NeuroImage 244 (2021) 118556

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Towards understanding how we pay attention in naturalistic visual search settings

Nora Turoman[a,b,c], Ruxandra I. Tivadar[a,d,e], Chrysa Retsa[a,f], Micah M. Murray[a,d,f,g], Pawel J. Matusz[a,b,g,*]

[a] The LINE (Laboratory for Investigative Neurophysiology), Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[b] MEDGIFT Lab, Institute of Information Systems, School of Management, HES-SO Valais-Wallis University of Applied Sciences and Arts Western Switzerland, Techno-Pôle 3, 3960 Sierre, Switzerland
[c] Working Memory, Cognition and Development lab, Department of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland
[d] Department of Ophthalmology, Fondation Asile des Aveugles, Lausanne, Switzerland
[e] Cognitive Computational Neuroscience group, Institute of Computer Science, Faculty of Science, University of Bern, Switzerland
[f] CIBM Center for Biomedical Imaging, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[g] Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Keywords:
Attentional control
Multisensory
Real-world
Semantic congruence
Temporal predictability
Context

ABSTRACT

Research on attentional control has largely focused on single senses and the importance of behavioural goals in controlling attention. However, everyday situations are multisensory and contain regularities, both likely influencing attention. We investigated how visual attentional capture is simultaneously impacted by top-down goals, the multisensory nature of stimuli, *and* the contextual factors of stimuli's semantic relationship and temporal predictability. Participants performed a multisensory version of the Folk et al. (1992) spatial cueing paradigm, searching for a target of a predefined colour (e.g. a red bar) within an array preceded by a distractor. We manipulated: 1) stimuli's goal-relevance via distractor's colour (matching vs. mismatching the target), 2) stimuli's multisensory nature (colour distractors appearing alone vs. with tones), 3) the relationship between the distractor sound and colour (arbitrary vs. semantically congruent) and 4) the temporal predictability of distractor onset. Reaction-time spatial cueing served as a behavioural measure of attentional selection. We also recorded 129-channel event-related potentials (ERPs), analysing the distractor-elicited N2pc component both canonically and using a multivariate electrical neuroimaging framework. Behaviourally, arbitrary target-matching distractors captured attention more strongly than semantically congruent ones, with no evidence for context modulating multisensory enhancements of capture. Notably, electrical neuroimaging of surface-level EEG analyses revealed context-based influences on attention to both visual and multisensory distractors, in how strongly they activated the brain and type of activated brain networks. For both processes, the context-driven brain response modulations occurred long before the N2pc time-window, with topographic (network-based) modulations at ~30 ms, followed by strength-based modulations at ~100 ms post-distractor onset. Our results reveal that both stimulus meaning and predictability modulate attentional selection, and they interact while doing so. Meaning, in addition to temporal predictability, is thus a second source of contextual information facilitating goal-directed behaviour. More broadly, in everyday situations, attention is controlled by an interplay between one's goals, stimuli's perceptual salience, meaning and predictability. Our study calls for a revision of attentional control theories to account for the role of contextual and multisensory control.

1. Introduction

Goal-directed behaviour depends on the ability to allocate processing resources towards stimuli important to current behavioural goals ("attentional control"). On the one hand, our current knowledge about attentional control may be limited to the rigorous, yet artificial, conditions in which it is traditionally studied. On the other hand, findings from studies assessing attentional control with naturalistic stimuli (audiostories, films) may be limited by confounds from other processes present in such settings. Here, we systematically tested how traditionally studied goal- and salience-based attentional control interact with more naturalistic, context-based control mechanisms.

* Corresponding author at: Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO Valais), Rue Technopole 3, 3960 Sierre, Switzerland.

https://doi.org/10.1016/j.neuroimage.2021.118556.
Received 30 July 2020; Received in revised form 31 August 2021; Accepted 3 September 2021
Available online 4 September 2021.

In the real world, the location of goal-relevant information is rarely known in advance. Since the pioneering visual search paradigm (Treisman and Gelade, 1980), we have known that in multi-stimulus settings, target attributes can be used to control attention. Here, research provided conflicting results as to whether primacy in controlling attentional selection lies in the task-relevance of objects' attributes (Folk et al., 1992) or their bottom-up salience (e.g. Theeuwes, 1991). Folk et al. (1992) used a version of the spatial cueing paradigm and revealed that attentional capture is elicited only by distractors that matched the target colour. Consequently, they proposed the 'task-set contingent attentional capture' hypothesis, whereby salient objects will capture attention only if they share features with the target and are thus potentially task-relevant. However, subsequently, mechanisms beyond goal-relevance were shown to serve as additional sources of attentional control, such as those based on spatiotemporal and semantic information within the stimulus and the environment where it appears (e.g., Chun and Jiang 1998; Peelen and Kastner, 2014; Summerfield et al., 2006; van Moorselaar and Slagter 2019; Press et al., 2020), and multisensory processes (Matusz and Eimer, 2011, 2013; Matusz et al., 2015a; Lunn et al., 2019; Soto-Faraco et al., 2019).

Some multisensory processes occur at early latencies (<100 ms post-stimulus), generated within primary cortices (e.g., Talsma and Woldroff, 2005; Raij et al., 2010; Cappe et al., 2010; reviewed in de Meo et al., 2015; Murray et al., 2016a). This enables multisensory processes to influence attentional selection in a bottom-up fashion, potentially independently of the observer's goals. This idea was supported by Matusz and Eimer (2011) who used a multisensory adaptation of Folk et al. (1992) task. The authors replicated the task-set contingent attentional capture effect and showed that visual distractors captured attention more strongly when accompanied by a sound, regardless of their goal-relevance. This demonstrated the importance of bottom-up multisensory enhancement for attentional selection of visual objects. However, interactions between such goals, multisensory influences on attentional control, and the stimuli's temporal and semantic context[1] remain unknown.

1.1. Top-down contextual factors in attentional control

The temporal structure of the environment is routinely used by the brain to build predictions. Attentional control uses such predictions to improve the selection of target stimuli (e.g., Correa et al., 2005; Coull et al., 2000; Green and McDonald, 2010; Miniussi et al., 1999; Naccache et al., 2002; Rohenkohl et al., 2014; Tivadar et al., 2021) and the inhibition of task-irrelevant stimuli (here, location- and feature-based predictions have been more researched than temporal predictions; e.g., reviewed in Noonan et al., 2018; van Moorselaar and Slagter 2020a). In naturalistic, multisensory settings, temporal predictions are known to improve language comprehension (e.g. Luo and Poeppel, 2007; ten Oever and Sack, 2015), yet their role as a source of attentional control is less known (albeit see, Zion Golumbic et al., 2012, for their role in the "cocktail party" effect). Semantic relationships are another basic principle of organising information in real-world contexts. Compared to semantically incongruent or meaningless (arbitrary) multisensory stimuli, semantically congruent stimuli are more easily identified and remembered (e.g.(Laurienti et al., 2004) Murray et al., 2004; Doehrmann and Naumer 2008; Chen and Spence, 2010; Matusz et al., 2015a; Tovar et al., 2020; reviewed in ten Oever et al. 2016; Murray et al., 2016b; (Matusz et al., 2017)Matusz et al., 2020) and also, more strongly attended (Matusz et al., 2015b,(Matusz et al., 2019c) , 2019b; reviewed in Soto-Faraco et al., 2019;(Matusz et al., 2019a)). For example, (Iordanenscu et al., 2008)demonstrated that search for naturalistic objects is faster when accompanied by irrelevant albeit congruent sounds.

What is unclear from existing research is the degree to which goal-based attentional control interacts with salience-driven (multisensory) mechanisms *and* such contextual factors. Researchers have been clarifying such interactions, but typically in a pair-wise fashion, between e.g., attention and semantic memory, or attention and predictions (reviewed in Summerfield and Egner 2009; (Gazzaley and Nobre, 2012); Press et al., 2020). However, in everyday situations these processes do not interact in an orthogonal, but, rather, a synergistic fashion, with multiple sources of control interacting simultaneously (ten Oever et al. 2016; Nastase et al., 2020). Additionally, in the real world, these processes operate on both unisensory and multisensory stimuli, where the latter are often more perceptually salient than the former (e.g., (Santangelo and Spence, 2007)Matusz and Eimer 2011). Thus, one way to create more complete and "naturalistic" theories of attentional control is by investigating how one's goals interact with *multiple* contextual factors in controlling attentional selection – and doing so in *multi-sensory* settings.

1.2. The present study

To shed light on how attentional control operates in naturalistic visual search settings, we investigated how visual and multisensory attentional control processes interact with distractors' temporal predictability and multisensory semantic relationship when all are manipulated simultaneously. We likewise set out to identify brain mechanisms supporting such complex interactions. To address these questions in a rigorous and state-of-the-art fashion, we employed a 'naturalistic laboratory' approach that builds on several methodological advances ((Matusz et al., 2019a)). First, we used a paradigm that isolates a specific cognitive process, i.e., Matusz and Eimer's (2011) multisensory adaptation of the Folk et al. (1992) task, where we now additionally manipulated distractors' temporal predictability and relationship between their auditory and visual features. In Folk et al.'s task, attentional control is measured via well-understood spatial cueing effects, where larger effects (e.g., for target-colour and audiovisual distractors) reflect stronger attentional capture. Notably, distractor-related responses have added value here as they isolate attentional from later, motor response-related, processes. Second, we measured a well-researched brain correlate of attentional object selection, the N2pc event-related potential (ERP) component. The N2pc is a negative-going voltage deflection starting at around 200 ms post-stimulus onset at posterior electrode sites contralateral to stimulus location (Luck and Hillyard, 1994a,b; Eimer, 1996; Girelli and Luck, 1997). Studies canonically analysing N2pc have provided strong evidence for task-set contingence of attentional capture (e.g., Kiss et al., 2008a,b; Eimer et al., 2009). Importantly, the N2pc is also sensitive to meaning (e.g., Wu et al., 2015) and predictions (e.g., Burra and Kerzel, 2013), whereas its sensitivity to multisensory enhancement is limited (van der Burg et al. 2011, but see below). This joint evidence makes the N2pc a valuable 'starting point' for investigating interactions between visual goals and more naturalistic sources of control. Third, analysing the traditional EEG markers of attention with advanced frameworks like electrical neuroimaging (e.g., Lehmann and Skrandies 1980; Murray et al., 2008; Tivadar and Murray 2019) might offer an especially robust, accurate and informative approach.

Briefly, an electrical neuroimaging framework encompasses multivariate, reference-independent analyses of global features of the electric scalp field. Its main added value is that it readily distinguishes the neurophysiological mechanisms that drive differences in ERPs across experimental conditions in *surface-level* EEG: 1) "gain control" mechanisms, modulating the strength of activation within an indistinguishable brain network, and 2) topographic (network-based) mechanisms, modulating the brain sources recruited for response (scalp EEG topography differences forcibly follow from changes in the underlying sources;

[1] Context has been previously defined as the "immediate situation in which the brain operates… shaped by external circumstances, such as properties of sensory events, and internal factors, such as behavioural goal, motor plan, and past experiences" (van Atteveldt et al., 2014).

Murray et al., 2008). Electrical neuroimaging overcomes interpretational limitations of canonical N2pc analyses. Most notably, a difference in mean N2pc amplitude can arise from both strength-based and topographic mechanisms (albeit it is assumed to signify gain control); it can also emerge from different brain source configurations (for a full discussion, see Matusz et al., 2019b).

We recently used this approach to better understand brain and cognitive mechanisms of attentional control. We revealed that distinct brain networks are active over the ~N2pc time-window during visual goal-based *and* multisensory bottom-up attention control (across the lifespan; Turoman et al., 2021a,b). However, these reflected spatially-selective, lateralised brain mechanisms, partly captured by the N2pc (via the contra- and ipsilateral comparison). There is little existing evidence to strongly predict how interactions between goals, stimulus salience and context can occur in the brain. Schröger et al. (2015) proposed that temporally unpredictable events attract attention more strongly (to serve as a signal to reconfigure the predictive model about the world), visible in larger behavioural responses and ERP amplitudes. Both predictions and semantic memory could be used to reduce attention to known (i.e., less informative) stimuli. Indeed, goal-based attention uses knowledge to facilitate visual and multisensory processing ((Summerfield et al., 2006) Iordanescu et al., 2008; Matusz et al., 2016; Sarmiento et al., 2016). However, several questions remain. Does knowledge affect attention to task-*irrelevant* stimuli the same way? Finally, how early do contextual factors influence stimulus processing here, if both processes are known to do so <150 ms post-stimulus (Summerfield and Egner, 2009; ten Oever et al. 2016). Do contextual processes operate through lateralised or non-lateralised brain mechanisms, and are they strenth-based and/or topographic in nature? Below we specify our hypotheses.

We expected to replicate the task-set contingent attentional capture (or TAC[2]) effect: In behaviour, visible as large behavioural capture for target-colour matching distractors and no capture for nontarget-colour matching distractors (e.g., Folk et al., 1992; Folk et al., 2002; Lien et al., 2008); in canonical EEG analyses visible as enhanced N2pc amplitudes for target-colour distractors over nontarget-colour distractors (Eimer et al., 2009). TAC should be modulated by both contextual factors: the predictability of distractor onset and the multisensory relationship between distractor features (semantic congruence vs. arbitrary pairing; Wu et al., 2015; Burra and Kerzel, 2013). However, as discussed above, we had no strong predictions how the contextual factors would modulate TAC (or if they interact while doing so), as these effects have never been tested systematically together, on audio-visual and task-irrelevant stimuli. For multisensory enhancement of attentional capture (MSE), we expected to replicate it behaviourally (Matusz and Eimer 2011), but without strong predictions about concomitant N2pc modulations (c.f. van der Burg et al. 2011). We expected MSE to be modulated by contextual factors, especially by multisensory relationship, based on the extensive literature on the role of semantic congruence in multisensory cognition (Doehrmann and Naumer, 2008; ten Oever et al. 2016). Again, we had no strong predictions as to the directionality of these modulations or interaction of their influences.

At the level of the brain, were primarily interested in whether interactions between visual goals (TAC), multisensory salience (MSE) and contextual processes are supported by strength-based (i.e., “gain”-like; i.e., one network is active more strongly for some and less strongly for other experimental conditions) and/or topographic (i.e., different networks are activated for different experimental conditions) brain mechanisms, as observable in *surface-level* EEG data when using multivariate analyses like electrical neuroimaging. The second aim of our study was to clarify if attentional and contextual control interactions are supported by lateralised (N2pc-like) or nonlateralised mechanisms. To this aim, we analysed if such interactions are captured by canonical N2pc analyses or electrical neuroimaging analyses of the lateralised distractor-elicited ERPs ~180–300 ms post-stimulus (N2pc-like time-window). These analyses would reveal the presence of strength- and topographic *spatially-selective* brain mechanisms contributing to attentional control. However, canonical analyses of the N2pc assume not only lateralised activity, but also symmetry; in brain anatomy, but also in scalp electrodes, detecting homologous brain activity over both hemispheres. This may prevent them from detecting other, less-strongly-lateralised brain mechanisms of attentional control. We have previously found nonlateralised mechanisms to play a role in attentional control in multisensory settings (Matusz et al., 2019b). Also, semantic information and temporal expectations (and feature-based attention) are known to modulate nonlateralised ERPs ((Saenz et al., 2003) ’ (Dell’Acqua et al., 2010) Dassanayake et al., 2016). Thus, as the third aim of our study, we investigated whether contextual control affects stages associated with attentional selection (reflected by the N2pc) or also earlier processing stages. We tested this by measuring strength- and/or topographic nonlateralised brain mechanisms across the whole post-stimulus time-period activity.

2. Materials and methods

2.1. Participants

Thirty-nine adult volunteers participated in the study (5 left-handed, 14 males, M_{age}: 27.5 years, *SD:* 4 years, range: 22–38 years). We conducted post-hoc power analyses for the two effects that have been previously behaviourally studied with the present paradigm, namely TAC and MSE. Based on the effect sizes in the original study of Matusz and Eimer (2011, Exp.2), the analyses revealed sufficient statistical power for both behavioural effects with the collected sample. For ERP analyses, we could calculate power analyses only for the TAC effect. Based on a purely visual ERP study (Eimer et al., 2009) we revealed there to be sufficient statistical power to detect TAC in the N2pc in the current study (all power calculations are available in the Supplementary Online Materials, SOMs). Participants had normal or corrected-to-normal vision and normal hearing and reported no prior or current neurological or psychiatric disorders. Participants provided informed consent before the start of the testing session. All research procedures were approved by the Cantonal Commission for the Ethics of Human Research (CER-VD; no. 2018-00241).

2.2. Task properties and procedures

General task procedures. The full experimental session consisted of participants completing four experimental Tasks. All the Tasks were close adaptations of the original paradigm of Matusz and Eimer (2011 Exp.2; that is, in turn, an adaptation of the spatial-cueing task of Folk et al. [1992]). Across all the Tasks, the instructions and the overall experimental set up were the same as in the study of Matusz & Eimer (1992, Exp.2; see Fig. 1A). Namely, participants searched for a target of a predefined colour (e.g., a red bar) in a 4-element array, and assessed the target’s orientation (vertical vs. horizontal). Furthermore, in all Tasks, the search array was always preceded by a distractor array containing colour elements. On each trial, one of those elements (distractor, "cue") always changed colour, to match either the target colour (red set of dots) or another, nontarget colour (blue set of dots). On 50% of all trials the colour distractors would be accompanied by a sound (audiovisual distractor condition). The distractor appeared in each of the four stimulus locations with equal probability (25%) and was thus not predictive of the location of the incoming target. Differences in response speed on trials where distractor and target appeared in the same vs. different locations were used to calculate behavioural cueing effects that were the basis of our analyses (see below). Like in the Matusz and Eimer (2011) study, across all Tasks, each trial consisted of the following sequence of arrays: base array (duration manipulated; see below), followed by distractor array (50 ms duration), followed by a fixation

[2] Please see Appendix 1 for the full list of abbreviations used in the manuscript.

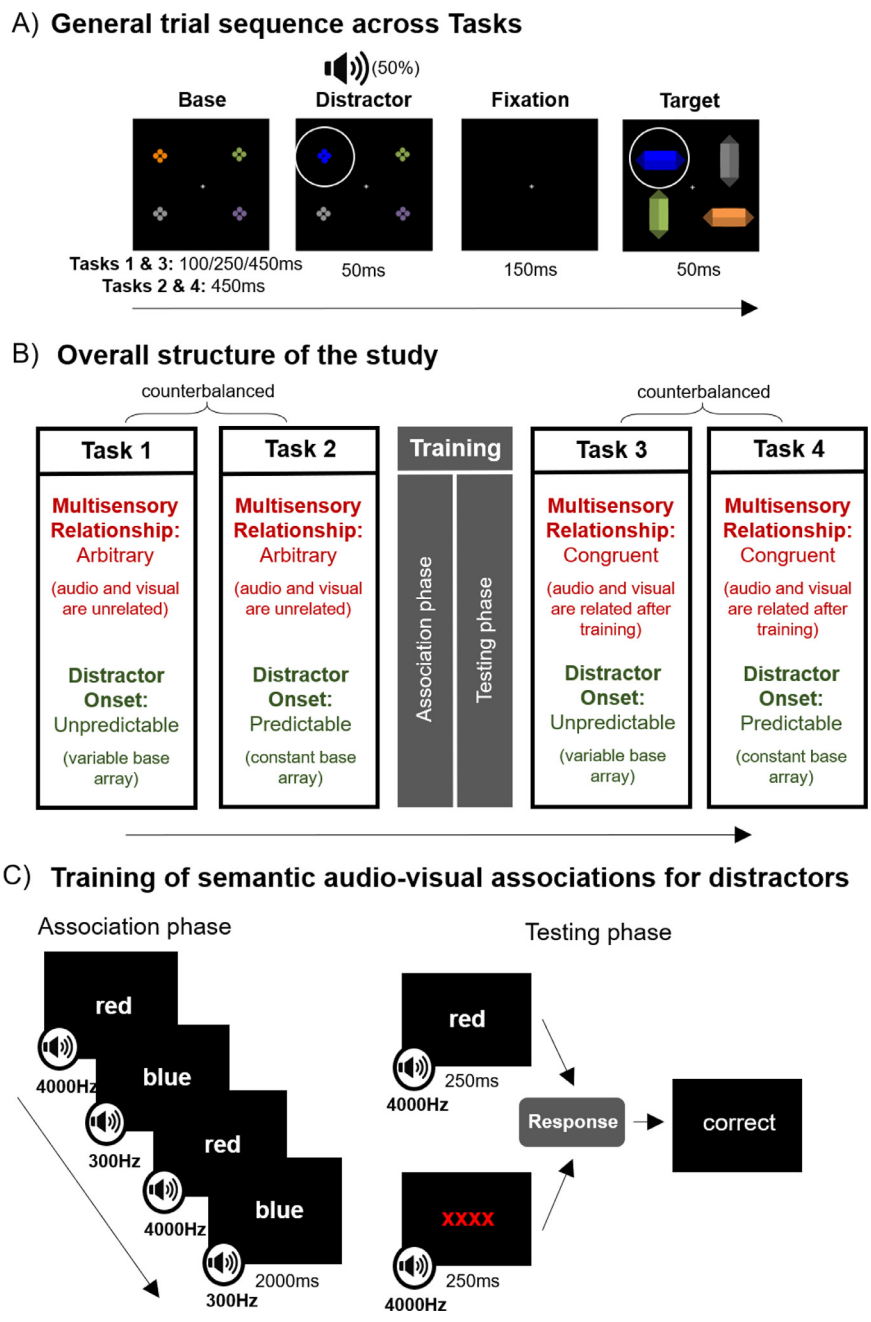

Fig. 1. A) An example trial of the general experimental "Task" is shown, with four successive arrays. The white circle around the target location (here the target is a blue diamond) and the corresponding distractor location serves to highlight, in this case, a target-matching distractor ("cue") colour condition, with a concomitant sound, i.e., TCCAV. B) The order of Tasks, with the corresponding conditions of Multisensory Relationship in red, and Distractor Onset in green, shown separately for each Task, in the successive order in which they appeared in the study. Under each condition, its operationalisation is given in brackets in the corresponding colour. Predictable and unpredictable blocks before and after the training (1 & 2 and 3 & 4, respectively) were counterbalanced across participants. C) Events that were part of the Training. *Association phase:* an example pairing option (red – high pitch, blue – low pitch) with trial progression is shown. *Testing phase:* the pairing learnt in the Association phase would be tested using a colour word or a string of x's in the respective colour. Participants had to indicate whether the pairing was correct via a button press, after which feedback was given. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

point (150 ms duration), and finally the target array (50 ms duration, see Fig. 1A).

The differences to the original study involved the changes necessary to implement the two new, contextual factors that were manipulated across the four Tasks (Fig. 1B).[3] To implement the *Multisensory Relationship* factor, after the first two Tasks, participants completed a training session (henceforth *Training*), after which they completed the remaining two Tasks. To implement the *Distractor Onset* factor, the predictability of the onset of the distractors was manipulated, being either stable (as in the original study, Tasks 2 and 4) or varying between three durations (Tasks 1 and 3). The setup involving four consecutive Tasks separated by the Training allowed a systematic comparison between the four levels of the two contextual factors. We now describe in more detail the procedures related to all Tasks, after which we provide more details on the different tasks themselves.

The base array contained four differently coloured sets of closely aligned dots, each dot subtending $0.1^\circ \times 0.1^\circ$ of visual angle. The sets of dots were spread equidistally along the circumference of an imaginary circle against a black background, at an angular distance of 2.1° from a central fixation point. Each set could be of one of four possible colours (according to the RGB scale): green (0/179/0), pink (168/51/166), gold (150/134/10), silver (136/136/132). In the distractor array, one of the base array elements changed colour to either a target-matching colour, or a target-nonmatching colour that was not present in any of the elements before. The remaining three distractor array elements did not change their colour. The distractors and the subsequent target diamonds could have either a blue (RGB values: 31/118/220) or red (RGB values:

[3] Compared to the original paradigm, we made two additional changes, to enable the Task 1 to serve as an adult control study in a developmental study (Turoman et al., 2021a). We reduced the number of elements in all arrays from 6 to 4, and targets were reshaped to look like diamonds rather than rectangles. Notably, despite these changes, we have replicated here the visual and multisensory attentional control effects.

224/71/52) colour. The target array contained four bars (rectangles), where one was always the colour-defined target. The target colour was counterbalanced across participants. Target orientation (horizontal or vertical) was randomly determined on each trial. The two distractor colours were randomly selected with equal probability before each trial, and the location of the colour change distractor was not spatially predictive of the subsequent target location (distractor and target location were the same on 25% of trials). On half of all trials, distractor onset coincided with the onset of a pure sine-wave tone, presented from two loudspeakers on the left and right sides of the monitor. Sound intensity was 80 dB SPL (as in Matusz and Eimer, 2011), measured using an audiometer placed at a position adjacent to participants' ears (CESVA SC160). Through manipulations of the in-/congruence between distractor and target colour and of the presence/absence of sound during distractor presentations, there were four types of distractors, across all the Tasks: visual distractors that matched the target colour (TCCV, short for *target-colour cue visual*), visual distractors that did not match the target colour (NCCV, *nontarget-colour cue visual*), audiovisual distractors that matched the target colour (TCCAV, *target-colour cue audiovisual*), and audiovisual distractors that did not match the target colour (NCCAV, *nontarget-colour cue, audiovisual*). We kept the term "cue" in the abbreviations of distractor conditions to keep them consistent with previous literature.

The experimental session consisted of 4 Tasks, each spanning 8 blocks of 64 trials. This resulted in 2048 trials in total (512 trials per Task). Participants were told to respond as quickly and accurately as possible to the targets' orientation by pressing one of two horizontally aligned round buttons (Lib Switch, Liberator Ltd.) that were fixed onto a tray bag on the participants' lap. If participants did not respond within 5000 ms of the target onset, next trial was initiated; otherwise the next trial was initiated immediately after the button press. Feedback on accuracy was given after each block, followed by a progress screen (*a treasure map*), which informed participants of the number of remaining blocks and during which participants could take a break. Breaks were also taken between Tasks, and before and after the Training. As a pilot study revealed sufficient proficiency at conducting the tasks after a few trials (over 50% accuracy), participants did not practice doing the Tasks before administration unless they had trouble following the task instructions. The experimental session took place in a dimly lit, sound-attenuated room, with participants seated at 90 cm from a 23″ LCD monitor with a resolution of 1080 × 1024 (60-Hz refresh rate, HP EliteDisplay E232). All visual elements were approximately equiluminant (~20 cd/m^2), as determined by a luxmeter placed at a position close to the screen, measuring the luminance of the screen filled with each respective element's colour. The averages of three measurement values per colour were averaged across colours and transformed from lux to cd/m^2 to facilitate comparison with the results of Matusz and Eimer (2011). The experimental session lasted <3 h in total, including an initial explanation and obtaining consent, EEG setup, administration of Tasks and Training, and breaks.

We now describe the details of the Tasks and Training, which occurred always in the same general order: Tasks 1 and 2, followed by the Training, followed by Tasks 3 and 4 (the order of Tasks 1 and 2 and, separately, the order of Tasks 3 and 4, was counterbalanced across participants). Differences across the four Tasks served to manipulate the two contextual factors (illustrated in Fig. 1B). The factor *Multisensory Relationship* represented the relation between the visual (the colour of the distractor) and the auditory (the accompanying sound) component stimuli that made up the distractors. These two stimuli could be related just by their simultaneous presentation (Arbitrary condition) or by additionally sharing meaning (Congruent condition). The factor *Distractor Onset* represented the temporal predictability of the distractors, i.e., whether their onset was constant within Tasks (Predictable condition), or variable (Unpredictable condition). The manipulation of the two context factors led to the creation of four contexts, represented by each of the Tasks 1–4 (i.e., Arbitrary Unpredictable, Arbitrary Predictable, Congruent Unpredictable, and Congruent Predictable). To summarise, the two within-task factors encompassing distractor colour and tone presence/absence, together with the two between-task factors resulted in a total of four factors in our analysis design: Distractor Colour (TCC vs. NCC), Distractor Modality (V vs. AV), Distractor Onset (Predictable vs. Unpredictable) and Multisensory Relationship (Arbitrary vs. Congruent).[4]

Tasks 1 and 2. As mentioned above, across Tasks 1 and 2, the colour of the distractor and the sound accompanying the colour distractor were related only by their simultaneous presentation. As such, trials from Tasks 1 and 2 made up the Arbitrary condition of the Multisensory Relationship factor. Sound frequency was always 2000 Hz (as in Matusz and Eimer, 2011). The main difference between Task 1 and Task 2 lied in the onset of the distractors in those tasks. Unbeknownst to participants, in Task 1, duration of the base array varied randomly on a trial-by-trial basis, between 100 ms, 250 ms and 450 ms, i.e., the distractor onset was unpredictable. In contrast, in Task 2, the base array duration was always constant, at 450 ms, i.e., the distractor onset was predictable. With this manipulation, considering the between-task factors: Task 1 represented Arbitrary (Multisensory Relationship) and Unpredictable (Distractor Onset) trials, and Task 2 - Arbitrary (Multisensory Relationship) and Predictable (Distractor Onset) trials.

Training. The Training served to induce in participants a semantic-level association between a specific distractor colour and a specific pitch. This rendered distractors in the Tasks following the Training semantically related (Congruent), and distractors in the preceding Tasks semantically unrelated (Arbitrary). The Training consisted of an Association phase followed by a Testing phase (both based on the association task in Sui et al., 2012; see also Sun et al., 2016).

I. Association phase. The Association phase served to induce the AV associations in participants. Participants were shown alternating colour word–pitch pairs, presented in the centre of the screen (the tone was presented from two lateral speakers, rendering it spatially diffuse and so appearing to also come from the centre of the screen). The words denoted one of two distractor colours (*red* or *blue*). The tone of either high (4000 Hz) or low (300 Hz) pitch. Both the colour word and sound were presented for 2 s, after which a central fixation cross was presented for 150 ms, followed by the next colour word–pitch pair. There could be two possible colour–pitch pairing options. In one, the high-pitch tone was associated with the word *red*, the low-pitch tone - with the word *blue*. In the second option, the high-pitch tone was associated with the word *blue*, the low-pitch tone with the word *red* (see Fig. 1C, Association phase). Pairing options were counterbalanced across participants. Thus, for participants trained with the first option, the pairing of word *red* and a high-pitch tone would be followed by the pairing of the word *blue* with a low-pitch tone, again followed by the *red*–high pitch pairing, etc. There were 10 presentations per pair, resulting in a total of 20 trials. Colour words were chosen instead of actual colours to ensure that the AV associations were based on meaning rather than lower-level stimulus features (for examples of such taught crossmodal correspondences see, e.g., Ernst, 2007). Also, colour words were shown in participants' native language (speakers: 19 French, 8 Italian, 5 German, 4 Spanish, 3 English). Participants were instructed to try to memorise the pairings as best as they could, being informed that they would be subsequently tested on how well they learnt the pairings.

II. Testing phase. The Testing phase served to ensure that the induced colour–pitch associations was strong. Now, participants were shown colour word–pitch pairings (as in the Association phase) but also

[4] As part of our stimulus design and alike Matusz and Eimer (2011), we manipulated a third within-task factor, i.e., whether the distractor and the upcoming target appeared in the same compared to a different location. This manipulation was necessary for us to compute behavioural attentional capture that were the bases of our complex 4-factor analyses However, to avoid confusing the reader, we have removed the descriptions of this factor from the main text and we only refer briefly to the manipulation in the *General task procedures*.

colour–pitch pairings (a string of x's in either red or blue, paired with a sound, Fig. 1C, *Testing phase* panel). Additionally, now, the pairings either matched or mismatched the type of associations induced in the Association phase, e.g., if the word *red* have been paired with a high-pitch tone in the Testing phase, the matching pair now would be a word *red* or red x's, paired with a high-pitch tone, and mismatching pair - the word *red* or red x's paired with a low-pitch tone. Participants had to indicate if a given pair was matched or mismatched by pressing one of two buttons (same button setup as in the Tasks). Participants whose accuracy was ≤50% had to repeat the testing.

The paradigm that Sui et al. (2012) have designed led to people being able to reliably associate low-level visual features (colours, geometric shapes) with abstract social concepts (themselves, their friend, a stranger). Following their design, in the Testing phase, each pairing was shown for 250 ms, of which 50 ms was the sound (instead of the stimulus duration of 100 ms that Sui et al. used, to fit our stimulus parameters). The pairing presentation was followed by a blank screen (800 ms), during which participants had to respond, and after each responses a screen with feedback on their performance appeared. Before each trial, a fixation cross was also shown, for 500 ms. Each participant performed three blocks of 80 trials, with 60 trials per possible combination (colour word – sound matching, colour word – sound mismatching, colour – sound matching, colour – sound mismatching). A final summary of correct, incorrect, and missed trials was shown at the end of Testing phase.

Tasks 3 and 4. Following the Training, in Tasks 3 and 4, the distractors' colour and the accompanying sound were now semantically related. Thus, the trials from these two Tasks made up the (semantically) Congruent condition of the Multisensory Relationship factor. Only congruent colour–pitch distractor pairings were now presented, as per the pairing option induced in the participants. That is, if the colour red was paired with a high-pitch tone in the Association phase, red AV distractors in Tasks 3 and 4 were always accompanied by a high-pitch tone. The pitch of sounds was now either 300 Hz (low-pitch condition; chosen based on Matusz and Eimer, 2013, where two distinct sounds were used) or 4000 Hz (high-pitch condition; chosen for its comparable perceived loudness in relation to the above two sound frequencies, as per the revised ISO 226:2003 equal-loudness-level contours standard; Spierer et al., 2013). As between Tasks 1 and 2, Task 3 and Task 4 differed in the predictability of distractor onsets, i.e., in Task 3, distractor onset was unpredictable, and in Task 4 - predictable. Therefore, Task 3 represented Congruent (Multisensory Relationship) and Unpredictable (Distractor Onset) trials, and Task 4 - Congruent (Multisensory Relationship) and Predictable (Distractor Onset) trials.

2.3. EEG acquisition and preprocessing

Continuous EEG data sampled at 1000 Hz was recorded using a 129-channel HydroCel Geodesic Sensor Net connected to a NetStation amplifier (Net Amps 400; Electrical Geodesics Inc., Eugene, OR, USA). Electrode impedances were kept below 50kΩ, and electrodes were referenced online to Cz. First, offline filtering involved a 0.1 Hz high-pass and 40 Hz low-pass as well as 50 Hz notch (all filters were second-order Butterworth filters with –12 dB/octave roll-off, computed linearly with forward and backward passes to eliminate phase-shift). Next, the EEG was segmented into peri-stimulus epochs from 100 ms before distractor onset to 500 ms after distractor onset. An automatic artefact rejection criterion of $\pm 100\mu V$ was used, along with visual inspection. Epochs were then screened for transient noise, eye movements, and muscle artefacts using a semi-automated artefact rejection procedure. Data from artefact contaminated electrodes were interpolated using three-dimensional splines (Perrin et al., 1987). Across all Tasks, 11% of epochs were removed on average and 8 electrodes were interpolated per participant (6% of the total electrode montage).

Cleaned epochs were averaged, baseline corrected to the 100 ms pre-distractor time interval, and re-referenced to the average reference. Next, to eliminate residual environmental noise in the data, a 50 Hz filter was applied.[5] All the above steps were done separately for ERPs from the four distractor conditions, and separately for distractors in the left and right hemifield. We next relabeled ERPs from certain conditions, as is done in traditional lateralised ERP analyses (like those of the N2pc). Namely, we relabelled single-trial data from all conditions where distractors appeared on the *left,* so that the electrodes over the left hemiscalp now represented the activity over the right hemiscalp, and electrodes over the right hemiscalp – represented activity over the left hemiscalp, thus creating "mirror distractor-on-the-right" single-trial data. Next, these mirrored data and the veridical "distractor-on-the-right" data from each of the four distractor conditions were averaged together, creating a single average ERP for each of the four distractor conditions. The contralaterality factor (i.e. contralateral vs. ipsilateral potentials) is normally represented by separate ERPs (one for contralateral activity, and one for ipsilateral activity; logically more pairs for pair-wise N2pc analyses). In our procedure, the lateralised voltage gradients across the whole scalp are preserved within each averaged ERP by simultaneous inclusion of both contralateral and ipsilateral hemiscalp activation. Such a procedure enabled us to fully utilise the capability of the electrical neuroimaging analyses in revealing both lateralised and non-lateralised mechanisms that support the interactions of attentional control with context control. As a result of the relabelling, we obtained four different ERPs: TCCV (targetcolour cue, Visual), NCCV (nontargetcolour cue, Visual), TCCAV (target-colour cue, AudioVisual), NCCAV (nontarget=colour cue, AudioVisual). Preprocessing and EEG analyses, unless otherwise stated, were conducted using CarTool software (available for free at www.fbmlab.com/cartool-software/; Brunet et al., 2011).

2.4. Data analysis design

Behavioural analyses. Like in Matusz and Eimer (2011), and because mean reaction times (RTs) and accuracy did not differ significantly between the four Tasks, the basis of our analyses was RT spatial cueing effects (henceforth "behavioural capture effects"). These were calculated by subtracting the mean RTs for trials where the distractor and target were in the same location from the mean RTs for trials where the distractor and the target location differed, separately for each of the four distractor conditions. Such spatial cueing data were analysed using the repeated-measures analysis of variance (rmANOVA). Error rates (%) were also analysed. As they were not normally distributed, we analysed error rates using the Kruskal–Wallis *H* test and the Durbin test. The former was used to analyse if error rates differed significantly between Tasks, while the latter was used to analyse differences between experimental conditions within each Task separately.

Following Matusz and Eimer (2011), RT data were cleaned by discarding incorrect and missed trials, as well as RTs below 200 ms and above 1000 ms. Additionally, to enable more direct comparisons with the developmental study for which current Task 1 served as an adult control (Turoman et al., 2021a), we have further removed trials with RTs outside 2.5SD of the individual mean RT. As a result, a total of 5% of trials across all Tasks were removed. Next, behavioural capture effects were submitted to a four-way 2 × 2 × 2 × 2 rmANOVA with factors: Distractor Colour (TCC vs. NCC), Distractor Modality (V vs. AV), Multisensory Relationship (Arbitrary vs. Congruent), and Distractor Onset (Unpredictable vs. Predictable). Due to the error data not fulfilling criteria for normality, we used Distractor-Target location as a factor in the analysis, conducting 3-way Durbin tests for each Task, with factors Distractor Colour, Distractor Modality, and Distractor-Target Location. All analyses, including post-hoc paired *t*-tests, were conducted using SPSS

[5] While filtering following epoch creation is normally discouraged (e.g., Widmann et al. 2015), control analyses we have carried out demonstrated that our filtering procedure was necessary and did not harm the data quality within our time-window of interest (for results of control analyses, see SOMs: Justification of filtering choices).

for Macintosh 26.0 (Armonk, New York: IBM Corporation). For brevity, we only present the RT results in the Results, and the error rate results can be found in SOMs.

ERP analyses. The preprocessing of the ERPs triggered by the visual and audiovisual distractors across the four different experimental blocks created ERP averages in which the contralateral versus ipsilateral ERP voltage gradients across the whole scalp were preserved. We first conducted a canonical N2pc analysis, as the N2pc is a well-studied and well-understood correlate of attentional selection in visual settings. However, it is unclear if the N2pc also indexes bottom-up attentional selection modulations by multisensory stimuli, or top-down modulations by contextual factors like multisensory semantic relationships (for visual-only study, see e.g., Wu et al., 2015) or stimulus onset predictability (for visual-only study, see e.g., Burra and Kerzel, 2013). N2pc analyses served also to bridge electrical neuroimaging analyses with the existing literature and EEG approaches more commonly used to investigate attentional control. Briefly, electrical neuroimaging encompasses a set of multivariate, reference-independent analyses of global features of the electric field measured at the scalp ((Koenig et al., 2014)Michel and Murray, 2012; Murray et al., 2008; Lehmann and Skrandies, 1980; Tivadar and Murray, 2019; Tzovara et al., 2012) that can detect spatiotemporal patterns in EEG across different contexts and populations (e.g., Neel et al., 2019; Matusz et al., 2018).

Canonical N2pc analysis. To analyse lateralised mechanisms using the traditional N2pc approach, we extracted mean amplitude values from, first, two electrode clusters comprising PO7/8 electrode equivalents (e65/90; an electrode pair most frequently used to analyse the N2pc), and, second, their six immediate surrounding neighbours (e58/e96, e59/e91, e64/e95, e66/e84, e69/e89, e70/e83), over the 180–300 ms post-distractor time-window (based on time-windows commonly used in traditional N2pc studies, e.g., Luck and Hillyard, 1994b; Eimer, 1996; including distractor-locked N2pc, Eimer and Kiss 2008; Eimer et al., 2009). Analyses were conducted on the mean amplitude of the N2pc difference waveforms, which were obtained by subtracting the average of amplitudes in the ipsilateral posterior-occipital cluster from the average of amplitudes in the contralateral posterior-occipital cluster. This step helped mitigate the loss of statistical power that could result from the addition of contextual factors into the design. N2pc means were thus submitted to a 4-way $2 \times 2 \times 2 \times 2$ rmANOVA with factors Distractor Colour (TCC vs. NCC), Distractor Modality (V vs. AV), Multisensory Relationship (Arbitrary vs. Congruent), and Distractor Onset (Unpredictable vs. Predictable), analogously to the behavioural analysis. Notably, the N2pc is not sensitive to the location of the stimulus of interest *per se,* but rather to the side of its presentation. As such, in canonical analyses of distractor-elicited N2pc, the congruence between distractor and target, unlike in behavioural analyses, is not considered (e.g., Lien et al., 2008; Eimer and Kiss 2008; Eimer et al., 2009). Consequently, in our N2pc analyses, target-location congruent and incongruent distractor ERPs were averaged, as a function of the side of distractor presentation.

Electrical Neuroimaging of the N2pc component. Our electrical neuroimaging analyses separately tested response strength and topography modulations in N2pc-like lateralised ERPs (see e.g. Matusz et al., 2019b for a detailed, tutorial-like description of how electrical neuroimaging measures can aid the study of attentional control processes). We assessed if interactions between visual goals, multisensory salience and contextual factors 1) modulated the distractor-elicited lateralised ERPs, and 2) if they did by altering the strength of responses within statistically indistinguishable brain networks and/or altering the recruited brain networks.

I. Lateralised analyses. To test for the involvement of strength-based spatially-selective mechanisms, we analysed Global Field Power (GFP) in lateralised ERPs. GFP is the root mean square of potential [μV] across the entire electrode montage (see Lehmann and Skrandies, 1980). To test for the involvement of network-related spatially-selective mechanisms, we analysed stable patterns in ERP topography characterising different experimental conditions using a clustering approach known as the Topographic Atomize and Agglomerate Hierarchical Clustering (TAAHC). This topographic clustering procedure generated sets of clusters of topographical maps that explained certain amounts of variance within the group-averaged ERP data. Each cluster was labelled with a 'template map' that represented the centroid of its cluster. The optimal number of clusters is one that explains the largest global variance in the group-averaged ERP data with the smallest number of template maps, and which we identified using the modified Krzanowski–Lai criterion (Murray et al., 2008). In the next step, i.e., the so-called fitting procedure, the group-averaged clustering results were 'fitted' back onto the single-subject data , such that each datapoint of each subject's ERP data over a chosen time-window was labelled by the template map with which it was best spatially correlated. This procedure resulted in a number of timeframes that a given template map was present over a given time-window, which durations (in milliseconds) we then submitted to statistical analyses described below.

In the present study, we conducted strength- and topographic analyses using the same 4-way repeated-measures design as in the behavioural and canonical N2pc analyses, on the lateralised whole-montage ERP data. Since the N2pc is a lateralised ERP, we first conducted an electrical neuroimaging analysis of lateralised ERPs in order to uncover the modulations of the N2pc by contextual factors. To obtain *global* electrical neuroimaging measures of *lateralised* N2pc effects, we computed a difference ERP by subtracting the voltages over the contralateral and ipsilateral hemiscalp, separately for each of the 4 distractor conditions. This resulted in a 59-channel difference ERP (as the midline electrodes from the 129-electrode montage were not informative). Next, this difference ERP was mirrored onto the other side of the scalp, creating a "fake" 129 montage (with values on midline electrodes now set to 0). It was on these mirrored "fake" 129-channel lateralised difference ERPs that lateralised strength-based and topographic electrical neuroimaging analyses were performed. Here, GFP was extracted over the canonical 180–300 ms N2pc time-window and submitted to a $2 \times 2 \times 2 \times 2$ rmANOVA with factors Distractor Colour (TCC vs. NCC), Distractor Modality (V vs. AV), as well as the two new factors, Multisensory Relationship (Arbitrary vs. Congruent), and Distractor Onset (Unpredictable vs. Predictable). Meanwhile, for topographic analyses, the "fake" 129-channel data across the 4 Tasks were submitted to a topographic clustering over the entire post-distractor period. Next, the data were fitted back over the 180–300 ms period. Finally, the resulting number of timeframes (in ms) was submitted to the same rmANOVA as the GFP data above.

II. Nonlateralised analyses. It remains unknown if the tested contextual factors modulate lateralised ERP mechanisms at all. Given evidence that semantic information and temporal expectations can modulate *nonlateralised* ERPs within the first 100–150 ms post-stimulus (e.g., Dell'Acqua et al., 2010; Dassanayake et al., 2016), we also investigated the influence of contextual factors on nonlateralised voltage gradients, in an exploratory fashion. It must be noted that ERPs are sensitive to the inherent physical differences in visual and audiovisual conditions. Specifically, on audiovisual trials, the distractor-induced ERPs would be contaminated by brain response modulations induced by sound processing, with these modulations visible in our data already at 40 ms post-distractor. Consequently, any direct comparison of visual-only and audiovisual ERPs would index auditory processing *per se* and not capture of attention by audiovisual stimuli. Such confounded sound-related activity is eliminated in the canonical N2pc analyses (and the electrical neuroimaging analyses based on them) through the contralateral-minus-ipsilateral subtraction. To eliminate this confound in our electrical neuroimaging analyses here, we calculated difference ERPs, first between TCCV and NCCV conditions, and then between TCCAV and NCCAV conditions. Such difference ERPs, just as the canonical N2pc difference waveforms, subtract out the sound processing confound in visually-induced ERPs. As a result of those difference ERPs, we removed factors Distractor Colour and Distractor Modality, and produced a new

factor, Target Difference (two levels: D_{AV} [TCCAV – NCCAV difference] and D_V [TCCV – NCCV difference]), that indexed the enhancement of visual attentional control by sound presence.

All nonlateralised electrical neuroimaging analyses involving context factors were based on the Target Difference ERPs. Strength-based analyses, voltage and GFP data were submitted to 3-way rmANOVAs with factors: Multisensory Relationship (Arbitrary vs. Congruent), Distractor Onset (Unpredictable vs. Predictable), and Target Difference (D_{AV} vs. D_V), and analysed using the STEN toolbox 1.0 (available for free at https://zenodo.org/record/1167723#.XS3lsi17E6h). Follow-up tests involved further ANOVAs and pairwise *t*-tests. To correct for temporal and spatial correlation of the signal (see Guthrie and Buchwald, 1991), we applied a temporal criterion of >15 contiguous timeframes, and a spatial criterion of >10% of the 129-channel electrode montage at a given latency for the detection of statistically significant effects at an alpha level of 0.05. As part of topographicanalyses, we segmented the ERP difference data across the post-distractor and pre-target onset period (0–300 ms from distractor onset). To isolate the effects related to each of the two cognitive processes and reduce the complexity of the performed analyses, we carried out two topographic clustering analyses. Topographic clustering on nonlinear mechanisms contributing to TAC was based on the visual Target Difference ERPs, while the clustering isolating MSE was based on further difference ERPs resulting now from the subtraction of D_{AV} and $D_{V\ ERPs.}$ Thus, 4 group-averaged ERPs were submitted to both clustering analyses, one for each of the context-related conditions. Next, the clutering results were fitted onto the canonical N2pc time-window (180–300 ms) as well as other, earlier time-periods, notably, also ones including time-periods highlighted by the GFP results as representing significant condition differences. The resulting map presence (in ms) over the given time-windows were submitted to 3-way rmANOVAs with factors: Multisensory Relationship (Arbitrary vs. Congruent), Distractor Onset (Unpredictable vs. Predictable), and Map (different numbers of maps, depending on the topographic clustering analyses and time-windows within each clustering analyses), followed by post-hoc *t*-tests. Maps with durations <15 contiguous timeframes were not included in the analyses. Unless otherwise stated in the Results, map durations were statistically different from 0 ms (as confirmed by post-hoc one-sample t-tests), meaning that they were reliably present across the time-windows of interest. Holm-Bonferroni corrections (Holm, 1979) were used to correct for multiple comparisons between map durations. Comparisons passed the correction unless otherwise stated.

3. Results

3.1. Behavioural analyses

3.1.1. Interaction of TAC and MSE with contextual factors

To shed light on attentional control in naturalistic settings, we first tested whether top-down visual control indexed by TAC interacted with contextual factors in behavioural measures. First, our 2 × 2 × 2 × 2 rmANOVA confirmed the presence of TAC, via a main effect of Distractor Colour, $F_{(1, 38)} = 340.4$, $p < 0.001$, $\eta_p^2 = 0.9$, with TCC distractors (42 ms), but not NCC distractors (−1 ms), eliciting reliable behavioural capture effects. Of central interest here, the strength of TAC was dependent on whether the multisensory relationship within the distractor involved mere simultaneity or semantic congruence. This was confirmed by a 2-way Distractor Colour × Multisensory Relationship interaction, $F_{(1, 38)} = 4.5$, $p = 0.041$, $\eta_p^2 = 0.1$ (Fig. 2). This effect was driven by behavioural capture effects elicited by TCC distractors being reliably larger for the Arbitrary (45 ms) than the Congruent (40 ms) condition, $t_{(38)} = 1.9$, $p = 0.027$. NCC distractors showed no evidence of a Multisensory Relationship modulation (Arbitrary vs. Congruent, $t_{(38)} = 1$, $p = 0.43$). Contrastingly, TAC showed no evidence of modulation by predictability of the distractor onset (no 2-way Distractor Colour × Distractor Onset interaction, $F_{(1, 38)} = 2$, $p = 0.16$). Thus, visual feature-based attentional control interacted with the contextual factor of distractor semantic congruence, but not distractor temporal predictability.

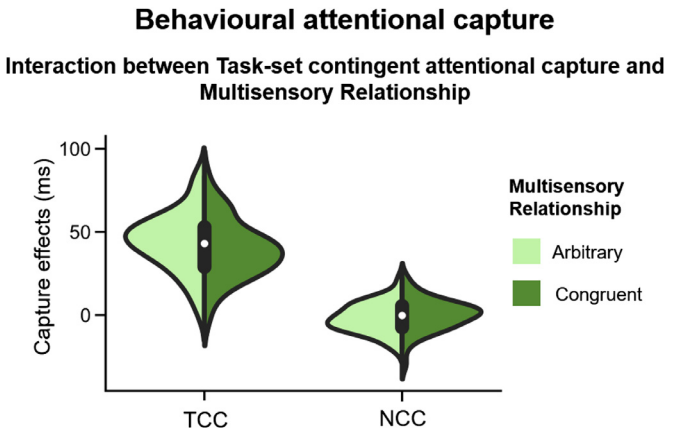

Fig. 2. The violin plots show the attentional capture effects (spatial cueing in milliseconds) for TCC and NCC distractors, and the distributions of single-participant scores according to whether Multisensory Relationship within these distractors was Arbitrary (light green) or Congruent (dark green). The dark grey boxes within each violin plot show the interquartile range from the 1st to the 3rd quartile, and white dots in the middle of these boxes represent the median. Larger values indicate *positive* behavioural capture effects (RTs faster on trials where distractor and target appeared in same vs. different location), while below-zero values – *inverted* capture effects (RTs slower on trials where distractor and target appeared in same vs. different location). Larger behavioural capture elicited by target-colour distractors (TCC) was found for arbitrary than semantically congruent distractors. Expectedly, regardless of Multisensory Relationship, attentional capture was larger for target-colour (TCC) distractors than for non-target colour distractors (NCC). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Next, we investigated potential interactions between multisensory attention enhancements and contextual factors. Expectedly, there was behavioural MSE (a significant main effect of Distractor Modality, $F_{(1, 38)}$=13.5, $p = 0.001$, η_p^2=0.3), where visually-elicited behavioural capture effects (18 ms) were enhanced on AV trials (23 ms). Unlike TAC, this MSE effect showed no evidence of interaction with either of the two contextual factors (Distractor Modality x Multisensory Relationship interaction, F<1; Distractor Modality x Distractor Onset interaction: *n.s.* trend, $F_{(1, 38)}$=3.6, $p = 0.07$, η_p^2= 0.1). Thus, behaviourally, MSE was not modulated by distractors' semantic relationship nor its temporal predictability. We have also observed other, unexpected effects, but as these were outside of the focus of the current paper, which aims to elucidate the interactions between visual (goal-based) and multisensory (salience-driven) attentional control and contextual mechanisms, we describe them only in SOMs.

3.2. ERP analyses

3.2.1. Lateralised (N2pc-like) brain mechanisms

We next investigated the type of brain mechanisms that underlie interactions between more traditional attentional control (TAC, MSE) and contextual control over attentional selection. Our analyses on the lateralised responses, spanning both a canonical and EN framework, revealed little evidence for a role of spatially-selective mechanisms in supporting the above interactions. Both canonical N2pc and electrical neuroimaging analyses confirmed the presence of TAC (see Fig. 3 for N2pc waveforms across the four distractor types). However, TAC did not interact with either of the two contextual factors. Lateralised ERPs also showed no evidence for sensitivity to MSE nor for interactions between MSE and any contextual factors. Not even the main effects of Multisensory

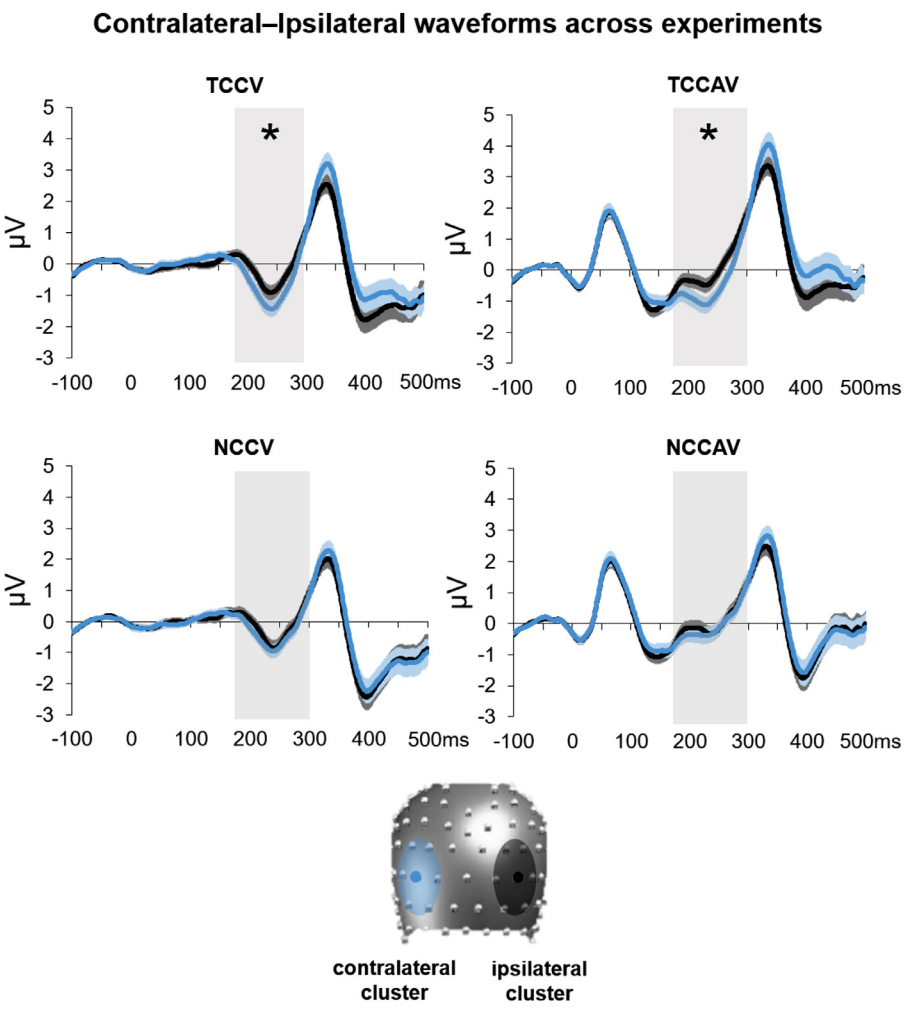

Fig. 3. Overall contra- and ipsilateral ERP waveforms representing a mean amplitude over electrode clusters (plotted on the head model at the bottom of the figure in blue and black), separately for each of the four distractor conditions , averaged across all four Tasks. The N2pc time-window of 180–300 ms following distractor onset is highlighted in grey, and significant contra-ipsi differences are marked with an asterisk ($p < 0.05$). As expected, only the TCC distractors elicited statistically significant contra-ipsi differences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Relationship and Distractor Onset[6] were present in lateralised responses (See SOMs for full description of the results of lateralised ERP analyses).

3.2.2. Nonlateralised brain mechanisms

A major part of our analyses focused on understanding the role of nonlateralised ERP mechanisms in the interactions between visual goals (TAC), multisensory salience (MSE) and contextual control. To remind the reader, to prevent nonlateralised ERPs from being confounded by the presence of sound on AV trials, we based our analyses here on the difference ERPs indexing visual attentional control under sound absence vs. presence. That is, we calculated ERPs of the difference between TCCV and NCCV conditions, and between TCCAV and NCCAV conditions (D_V and D_{AV} levels, respectively, of the Target Difference factor). We focus the description of these results on the effects of interest (see SOMs for full description of results).

The 2 × 2 × 2 (Multisensory Relationship × Distractor Onset × Target Difference) rmANOVA on electrode-wise voltage analyses revealed a main effect of Target Difference at 53–99 ms and 141–179 ms, thus both at early, perception-related, and later, attentional selection-related latencies (reflected by the N2pc). Across both time-windows, amplitudes were larger for D_{AV} (TCCAV – NCCAV difference) than for D_V (TCCV – NCCV difference). This effect was further modulated, evidenced by a 2-way Target Difference × Multisensory Relationship interaction, at the following time-windows: 65–103 ms, 143–171 ms, and 194–221 ms (all p's < 0.05). The interaction was driven by Congruent distractors showing larger amplitudes for D_{AV} than D_V within all 3 time-windows (65–97 ms, 143–171 ms, and 194–221 ms; all p's < 0.05). No similar differences were found for Arbitrary distractors, and there were no other interactions that passed the temporal and spatial criteria for multiple comparisons of >15 contiguous timeframes and >10% of the 129-channel electrode montage.

3.2.2.1. Interaction of TAC with contextual factors. We next used electrical neuroimaging analyses to investigate the contribution of the strength- and topographic nonlateralised mechanisms to the interactions between TAC and contextual factors.

Strength-based brain mechanisms. A 2 × 2 × 2 Target Difference × Multisensory Relationship × Distractor Onset rmANOVA on the GFP mirrored the results of the electrode-wise analysis on ERP voltages by showing a main effect of Target Difference spanning a large part of the first 300 ms post-distractor both before and in N2pc-like time-windows (19–213 ms, 221–255 ms, and 275–290 ms). Like in the voltage waveform analysis, the GFP was larger for D_{AV} than D_V (all p's < 0.05). In GFP, Target Difference interacted both with Multisensory Relationship (23–255 ms) and separately with Distractor Onset (88–127 ms; see SOMs for full description). Notably, there was a 3-way Target Difference × Multisensory Relationship × Distractor Onset interaction, spanning 102–124 ms and 234–249 ms. We followed up this interaction with a series of post-hoc tests to gauge the modulations of TAC (and MSE, see below) by the two contextual factors.

In GFP, Multisensory Relationship and Distractor Onset interacted independently of Target Difference in the second time-window, which

[6] Any ERP results related to Distractor Onset are unlikely to be confounded by shifted baseline due to potential dominance of one ISI type (100ms, 250ms, 450ms) over others, as no such dominance was identified in a subsample of data.

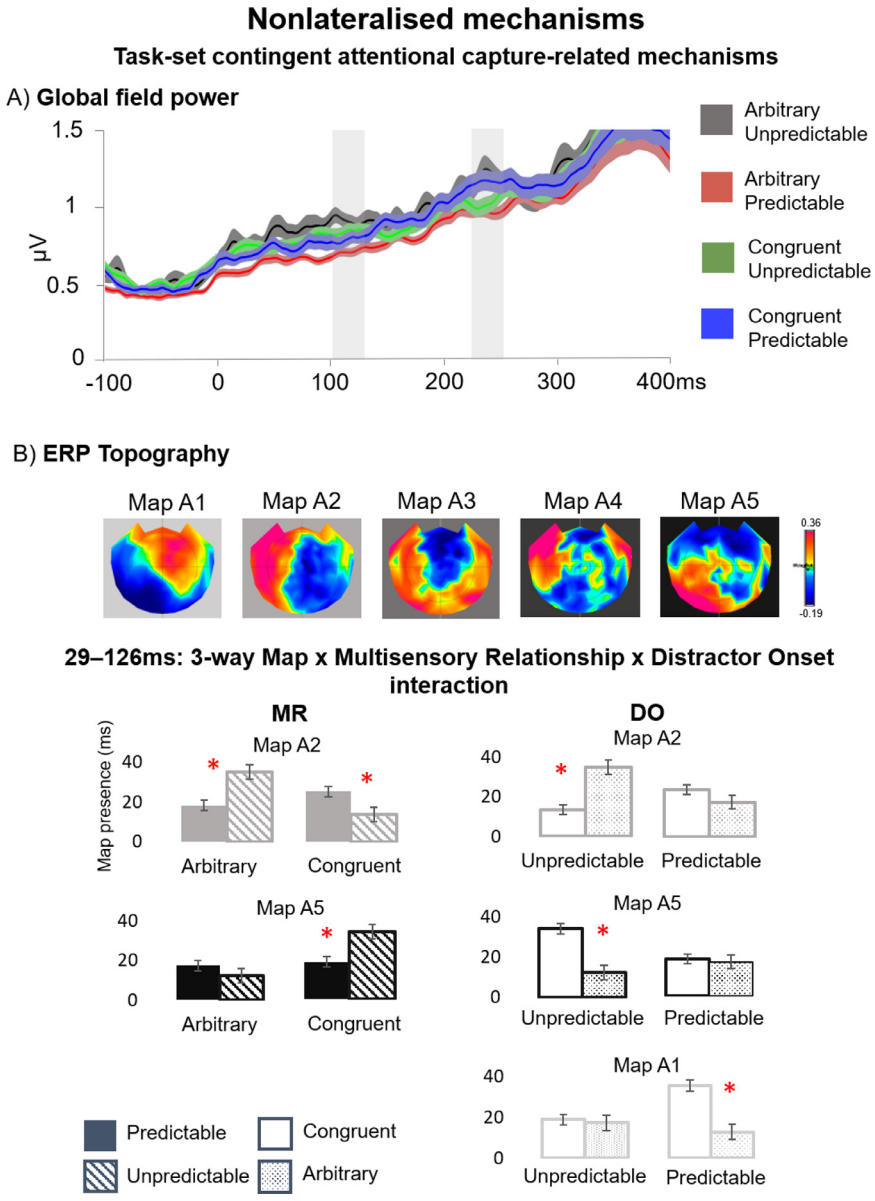

Fig. 4. Nonlateralised GFP and topography results for the visual only difference ERPs (D_V condition of Target Difference), as a proxy for TAC. A) Mean GFP over the post-distractor and pre-target time-period across the 4 Tasks (as a function of the levels of Multisensory Relationship and Distractor Onset that they represent), as denoted by the colours on the legend. The time-windows of interest (102–124 ms and 234–249 ms) are highlighted by grey areas. B) Template maps over the post-distractor time-period as revealed by the topographic clustering (Maps A1 to A5) are shown in top panels. In lower panels are the results of the fitting procedure over the 29–126 ms time-window. The results displayed here are the follow-up tests of the 3-way Map x Multisensory Relationship x Distractor Onset interaction as a function of Multisensory Relationship (MR; leftward panel) and of Distractor Onset (DO; rightward panel). Bars are coloured according to the template maps that they represent. Conditions are represented by full colour or patterns per the legend. Error bars represent standard errors of the mean.

results we describe in SOMs. To gauge differences in the strength of TAC in GFP across the 4 contexts (i.e., Arbitrary Unpredictable, Arbitrary Predictable, Congruent Unpredictable, and Congruent Predictable), we focused the comparisons on only visually-elicited target differences (to minimise any potential confounding influences from sound processing) across the respective levels of the 2 contextual factors. The weakest GFPs were observed for Arbitrary Predictable distractors (Fig. 4A). They were weaker than GFPs elicited for Arbitrary Unpredictable distractors (102–124 ms and 234–249 ms), and Predictable Congruent distractors (only in the later window, 234–249 ms).

Topography-based brain mechanisms. We focused the clustering of the TAC-related topographic activity on the whole 0–300 ms post-distractor time-window (before the target onset), which revealed 10 clusters that explained 82% of the global variance within the visual-only ERPs. This time-window of 29–126 ms post-distractor was selected based on the GFP peaks, which are known to correlate with topographic stability (Lehmann 1987; Brunet et al., 2011), and in some conditions, based on the fact that specific templates dominated responses in group-averaged data for some context conditions, e.g., Arbitrary Unpredictable and Congruent Unpredictable conditions, but not for other conditions. This was confirmed by our statistical analyses, with a 2 × 2 × 5 rmANOVA over the 29–126 ms post-distractor time-window, which revealed a 3-way Multisensory Relationship × Distractor Onset × Map interaction, $F_{(3.2,122)} = 5.3$, $p = 0.002$, $\eta_p^2 = 0.1$.

Follow-up tests in the 29–126 ms time-window focused on maps differentiating between the 4 contexts as a function of the two contextual factors (results of follow-up analyses as a function of Multisensory Relationship and Distractor Onset are visible in Fig. 4B in leftward panel and rightward panel, respectively). These results confirmed that context altered the processing of distractors from early on. The results also confirmed that the context did so by engaging different networks for the majority of the different combinations of levels of Multisensory Relationship and Distractor Onset: Arbitrary Predictable - Map A1, Arbitrary Unpredictable - Map A2, and Congruent Unpredictable - Map A5, (no map was predominantly involved in the responses for Congruent Predictable).

Arbitrary Predictable distractors, which elicited the weakest GFP, recruited predominantly Map A1 (37 ms) during processing. This map was more involved in the processing of those distractors vs. Congruent

Predictable distractors (21 ms), $t_{(38)} = 2.7$, $p = 0.013$ (Fig. 4B bottom panel).

Arbitrary Unpredictable distractors largely recruited Map A2 (35 ms) during processing. This map was more involved in the processing of these distractors vs. Arbitrary Predictable distractors (18 ms), $t_{(38)} = 2.64$, $p = 0.012$ (Fig. 4B top leftward panel), as well as vs. Congruent Unpredictable distractors (14 ms), $t_{(38)} = 3.61$, $p < 0.001$ (Fig. 4B top rightward panel).

Congruent Unpredictable distractors principally recruited Map A5 (34 ms) during processing, which was more involved in the processing of these distractors vs. Congruent Predictable distractors (19 ms) distractors, $t_{(38)} = 2.7$, $p = 0.039$ (Fig. 4B middle leftward panel), as well as vs. Arbitrary Unpredictable (12 ms) distractors, $t_{(38)} = 3.7$, $p < 0.001$ (Fig. 4B middle rightward panel).

Congruent Predictable distractors recruited different template maps during processing, where Map A2 was more involved in responses to those distractors (25 ms) vs. Congruent Unpredictable distractors (14 ms), $t_{(38)} = 2.17$, $p = 0.037$, but not other distractors, p's>0.2 (Fig. 4B top leftward panel).

3.2.2.2. Interaction of MSE with contextual factors. We next analysed the strength- and topographic nonlateralised mechanisms contributing to the interactions between MSE and contextual factors.

Strength-based brain mechanisms. To gauge the AV-induced enhancements between D_{AV} and D_V across the 4 contexts, we explored the above-mentioned 2 × 2 × 2 GFP interaction using a series of simple follow-up post-hoc tests. We first tested if response strength between D_{AV} and D_V was reliably different within each of the 4 contextual conditions. AV-induced ERP responses were enhanced (i.e., larger GFP for D_{AV} than D_V distractors) for both Predictable and Unpredictable Congruent distractors, across both earlier and later time-windows. Likewise, AV enhancements were also found for Arbitrary Predictable distractors, but only in the earlier (102–124 ms) time-window. Unpredictable distractors showed similar GFP across D_{AV} and D_V trials. Next, we compared the AV-induced MS enhancements across the 4 contexts, by creating further (D_{AV} minus D_V) difference ERPs for each context. AV-induced enhancements were weaker for Predictable Arbitrary distractors than Predictable Congruent distractors (102–124 ms and 234–249 ms; Fig. 5A).

Topography-based brain mechanisms. We then used the difference (D_{AV} minus D_V) ERPs (as in the second part of the GFP analyses) to focus the clustering selectively on the MSE-related topographic activity. This clustering, carried out on the 0–300 ms post-distractor and pre-target time-window, revealed 7 clusters that explained 78% of the global variance within the AV-V target difference ERPs.

In this topographic clustering there were multiple GPF peaks, with elongated near-synchronous periods of time where different maps were suggested to be present across the four context conditions in the group-averaged data. One of those maps (Map B3) was first present in the two congruent distractor conditions, to then become absent and reappear again. In the view of this patterning, we decided to fit the group-average data from these three subsequent time-windows to single-subject data: 35–110 ms, 110– 190 ms, and 190– 300 ms. To foreshadow the results, in the first and third time-windows the MSE-related template maps were modulated only by Multisensory Relationship, while in the middle time-window – by both Multisensory Relationship and Distractor Onset.

In the first, 35–110 ms time-window, the modulation of map presence by Multisensory Relationship was evidenced by a 2-way Map × Multisensory Relationship interaction, $F_{(2.1,77.9)} = 9.2$, $p < 0.001$, $\eta_p^2 = 0.2$. This effect was driven by one map (map B3) that, in this time-window, dominated responses to Congruent (42 ms) vs. Arbitrary (25 ms) distractors, $t_{(38)} = 4.3$, $p = 0.02$, whereas another map (map B5) dominated responses to Arbitrary (33 ms) vs. Congruent (18 ms) distractors, $t_{(38)} = 4$, $p = 0.01$ (Fig. 5B top and upper leftward panels, respectively).

In the second, 110–190 ms time-window, map presence was modulated by both contextual factors, with a 3-way Map × Multisensory Relationship × Distractor Onset interaction, $F_{(2.6,99.9)} = 3.7$, $p = 0.02$, $\eta_p^2 = 0.1$ (just as it did for TAC). We focused follow-up tests in that time-window again on maps differentiating between the 4 context conditions, as we did for the 3-way interaction for TAC (results of follow-ups as a function of Multisensory Relationship and Distractor Onset are visible in Fig. 5B, middle upper and lower panels, respectively). Context processes again interacted to modulate the processing of distractors, although now they did so after the first 100 ms. They did so again by engaging different networks for different combinations of Multisensory Relationship and Distractor Onset: Arbitrary Predictable - Map B1, Arbitrary Unpredictable - Map B5, Congruent Unpredictable - Map B6, and now also Congruent Predictable - Map B3.

Arbitrary Predictable distractors, which again elicited the weakest GFP, during processing mainly recruited Map B1 (35 ms). This map dominated responses to these distractors vs. Arbitrary Unpredictable distractors (18 ms, $t_{(38)} = 2.8$, $p = 0.01$; Fig. 5B upper panel), as well as vs. Congruent Predictable distractors (17 ms, $t_{(38)} = 2.8$, $p = 0.006$; Fig. 5B lower panel).

Arbitrary Unpredictable distractors largely recruited during processing one map, Map B5 (33 ms). Map B5 was more involved in responses to these distractors vs. Arbitrary Predictable distractors (17 ms, $t_{(38)} = 2.6$, $p = 0.042$; Fig. 5B upper panel), as well as vs. Congruent Unpredictable distractors (13 ms, $t_{(38)} = 3.4$, $p = 0.002$; Fig. 5B bottom panel).

Congruent Unpredictable distractors principally recruited during processing Map B6 (37 ms). Map B6 was more involved in responses to these distractors vs. Congruent Predictable distractors (21 ms, $t_{(38)} = 2.5$, $p = 0.02$), and vs. Arbitrary Unpredictable distractors (24 ms, $t_{(38)} = 2.3$, $p = 0.044$).

Congruent Predictable distractors mostly recruited during processing Map B3 (25 ms). Map B3 was more involved in responses to these distractors vs. Predictable Arbitrary distractors (8 ms, $t_{(38)} = 2.2$, $p = 0.005$), and, at statical-significance threshold level, vs. Congruent Unpredictable distractors (12 ms, $t_{(38)} = 2.2$, $p = 0.0502$).

In the third, 190–300 ms time-window, the 2-way Map × Multisensory Relationship interaction was reliable at $F_{(3.2,121.6)} = 3.7$, $p = 0.01$, $\eta_p^2 = 0.1$. Notably, the same map as before (map B3) was more involved, at a non-statistical trend level, in the responses to Congruent (50 ms) vs. Arbitrary distractors (33 ms), $t_{(38)} = 3.6$, $p = 0.08$, and another map (map B1) predominated responses to Arbitrary (25 ms) vs. Congruent (14 ms) distractors, $t_{(38)} = 2.3$, $p = 0.02$ (Fig. 5B rightward panel).

4. Discussion

Attentional control is necessary to cope with the multitude of stimulation in everyday situations. However, in such situations, the observer's goals and stimuli's salience routinely interact with contextual processes, yet such multi-pronged interactions between control processes have never been studied. Below, we discuss our findings on how visual and multisensory attentional control interact with distractor temporal predictability and semantic relationship. We then discuss the spatiotemporal dynamics in nonlateralised brain mechanisms underlying these interactions. Finally, we discuss how our results enrich the understanding of attentional control in real-world settings.

4.1. Interaction between task-set contingent attentional capture and contextual control

Visual control interacted most robustly with stimuli's semantic relationship. Behaviourally, *target-matching* visual distractors captured attention more strongly when they were arbitrarily connected than semantically congruent. This was accompanied by a cascade of modulations of nonlateralised brain responses, spanning both the attentional selection, N2pc-like stage and much earlier, perceptual stages. Arbitrary distractors, but only predictable ones, first recruited one particular brain network (Map A1), to a larger extent than predictable semantically congruent distractors, and did so early on (29–126 ms post-distractor).

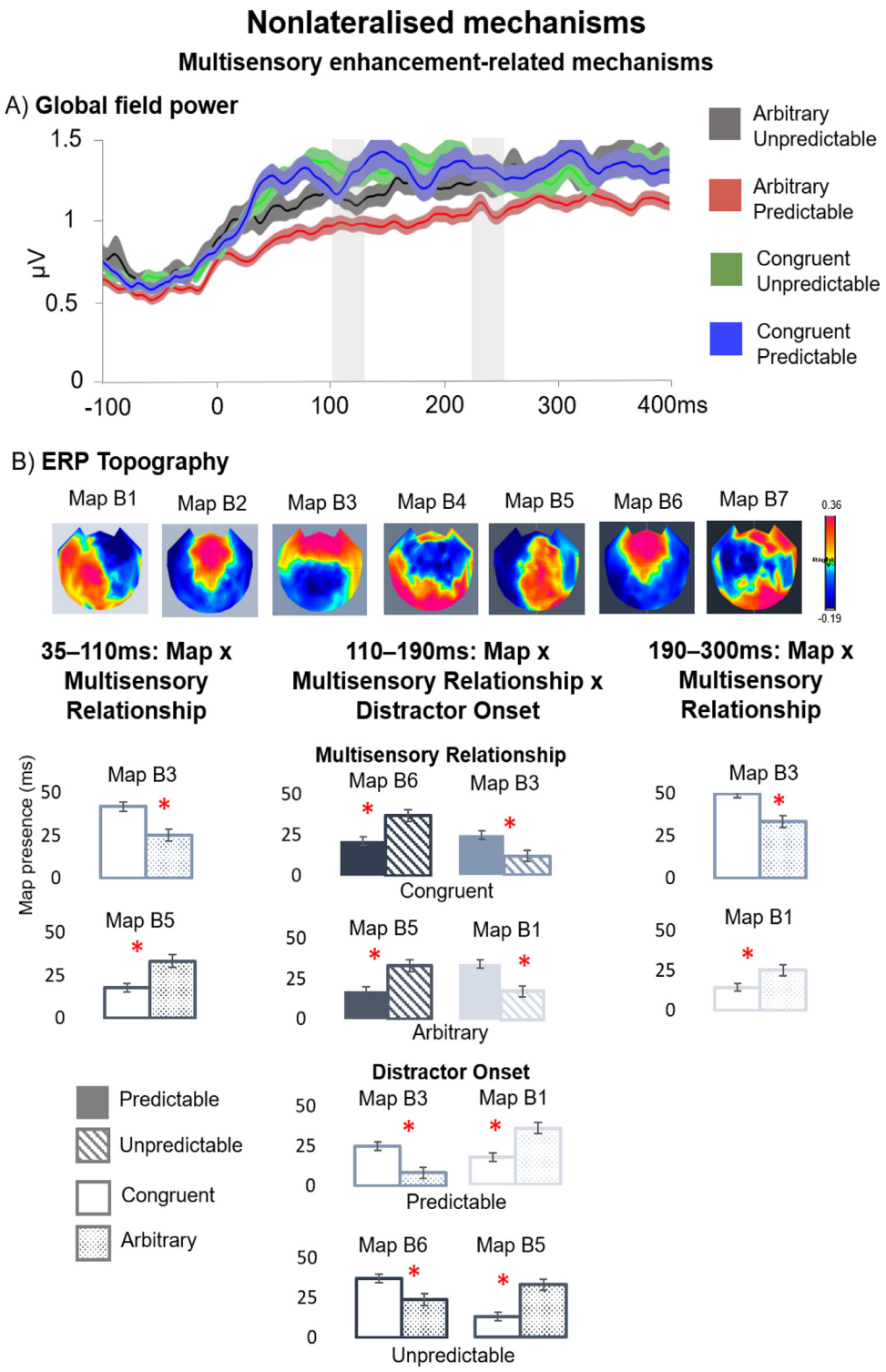

Fig. 5. Nonlateralised GFP and topography results for the difference ERPs between the DAV and DV conditions of Target Difference, as a proxy for MSE. A) Mean GFP over the post-distractor and pre-target time-period across the 4 experimental tasks (as a function of the levels of Multisensory Relationship and Distractor Onset that they represent), as denoted by the colours on the legend. The time-windows of interest (102–124 ms and 234–249 ms) are highlighted by grey bars. B) Template maps over the post-distractor time-period as revealed by the topographic clustering (Maps B1 to B7) are shown on top. Below are the results of the fitting procedure over the three time-windows: 35–110 ms, 110–190 ms, and 190–300 ms . Here we display the follow-ups of the interactions observed in each time-window: in 35–110 ms and 190–300 ms time-windows, the 2-way Map x Multisensory Relationship interaction (leftward and rightward panels, respectively), and in the 110–190 ms time-window, follow-ups of the 3-way Map x Multisensory Relationship x Distractor Onset interaction as a function of Multisensory Relationship and of Distractor Onset (middle panels). Bars are coloured according to the template maps that they represent. Conditions are represented by full colour or patterns per the legend. Error bars represent standard errors of the mean.

Arbitrary predictable distractors also elicited suppressed responses, in the later part of this early time-window (102–124 ms; where they elicited the weakest responses). In the later, N2pc-like (234–249 ms) time-window, responses to arbitrary predictable distractors were again weaker, now compared to semantically congruent predictable distractors.

This cascade of network- and strength-based modulations of non-lateralised brain responses might epitomise a potential brain mechanism for interactions between visual top-down control and multiple sources of contextual control, as they are consistent with the existing literature. The discovered early (~30–100 ms) topographic modulations for predictable (compared to unpredictable) target-matching distractors is consistent with findings that predictions attenuate the earliest visual perceptual stages (C1 component, ~50–100 ms post-stimulus; Dassanayake et al., 2016). The subsequent, mid-latency response suppressions (102–124 ms, where we found also topographic modulations) for predictable distractors are in line with N1 attenuations for self-generated sounds (Baess et al., 2011; Klaffehn et al., 2019), and the

latencies where the brain might promote the processing of unexpected events (Press et al., 2020). Notably, these latencies are also in line with the onset (~115 ms post-stimulus) of goal-based suppression of salient visual distractors (there: presented simultaneously with targets), i.e., distractor positivity (Pd; Sawaki and Luck 2010). Finally, the response suppressions we found at later, N2pc-like, attentional selection stages (234–249 ms), are also consistent with some extant (albeit scarce) literature. Van Moorselaar and Slagter (2019) showed that when salient visual distractors appear in predictable locations, they elicit the N2pc but no longer a (subsequent, post-target) Pd, suggesting that once the brain learns the distractors' location, it can suppress them without the need for active inhibition. More recently, van Moorselaar et al. (2020) showed that the representation of the predictable distractor feature could be decoded already from pre-stimulus activity. While our paradigm was not optimised for revealing such effects, pre-stimulus mechanisms could indeed explain our early-onset (~30 ms) context-elicited neural effects. The robust response suppressions for predictable stimuli are also consistent with recent proposals for interactions between predictions and auditory attention. Schröger et al. (2015) suggested that greater attention is deployed to more "salient" stimuli, i.e., those for which a prediction is missing, so that the predictive model can be reconfigured to encompass such predictions in the future. This reconfiguration, in turn, requires top-down goal-based attentional control. Our results extend this model to the visual domain. Also, the fact that we observed the response modulation cascade and behavioural benefits may also support the Schröger et al.'s tenet that different, but connected, predictive models exist at different levels of the cortical hierarchy.

Existing findings jointly strengthen our interpretations that goal-based top-down control utilises contextual information to alter visual processing . Our findings also extend the extant ideas in several ways. First, they show that in context-rich settings (i.e., involving multiple sources of contextual control), goal-based control will use both stimulus-related predictions and stimulus meaning to facilitate task-relevant processing. Second, context information modulates not only early, pre-stimulus and late, attentional stages, but also early *stimulus-elicited* responses. Third, our findings also suggest candidate mechanisms for supporting interactions between goal-based control and multiple sources of contextual information. Namely, context will modulate the early stimulus processing by recruiting distinct brain networks for stimuli representing different contexts, e.g., the brain networks recruited by predictable distractors differed between arbitrarily linked and semantically congruent stimuli (Map A1 and A2, respectively). Also, this distinct network recruitment might lead to the suppressed (potentially more efficient; c.f. repetition suppression, Grill-Spector et al., 2006) brain responses. These early response attenuations can extend to later stages, associated with attentional selection. Thus, it is the early differential brain network recruitment that might trigger a cascade of spatiotemporal brain dynamics leading effectively to the stronger behavioural capture, here for predictable (arbitrary) distractors. However, for distractors, these behavioural benefits may be most robust for arbitrary target-matching stimuli (as opposed to semantically congruent), with prediction-based effects being less apparent.

4.2. Interaction between multisensory enhancement of attentional capture and contextual control

Across brain responses, multisensory-induced processes in our study interacted with both contextual processes. To measure effects related to multisensory-elicited modulations and to their interactions with contextual information, we analysed AV–V differences within the Target Difference ERPs.

The interactions between multisensory modulations and context processes were instantiated, like those for visual attentional control, via an early-onset cascade of strength- and topographic (network-based) non-lateralised brain mechanisms. This cascade again started early (now 35–110 ms post-distractor). A separate topographic clustering analysis revealed that in the multisensory-modulated responses the brain first distinguished only between semantically congruent and arbitrarily linked distractors. These distractors recruited predominantly different brain networks (Map B3 and B5, respectively). Around the end of the time-window of these topographic, network-based modulations, at 102–124 ms, multisensory-elicited brain responses were also modulated in their strength. Arbitrary predictable distractors again triggered weaker responses, now compared to semantically congruent predictable distractors. Multisensory-elicited responses predominantly recruited distinct brain networks for the four context conditions from 110 ms until 190 ms post-distractor, thus spanning stages linked to perception and attentional selection. Here, Maps B3 and B5 were now recruited for responses to semantically congruent predictable and arbitrary unpredictable distractors, respectively. Meanwhile, Maps B1 and B6 were recruited for arbitrary predictable and semantically congruent unpredictable distractors, respectively. In the subsequent time-window (190–300 ms) that mirrors the time-window used in the canonical N2pc analyses, multisensory-related responses again recruited different brain networks. There, Map B3 (previously: congruent predictable distractors) again was predominantly recruited by semantically congruent over arbitrary distractors, and now Map B1 (previously: arbitrary predictable distractors) - by arbitrary distractors over congruent ones. In the middle of this time-window (234–249 ms), responses differed in their strength, with predictable arbitrary distractors eliciting weaker responses compared to semantically congruent predictable distractors.

To summarise, distractors' semantic relationship played a dominant (but not absolute) role in interactions between multisensory-elicited and contextual processes. The AV–V difference ERPs were modulated exclusively by multisensory relationships both in the earliest, perceptual (35–110 ms) time-window and the latest, N2pc-like (190–300 ms) time-window linked to attentional selection. At both stages, distinct brain networks predominated responses to semantically congruent and arbitrary distractors. These results suggest that from early perceptual stages the brain "relays" the processing of (multisensory) stimuli as a function of them containing meaning (vs. lack thereof, to the observer) up to stages of attentional selection. Notably, the same brain network (Map B3) supported multisensory processing of semantically congruent distractors across both time-windows, while different networks were recruited by arbitrarily linked distractors.

Thus, a single network might be recruited for processing (minimally) meaningful multisensory stimuli. In the light of our behavioural results, this brain network could be involved in suppressing behavioural attentional capture for familiar, semantically congruent (over arbitrarily linked) distractors via top-down goal-driven attentional control. This idea is supported by the interactions between distractors' multisensory-driven modulations, their multisensory relationship, and their temporal predictability in the second, 110–190 ms time-window. Therein, the same "semantic" Map B3 was still present, albeit now recruited duing responses to semantically congruent (over arbitrary) *predictable* distractors. Based on the existing evidence that predictions are used in service of goal-based behaviour (Schröger et al., 2015; van Moorselaar et al. 2020; Matusz et al., 2016(Retsa et al., 2018, 2020)), one could argue that the brain network reflected by Map B3 might play a role in integrating contextual information across both predictions and meaning (though mostly meaning, as it remained recruited by semantically congruent distractors throughout the distractor-elicited response). The activity of this network might have contributed to the overall stronger brain responses (indicated by GFP results) to semantically congruent multisensory stimuli, which in turn contributed to the suppression of multisensory enhancements of behavioural attentional capture by those distractors. While these are the first results of this kind, they open an exciting possibility that surface-level EEG/ERP studies can reveal the network- and strength-related brain mechanisms (potentially a single network for "gain control" up-modulation) by which goal-based processes control (i.e., suppress) multisensorily-driven enhancements of visual attentional capture.

4.3. Towards understanding how we pay attention in naturalistic settings

It is now relatively well-established that the brain facilitates goal-directed processing (from perception to attentional selection) via processes based on observer's goals (e.g. Folk et al., 1992; Desimone and Duncan 1995), predictions about the outside world (Summerfield and Egner 2009; Schröger et al., 2015; Press et al., 2020), and long-term memory contents (Summerfield et al., 2006; Peelen and Kastner 2014). Also, multisensory processes are increasingly recognised as an important source of bottom-up, attentional control (e.g. (Santangelo and Spence, 2007) Matusz and Eimer 2011;(Turoman et al., 2021a) ; Fleming et al., 2020). By studying these processes largely in isolation, researchers have clarified how they support goal-directed behaviour. However, in the real world, observer's goals interact with multisensory processes and multiple types of contextual information. Our study sheds the first light on this "naturalistic attentional control".

Understanding of attentional control in the real world has been advanced by research on feature-related mechanisms (Theeuwes 1991; Folk et al., 1992; Desimone and Duncan 1995; Luck et al., 2020), which support attentional control where target location information is often missing. Here, we aimed to increase the ecological validity of this research by investigating how visual feature-based attention (as indexed by TAC) transpires in context-rich, multisensory settings (see SOMs for a discussion of our replication of TAC). Our findings of reduced capture for semantically congruent than artificially linked target-colour matching distractors is novel and important, as they suggest stimuli's meaning is also utilised to suppress attention (to distractors). Until now, known benefits of meaning were limited to target selection (Thorpe et al., 1996; Iordanescu et al., 2008; Matusz et al., 2019a). Folk et al. (1992) famously demonstrated that attentional capture by distractors is sensitive to the observer's goals; we reveal that distractor's meaning may serve as a second source of goal-based attentional control. This provides a richer explanation for how we stay focused on tasks in everyday situations, despite many objects matching attributes of our current behavioural goals.

To summarise, in the real world, attention should be captured more strongly by stimuli that are unpredictable (Schröger et al., 2015), but also by those unknown or without a clear meaning. On the other hand, stimuli with high strong spatial and/or temporal alignment across the senses (and thus stronger bottom-up salience) may be more resistant to such goal-based attentional control (suppression), as we have shown here (multisensory enhancement of attentional capture; see also (Santangelo and Spence, 2007) Matusz and Eimer 2011; van der Burg et al. 2011; Turoman et al., 2021a; Fleming et al., 2020). As multisensory distractors captured attention more strongly even in current, context-rich settings, this confirms the importance of multisensory salience as a *potential* source of bottom-up attentional control in naturalistic environments (see SOMs for a short discussion of this replication).

The investigation of brain mechanisms underlying known EEG/ERP correlates (N2pc, for TAC) via advanced multivariate analyses has enabled us to provide a comprehensive, novel account of attentional control in a multi-sensory, context-rich setting. Our results jointly support the primacy of goal-based control in naturalistic settings. Multisensory semantic congruence reduced behavioural attentional capture by target-matching colour distractors compared to arbitrarily linked distractors. Context modulated nonlateralised brain responses to target-related (TAC) distractors via a cascade of strength- and topographic mechanisms from early (~30 ms post-distractor) to later, attentional selection stages. While these results are first of this kind and need replication, they suggest that context-based goal-directed modulations of distractor processing "snowball" from early stages (potentially involving pre-stimulus processes, e.g. van Moorselaar and Slagter, 2020) to control behavioural attentional selection. Responses to predictable arbitrary (target-matching) distractors revealed by our electrical neuroimaging analyses might have driven the larger behavioural capture by arbitrary than semantically congruent distractors. The former engaged distinct brain networks and triggered the weakest and so potentially most efficient (Grill-Spector et al., 2006) responses. One potential reason for the absence of such effects in behavioural measures is the small magnitude of behavioural effects: while the TAC effect has a magnitude of ~50 ms, both the MSE effect and semantically-driven suppression were small, at around ~5 ms. This may also be the reason why context-driven effects were absent in the behavioural measures of multisensory enhancement of attentional capture, despite involving a complex, early-onsetting cascade of strength- and topographic modulations.

Our results point to a potential brain mechanism by which semantic relationships influence goal-directed behaviour towards task-irrelevant information. Namely, our electrical neuroimaging analyses of surface-level EEG identified a brain network that is recruited by semantically congruent stimuli at early, perceptual stages, and that remains active at N2pc-like, attentional selection stages. While remaining cautious when interpreting our results, this network might have contributed to the prolonged enhanced AV-induced responses that we observed for semantically congruent multisensory distractors. These enhanced brain responses, together with the concomitant *suppressed behavioural attention* capture effects, are consistent with a "gain control" mechanism, in the context of distractor processing (e.g. Sawaki and Luck 2010; Luck et al., 2020). Our results reveal that such "gain control", at least in some cases, operates by relaying processing of certain stimuli to distinct brain networks. We have purported the existence of such a "gain control" mechanism in a different study on (top-down) multisensory attentional control (e.g.(Matusz et al., 2019b)). While these are merely speculations that would require source estimations to be supported, the enhanced responses to meaningful distractors may thus reflect enhanced goal-based control over those stimuli. Such a process could potentially recruit a network involving the anterior hippocampus and putamen, which help maintain active representations of task-relevant information while updating the representation of to-be-suppressed information (McNab and Klingberg 2008; Jiang et al., 2015(Sadeh et al., 2011)). Our electrical neuroimaging analyses of the surface-level N2pc data (see also ;(Matusz et al., 2019b) Turoman et al., 2021a) might have potentially revealed when and how such memory-related brain networks modulate attentional control over task-irrelevant stimuli.

4.4. N2pc as an index of attentional control

We have previously discussed the limitations of canonical N2pc analyses in capturing neurocognitive mechanisms by which visual top-down goals and multisensory bottom-up salience simultaneously control attention selection (Matusz et al., 2019b). The mean N2pc amplitude modulations are commonly interpreted as "gain control", but they can be driven by both strength- (i.e., "gain") and topographic (network-based) mechanisms. Canonical N2pc analyses cannot distinguish between those two brain mechanisms. Contrastingly, Matusz et al. (2019b) have shown evidence for both brain mechanisms underlying N2pc-like responses. These and other results of ours (Turoman et al., 2021a,(Turoman et al., 2021b)) provided evidence from surface-level EEG data for different brain sources contributing to the N2pc's, a finding that has been previously shown only in *source*-level data (Hopf et al., 2000). These results point to a certain limitation of the N2pc (canonically analysed), which is an EEG *correlate* of attentional selection, but where other analytical approaches are necessary to reveal brain mechanisms of attentional selection.

Here, we have shown that the lateralised, spatially-selective brain mechanisms, approximated by the N2pc and revealed by electrical neuroimaging analyses, are limited in how they contribute to attentional control in some settings. Rich, multisensory, and context-laden influences over goal-based top-down attention are, in our current paradigm, not captured by such lateralised mechanisms. In contrast, nonlateralised (or at least *relatively less* lateralised, see Figs. 4 and 5) brain networks seem to support such interactions for visual and multisensory distractors - from early on, up to stages of attentional selection. We nevertheless want to reiterate that paradigms that can gauge N2pc offer an important

starting point for studying attentional control in less traditional multisensory and/or context-rich settings. There, multivariate analyses, and an electrical neuroimaging framework in particular, might be useful in readily revealing new mechanistic insights into attentional control.

4.5. Broader implications

Our findings are important to consider when aiming to study attentional control, and information processing more generally, in naturalistic settings (e.g., while viewing movies, listening to audiostories) and veridical real-world environments (e.g. the classroom or the museum). Additionally, conceptualisations of ecological validity ((Peelen and Kastner, 2014) Shamay-Tsoory and Mendelsohn 2019; Vanderwal et al., 2019; Eickhoff et al., 2020; Cantlon 2020) should go beyond traditionally invoked components (e.g., observer's goals, context, socialness) to encompass contribution of multisensory processes. For example, naturalistic studies should compare unisensory and multisensory stimulus/material formats, to measure/estimate the contribution of multisensory-driven bottom-up salience to the processes of interest. More generally, our results highlight that hypotheses about how neurocognitive functions operate in everyday situations can be built already in the laboratory, if one manipulates systematically, together and across the senses, goals, salience, and context (van Atteveldt et al. 2018; Matusz et al., 2019c). Such a cyclical approach (Matusz et al., 2019a; see also Naumann et al., 2020 for a new tool to measure ecological validity of a study) involving testing of hypotheses across laboratory and veridical real-world settings could be highly promising for successfully bridging the two, typically separately pursued types of research. As a result, such an approach could create more complete theories of naturalistic attentional control.

AuthorCredit

Author contribution is stated at the end of the revised manuscript file

DataStatement

Unfortunately, the current data set was acquired before both researchers and the broader public began to understand the scientific importance of data sharing. As we had not consented participants with an explicit data sharing statement, we cannot upload the current data set to an open data repository. Researchers interested in analysing the current data set are very welcome to contact the Corresponding Author (pawel.matusz@hevs.ch) for data sharing.

Credit authorship contribution statement

Nora Turoman: Investigation, Formal analysis, Data curation, Software, Visualization, Writing – original draft, Writing – review & editing. **Ruxandra I. Tivadar:** Software, Writing – review & editing. **Chrysa Retsa:** Software, Writing – review & editing. **Micah M. Murray:** Funding acquisition, Methodology, Resources, Formal analysis, Software, Supervision, Writing – review & editing. **Pawel J. Matusz:** Conceptualization, Funding acquisition, Methodology, Resources, Formal analysis, Software, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

We thank The EEG Brain Mapping Core of the center for Biomedical Imaging (CIBM) for providing the infrastructure. This project was supported by the Pierre Mercier Foundation to P.J.M. Financial support was likewise provided by the Swiss National Science Foundation (grants: 320030_149982 and 320030_169206 to M.M.M., PZ00P1_174150 to P.J.M., the National Centre of Competence in research project "SYNAPSY, The Synaptic Bases of Mental Disease" [project 51AU40_125759]), and grantor advised by Carigest SA (232920) to M.M.M.. P.J.M. and M.M.M. are both supported by Fondation Asile des Aveugles.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118556.

Appendix 1. Abbreviations

N2pc – the N2pc event-related component
EEG – Electroencephalography
ERPs – Event-Related Potentials
TAC – Task-set Contingent Attentional Capture
MSE – Multisensory Enhancement of Attentional Capture
SOMs – Supplementary Online Materials
TCCV – target-colour cue visual
NCCV – nontarget-colour cue visual
TCCAV – target-colour cue audiovisual
NCCAV – nontarget-colour cue audiovisual
rmANOVA – repeated-measures analysis of variance
GFP – Global Field Power
TAAHC – Topographic Atomize and Agglomerate Hierarchical Clustering
D_{AV} – Target Difference, difference between TCCAV and NCCAV conditions
D_{V} – Target Difference, difference between TCCV and NCCV conditions
DO – Distractor Onset
MR – Multisensory Relationship

References

Baess, P., Horváth, J., Jacobsen, T., Schröger, E., 2011. Selective suppression of self-initiated sounds in an auditory stream: an ERP study. Psychophysiology 48 (9), 1276–1283.
Brunet, D., Murray, M.M., Michel, C.M., 2011. Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput. Intell. Neurosci. 2011.
Burra, N., Kerzel, D., 2013. Attentional capture during visual search is attenuated by target predictability: evidence from the N2pc, Pd, and topographic segmentation. Psychophysiology 50 (5), 422–430.
Cantlon, J.F., 2020. The balance of rigor and reality in developmental neuroscience. Neuroimage 216, 116464.
Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. J. Neurosci. 30 (38), 12572–12580.
Chen, Y.C., Spence, C., 2010. When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. Cognition 114 (3), 389–404.
Chun, M.M., Jiang, Y., 1998. Contextual cueingcueing: implicit learning and memory of visual context guides spatial attention. Cognit. Psychol. 36 (1), 28–71.
Correa, Á., Lupiáñez, J., Tudela, P., 2005. Attentional preparation based on temporal expectancy modulates processing at the perceptual level. Psychon. Bull. Rev. 12 (2), 328–334.
Coull, J.T., Frith, C.D., Büchel, C., Nobre, A.C., 2000. Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. Neuropsychologia 38 (6), 808–819.
Dassanayake, T.L., Michie, P.T., Fulham, R., 2016. Effect of temporal predictability on exogenous attentional modulation of feedforward processing in the striate cortex. Int. J. Psychophysiol. 105, 9–16.
De Meo, R., Murray, M.M., Clarke, S., Matusz, P.J., 2015. Top-down control and early multisensory processes: chicken vs. egg. Front. Integr. Neurosci. 9 (17), 1–6.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., Grainger, J., 2010. ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. Front. Psychol. 1, 177.
Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18 (1), 193–222.
Doehrmann, O., Naumer, M.J., 2008. Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res. 1242, 136–150. doi:10.1016/J.BRAINRES.2008.03.071.
Eickhoff, S.B., Milham, M., Vanderwal, T., 2020. Towards clinical applications of movie fMRI. Neuroimage, 116860.
Eimer, M., 1996. The N2pc component as an indicator of attentional selectivity. Electroencephalogr. Clin. Neurophysiol. 99 (3), 225–234.
Eimer, M., Kiss, M., 2008. Involuntary attentional capture is determined by task set: evidence from event-related brain potentials. J. Cognit. Neurosci. 20 (8), 1423–1433.
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. J. Exp. Psychol. 35 (5), 1316–1328.
Ernst, M.O., 2007. Learning to integrate arbitrary signals from vision and touch. J. Vis. 7 (5) 7-7.
Fleming, J.T., Noyce, A.L., Shinn-Cunningham, B.G., 2020. Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus. Neuropsychologia 146, 107530.
Folk, C.L., Leber, A.B., Egeth, H.E., 2002. Made you blink! Contingent attentional capture produces a spatial blink. Percept. Psychophys. 64 (5), 741–753.
Folk, C.L., Remington, R.W., Johnston, J.C., 1992. Involuntary covert orienting is contingent on attentional control settings. J. Exp. Psychol. 18 (4), 1030–1044.
Gazzaley, A., Nobre, A.C., 2012. Top-down modulation: bridging selective attention and working memory. Trends Cognit. Sci. 16 (2), 129–135.
Girelli, M., Luck, S.J., 1997. Are the same attentional mechanisms used to detect visual search targets defined by color, orientation, and motion? J. Cognit. Neurosci. 9 (2), 238–253.
Golumbic, E.M.Z., Poeppel, D., Schroeder, C.E., 2012. Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang. 122 (3), 151–161.
Green, J.J., McDonald, J.J., 2010. The role of temporal predictability in the anticipatory biasing of sensory cortex during visuospatial shifts of attention. Psychophysiology 47 (6), 1057–1065.
Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cognit. Sci. 10 (1), 14–23.
Guthrie, D., Buchwald, J.S., 1991. Significance testing of difference potentials. Psychophysiology 28 (2), 240–244.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6 (2), 65–70.
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. Cereb. Cortex 10, 1233–1241.
Iordanenscu, L., Guzman-Martinez, E., Grabowecky, M., Suzuki, S., 2008. Characteristic sounds facilitate visual search. Psychonomic Bulletin & Review 15, 548–554.
Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., Suzuki, S., 2008. Characteristic sounds facilitate visual search. Psychon. Bull. Rev. 15 (3), 548–554.
Jiang, J., Brashier, N.M., Egner, T., 2015. Memory meets control in hippocampal and striatal binding of stimuli, responses, and attentional control states. J. Neurosci. 35, 14885–14895.
Kiss, M., Jolicœur, P., Dell'Acqua, R., Eimer, M., 2008a. Attentional capture by visual singletons is mediated by top-down task set: new evidence from the N2pc component. Psychophysiology 45 (6), 1013–1024.
Kiss, M., Van Velzen, J., Eimer, M., 2008b. The N2pc component and its links to attention shifts and spatially selective visual processing. Psychophysiology 45, 240–249.
Klaffehn, A.L., Baess, P., Kunde, W., Pfister, R., 2019. Sensory attenuation prevails when controlling for temporal predictability of self-and externally generated tones. Neuropsychologia 132, 107145.
Koenig, T., Stein, M., Grieder, M., Kottlow, M., 2014. A tutorial on data-driven methods for statistically assessing ERP topographies. Brain Topogr. 27 (1), 72–83.
Laurienti, P. J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., Wallace, M.T., 2004. Semantic congruence is a critical factor in multisensory behavioral performance.. Experimental Brain Research 4, 405–414.
Lehmann, D., 1987. Principles of spatial analysis. In: Gevins, A.S., Remont, A. (Eds.), Methods of Analysis of Brain Electrical and Magnetic Signals. Elsevier, Amsterdam, The Netherlands, pp. 309–354.
Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard evoked multichannel potential fields. Electroencephalogr. Clin. Neurol. 48, 609–621.
Lien, M.C., Ruthruff, E., Goodin, Z., Remington, R.W., 2008. Contingent attentional capture by top-down control settings: converging evidence from event-related potentials. J. Exp. Psychol. 34 (3), 509.
Luck, S.J., Gaspelin, N., Folk, C.L., Remington, R.W., Theeuwes, J., 2020. Progress toward resolving the attentional capture debate. Vis. Cognit. 1–21. doi:10.1080/13506285.2020.1848949.
Luck, S.J., Hillyard, S.A., 1994a. Electrophysiological correlates of feature analysis during visual search. Psychophysiology 31, 291–308.
Luck, S.J., Hillyard, S.A., 1994b. Spatial filtering during visual search: evidence from human electrophysiology. J. Exp. Psychol. 20 (5), 1000–1014.
Lunn, J., Sjoblom, A., Ward, J., Soto-Faraco, S., Forster, S., 2019. Multisensory enhancement of attention depends on whether you are already paying attention. Cognition 187, 38–49.
Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010.
Matusz, P.J., Eimer, M., 2013. Top-down control of audiovisual search by bimodal search templates. Psychophysiology 50 (10), 996–1009.
Matusz, P.J., Eimer, M., 2011. Multisensory enhancement of attentional capture in visual search. Psychon. Bull. Rev. 18 (5), 904.
Matusz, P.J., Wallace, M.T., Murray, M.M., 2020. Multisensory contributions to object recognition and memory across the life span. In: Multisensory Perception. Academic Press, pp. 135–154.
Matusz, P.J., Dikker, S., Huth, A.G., Perrodin, C., 2019a. Are we ready for real-world neuroscience? J. Cognit. Neurosci. 31 (3), 327.
Matusz, P.J., Turoman, N., Tivadar, R.I., Retsa, C., Murray, M.M., 2019b. Brain and cognitive mechanisms of top–down attentional control in a multisensory world: benefits of electrical neuroimaging. J. Cognit. Neurosci. 31 (3), 412–430.
Matusz, P.J., Merkley, R., Faure, M., Scerif, G., 2019c. Expert attention: attentional allocation depends on the differential development of multisensory number representations. Cognition 186, 171–177.
Matusz, P.J., Key, A.P., Gogliotti, S., Pearson, J., Auld, M.L., Murray, M.M., Maitre, N.L., 2018. Somatosensory plasticity in pediatric cerebral palsy following constraint-induced movement therapy. Neural Plast. 2018.
Matusz, P.J., Wallace, M.T., Murray, M.M., 2017. A multisensory perspective on object memory. Neuropsychologia 105, 243–252.
Matusz, P.J., Retsa, C., Murray, M.M., 2016. The context-contingent nature of cross-modal activations of the visual cortex. Neuroimage 125, 996–1004.
Matusz, P.J., Thelen, A., Amrein, S., Geiser, E., Anken, J., Murray, M.M., 2015a. The role of auditory cortices in the retrieval of single-trial auditory–visual object memories. Eur. J. Neurosci. 41 (5), 699–708.
Matusz, P.J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R., Scerif, G., 2015b. Multi–modal distraction: insights from children's limited attention. Cognition 136, 156–165.
McNab, F., Klingberg, T., 2008. Prefrontal cortex and basal ganglia control access to working memory. Nat. Neurosci. 11, 103–107.
Michel, C.M., Murray, M.M., 2012. Towards the utilization of EEG as a brain imaging tool. Neuroimage 61 (2), 371–385.
Miniussi, C., Wilding, E.L., Coull, J.T., Nobre, A.C., 1999. Orienting atten- tion in the time domain: modulation of potentials. Brain 122, 1507–1518.
Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20 (4), 249–264.
Murray, M.M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., Matusz, P.J., 2016a. The multisensory function of the human primary visual cortex. Neuropsychologia 83, 161–169.
Murray, M.M., Lewkowicz, D.J., Amedi, A., Wallace, M.T., 2016b. Multisensory processes: a balancing act across the lifespan. Trends Neurosci. 39 (8), 567–579.
Murray, M.M., Michel, C.M., De Peralta, R.G., Ortigue, S., Brunet, D., Andino, S.G., Schnider, A., 2004. Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage 21 (1), 125–135.
Naccache, L., Blandin, E., Dehaene, S., 2002. Unconscious masked priming depends on temporal attention. Psychol. Sci. 13 (5), 416–424.
Nastase, S.A., Goldstein, A., Hasson, U., 2020. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. Neuroimage . In print.
Naumann, S., Byrne, M.L., de la Fuente, L.A., Harrewijn, A., Nugiel, T., Rosen, M.L., ... & Matusz, P.J. (2020). Assessing the degree of ecological validity of your study: introducing the Ecological Validity Assessment (EVA) Tool. PsyArXiv. DOI: 10.31234/osf.io/qb9tz.
Neel, M.L., Yoder, P., Matusz, P.J., Murray, M.M., Miller, A., Burkhardt, S., ..., Maitre, N.L., 2019. Randomized controlled trial protocol to improve multisensory neural processing, language and motor outcomes in preterm infants. BMC Pediatr. 19 (1), 1–10.
Noonan, M.P., Crittenden, B.M., Jensen, O., Stokes, M.G., 2018. Selective inhibition of distracting input. Behav. Brain Res. 355, 36–47.
Peelen, M.V., Kastner, S., 2014. Attention in the real world: toward understanding its neural basis. Trends Cognit. Sci. 18 (5), 242–250.
Perrin, F., Pernier, J., Bertnard, O., Giard, M.H., Echallier, J.F., 1987. Mapping of scalp potentials by surface spline interpolation. Electroencephalogr. Clin. Neurophysiol. 66 (1), 75–81.
Press, C., Kok, P., Yon, D., 2020. The perceptual prediction paradox. Trends Cognit. Sci. 24 (1), 13–24.
Raij, T., Ahveninen, J., Lin, F.H., Witzel, T., Jääskeläinen, I.P., Letham, B., ..., Hämäläinen, M., 2010. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur. J. Neurosci. 31 (10), 1772–1782.
Retsa, C., Matusz, P.J., Schnupp, J.W., Murray, M.M., 2018. What's what in auditory cortices? Neuroimage 176, 29–40.
Retsa, C., Matusz, P.J., Schnupp, J.W., Murray, M.M., 2020. Selective attention to sound features mediates cross-modal activation of visual cortices. Neuropsychologia 144, 107498.
Rohenkohl, G., Gould, I.C., Pessoa, J., Nobre, A.C., 2014. Combining spatial and temporal expectations to improve visual perception. J. Vis. 14 (4) 8-8.
Sadeh, T., Shohamy, D., Levy, D.R., Reggev, N., Maril, A., 2011. Cooperation between the hippocampus and the striatum during episodic encoding. J. Cognit. Neurosci. 23 (7), 1597–1608.
Saenz, M., Buraĉas, G.T., Boynton, G.M., 2003. Global feature-based attention for motion and color. Vision Research 43, 629–637.
Santangelo, V., Spence, C., 2007. Multisensory cues capture spatial attention regardless of perceptual load.. Journal of Experimental Psychology: Human Perception and Performance 33, 1311.
Sarmiento, B.R., Matusz, P.J., Sanabria, D., Murray, M.M., 2016. Contextual factors multiplex to control multisensory processes. Hum. Brain Mapp. 37 (1), 273–288.
Sawaki, R., Luck, S.J., 2010. Capture versus suppression of attention by salient singletons: electrophysiological evidence for an automatic attend-to-me signal. Atten. Percept. Psychophys. 72 (6), 1455–1470.

Schröger, E., Marzecová, A., SanMiguel, I., 2015. Attention and prediction in human audition: a lesson from cognitive psychophysiology. Eur. J. Neurosci. 41 (5), 641–664.
Shamay-Tsoory, S.G., Mendelsohn, A., 2019. Real-life neuroscience: an ecological approach to brain and behavior research. Perspect. Psychol. Sci. 14 (5), 841–859.
Soto-Faraco, S., Kvasova, D., Biau, E., Ikumi, N., Ruzzoli, M., Morís-Fernández, L., Torralba, M., 2019. Multisensory Interactions in the Real World. Cambridge University Press.
Spierer, L., Manuel, A.L., Bueti, D., Murray, M.M., 2013. Contributions of pitch and bandwidth to sound-induced enhancement of visual cortex excitability in humans. Cortex 49 (10), 2728–2734.
Sui, J., He, X., Humphreys, G.W., 2012. Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. J. Exp. Psychol. 38 (5), 1105.
Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. Trends Cognit. Sci. 13 (9), 403–409.
Summerfield, J.J., Lepsien, J., Gitelman, D.R., Mesulam, M.M., Nobre, A.C., 2006. Orienting attention based on long-term memory experience. Neuron 49 (6), 905–916.
Sun, Y., Fuentes, L.J., Humphreys, G.W., Sui, J., 2016. Try to see it my way: embodied perspective enhances self and friend-biases in perceptual matching. Cognition 153, 108–117.
Talsma, D., Woldorff, M.G., 2005. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J. Cognit. Neurosci. 17, 1098–1114.
Ten Oever, S., Sack, A.T., 2015. Oscillatory phase shapes syllable perception. Proc. Natl. Acad. Sci. 112 (52), 15833–15837.
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. Exp. Brain Res. 234 (5), 1307–1323.
Theeuwes, J., 1991. Cross-dimensional perceptual selectivity. Percept. Psychophys. 50 (2), 184–193.
Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381 (6582), 520–522.
Tivadar, R.I., Murray, M.M., 2019. A primer on electroencephalography and event-related potentials for organizational neuroscience. Organ. Res. Methods 22 (1), 69–94.
Tivadar, R.I., Knight, R.T., Tzovara, A., 2021. Automatic sensory predictions: a review of predictive mechanisms in the brain and their link to conscious processing. In: Frontiers in Human Neuroscience, p. 438.
Tovar, D.A., Murray, M.M., Wallace, M.T., 2020. Selective enhancement of object representations through multisensory integration. J. Neurosci. doi:10.1523/JNEUROSCI.2139-19.2020, In press.
Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. Cognit. Psychol. 12 (1), 97–136.
Turoman, N., Tivadar, R.I., Retsa, C., Maillard, A.M., Scerif, G., Matusz, P.J., 2021a. The development of attentional control mechanisms in multisensory environments. Dev. Cognit. Neurosci. 48, 100930.
Turoman, N., Tivadar, R.I., Retsa, C., Maillard, A.M., Scerif, G., Matusz, P., 2021b. Uncovering the mechanisms of real-world attentional control over the course of primary education. Mind Brain Educ. In press.
Tzovara, A., Murray, M.M., Michel, C.M., De Lucia, M., 2012. A tutorial review of electrical neuroimaging from group-average to single-trial event-related potentials. Dev. Neuropsychol. 37 (6), 518–544.
Van Atteveldt, N., Murray, M.M., Thut, G., Schroeder, C.E., 2014. Multisensory integration: flexible use of general operations. Neuron 81 (6), 1240–1253.
van Atteveldt, N., van Kesteren, M.T.R., Braams, B., Krabbendam, L., 2018. Neuroimaging of learning and development: improving ecological validity. Frontline Learn. Res. 6 (3), 186–203. doi:10.14786/flr.v6i3.366.
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.
van Moorselaar, D., Slagter, H.A., 2019. Learning what is irrelevant or relevant: expectations facilitate distractor inhibition and target facilitation through distinct neural mechanisms. J. Neurosci. 39 (35), 6953–6967.
van Moorselaar, D., Slagter, H.A., 2020. Inhibition in selective attention. Ann. N. Y. Acad. Sci. 1464 (1), 204.
van Moorselaar, D., Daneshtalab, N., & Slagter, H. (2020). Neural mechanisms underlying distractor inhibition on the basis of feature and/or spatial expectations. bioRxiv.
Vanderwal, T., Eilbott, J., Castellanos, F.X., 2019. Movies in the magnet: naturalistic paradigms in developmental functional neuroimaging. Dev. Cognit. Neurosci. 36, 100600.
Widmann, A., Schröger, E., Maess, B., 2015. Digital filter design for electrophysiological data–a practical approach. J. Neurosci. Methods 250, 34–46.
Wu, R., Nako, R., Band, J., Pizzuto, J., Ghoreishi, Y., Scerif, G., Aslin, R., 2015. Rapid attentional selection of non-native stimuli despite perceptual narrowing. J. Cognit. Neurosci. 27 (11), 2299–2307.