes. *Human Brain Mapping*, *37*(1). https://doi.org/10.1002/hbm.23030
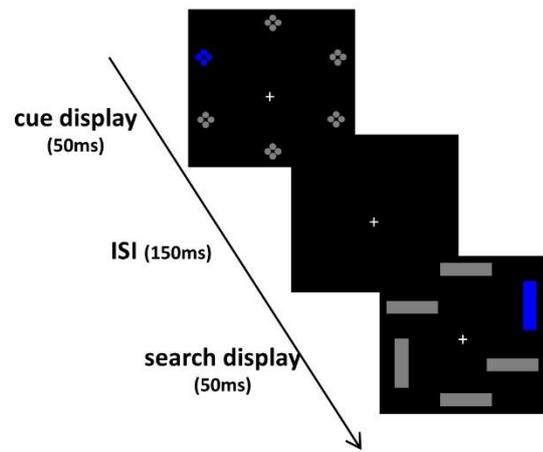
Scerif, G., Longhi, E., Cole, V., Karmiloff-Smith, A., & Cornish, K. (2012). Attention across modalities as a longitudinal predictor of early outcomes: The case of fragile X syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(6), 641–650. https://doi.org/10.1111/j.1469-7610.2011.02515.x

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*(3), 445—463. https://doi.org/10.1002/cne.903430308

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cerebral Cortex*, *14*(12), 1346–1357. https://doi.org/10.1093/cercor/bhh095

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Shulman, G. L., D'Avossa, G., Tansy, A. P., & Corbetta, M. (2002). Two attentional processes in the parietal lobe. *Cerebral Cortex*, *12*(11), 1124–1131. https://doi.org/10.1093/cercor/12.11.1124

Stein, B. (2012). *The New Handbook of Multisensory Processing*. Massachusetts (MA), USA: 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. https://doi.org/10.1038/nrn2331

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 of the United States of America*, *103*(21), 8239–8244. https://doi.org/10.1073/pnas.0509704103

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*(5). https://doi.org/10.1007/s00221-016-4590-z

Thelen, A., Matusz, P. J., & Murray, M. M. (2014). Multisensory context portends object memory. *Current Biology*, *24*(16), R734–R735. https://doi.org/10.1016/j.cub.2014.06.040

Tivadar, R. I., & Murray, M. M. (2018). A primer on electroencephalography and event-related potentials for organizational neuroscience. *Organizational Research Methods*, *22*(1).

Townsend, J. T., & Ashby, F. G. (1978). Methods of Modeling Capacity in Simple Processing Systems. *Cognitive Theory (Volume 3)*.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

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), 1–15. https://doi.org/10.1038/s41598-018-22679-x

van Atteveldt, N. M., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory Integration: Flexible Use of General Operations. *Neuron*, *81*(6), 1240–1253. https://doi.org/10.1016/J.NEURON.2014.02.044

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. https://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2010.12.068

Wojciulik, E., Kanwisher, N., & Driver, J. (2009). Ewa Wojciulik, Nancy Kanwisher and Jon Driver. *Journal of*

Neurophysiology, 1574–1578.

A. Trial sequence

cue display
(50ms)

ISI (150ms)

search display
(50ms)

B. Search task conditions

Colour task

Colour-Tone task

V+

V+A+

V-

V-A+

V+A-

V-A-

Figure 1

1 Canonical N2pc analyses

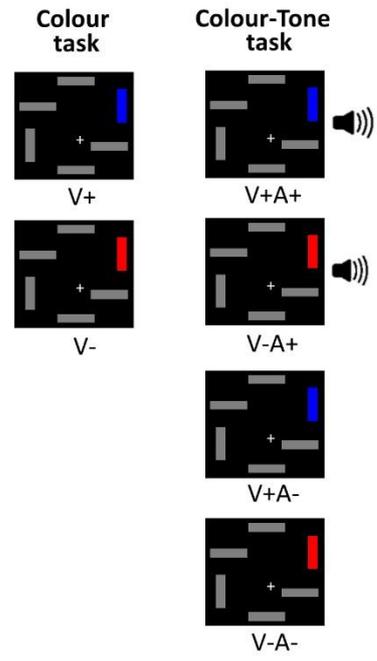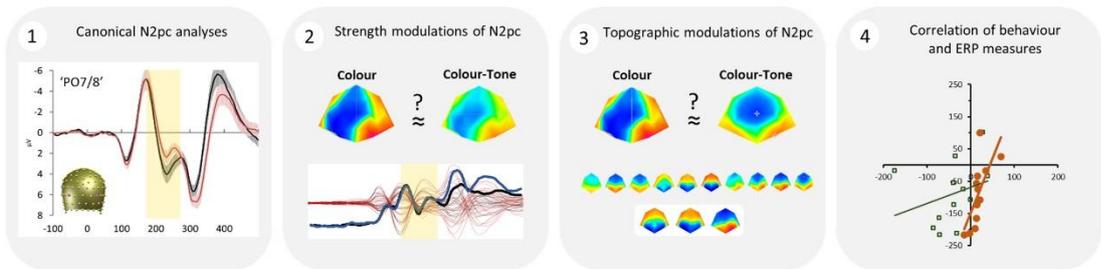**2** Strength modulations of N2pc
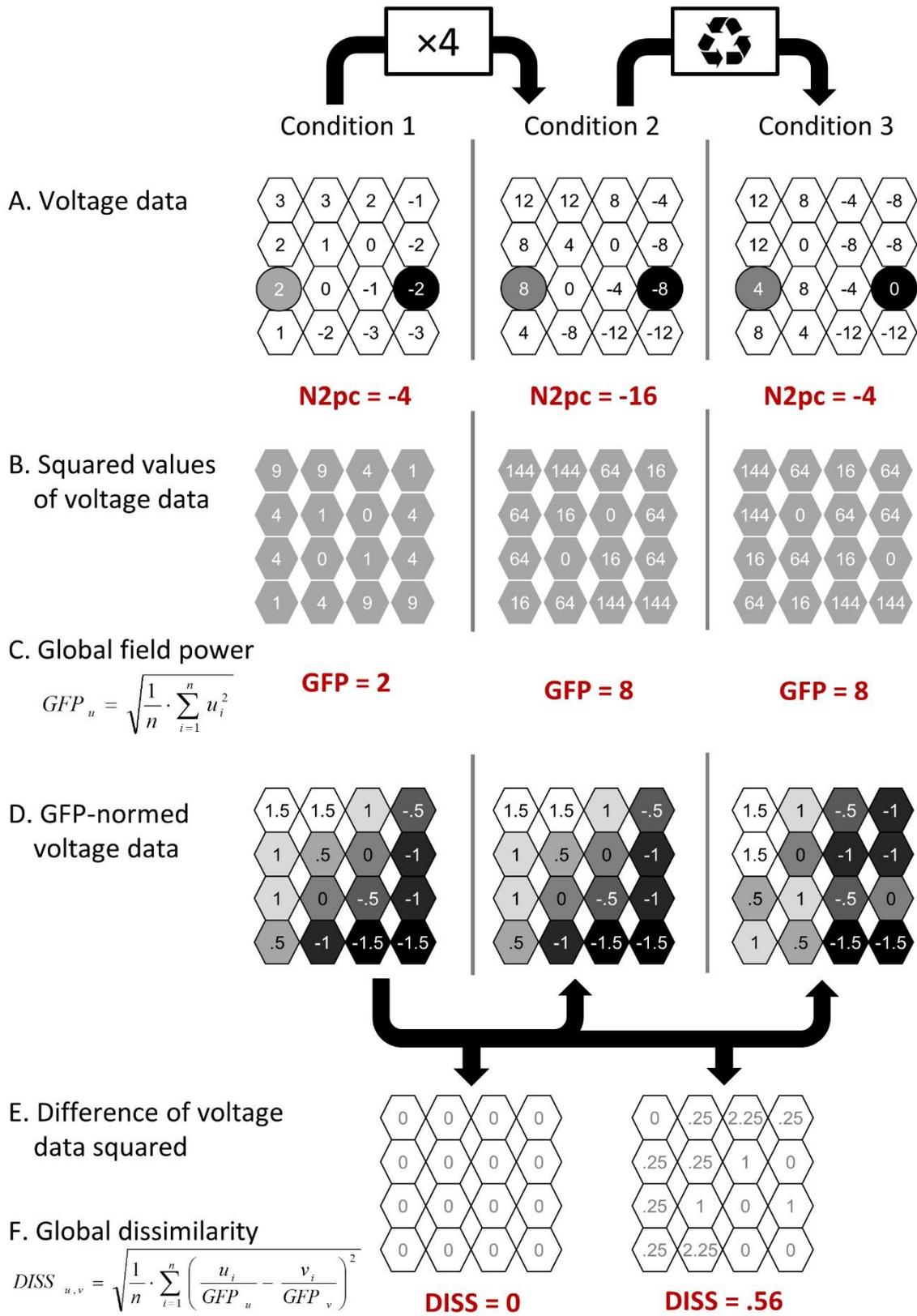
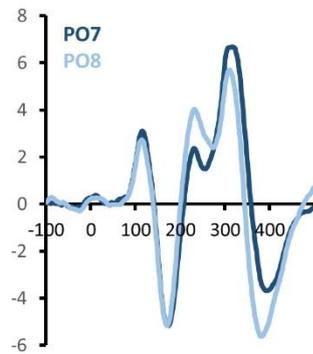**3** Topographic modulations of N2pc

4 Correlation of behaviour and ERP measures
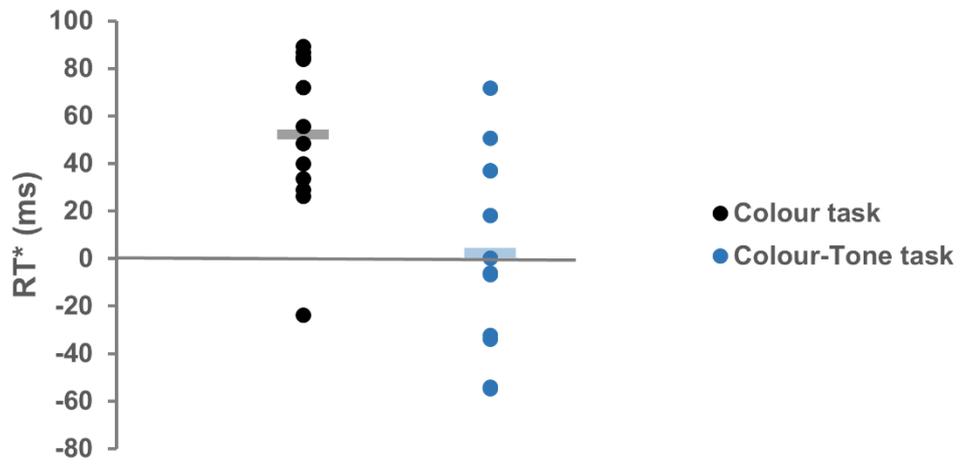
Figure 2

Figure 3

Figure 4

A. Colour Task
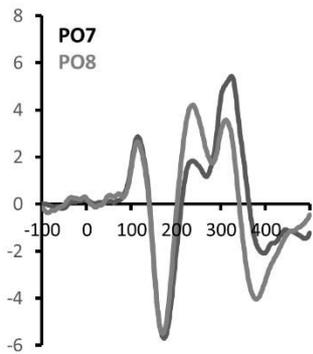
B. Colour-Tone Task

C. PO7/PO8 Difference

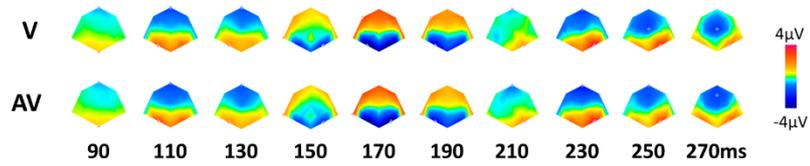N2pc mean amplitude

D. Global Field Power

% Behavioural difference

% ERP difference

GFP mean amplitude
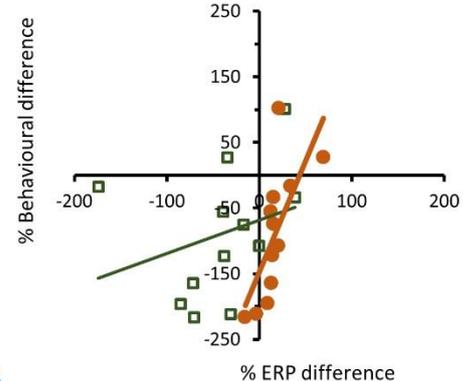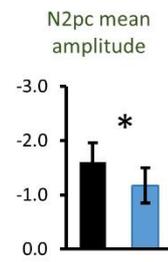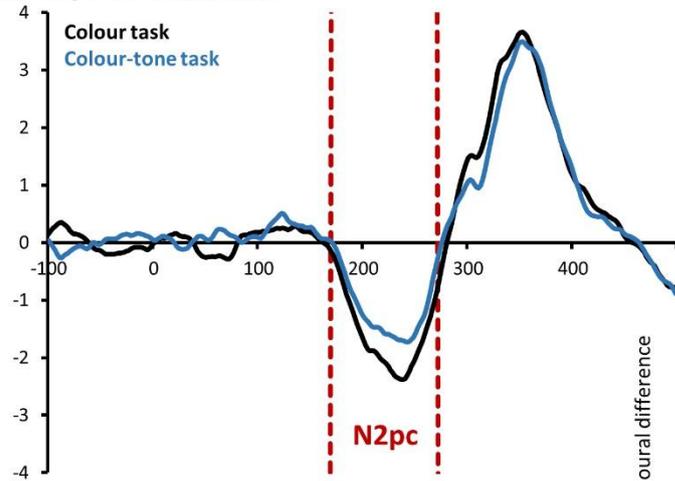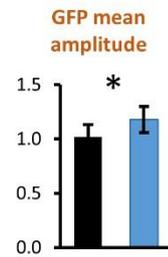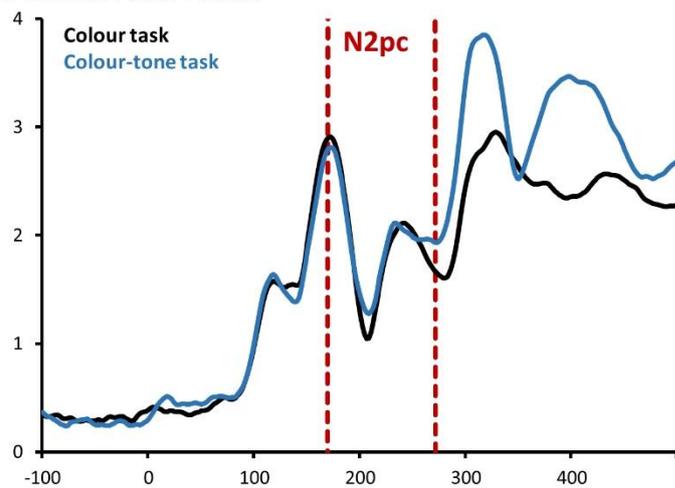
Figure 5

A. Group-Averaged ERP topographies

Hierarchical Topographic Clustering
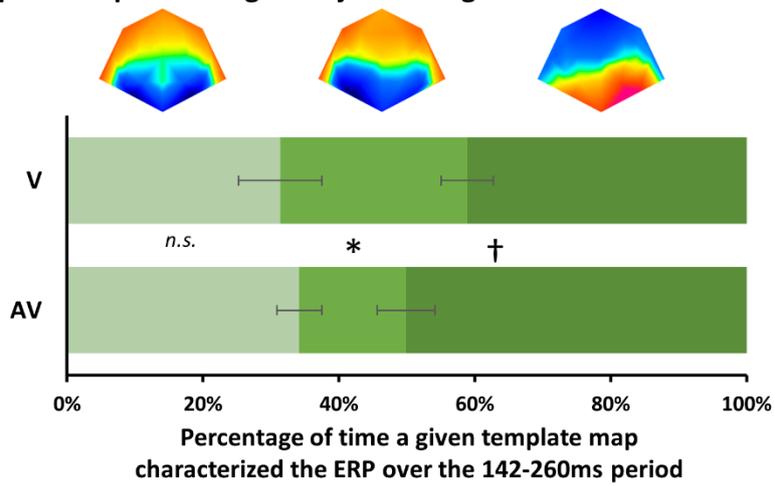
B. Template Maps and Single-subject Fitting

Percentage of time a given template map characterized the ERP over the 142-260ms period

Figure 6