How we pay attention in naturalistic settings

Nora Turoman[1,4], Ruxandra I. Tivadar[1,2], Chrysa Retsa[1], Micah M. Murray[1-4], Pawel J. Matusz[1,3,5*]

[1] The LINE (Laboratory for Investigative Neurophysiology), Department of Radiology and Clinical Neurosciences, University Hospital Center and University of Lausanne, Lausanne, Switzerland

[2] Department of Ophthalmology, Fondation Asile des Aveugles, Lausanne, Switzerland

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

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

[5] Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO Valais), Sierre, Switzerland

* Corresponding author:
Pawel Matusz
Information Systems Institute
University of Applied Sciences Western Switzerland (HES-SO Valais)
Rue Technopole 3
3960 Sierre, Switzerland
Tel: +41 27 606 9060

Abstract

Traditional research on attentional control has largely focused on single senses and the importance of one's behavioural goals in controlling attentional selection, thus limiting its generalizability to real-world contexts. These contexts are both inherently multisensory and contain regularities that also contribute to attentional control. To get a better understanding of how attention is controlled in the real world, we investigated how visual attentional capture was impacted by top-down goals (indexed by task-set contingent attentional capture) and the multisensory nature of stimuli, as well as top-down contextual factors such as semantic relationships and temporal predictability of stimulus onset. Participants performed a multisensory version of Folk et al. (1992) spatial cueing paradigm, while their 129-channel event-related potentials (ERPs) were recorded. Reaction-time spatial cueing served as a behavioural measure of attentional control, while the N2pc ERP component was analysed both canonically and using a multivariate electrical neuroimaging (EN) framework. Behaviourally, target-congruent colour distractors captured attention more strongly when they were simultaneous than semantically congruent (nontarget-congruent colour distractors failed to capture attention), with no behavioural evidence for context modulating multisensory enhancements of capture. However, our EN analyses revealed context-based influences on attention to both visual and multisensory distractors, on how strongly they activated brain networks and in the type of activated brain networks. In both cases, these context-driven brain response modulations occurred early on (long before the traditional N2pc time-window), with network-based modulations at ~30ms post-distractor, followed by strength-based modulations at ~100ms post-distractor. Our findings revealed that in naturalistic settings, meaning, next to predictions (spatial, temporal etc.) might be a second important source of contextual information utilised to facilitate goal-directed attention. Therein, attentional selection is controlled by an interaction of one's goals, stimulus' perceptual (multisensory-driven) salience *and* an interaction of stimulus' meaning and its predictability. Our study demonstrates how investigating more traditional, lab-studied control mechanisms and processes more typical for everyday life reveals a complex interplay between goal-, stimulus- and context-based processes in attentional control. As such, our findings call for a revision of traditional models of visual attentional control to account for the role of both contextual and multisensory control mechanisms.

Keywords: attentional control, multisensory, real-world, semantic congruence, temporal predictability, context

Introduction

Goal-directed behaviour depends on the ability to allocate processing resources towards those stimuli that are important to the current behavioural goals of the individual – a set of skills jointly known as "attentional control". On the one hand, our current knowledge about attentional control may be limited to the rigorous yet highly artificial conditions in which it is traditionally studied. On the other hand, findings from studies that assess attentional control with naturalistic stimuli (audio recordings, films, etc.) may be limited in their validity by confounds from other processes present in fully naturalistic settings. Here, we systematically tested how traditionally studied goal- and salience-based attentional control mechanisms interact with more naturalistic, context-based mechanisms using a 'naturalistic laboratory' approach that combines the rigour of lab-based studies with variabilities present in everyday situations (Matusz et al., 2019a).

In the real world, the location of goal-relevant information is rarely known in advance. For example, when searching for a friend on a crowded city street, we may know what our 'target' looks like, but not necessarily where it is located. For such cases, the visual search paradigm (Treisman & Gelade, 1980) has provided a model for how attentional control operates; it demonstrated that a target's stimulus attributes can be effective in controlling attentional selection of objects in multi-stimulus settings. Subsequent research scrutinised whether stimulus-based or goal-driven mechanisms have primacy in controlling attention. While some have argued for the priority of bottom-up stimulus salience (e.g. Theeuwes, 1991), others have proposed that the individual's current behavioural goals guide attentional selection (Folk et al., 1992). Folk and colleagues (1992) employed a version of spatial cueing paradigm, where participants searched for a colour-defined target in a search array and the search array was always preceded by a distractor array, where one element was either a colour singleton or an abrupt onset. In this colour-defined search, attentional capture - indexed by spatial cueing (a difference in response speed to targets in cued versus uncued locations) - was elicited only by colour-defined distractors. This prompted them to propose the 'task-set contingent attentional capture' (or TAC) hypothesis, whereby salient but irrelevant objects will only capture attention if they share features with the target and are therefore potentially task-relevant. In other words, top-down goal-based factors had primacy in controlling attentional selection.

However, it has quickly become clear that to understand how we pay attention in the real world, we must expand study designs to include components of real-world situations, such as the spatiotemporal or semantic information contained in the stimulus and/or the environment within which it occurs (e.g., Chun & Jiang 1998; Peelen & Kastner, 2014; Summerfield et al., 2006; van Moorselaar & Slagter 2019; Press et al. 2020) as well as the multisensory nature of objects in those environments (Matusz & Eimer, 2011, 2013; Matusz et al. 2015; Lunn et al. 2019; Soto-Faraco et al. 2019). One step towards increasing the ecological validity of the dominating models of attentional control is to systematically test how visually-derived goal- and context-based mechanisms interact with processes that are based on integration of information across the senses (henceforth, multisensory processes). The integration of simultaneous inputs across multiple senses, has been observed within the first 100ms post-stimulus onset, both in humans, as well as alert and even anaesthetized animals, and regardless of task-relevance (e.g., Fort et al. 2002; Ghazanfar et al. 2005; Giard and Peronnet 1999; Kayser et al. 2008; Rowland & Stein, 2007; Talsma & Woldroff, 2005; reviewed in de Meo et al., 2015). The early latency and putative source in low-level cortices (e.g., Cappe et al., 2010; Murray et al., 2016a; Raij et al., 2010) of such multisensory integration based on pure simultaneity detection suggests that multisensory processes may influence attentional selection in a bottom-up fashion, perhaps even independently of top-down goal-based attentional control (Matusz & Eimer, 2011). Matusz and Eimer (2011) adapted the Folk et al.'s (1992) spatial cueing paradigm to multisensory settings and demonstrated that indeed visual distractors capture attention when they match target's feature, such as colour, replicating the TAC effect. However, they also showed the same visual distractors captured attention more strongly when accompanied by a sound, regardless of their goal-relevance. This study revealed the importance of such bottom-up multisensory enhancement (MSE) for attentional selection of visual objects even under strong top-down goal-based visual control. However, relevance to behavioural goals is just one of many top-down attentional control mechanisms impacting behaviour in the outside world. Many such mechanisms have been studied in the visual domain (Summerfield & Egner, 2009; Gazzaley & Nobre, 2012; Peelen & Kastner, 2014). Two

such factors that deserve special attention are the temporal and semantic organisation of stimuli within a given 'context'[1].

Top-down contextual factors of attentional control

Temporal structure of the environment is routinely used to build predictions. In attentional contexts, attentional control will utilise those predictions to improve selection of target stimuli (e.g., Correa et al., 2005; Coull et al., 2000; Green & McDonald, 2010; Miniussi et al., 1999; Naccache et al., 2002; Rohenkohl et al., 2014). Predictions are likewise utilised in service of inhibition of task-irrelevant stimuli albeit in that area the role of location or feature-based predictions is more researched (reviewed in Noonan et al. 2018; van Moorselaar & Slagter 2020a). In naturalistic, multisensory settings, temporal predictions have been shown to be particularly important, where they aid, e.g., decoding meaning from syllables (ten Oever & Sack, 2015), sentences (Luo & Poeppel, 2007) and across competing streams of information (Zion Golumbic, Poeppel, & Schroeder, 2012). What remains unknown is whether temporal predictions in a similar vein facilitate top-down goal-based attentional control in real-world multisensory settings. In one recent example, we demonstrated how spatial predictability can be effectively utilised to suppress the processing of unisensory goal-irrelevant auditory distractors (Matusz et al. 2016).

Apart from low-level stimulus regularities like temporal predictability, semantic relationships embedded in real-world environments are another basic organising principle of information in such contexts (Amedi et al., 2005; Murray et al., 2016b; ten Oever et al. 2016; Soto-Faraco et al., 2019). There has been plentiful evidence that semantic relationships influence the processing of multisensory objects, in that semantically congruent audiovisual stimuli are more effectively identified and searched, compared to over incongruent stimuli (e.g., Chen & Spence, 2010; Iordanescu et al., 2008; Tovar et al. 2020; for review see Doehrmann & Naumer, 2008). Since real-world contexts are composed of meaningful objects, the brain likely has adapted to benefit from the redundancy of semantic information within multisensory objects in external environments, perhaps even more so than from redundancies in spatio-temporal information (Alais et al. 2010; ten

[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).

Oever et al. 2016). For example, Murray and colleagues have showed that semantically congruent multisensory pairings are better remembered than arbitrary (simultaneous) multisensory pairings (Murray et al., 2004; 2005; Matusz et al., 2015; Thelen et al. 2015, reviewed in Matusz, Wallace & Murray, 2017, 2020). It remains unclear, however, if semantic congruence would likewise enhance attentional control over multisensory stimuli over and above the enhancement engendered by multisensory simultaneity.

Taken together, the evidence suggests that predictions about stimulus onset and meaning generated through experience with the environment are commonly employed to facilitate achieving one's current behaviour goals. In line with Bayesian frameworks, the brain utilises such to bias interpretation of sensory information (and action) *towards* those predictions (e.g. Summerfield & Egner, 2009). Alternatively, the brain is thought to suppress expected inputs on behalf of processing of the most informative, "surprising" events (e.g. Richter et al. 2018) or do so when the stimulus-driven surprise / informativeness is particularly high (e.g. Press et al. 2020). Notwithstanding, the interplay between simultaneously occurring behavioural goals, multisensorily-generated perceptual salience and spatiotemporal/ semantic predictions in real-world multisensory settings remains practically unresearched (see Ten Oever et al., 2016; Matusz et al. 2019a, 2019b). In order to ascertain the various, potentially interacting, influences that could affect attentional control in multisensory settings, the present study employed an experimental paradigm that emulates real-world conditions while retaining a sufficient degree of experimental control to systematically manipulate all the factors of interest.

The present study

In the present study, we investigated whether contextual factors influence attentional control processes such as visual top-down control and bottom-up multisensory control. Specifically, we asked whether attentional control over visual and audiovisual distractors is influenced by the temporal predictability of their onset or the multisensory relationship between auditory and visual distractor features.

For this purpose, we used the multisensory version of Folk's spatial cueing paradigm developed by Matusz and Eimer (2011). This multisensory paradigm allows a direct and systematic comparison of well-known visual control processes and lesser known multisensory control processes. Top-down visual attentional control was behaviourally

189 assessed through the presence of TAC. Bottom-up control of attention by multisensory
190 stimuli was assessed through the presence of MSE. We were interested in whether the
191 predictability of the onset, on the one hand, and semantic relationships between visual and
192 auditory distractor features would modulate the TAC and MSE effects observed for visual
193 and audiovisual distractors.

194 The behavioural paradigm was combined with a simultaneous EEG recording. Event-
195 related potentials (ERPs) were time-locked to distractor onset and were analysed both with
196 canonical measures and using a multivariate whole-scalp electrical neuroimaging (EN)
197 framework. The canonical measure was the N2pc component; a negative-going voltage
198 deflection around 200ms post-stimulus onset at posterior electrode sites contralateral to
199 stimulus location (Luck & Hillyard, 1994a, 1994b; Eimer, 1996; Girelli & Luck, 1997) and a
200 classic ERP correlate of the attentional selection of potential visual targets (Eimer et al.,
201 2009; Eimer & Kiss, 2008; Kiss et al., 2008a; 2008b). The N2pc has been shown to be
202 sensitive to some contextual influences, especially match with semantic category (e.g. Wu
203 et al., 2015), with one notable study showing also prediction-based influences (Burra &
204 Kerzel, 2013). However, there is little evidence that N2pc is suitable to measure
205 behaviourally-observed bottom-up enhancements of attentional capture by multisensory
206 distractors (e.g. van der Burg et al. 2011). Notwithstanding, canonical N2pc analyses tend to
207 be limited in their neuropsychological interpretability (for a more detailed discussion, see
208 Methods, and Matusz et al., 2019b), which are readily overcome with the use of EN. It is for
209 this reason that we combined EN measures with canonical N2pc measures in an attempt to
210 elucidate the mechanisms of attentional selection in real world settings.

211 We expected to replicate the well-established TAC effect. In behaviour, this
212 encompassed reliable spatial cueing effects for target-congruent colour distractors, and
213 zero- or negative cueing, indicating suppression of capture, for nontarget-congruent
214 distractors (as in e.g., Folk et al., 1992; Folk, et al., 2002; Lien et al., 2008). Likewise, we
215 expected enhanced N2pc amplitudes for target-colour congruent visual distractors
216 compared to nontarget-colour congruent distractors (as in, e.g., Eimer et al., 2009). We
217 expected that as a reflection of the brain's strength of top-down visual attentional control,
218 TAC would be reduced for temporally predictable and/or semantically congruent (i.e.
219 embedded with meaning following the Training phase of the study) distractors. For
220 example, *spatially* predictable visual distractors were recently shown to be effectively

221 inhibited and this inhibition is modulated by feature-based predictions (e.g., Van
222 Moorselaar & Slagter 2019, 2020b). In turn, semantic relationships have been previously
223 shown to enhance visual target selection (e.g. Wu et al., 2015), indicating reliance on goal-
224 based mechanisms. As for MSE, we expected to replicate the behavioural enhancement
225 observed by Matusz and Eimer (2011), similarly expecting that at least one of the contextual
226 factors would affect MSE, where we expected it to reduce MSE. For example, predictability
227 of location of sound distractors is known to eliminate sound-induced activations of occipital
228 responses (Matusz et al. 2016). In turn, similarly to visual symbols, also multisensory
229 semantic relationships are known to enhance visual target selection across the senses (e.g.
230 Iordanescu et al., 2008).

231 We did not have strong predictions whether the interactions with context-based
232 processes on behavioural capture effects will be accompanied by concomitant N2pc
233 modulations. For TAC, known to be well-reflected in N2pc (e.g. Eimer et al. 2009), we
234 predicted it may be further modulated by temporal predictability (Burra & Kerzel, 2013) or
235 semantic relationships (Wu et al. 2015) albeit existing research focused on target-elicited
236 N2pc's and did not measure context-based modulations over the difference between target-
237 colour vs. nontarget-colour N2pcs; this leaves the question of context-based modulations of
238 TAC on distractor-elicited N2pc's open. With respect to MSE, there is little evidence for
239 sensitivity of N2pc to perceptual salience, visual or otherwise, of distractors (cf. Eimer et al.
240 2009; van der Burg et al. 2011), thus we did not have strong predictions whether distractor-
241 elicited N2pc's would be stronger for audiovisual than visual distractors, or whether this
242 MSE would interact with context-based processes. We likewise had no strong expectations
243 as to whether the two contextual factors would interact in modulating, either TAC or MSE.

244 Finally, with limited prior knowledge on the potential contextual influences on TAC
245 and MSE, we likewise investigated the role of lateralised, N2pc-like and nonlateralised
246 mechanisms in supporting the interactions between TAC and MSE on the one hand, and
247 contextual factors on the other hand. As before (Matusz et al. 2019b), we have used the
248 well-researched N2pc component to guide our EN analyses of the contextual influences over
249 TAC and MSE. That is, in EN, we analysed lateralised ERPs at ~180–300ms after stimulus
250 onset that are captured by the canonical N2pc component analyses, to study the
251 importance of strength- and network-based lateral brain mechanisms in supporting the
252 studied interactions. However, given the exploratory nature of these EN analyses, and

253 previous research that has demonstrated nonlateralised ERP modulations by semantic
254 information and temporal expectations (e.g., Dell'Acqua et al., 2010; Dassanayake et al.,
255 2016), we also analysed nonlateralised ERPs within an EN framework and did so across the
256 whole post-stimulus time-period.
257
258 **Materials and Methods**
259
260 **Participants**
261 Thirty-nine adult volunteers participated in the study (5 left-handed, 14 male, M_{age}:
262 27.5years, *SD:* 4years, range: 22–38years). Participants had normal or corrected-to-normal
263 vision and normal hearing and reported no prior or current neurological or psychiatric
264 disorders. Participants provided informed consent before the start of the experiment. All
265 research procedures were approved by the Cantonal Commission for the Ethics of Human
266 Research (CER-VD; no. 2018-00241).
267
268 **Task properties and procedures**
269 Each participant took part in one testing session consisting of four experimental tasks
270 (henceforth referred to as 'experiments', Figure 1A), where the first two experiments were
271 followed by a training task (henceforth referred to as 'training', Figure 1C). The first two
272 tasks involved non-semantically related colour-pitch combinations as in the original study,
273 while the last two involved colour-pitch combinations that were semantically congruent (the
274 factor of Multisensory Relationship in Figure 1B). Such a semantic relationship between
275 colours and sounds was created using a training task that was based on a semantic
276 association task (described in detail below; Sui et al. 2012; see also Sun et al., 2016).
277 Further, Experiments 1 and 3 involved distractor onsets variable in duration, while
278 Experiments 2 and 4 involved cues that had a constant onset (the factor of Distractor Onset
279 in Figure 1B). As a pilot study revealed sufficient proficiency at the experimental task (over
280 50% accuracy) after a few trials, participants did not practice the task before its
281 administration.
282 Experimental and training tasks were conducted in a dimly lit, sound-attenuated
283 room, with participants seated at a distance of 90 cm from a 23" LCD monitor with a
284 resolution of 1080 × 1024 (60-Hz refresh rate, HP EliteDisplay E232). All visual elements

were approximately equiluminant (~20cd/m^2), as determined by a luxmeter (model SC160, CESVA Instruments) placed at a position adjacent to participants' eyes, 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 in order to facilitate comparison with the results of Matusz & Eimer (2011). The testing session lasted no longer than 3h in total, including an initial explanation and obtaining consent, EEG setup, experiments and training, and breaks.

Experiments. Across the experimental tasks, participants searched for a colour predefined target (e.g., a red bar) in a search array, and assessed the target's orientation (vertical vs. horizontal). The search array was always preceded by an array containing distractors. Distractors were visual and audiovisual stimuli that could match the target colour (red set of dots) or not match the target colour (blue set of dots), as in the original Matusz and Eimer (2011) study (Exp.2).

Like in the former study, each experimental trial consisted of the following sequence of arrays: base array, followed by distractor array, followed by a fixation point, and finally a target array (Figure 1A). However, compared to the original Matusz and Eimer (2011) paradigm, the number of elements was reduced from 6 to 4 and targets were reshaped to look like diamonds rather than rectangles, as Experiment 1 served as an adult control for a different, developmental study (Turoman et al., in prep.). Thus, the base array contained four differently coloured sets of closely aligned dots, each dot subtending 0.1° × 0.1° of visual angle. Elements 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 element could be 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-congruent colour, or a nontarget-congruent colour that was not present in any of the elements before. The remaining three distractor array elements did not change their colour. The cues and the subsequent target "diamonds" could have either a blue (RGB values: 31/118/220) or red (RGB values: 224/71/52) colour. The target array contained four bars (rectangles) where one was always the colour-defined target. 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

317 trial, and the colour change was not spatially predictive of the subsequent target location
318 (same distractor– target location on 25% of trials). On half of all trials, distractor onset
319 coincided with the onset of a pure sine-wave tone, presented from two loudspeakers on the
320 left and right sides of the monitor. Sound intensity was 80 dB SPL, as measured using an
321 audiometer placed at a position adjacent to participants' ears.

322 Manipulations of distractors' target-colour congruent and the presence/absence of
323 sound resulted in four general distractor conditions: TCCV (target-congruent colour
324 distractor visual), NCCV (nontarget-congruent colour, Visual), TCCAV (target-congruent
325 colour distractor, AudioVisual), NCCAV (nontarget-congruent colour, AudioVisual). These
326 conditions translated into 3 factors that were comparable to Matusz and Eimer's original
327 design: Distractor Colour (target-congruent colour - TCC vs. nontarget-congruent colour -
328 NCC), Distractor Modality (Visual - V vs. AudioVisual - AV) and Cue-Target Location (Same vs.
329 Different). Further manipulations included the introduction of two contextual factors:
330 Distractor Onset and Multisensory Relationship, which were manipulated across the four
331 experimental tasks. For the Distractor Onset factor, in Experiments 2 and 4, base array
332 duration (and therefore distractor onset) was kept constant at 450ms, as in the original
333 Matusz and Eimer (2011) paradigm. Meanwhile in Experiments 1 and 3, base array duration
334 (and therefore distractor onset) was varied between 100, 250 and 450ms. This way, the
335 strength of attentional capture by temporally predictable distractors could be compared
336 with the capture elicited by unpredictable distractors, and if and how these visual and
337 audiovisual distractors differ on that dimension. For the Multisensory Relationship factor,
338 the sound frequency was set to 2000Hz in Experiments 1 and 2, as in the Matusz and Eimer
339 (2011) paradigm, and alternated between 4000Hz (high-pitch) or 300Hz (low-pitch) in
340 Experiments 3 and 4. Then, a Training was presented after Experiment 2, in order to induce
341 in participants a semantic-level association between a specific distractor colour and a
342 specific pitch (Figure 1C). Specifically, following the Training, on AV trials, tones were always
343 congruent ('matched') with the colour of the cue that they were presented with. (e.g.,
344 participants who were taught to associate red and high-pitch sounds and blue colour and
345 low-pitch sounds would, after the Training, only be presented with those AV pairings). This
346 way, the strength of attentional capture by colour-pitch combinations characterised only by
347 their simultaneous presentation could be compared with the capture elicited by colour-
348 pitch combinations characterised by semantic congruence.

To summarise, the study design included 5 different factors. These were Distractor Colour, Distractor Modality, and Cue-Target Location, and two new factors: Distractor Onset (DO; Predictable vs. Unpredictable) and Multisensory Relationship (MR; Simultaneity vs. Congruence). The full experimental session consisted of 8 blocks of 64 trials each, for each of the 4 experiments, resulting in 2,048 trials in total (512 trials per experiment). 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 5000ms of the target presentation, the next trial was initiated, otherwise the next trial was initiated immediately after a button press. Feedback on accuracy was given after each block, followed by a 'progress (treasure) map' which informed participants of the number of blocks remaining until the end, and during which participants could take a break. Breaks were also taken between each experimental task.

Training. The Training procedure consisted of an Association phase and a Testing phase. In the Association phase, participants were shown alternating colour word–pitch pairs. Specifically, each pair consisted of a word, denoting a distractor colour, that was presented on the centre of the screen at the same time as a spatially diffuse pure tone that was either high (4000Hz) or low (300Hz) in pitch. Both the colour word and sound were presented for 2 seconds, after which a central fixation cross was presented for 150ms, followed by the next colour word-pitch pair. Colour words were paired with sounds according to two possible pairing options. In one pairing option, the high-pitch tone was associated with the word 'red' and the low-pitch tone with the word 'blue', and in another pairing option, the high-pitch tone was associated with the word 'blue' and the low-pitch tone with the word 'red' (Figure 1C, Association phase). Pairing options were counterbalanced across participants. Therefore, if the first pairing option was selected, a presentation of the word 'red' with a high-pitch tone would be followed by a presentation of the word 'blue' with a low-pitch tone, which would again be followed by the former pair, etc. There were ten presentations per pair, resulting in a total of 20 trials. Colour words were chosen instead of actual colours to ensure that associations were based on semantic meaning rather than a linking of basic stimulus features (for examples of such taught crossmodal correspondences see e.g., Ernst, 2007). Colour words were shown in participants' native language (speakers: 19 French, 8 Italian, 5 German, 4 Spanish, 3 English).

Participants were instructed to observe and try to memorise the pairings as best as they could, as they would be subsequently tested on how well they learnt the pairings.

The strength of colour-pitch associations was assessed in the Testing phase. Here, participants were shown colour word-pitch pairings (as in the training) as well as colour-pitch pairings (a string of x's in either red or blue paired with a sound, Figure 1C, Testing phase). Based on the pairing option that participants were 'taught' in the Association phase, pairings could be either matched or mismatched. For example, if 'red' was paired with a high-pitch tone in the Association phase, in the Testing phase, the word 'red' (or red x's) paired with a high-pitch tone would match, while the word 'red' (or red x's) paired with a low-pitch tone would be mismatched. Participants had to indicate whether a given pair was matched or mismatched by pressing an appropriate button on the same response setup as in the experiments. In a similar paradigm used by Sui et al. (2012), people were able to reliably associate low-level visual features (colours and geometric shapes) with abstract social concepts such as themselves, their friend, and a stranger. Following their design, in the Testing phase each pairing was shown for 250ms, of which 50ms was the sound (instead of the stimulus duration of 100ms that Sui et al. (2012) used, to fit our stimulus parameters), followed by an 800ms blank screen where choices were to be made, and feedback on performance after each answer was given. Before each trial, a fixation cross was shown for 500ms. Each participant performed three blocks of 80 trials, with 60 trials per possible combination (colour word – sound matching, colour word – sound nonmatching, colour – sound matching, colour – sound nonmatching). A final summary of correct, incorrect, and missed trials was shown at the end of testing phase.

EEG acquisition and preprocessing

Continuous EEG data sampled at 1000Hz 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 0.1Hz high-pass and 40Hz low-pass as well as 50Hz notch and a second-order Butterworth filter (–12dB/octave roll-off, computed linearly with forward and backward passes to eliminate phase-shift). Next, the EEG was segmented into peri-stimulus epochs from 100ms before distractor onset to 500ms after distractor onset.

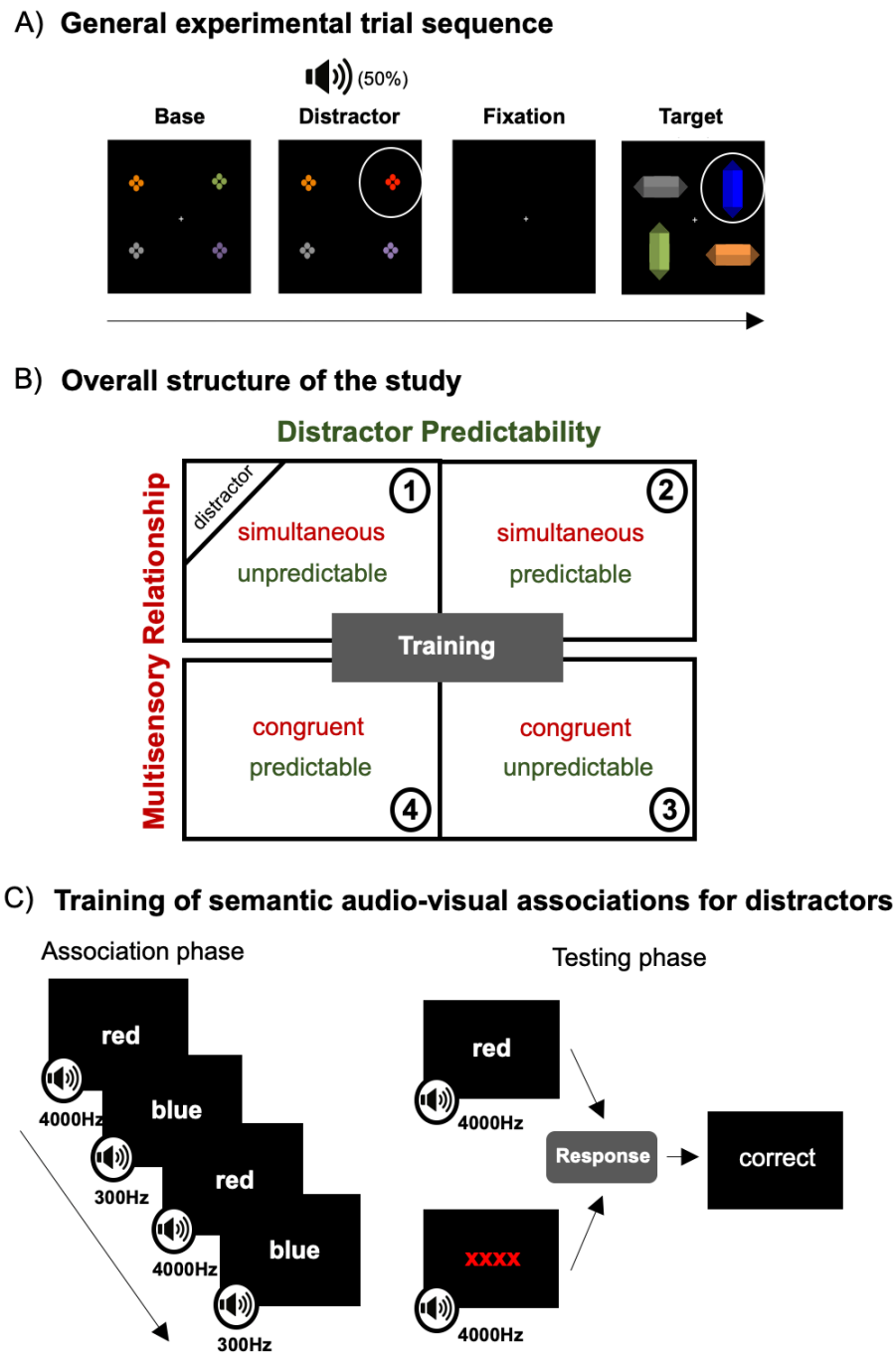


Figure 1. A) An example trial of the General experimental task is shown, with four successive arrays. The white circle around the target location (here the target is a blue diamond) and the corresponding distractor location serves to highlight, in this case, a nontarget-colour distractor condition, with a concomitant sound, i.e., NCCAV. **B)** The order of tasks, with the corresponding conditions of Multisensory Relationship (MR) in red, and Distractor Onset (DO) in green, shown

separately for each experiment. 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. Following the Training participants always encountered 'semantically congruent' (associated during the Training) audiovisual pairings, e.g. red – high-pitch and blue – low-pitch AV pairings.

An automatic artefact rejection criterion of ±100µV 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 experiment, 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 100ms pre-distractor time interval, and re-referenced to the average reference. To eliminate residual environmental noise in the data, 50Hz filter was applied. All of the above steps were done separately for ERPs from the four distractor conditions, separately for cues in the left and right hemifield. To analyse cue-elicited lateralised ERPs, single-trial data from all conditions with cues presented on the left were relabelled to have electrodes over the left hemiscalp represent activity over the right hemiscalp, and vice versa. After relabelling, the "mirror cue-on-the-right" single-trial data and the veridical "cue-on-the-right" data were averaged together, creating a single average ERP for each of the 4 distractor conditions. This way, the contralaterality that would normally be represented by separate ERPs, was preserved within each averaged ERP, with contralateral and ipsilateral hemiscalp activations present within each averaged ERP. As a result, we obtained 4 different ERPs: TCCV (target colour-cue, Visual), NCCV (nontarget colour-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, Murray, & Michel, 2011).

452 **Data analysis design**

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

465 Following Matusz and Eimer (2011), RT data were cleaned by discarding incorrect
466 and missed trials, as well as RTs below 200ms and above 1000ms. Additionally, to enable
467 more direct comparisons with the developmental study for which current Experiment 1
468 served as an adult control (Turoman et al. 2020), we have further removed trials with RTs
469 outside 2.5SD of the individual mean RT. As a result, a total of 5% of trials across all
470 experiments were removed. Next, behavioural capture effects were submitted to a four-
471 way 2 × 2 × 2 × 2 rmANOVA with factors: Distractor Colour (TCC vs. NCC), Distractor
472 Modality (V vs. AV), Multisensory Relationship (MR; Simultaneity vs. Congruence), and
473 Distractor Onset (DO; Unpredictable vs. Predictable). Due to the error data not fulfilling
474 criteria for normality, we used Cue-Target location as a factor in the analysis, conducting 3-
475 way Durbin tests for each experiment, with factors Distractor Colour, Distractor Modality,
476 and Cue-Target Location. All analyses, including post-hoc paired *t*-tests, were conducted
477 using SPSS for Macintosh 26.0 (Armonk, NY: IBM Corp). For brevity, we only present the RT
478 results in the Results, and the error rate results can be found in Supplemental Online
479 Materials ("SOMs" henceforth).

480 **ERP analyses**. The preprocessing of the ERPs triggered by the visual and audiovisual
481 cues across the 4 different experimental blocks created ERP averages in which the
482 contralateral versus ipsilateral ERP voltage gradients across the whole scalp were preserved.
483 We first conducted a canonical N2pc analysis, as the N2pc is a well-studied and well-

484 understood correlate of attentional selection in visual settings. However, it is unclear if the
485 N2pc also indexes bottom-up attentional selection modulations by multisensory stimuli, or
486 top-down modulations by contextual factors like multisensory semantic relationships (for
487 visual-only study, see e.g., Wu et al. 2015) or stimulus onset predictability (for visual-only
488 study, see e.g., Burra & Kerzel, 2013). N2pc analyses served also to bridge EN analyses with
489 the existing literature and EEG approaches more commonly used to investigate attentional
490 control. Briefly, EN encompasses a set of multivariate, reference-independent analyses of
491 global features of the electric field measured at the scalp (Biasiucci et al., 2019; Koenig et
492 al., 2014; Michel & Murray, 2012; Murray, Brunet, & Michel, 2008; Lehmann & Skrandies,
493 1980; Tivadar & Murray, 2019; Tzovara et al., 2012). Below we describe only the analyses
494 conducted as part of the present study. EN analyses offer two key advantages over
495 canonical N2pc analyses. First, they consider signal from the entire electrode montage,
496 whereas N2pc analyses only do so for two electrodes in a montage, potentially overlooking
497 relevant modulations occurring in other areas of the scalp electric field. Second, N2pc
498 analyses can only reveal average amplitude differences between experimental conditions,
499 but not the brain mechanisms underlying these differences. In contrast, EN analyses can
500 reveal whether brain response modulations between experimental conditions arise from: 1)
501 differences in response strength within a single brain network, or 2) differential brain
502 networks being active for different conditions. We focused our analyses on EN analyses of
503 lateralised ERPs (which are the centre of the canonical N2pc analyses), but instead of using a
504 pair of contralateral channels, we analysed modulations across the entire 129-channel
505 montage.

506 *Canonical N2pc analysis.* To analyse N2pc in traditional way, we extracted mean
507 amplitude values from, first, two electrode clusters comprising PO7/8 electrode equivalents
508 (e65/90; most frequent electrode pair used to analyse the N2pc), and, second, their six
509 immediate surrounding neighbours (e58/e96, e59/e91, e64/e95, e66/e84, e69/e89,
510 e70/e83), over the 180–300ms post-distractor time-window. Analyses were conducted on
511 the mean amplitude of the N2pc difference waveforms, which were obtained by subtracting
512 the average of amplitudes in the ipsilateral posterior-occipital cluster from the average of
513 amplitudes in the contralateral posterior-occipital cluster. This step helped mitigate the loss
514 of statistical power that could result from the addition of contextual factors into the design.
515 N2pc means were thus submitted to a 4-way 2 × 2 × 2 × 2 rmANOVA with factors Distractor

516 Colour (TCC vs. NCC), Distractor Modality (V vs. AV), MR (Simultaneity vs. Congruence), and
517 Distractor Onset (DO; Unpredictable vs. Predictable), analogously to the behavioural
518 analysis.

519 *Electrical Neuroimaging of the N2pc component.* EN analyses separately tested
520 response strength and topography (see e.g. Matusz et al., 2019b for a tutorial-like
521 description of how EN measures can be applied to study attentional control processes). We
522 assessed whether top-down and bottom-up factors modulated the cue-elicited lateralised
523 ERPs by altering the strength of responses within statistically indistinguishable brain
524 networks. For this purpose, we analysed Global Field Power (GFP), which is the root mean
525 square of potential [μV] across the entire electrode montage (see Lehmann & Skrandies,
526 1980).

527 Subsequently, we analysed stable patterns of ERP topography that characterise
528 experimental conditions using a clustering approach known as the Topographic Atomize and
529 Agglomerate Hierarchical Clustering (TAAHC). This clustering (also known as
530 "segmentation") procedure generates sets of clusters of topographical maps that predict
531 the largest variance within the group-averaged ERP data. Each cluster is labelled with a
532 'template map' that represents the centroid of its cluster. The optimal number of clusters is
533 one that explains the largest global explained variance (GEV) in the group-averaged ERP
534 data with the smallest number of template maps, and which we identified using the
535 modified Krzanowski–Lai criterion (Murray et al., 2008). In the next step, i.e., the so-called
536 fitting procedure, the single-subject data was 'fitted' back onto the segmentation results,
537 such that each datapoint of each subject's ERP data over a chosen time-window was
538 labelled by the template map with which it was best spatially correlated. This procedure
539 resulted in a number of timeframes that a given template map was present over a given
540 time-window, which durations (in milliseconds) we then submitted to statistical analyses
541 described below. More generally, we hasten to remind the reader that if two ERPs differ in
542 topography, they also differ in the configurations of underlying brain sources (Lehmann,
543 1987; Vaughan Jr, 1982).

544 In the present study, we conducted strength-based and topography-based analyses
545 using the same 4-way repeated measures design as in the behavioural and N2pc analyses,
546 on both lateralised and nonlateralised whole-montage ERP data. Since the N2pc is a
547 lateralised ERP, we first conducted an EN analysis of lateralised ERPs in order to uncover the

548 modulations of the N2pc by contextual factors. Thus, in order to obtain *global* EN measures
549 of *lateralised* N2pc effects, we computed a difference ERP by subtracting the voltages over
550 the contralateral and ipsilateral hemiscalp, separately for each of the 4 distractor
551 conditions. This resulted in a 59-channel difference ERP (as the midline electrodes from the
552 129-electrode montage were not informative). Next, this difference ERP was mirrored onto
553 the other side of the scalp, recreating a "fake" 129 montage (with values on midline
554 electrodes now set to 0). It was on these mirrored "fake" 129-channel lateralised difference
555 ERPs that lateralised strength-based and topography-based EN analyses were performed.
556 Here, GFP was extracted over the canonical 180–300ms N2pc time-window and submitted
557 to a 2 × 2 × 2 × 2 rmANOVA with factors Distractor Colour (TCC vs. NCC), Distractor Modality
558 (V vs. AV), as well as the two new factors, MR (Simultaneity vs. Congruence), and Distractor
559 Onset (DO; Unpredictable vs. Predictable). Meanwhile, for topographic analyses, the "fake"
560 129-channel data across the 4 experiments were submitted to a segmentation over the
561 entire post-distractor period. Next, the data were fitted back over the 180–300ms period.
562 Finally, the resulting number of timeframes (in ms) was submitted to the same rmANOVA as
563 the GFP data above.

564 Even though the N2pc is lateralised, it remains unknown whether the employed
565 contextual factors impact the N2pc at all. Given evidence that semantic information and
566 temporal expectations can modulate *nonlateralized* ERPs within the first 100-150ms post-
567 stimulus (e.g., Dell'Acqua et al., 2010; Dassanayake et al., 2016), we also investigated the
568 influence of contextual factors on nonlateralized voltage gradients, in an exploratory
569 fashion. It must be noted that ERPs are sensitive to the inherent physical differences in
570 visual and audiovisual conditions. Specifically, on audiovisual trials, the cue-induced ERPs
571 would be contaminated by brain response modulations induced by sound processing, with
572 these modulations visible in our data already at 40ms post-cue. Consequently, any direct
573 comparison of visual-only and audiovisual ERPs would index auditory processing per se and
574 capture of attention by audiovisual stimuli. Such confounded sound-related activity is
575 eliminated in the canonical N2pc analyses through the contralateral-minus-ipsilateral
576 subtraction. Here, to eliminate this confound in our exploratory EN analyses, we calculated
577 new difference ERPs, first between TCCAV and NCCAV conditions, and then between TCCV
578 and NCCV conditions. Such difference ERPs, just as the canonical N2pc difference waveform,
579 subtracted out the sound processing confound in visually-induced ERPs. These difference

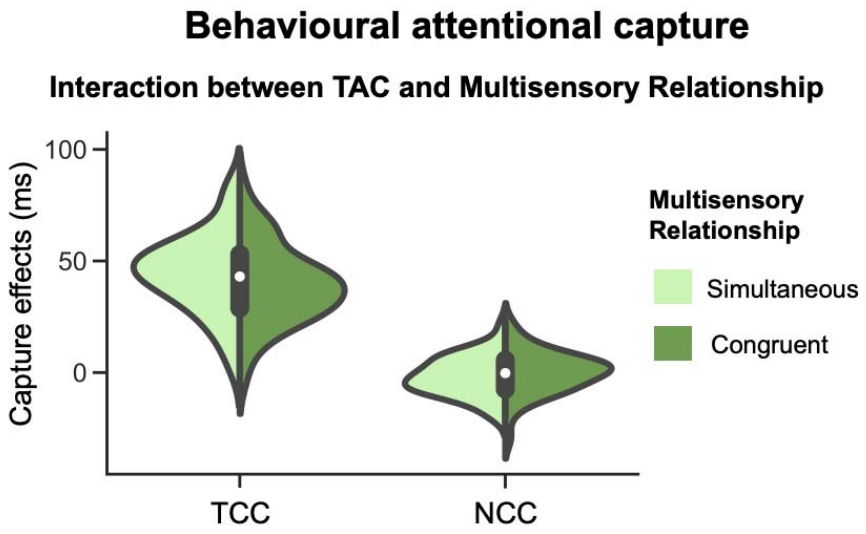
580 ERPs produced a new factor that indexed the enhancement of visual attentional control by
581 sound presence that we henceforth refer to as Target Difference. All nonlateralised EN
582 analyses were conducted on these difference ERPs and included the factor Target Difference
583 with two levels: D_{AV} (TCCAV – NCCAV difference) and D_V (TCCV – NCCV difference).
584 As part of strength-based analyses, voltage and GFP data were analysed using the
585 STEN toolbox 1.0 (available for free at https://zenodo.org/record/1167723#.XS3lsi17E6h),
586 and submitted to 3-way rmANOVAs with factors: MR (Simultaneity vs. Congruence), DO
587 (Unpredictable vs. Predictable), and Target Difference (D_{AV} vs. D_V). Follow-up tests involved
588 further ANOVAs and pairwise *t*-tests. To correct for temporal and spatial correlation (see
589 Guthrie & Buchwald, 1991), we applied a temporal criterion of >15 contiguous timeframes,
590 and a spatial criterion of >10% of the 129- channel electrode montage at a given latency for
591 the detection of statistically significant effects at an alpha level of 0.05. As part of
592 topography-based analyses, we segmented the ERP difference data across the post-
593 distractor and pre-target onset period (0 – 300ms from distractor onset) and conducted
594 clustering of the data to obtain template maps. Next, the data were fitted onto the
595 canonical N2pc time-window (180–300ms), as well as time-periods that were highlighted by
596 the GFP data as representing significant condition differences. The resulting map presence
597 (in ms) over the given time-windows were submitted to 4-way rmANOVAs with factors: MR
598 (Simultaneity vs. Congruence), DO (Unpredictable vs. Predictable), Target Difference (D_{AV} vs.
599 D_V), and Map (different numbers of maps for different time-windows), followed by post-hoc
600 *t*-tests. Maps with durations <15 contiguous timeframes were not included in the analyses.
601 Unless otherwise stated in the Results, map durations were statistically different from 0ms
602 (as confirmed by post-hoc one-sample t-tests), meaning that they were reliably present
603 across the time-windows of interest. Holm-Bonferroni corrections (Holm, 1979) were used
604 to correct for multiple comparisons between map durations. Comparisons passed the
605 correction unless otherwise stated.
606
607 **Results**
608
609 **Behavioural analyses**
610 ***Interaction of TAC with contextual factors***

611 Our 2 × 2 × 2 × 2 rmANOVA revealed a number of main effects and interactions, both
612 expected and unexpected (full description of the results in SOMs). As expected, behavioural
613 capture effects showed the TAC, evidenced by a significant main effect of Distractor Colour,
614 $F_{(1, 38)} = 340.4$, $p < 0.001$, $\eta_p^2 = 0.9$. Specifically, TCC distractors (42ms), but not NCCs (-1ms),
615 elicited reliable behavioural capture effects.

616 The strength of top-down visual attention indexed by TAC was dependent on the
617 multisensory relationship, demonstrated by a 2-way Distractor Colour × MR interaction, $F_{(1,}$
618 $_{38)} = 4.5$, $p = 0.041$, $\eta_p^2 = 0.1$ (Figure 2). This effect was driven by behavioural capture effects
619 elicited by TCC distractors being reliably larger under simultaneity (45ms) than congruence
620 (40ms), $t_{(38)} = 1.9$, $p = 0.027$. Capture effects elicited by NCC distractors showed no evidence
621 of a modulation by MR (Simultaneity vs. Congruence: -2ms vs. -0.2ms, $t_{(38)} = 1$, $p = 0.43$). In
622 contrast, there was no evidence for the strength of TAC to be dependent on predictability of
623 the distractor onset, with no 2-way Distractor Colour × DO interaction, $F_{(1, 38)} = 2$, $p = 0.16$.

624

625



626

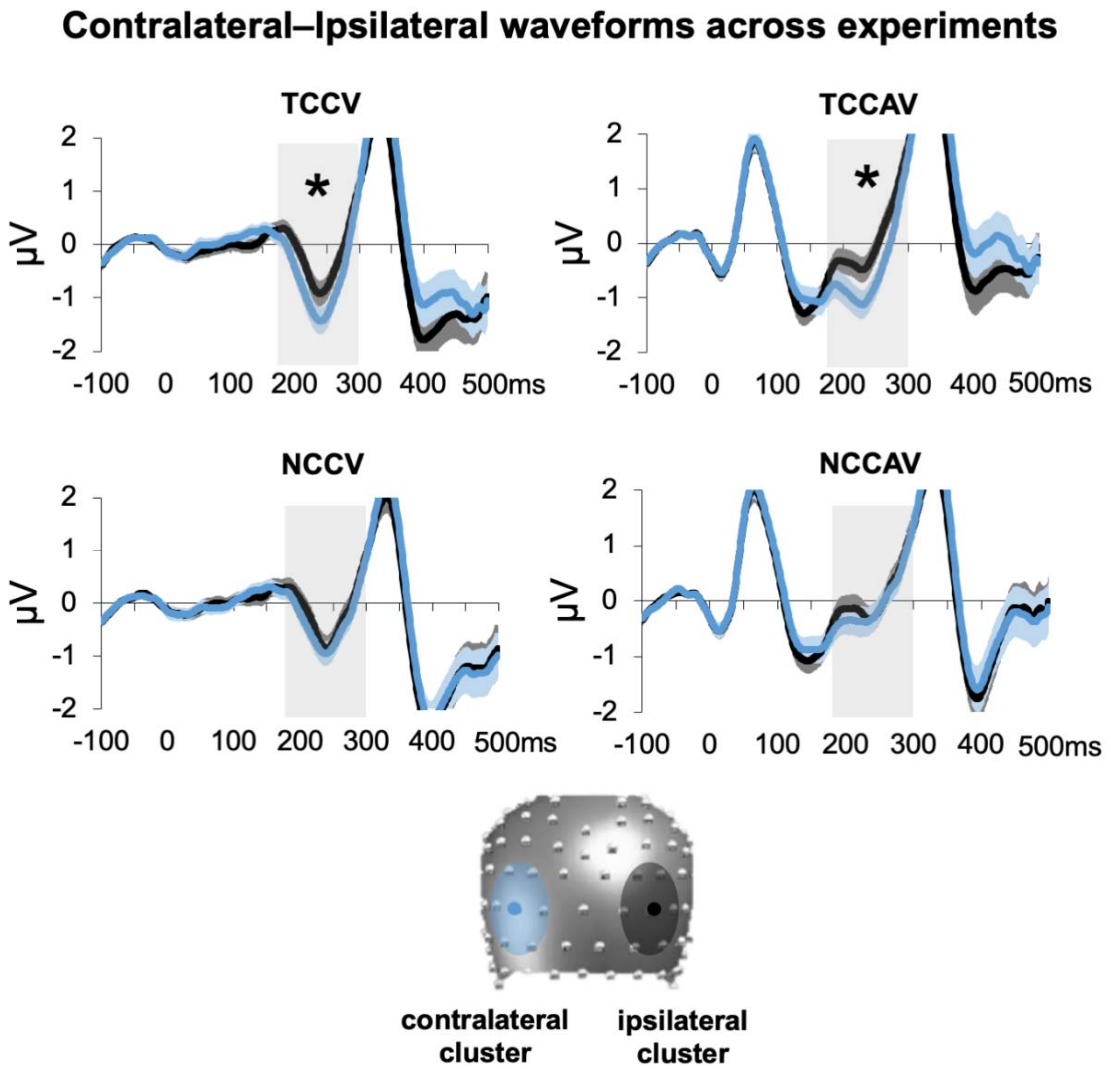
627 **Figure 2.** Larger behavioural capture elicited by target-colour cue (TCC) distractors were found for
628 simultaneous than for semantically congruent distractors. Expectedly, regardless of Multisensory Relationship,
629 attentional capture was larger for target-colour (TCC) cues than for non-target colour cues (NCC).

630

631

632 ***Interaction of MSE with contextual factors***

633 Behavioural capture effects likewise showed MSE, evidenced by a significant main effect of
634 Distractor Modality, $F_{(1, 38)} = 13.5$, $p = 0.001$, $\eta_p^2 = 0.3$, where visually-elicited behavioural
635 capture effects (18ms) were enhanced on audiovisual trials (23ms). Unlike TAC, and
636 contrary to our predictions, the MSE effect was not further modulated by either contextual
637 factor, with no 2-way interaction with MR, $F < 1$, and the 2-way interaction with DO showing
638 only a statistically nonsignificant trend, $F_{(1, 38)} = 3.6$, $p = 0.07$, $\eta_p^2 = 0.1$.
639



640
641 **Figure 3.** Overall contra- and ipsilateral ERP waveforms representing a mean amplitude over electrode clusters
642 (plotted on the head model at the bottom of the figure in blue and black [in light blue and dark grey,
643 respectively, ms-by-ms error areas representing standard error of the mean, SEM]), separately for each of the
644 four experimental conditions (Distractor Colour x Distractor Modality), averaged across all four adult
645 experiments. The N2pc time-window of 180–300ms 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.

Unexpectedly, the overall extent of multisensory enhancement depended on the colour of the distractor, evidenced by a 2-way Distractor Colour × Distractor Modality interaction, $F_{(1, 38)} = 6.4$, $p = 0.015$, $\eta_p^2 = 0.2$. This effect was driven by behavioural capture effects elicited by TCC distractors to be larger on AV (38ms) than V trials (46ms), $t_{(38)} = 4.3$, $p < 0.001$; there was no evidence for similar effects for NCC distractors, $t_{(38)} = .3$ $p = 0.78$; see Supplemental Figure 1). However, this pattern was not observed within each experiment individually (all *F*'s < 3, all *p*'s > 0.1; see SOMs for full details). Behavioural capture effects were also jointly modulated by the two contextual factors independently of TAC and MSE, $F_{(1, 38)} = 8.3$, $p = 0.007$, $\eta_p^2 = 0.2$. As these interactions were outside of our main research questions focusing on TAC- and MSE-bound interactions with the contextual factors, we describe them in full in the SOMs.

ERP analyses

The role of lateralised (N2pc-like) brain mechanisms

The role of lateralised, N2pc-like brain mechanisms in the interactions between TAC and MSE on the one hand and the two studied contextual factors on the other hand was limited, even with the added use of EN analyses. The full discussion of these results can be found in the SOMs, while we only briefly summarise them here. Canonical N2pc analyses on the mean amplitudes confirmed TAC, with larger mean N2pc for TCCs (-0.08µV) than for NCCs (-0.02µV; plotted for V and AV conditions for reference, Figure 3). Both GFP and topographic analyses of the lateralised ERPs also showed evidence of TAC. However, there was no evidence for the modulation of TAC by contextual factors (in N2pc, *F*'s < 3, *p*'s > 0.1, in GFP and topography, all *F*'s < 1). For MSE, behavioural modulations were not mirrored in either canonical N2pc analyses, GFP or topographic analyses; similarly, there was no evidence for modulation of MSE by contextual factors in any of these measures. Neither the main effects of MR and DO, nor the 2-way MR x DO interaction were significant in the lateralised analyses of N2pc, GFP, or ERP topography (all *F*'s < 1). Map presence in the N2pc time-window was modulated jointly by the two contextual factors; we describe those results in SOMs.

678

679 ***The role of nonlateralised brain mechanisms***

680 The major part of our analyses focused on investigating the role of nonlateralised ERP
681 mechanisms in contributing to the interactions between the multisensory, stimulus-driven
682 and visual, goal-based attentional control and contextual control. To remind the reader, to
683 prevent these nonlateralised ERPs from being confounded by the presence of sound on AV
684 trials, we based the current analyses on difference ERPs indexing visual attentional control
685 under sound presence vs. absence. Specifically, these ERPs reflected the difference between
686 TCCAV and NCCAV conditions, and between TCCV and NCCV conditions (D_{AV} and D_V levels,
687 respectively, of the factor Target Difference). We focus the description of these results on
688 the effects of interest (see SOMs for full description of results).

689 The 2 × 2 × 2 rmANOVA on electrode-wise voltage analyses revealed that the main
690 effect of Target Difference was significant at 53–99ms and 141–179ms, where amplitudes
691 were larger D_{AV} (TCCAV – NCCAV difference) than for D_V (TCCV – NCCV difference). This
692 effect was further modulated by the multisensory relationship within the distractors, with a
693 2-way Target Difference × MR interaction, at the following time-windows: 65–103ms, 143–
694 171ms, and 194–221ms (all p's < 0.05). This effect was driven by semantically congruent
695 distractors showing larger amplitudes for D_{AV} than D_V within all 3 time-windows (65–97ms,
696 143–171ms, and 194–221ms; all p's < 0.05). No similar differences were found for
697 simultaneous distractors, and there were no other interactions.

698

699 *Interaction of TAC with contextual factors.* We next used EN analyses to investigate the
700 contribution of the strength- and topography-based nonlateralised mechanisms to the
701 interactions between TAC and contextual factors.

702 *Strength-based brain mechanisms.* A 2 × 2 × 2 rmANOVA on the GFP, mirroring the
703 electrode-wise analysis on ERP voltages, also showed the main effect of Target Difference
704 spanning a large part of the first 300ms post-distractor (19–213ms, 221–255ms, and 275–
705 290ms), where, just like the voltages, also GFP was larger for D_{AV} than D_V (all p's < 0.05).

706 In GFP, Target Difference interacted both with MR (23–255ms) and separately with
707 DO (88–127ms; see SOMs for full description), but most notable was the 3-way Target
708 Difference × MR × DO interaction, spanning 102–124ms and 234–249ms. We then followed
709 up this interaction with series of post-hocs to gauge the modulations of TAC (and MSE, see

710 below) by the two contextual factors. In GFP, MR and DO interacted independently of
711 Target Difference in the second time-window, which results we describe in the SOMs.
712 To gauge differences in the strength of TAC in GFP across the 4 contexts, we then
713 compared directly only the visually-elicited target differences across respective levels of the
714 2 contextual factors. Weakest GFPs were observed for simultaneous predictable distractors
715 (Figure 4A). They were weaker than GFPs elicited by simultaneous unpredictable distractors
716 (102–124ms and 234–249ms), and predictable congruent distractors (only in the later
717 window, 234–249ms).

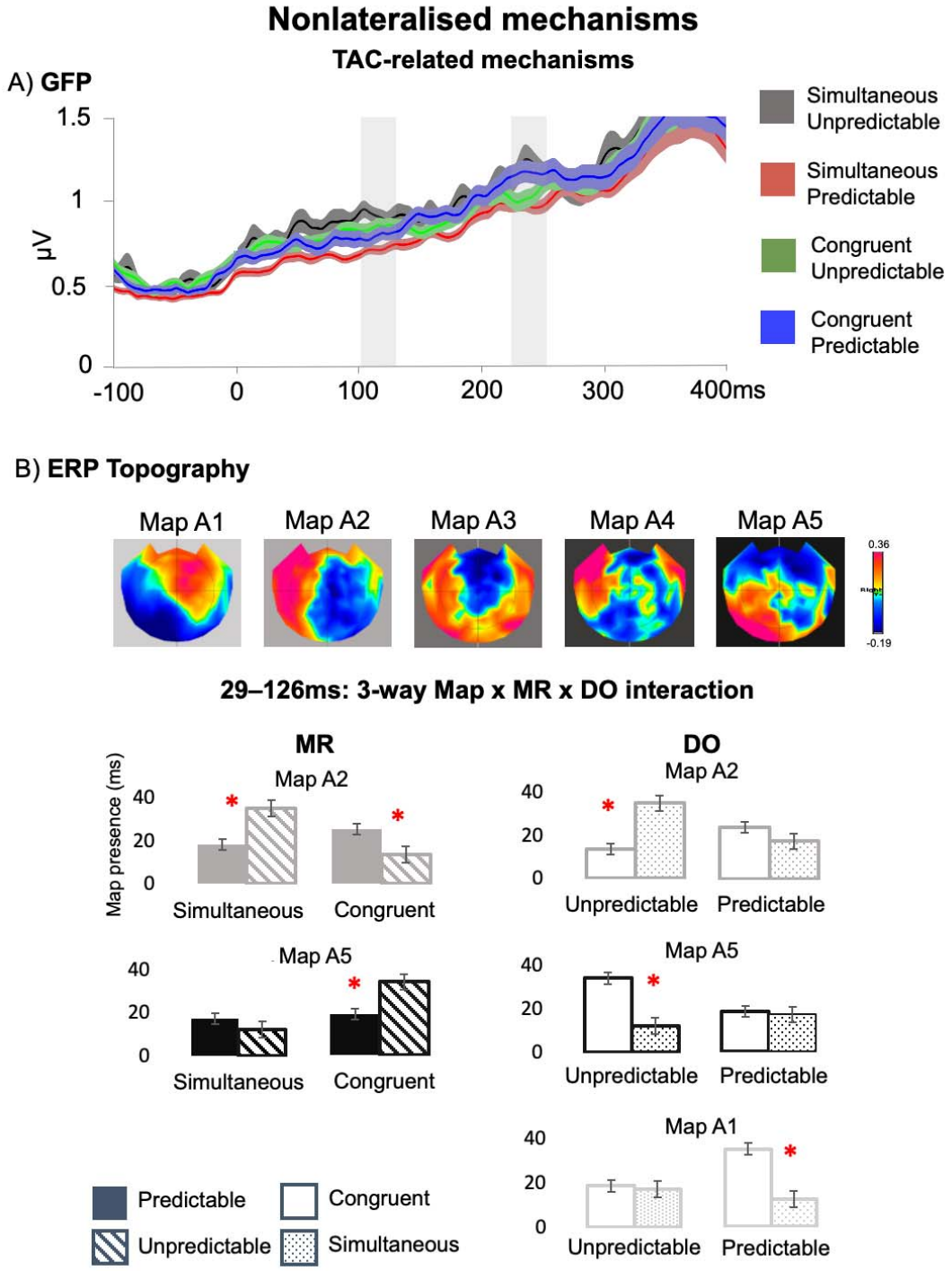


Figure 4. Nonlateralised GFP and topography results for the visual only difference ERPs (DV condition of Target Difference), as a proxy for TAC. **A)** Mean GFP over the post-distractor and pre-target time-period across the 4 experimental tasks (as a function of the levels of MR and DO that they represent), as denoted by the colours on the legend. The time-windows of interest (102–124ms and 234–249ms) are highlighted by grey areas. **B)** Template maps over the post-distractor time-period as revealed by the segmentation (Maps A1 to A5) are

shown in top panels. In lower panels are the results of the fitting procedure over the 29–126ms time-window. The results displayed here are the follow-up tests of the 3-way Map x MR x DO interaction as a function of MR (leftward panel) and of 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 SEMs.

Topography-based brain mechanisms. We focused the segmentation of the TAC-related topographic activity on the 0–300ms post-distractor time-window (before the target onset), which revealed 10 clusters that explained 82% of the global explained variance within the visual-only ERPs. This time-window was largely composed of distinct template maps across the 4 context conditions, except a time-window of 29–126ms post-distractor where segmentation revealed template maps shared across conditions. A 5 × 2 × 2 rmANOVA on the map presence over the 29–126ms post-distractor time-window revealed a 3-way Map x MR × DO interaction, $F_{(3.2,122)} = 5.3$, $p = 0.002$, $\eta_p^2 = 0.1$.

Follow-up tests focused on maps differentiating between the 4 conditions, first as a function of MR. Simultaneous predictable distractors, which elicited the weakest GFP, activated one map (map A2), and did so for reliably shorter time (18ms) compared to simultaneous unpredictable (35ms) distractors, $t_{(38)} = 2.96$, $p < .01$. The same map showed the opposite pattern for semantically congruent distractors, where it instead predominated responses to predictable (25ms) vs. unpredictable (14ms) distractors, $t_{(38)} = 2$, $p = 0.04$ (Figure 4B, leftward panel). In contrast with this pattern, a second map (map A5) predominated responses to semantically congruent unpredictable (34ms) vs. semantically congruent predictable (19ms) distractors, $t_{(38)} = 2.7$, $p = 0.04$ (Figure 4B, leftward panel).

Follow-ups conducted as a function of DO revealed that the simultaneous predictable distractors activated a third map (map A1), which predominated responses to these distractors (37ms) vs. semantically congruent predictable distractors (21ms), $t_{(38)} = 2.7$, $p < 0.01$ (Figure 4B, rightward panel). Map A2 also predominated responses to unpredictable simultaneous distractors (35ms) vs. unpredictable semantically congruent (14ms) distractors, $t_{(38)} = 3.7$, $p = 0.001$ (Figure 4B, rightward panel). And Map A5 again predominated responses to unpredictable semantically congruent (34ms), now compared to unpredictable simultaneous (12ms) distractors, $t_{(38)} = 3.7$, $p < 0.001$.

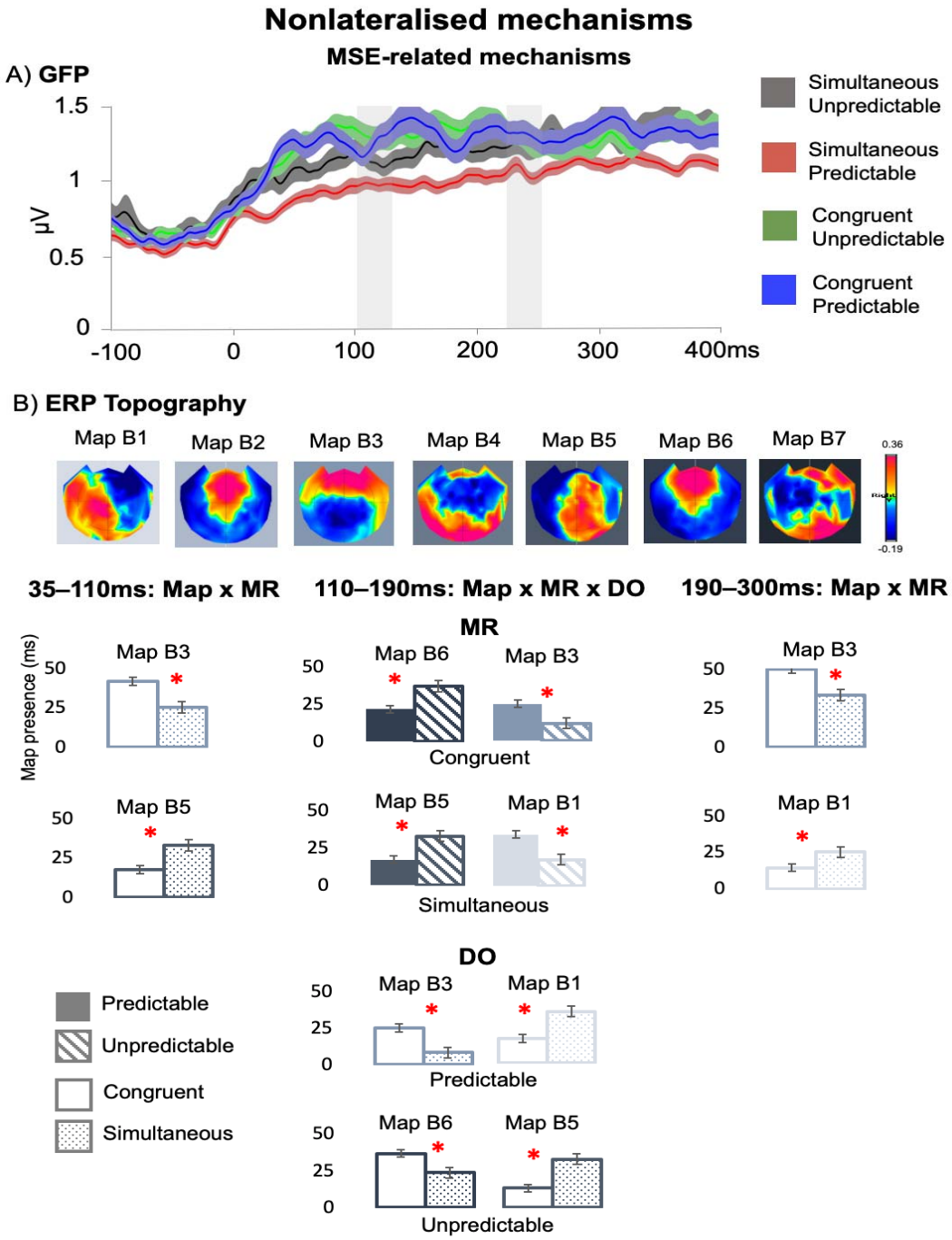


Figure 5. Nonlateralised GFP and topography results for the difference ERPs between the D_{AV} and D_V conditions of Target Difference, as a proxy for MSE. **A)** Mean GFP over the post-distractor and pre-target time-period across the 4 experimental tasks (as a function of the levels of MR and DO that they represent), as denoted by the colours on the legend. The time-windows of interest (102–124ms and 234–249ms) are highlighted by grey bars. **B)** Template maps over the post-distractor time-period as revealed by the segmentation (Maps A1 to A7) are shown on top. Below are the results of the fitting procedure over the three

764 time-windows: 35–110, 110–190, and 190–300 time-window. Here we display the follow-ups of the
765 interactions observed in each time-window: in 35–110 and 190–300 time-windows, the 2-wau Map x MR
766 interaction (leftward and rightward panels, respectively), and in the 110–190 time-window, follow-ups of the
767 3-way Map x MR x DO interaction as a function of MR and of DO (middle panel). Bars are coloured according
768 to the template maps that they represent. Conditions are represented by full colour or patterns per the
769 legend. Error bars represent SEMs.
770
771
772 *Interaction of MSE with contextual factors.* We then analysed the strength- and topography-
773 based nonlateralised mechanisms contributing to the interactions between MSE and
774 contextual factors.
775 *Strength-based brain mechanisms.* To gauge the AV-induced MS enhancements
776 between D_{AV} and D_V across the 4 contexts, we explored the abovementioned 2 × 2 × 2 GFP
777 interaction using a series of simple follow-up post-hocs. We first tested if the response
778 strength between D_{AV} and D_V was reliably different within each of the 4 contextual
779 conditions. We found that enhanced AV-induced target ERP responses characterised
780 primarily semantic distractors. That is, for both predictable and unpredictable semantically
781 congruent distractors, GFP were larger for D_{AV} than D_V across both earlier and later time-
782 windows. However, for simultaneous distractors, D_{AV} and D_V were larger only in predictable
783 contexts and only in the earlier (102–124ms) time-window; no differences were found at all
784 for unpredictable distractors.
785 Next, to gauge the AV-induced MS enhancements between D_{AV} and D_V across the 4
786 contexts, we calculated the subtraction between D_{AV} and D_V for each context, and
787 compared these across the respective levels of the 2 factors. AV-induced enhancements
788 were weaker for predictable simultaneous distractors than predictable semantically
789 congruent distractors (102–124ms and 234–249ms;
790 Figure 5A). The AV-enhancements for predictable simultaneous were weaker also
791 compared to unpredictable simultaneous distractors in both time-windows but these did
792 not meet the temporal criteria for statistical significance.
793 *Topography-based brain* mechanisms. We focused the segmentation of the MSE-
794 related topographic activity in the 0–300ms post-distractor time-window, which revealed 7
795 clusters that explained 78% of the GEV within the AV-V target difference ERPs. We fit the
796 data in three subsequent time-windows where there were clear topographic patterns, 35–

110ms, 110– 190ms, and 190– 300ms. To foreshadow the results, in the first and third time-window the MSE-related templates were modulated only by MR, while in the middle time-window – by both contextual factors.

In the first, 35–110ms time-window, the modulation of map presence by MR was evidenced by a 2-way Map × MR 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 predominated responses to semantically congruent (42ms) vs. simultaneous (25ms) distractors, $t_{(38)} = 4.3$, $p = 0.02$, whereas another map (map B5) predominated responses to simultaneous (33ms) vs. semantically congruent (18ms) distractors, $t_{(38)} = 4$, $p = 0.01$ (

Figure 5B leftward panel).

In the second, 110– 190ms time-window, map presence was modulated by both contextual factors, with a 3-way Map × MR × DO interaction, $F_{(2.6,99.9)} = 3.7$, $p = 0.02$, $\eta_p^2 = 0.1$.

Map B6 predominated responses to *semantically congruent unpredictable* (37ms) vs. semantically congruent predictable distractors (21ms, $t_{(38)} = 2.5$, $p = 0.02$; Figure **5**B middle top panel), and vs. simultaneous unpredictable distractors (24ms, $t_{(38)} = 2.3$, $p = 0.04$; Figure **5**B bottom middle panel). Map B5 predominated responses to *simultaneous unpredictable* (33ms) vs. simultaneous predictable distractors (17ms, $t_{(38)} = 2.6$, $p = 0.04$; Figure **5**B middle upper panel), and vs. semantically congruent unpredictable distractors (13ms, $t_{(38)} = 3.4$, $p = 0.002$; Figure **5**B middle lower panel). Map B3 predominated responses to *semantically congruent predictable* (25ms) vs. semantically congruent unpredictable distractors (12ms, $t_{(38)} = 2.2$, $p = 0.05$; Figure **5**B middle top panel), and vs. predictable simultaneous distractors (8ms, $t_{(38)} = 2.2$, $p = 0.005$; Figure **5**B middle lower panel). Finally, map B1 predominated responses to *simultaneous predictable* (35ms; Figure **5**B middle upper panel) vs. simultaneous unpredictable distractors (18ms, $t_{(38)} = 2.8$, $p = 0.01$), and vs. semantically congruent predictable distractors (17ms, $t_{(38)} = 2.8$, $p = 0.01$; Figure **5**B middle lower panel).

In the third, 190–300ms time-window, the 2-way Map × MR 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) predominated responses to semantically congruent (50ms) vs. simultaneous distractors (33ms), $t_{(38)} = 3.6$, $p = 0.08$, and another map (map B1) predominated responses to simultaneous (25ms) distractors vs. semantically congruent (14ms) distractors, $t_{(38)} = 2.3$, $p = 0.02$ (

Figure 5B rightward panel).

Discussion

Real-world environments are cluttered with competing information and necessitate effective attentional control in order to focus on relevant inputs and ignore unimportant inputs. In such environments, goal-based top-down and bottom-up processes interact with context-based processes of attentional control. Yet these types of control processes have traditionally been studied in isolation. Here, for the first time, we studied these control processes together, as they interact in multisensory settings. First, we studied how visual-top down control (as indexed by task-set contingent attentional capture, or TAC) interacts with contextual factors, like semantic relationships within distractor features and their temporal predictability. Second, we investigated how such interactions transpire for multisensory bottom-up attentional control (as indexed by multisensory enhancement of visual attentional capture, or MSE). Finally, we discuss our findings in the context of theoretical proposals regarding how contextual factors interact with visual attentional control, suggesting how these the interactions may occur in more naturalistic, real-world settings.

Interaction of task-set contingent attentional capture with contextual control

We revealed several new findings regarding how visual attentional control operates in settings that are rich in contextual and multisensory information. Most importantly, we found that visual top-down goal-based attentional control indexed by task-set contingent attentional capture (TAC) interacted with both contextual factors. At the level of behaviour, these interactions involved multisensory relationship (MR) and transpired by target-colour distractors capturing attention more strongly when they were simultaneous rather than semantically congruent. In turn, multisensory relationship did not modulate behavioural capture by nontarget-colour distractors. This pattern of results suggests that the brain utilises context-related information for goal-directed behaviour also in multisensory contexts.

The EN analyses of the nonlateralised ERPs across the whole post-distractor time-window both confirmed and enriched our behavioural findings. Notably, this contrasted with the canonical lateralised, N2pc-like brain mechanisms, which revealed no modulations of TAC by either contextual factor. First, in terms of findings in lateralised brain mechanisms, target difference ERPs from visual-only conditions (reflecting the TAC) showed especially attenuated responses (indexed by lower GFP) to simultaneous distractors that were predictable. They were reliably weaker than simultaneous *unpredictable* distractors and *semantically congruent* predictable distractors. Second, it is notable that these TAC–context interactions occurred both in the time-window of the N2pc, i.e., during the 'attentional selection of candidate target objects' (Eimer, 1996, 2014), but also earlier. Namely, the suppression of GFP to simultaneous predictable vs. simultaneous unpredictable distractors were observed already in the early, 102–124ms (and later, N2pc-like, 234–249ms time-window) time-window. Third, segmentation analyses showed that simultaneous predictable distractors activated two distinct configurations of brain sources (reflected by two distinct templates) already ~30ms post-distractor. One map (map A1) predominated responses to these distractors, compared to predictable semantically congruent distractors; notably, the second map (map A2) activated by simultaneous predictable distractors was involved in their processing to reliably lesser extent than that of simultaneous *unpredictable* distractors.

To summarise, target-colour visual distractors capture attention more strongly when they are simultaneous than semantically congruent. This effect co-occurred with a range of findings in nonlateralised brain mechanisms spanning both the attentional selection, N2pc-like time-window as well as much earlier time-windows. Simultaneous distractors, but only predictable ones, activated one network in the early time-window for longer period of time, and another network for reliably shorter period of time, compared to semantically congruent distractors. These network-based processes were followed by strength-related processes, where the same, simultaneous (again, predictable) distractors elicited the most attenuated responses (compared to both semantically congruent predictable and simultaneous unpredictable distractors). This cascade of modulations over nonlateralised brain mechanisms might offer a potential mechanisms whereby goal-directed attentional control utilises contextual information. While in our current paradigm, behaviourally these effects were more prominent for the semantic relationships within the distractors, our

multivariate EN analyses of the nonlateralised mechanisms revealed the importance of both semantic and temporal prediction-based processes, which transpire across both strength-based and network-based mechanisms, and from early on.

To our knowledge, present findings are the first demonstration of how goal-driven feature-based attentional control interacts with multiple sources of contextual control. They both confirm and extend existing understanding of how predictions affect perception and action, as well as attentional control. For one, the early (~30ms) suppressed responses to predictable target-colour (compared to unpredictable) stimuli are in line with the early-onset (~50ms post-stimulus) suppressions of visual ERPs to temporally predictable stimuli (C1; Dassanayake et al. 2016). Our EN analyses revealed that the earliest suppression at least in some contexts might be driven by weaker activation of relevant brain networks, which at later stage (~100ms post-stimulus) also involve suppressed brain responses. The latency of those later, strength-based suppressions are in line with attenuations of N1 activity for self-generated sounds (Baess et al. 2011; Klaffehn et al. 2019), and around the time when the brain promotes the processing of unexpected events according to recent theories (e.g., Press et al. 2020). Notably, the latency of our later, strength-based effects is also in line with the earliest latency (~115ms post-stimulus) of the goal-based suppression of salient (i.e., presented at the same time as targets) visual distractors (distractor positivity, Pd; Sawaki & Luck 2010). Van Moorselaar and Slagter (2019) showed that when such salient visual distractors appear in predictable locations, they elicit N2pc but no longer a Pd, suggesting that learning the location of the distractor enables the brain to effectively ignore it, without the need for active inhibition (see Matusz et al. 2016 for similar findings showing spatial predictability inhibiting attention shifts to irrelevant sounds). More recently, van Moorselaar et al. (2020b) showed that two types of predictions regarding the distractor, such as that of its location and of its feature, interact, with feature-based prediction attenuating location-based suppression. Notably, representation of the predictable distractor feature could be decoded already from pre-stimulus activity. While the two paradigms differ from each other, our results might potentially enrich those findings. Our task is not optimised for measuring Pd (as distractors appear at irrelevant moments in time), but the target-colour distractors therein reliably capture attention (Folk et al. 1992). As we have shown, semantic relationships might be a mechanism through which this capture can be at least partly suppressed. Thus, our findings suggest that, attentional control will flexibly

utilise all available contextual information - different types of predictions (spatial, feature-based, temporal, etc.; e. van Moorselaar & Slagter 2020b) and semantic relationships - to facilitate goal-directed processing.

While this was not the main aim of the present study, it is notable that we observed the task-set contingent attentional capture despite the addition to the paradigm of both the multisensory manipulation (as in Matusz & Eimer 2011) and the two contextual factors. As in traditional literature, target-colour distractors triggered reliable behavioural capture effects and enhanced mean N2pc amplitudes compared to nontarget-colour distractors (Eimer et al., 2009, Eimer & Kiss, 2008, Kiss et al., 2008a; 2008b). This suggests that this form of visual top-down attentional control of object attentional selection, based on target features (like colour), is an effective, robust form of attentional control, interacting with both multisensory and contextual mechanisms.

Interaction of multisensory enhancement of attentional capture with contextual control

Our results offer an important replication of the multisensory enhancement of attentional capture (MSE) effect demonstrated in a rigorous paradigm by Matusz and Eimer (2011). In the current setup and unlike the original study (Matusz & Eimer 2011), multisensory enhancement of attentional capture did interact with colour cue. Given the strong feature-specific visual top-down attentional demands of the task, the evidence of behavioural multisensory enhancement supports the idea put forward by Matusz and Eimer (2011) that multisensory enhancements of visual attentional capture can occur involuntarily and independently of task demands (see also Matusz et al. 2015). We did indeed observe again the two-way interaction that let the authors make such a conclusion. In contrast, here, this effect was indeed dependent on the colour matching between target and the cue. Namely, the MSE of attentional behavioural capture was found for target-colour distractors, but not for the nontarget-colour distractors. However, when we tested for this 3-way interaction within each of the experiments, this interaction was no longer present. It was not present even in Experiment 2 that was the closest to the original study, as in both studies distractors were temporally predictable and simultaneous. While care is needed when interpreting these null results, one possibility is that in fact in the real world bottom-up multisensory and top-down goal-directed attentional control processes may interact - where goal-based attentional control processes would also utilise the increased salience of simultaneous

multisensory objects. However, it is possible that such interactions are subtle and/or the current paradigm is not optimal for gauging them.

While we again found multisensory enhancement of visual attentional capture (which interacted with TAC), its interactions with the two contextual factors did not transpire at the level of behaviour. Notably, the lateralised ERPs that are traditionally analysed in attentional control research and the canonical N2pc analyses that gauge them appear not to be sensitive to multisensory enhancements altogether. To date, there is weak evidence for distractor-based bottom-up effects of multisensory enhancement in lateralised ERPs (van der Burg et al. 2011). Lateralised ERPs (and N2pc in particular) more readily reveal influences of top-down goal-based multisensory processes (Matusz & Eimer 2013; Matusz et al. 2019). The contribution of lateralised brain mechanisms to the interactions between multisensory enhancements of visual attention capture and contextual processes might be equally limited. As in the case of top-down visual attentional control, nonlateralised brain processes contribute to those interactions more prominently. We discuss those findings below in detail.

First, similarly to visual attentional control, nonlateralised brain mechanisms were involved in multisensory processing from early on after distractor onset and included both strength-based and network-based (nonlateralised) brain mechanisms. Specifically, the Target Difference factor that was part of the 3-way interaction with the two contextual factors, contained both visual and multisensory target-difference (TCC and NCC difference). When we analysed the AV–V difference ERPs as a proxy to measure processing related to multisensory enhancement, we identified reliably stronger (indexed by GFP) multisensory than visual ERPs for simultaneous predictable (but not unpredictable) distractors (in the earlier 102-124ms time-window), and for both predictable and unpredictable semantically congruence distractors, across both earlier and later (234-249ms) time-windows. Notably, in both time-windows audiovisually-induced enhancement of ERPs were reliably larger for predictable and semantically congruent than predictable and simultaneous distractors. This pattern of results provides a novel insight into attentional selection of audio-visual objects. Specifically, even when such objects are irrelevant to the task, multisensory relationships modulate their processing, they do so from earlier, perceptual to later, cognitive stages, and through interactions with other contextual factors, such as stimulus (temporal) predictability.

994 Second, network-related brain mechanisms contributed to the interactions between
995 multisensory enhancements and the two contextual factors and did so throughout the post-
996 distractor time-period. Semantic relationship played a predominant (but not absolute) role
997 in those nonlateralised response interactions. The AV–V difference ERPs in the first, early
998 (35–110ms time-window) and late, N2pc-like (190–300ms) time-windows were both
999 modulated exclusively by multisensory relationships. Therein, the same map predominated
1000 the responses to semantically congruent as compared to simultaneous distractors, while a
1001 different map in the two time-windows was predominated those to simultaneous
1002 distractors. This result suggests that both early perceptual processing and later, attentional
1003 selection processing is modulated by meaning (vs lack thereof). The fact that this was the
1004 same template map (map B3) suggests that a single network of brain sources was activated
1005 for a longer time period at both stages to promote the processing of meaningful
1006 multisensory stimuli. In contrast, in the second, 110–190ms time-window, the AV–V
1007 difference ERPs was modulated by both multisensory relationship and distractor
1008 predictability. The same "semantic" template map was present also in this mid-latency time-
1009 window, now predominating the response for semantically congruent predictable versus
1010 unpredictable distractors. Notably, the two maps that were present in the mid-latency time-
1011 window for simultaneous distractors and modulated further as a function of DO, were also
1012 present in the early time-window (map B5) and in the late time-window (map B1). Together,
1013 these findings suggest that from early on distinct types of networks are activated for
1014 meaningful and meaningless stimuli, even when those stimuli are not relevant to the
1015 current task and goals. We know that semantic relationships can improve object recognition
1016 (e.g. animals) even when only one of the senses is task-relevant (Yuval-Greenberg &
1017 Deouell, 2007). Similar benefits are found in unisensory (visual, auditory) in a continuous
1018 old/new recognition task (Murray et al. 2004; Matusz et al. 2015a; Thelen et al. 2015),
1019 where benefits from mere multisensory simultaneity shows strong inter-individual
1020 variability (Thelen et al. 2014). The current results enrich those findings by showing that in
1021 naturalistic settings, object processing (attentional selection) will be jointly affected by
1022 bottom-up multisensory processes and one's goals, but also contextual factors, such as
1023 stimulus predictability. Furthermore, these influences likely transpire across both task-
1024 relevant and task-irrelevant objects, contrary to what traditional research focused
1025 predominantly on target processing might suggest.

1026

1027 **How we pay attention in naturalistic settings**

1028 It is well-recognised that the brain can process and act on a very limited portion of the
1029 incoming stimuli. One mechanism to aid resolution of this challenge is goal-based
1030 attentional control that promotes the processing of stimuli matching our currently set
1031 behavioural goals, based on stimulus features or their location in space or time (Folk et al.
1032 1992; Gazzeley & Nobre 2012; Noonan et al. 2018). It is now well-established that the brain
1033 likewise utilises regularities in the outside world to help itself reduce the complexity of the
1034 external stimulation, by creating expectations that constrain the potential interpretation of
1035 the incoming stimuli (Summerfield & Egner 2009; Press et al. 2020). It is likewise
1036 increasingly recognised that the semantic relationships between objects and events in
1037 everyday situations stored in the long-terms semantic memory impact stimulus processing
1038 and attentional selection (Summerfield et al. 2006; Peelen & Kastner 2014). While these
1039 theoretical accounts have been invaluable in clarifying the mechanisms that support goal-
1040 directed behaviour in everyday situations, these mechanisms (and their interactions with
1041 goal-based attentional control) have been studied systematically in isolation in the visual
1042 (Summerfield & Egner 2009; van Moorselaar & Slagter 2019; van Moorselaar et al. 2020b),
1043 or the auditory (e.g., Chennu et al. 2013) domain. Here, we showed that more traditionally
1044 studied attentional control mechanisms, such as feature-based goal-driven attention control
1045 (indexed by TAC) and salience-based (here, multisensory) attention control mechanisms
1046 interact with contextual mechanisms including both stimulus predictability and semantic
1047 relationships.

1048 In line with existing visual literature, visual goal-based attentional control interacted,
1049 behaviourally, with semantic relationships induced in the audio-visual stimuli half-way
1050 through the experiment. Namely, distractors that matched the colour of the target and so
1051 the goals captured attention *less* when they contained meaning, compared to when only
1052 simultaneity joined visual and auditory distractor stimuli. This is an important finding as it
1053 suggests that goal-based control by stimulus feature (like colour) may, indeed, involuntarily
1054 extend to even task-irrelevant stimuli, as postulated originally by Folk and colleagues (1992)
1055 and replicated numerous times. Here, we revealed that equipping task-irrelevant stimuli
1056 with meaning may provide the brain with an otherwise unavailable mechanism to improve
1057 goal-based attentional control. In our task, distractors appear at task-irrelevant moments in

time and uninformative about subsequent spatial location of the target, and so it is beneficial to inhibit their processing if possible. Thus, more generally, in everyday situations, where many objects likely match features with the objects we are currently searching, their ability to capture our attention might be *de facto* weaker than what the traditional literature on visual attentional control suggests.

Our EEG results demonstrated that semantically congruent and simultaneous stimuli preferentially activate distinct brain networks as early as 30ms after onset of the distractors. Therein, the temporal predictability further differentiated the brain networks activated by semantically congruent and simultaneous distractors; however, modulations stemming from predictions were not reliably visible in our behavioural measures of attentional capture. It is important to note that context-based interactions of meaning and predictability influenced the brain networks activated by target-congruent and nontarget-congruent visual distractors as early as ~30ms – before any attentional selection-related activity is expected to happen. While our analyses prevent strong conclusions, it is possible that these differently activated early brain networks led to the stronger brain responses at both 100ms and 250ms post-distractor, the latter effects matching (later parts of) the time-window of spatially-selective responses indexed by the N2pc. These findings suggest that, in everyday environments, attentional control will utilise all available contextual information about visual stimuli and their context (semantic relationships; predictions) in the service of goal-related processing. These influences will occur already during early stages of brain activity, with at least some of those modulations transpiring also at the level of behaviour. In other words, meaning may be an important - yet strongly under-researched - type of contextual information utilised by goal-based attentional control in both unisensory (Chennu et al. 2013; Peelen & Kastner 2014; Retsa et al. 2020) and multisensory (Matusz et al. 2016; Sarmiento et al. 2016; Retsa et al. 2018) settings. Such findings provide important new insights into how attentional control operates in real-world environments, where space- and feature-based attentional control interact routinely with contextual processes. Notably, all these influences that we have observed were measured on task-*irrelevant* stimuli, suggesting an even more complex interplay between goal-based attentional control and distraction than one proposed by Folk and colleagues (1992), but one where both task-based and contextual information is utilised simultaneously to control our attention to both

task-relevant and task-irrelevant objects (ten Oever et al. 2016; Noonan et al. 2018; van Moorselaar & Slagter 2020a).

Our nonlateralised global EEG analyses revealed robust interactions between bottom-up multisensory attentional control and contextual influences. Distinct brain networks were involved predominantly in attention towards semantically congruent and simultaneous multisensory stimuli. These modulations began early on during distractor processing and transpired throughout the time-window of N2pc-like attentional processes. While these interactions were not visible at the level of behaviour, they are nevertheless notable. This is because these interactions occurred despite the paradigm involving stringent, visual feature-specific attentional control settings and thus was more optimised for abolishing multisensory enhancement of visual attentional capture. Not only was multisensory enhancement not eliminated, it was replicated (Matusz & Eimer 2011) and shown to interact with both stimulus meaning and stimulus predictability. This pattern of results suggest that it is a robust finding and thus needs to be considered when aiming to study attentional control and information processing more generally in naturalistic settings (e.g. such as during viewing movies or listening to audiostories) or veridical real-world environments (e.g. the classroom or the museum). Our results suggest that in such contexts person's goals will interact with the contextual information (stimulus meaning, different forms of prediction) but also with dynamic changes in stimulus perceptual salience driven by moment-to-moment bottom-up multisensory enhancement. The strength of these modulations in everyday situations can be at least partly gauged through naturalistic studies in the laboratory (van Atteveldt et al. 2018; Matusz et al. 2019a), such as the present one, where task-relevance, attention demands and context influences are systematically varied together with manipulations of the engaged senses. In more naturalistic settings, the same effects could be measured, and thus controlled, by systematically comparing, where possible, unisensory and multisensory stimulus or event presentations. Such a cyclical (Matusz et al. 2019a) approach involving systematic testing of hypotheses across laboratory and outside settings is most likely to bring us closer to successfully bridging the laboratory-developed (typically visually-derived) hypotheses and theories of attentional control and other cognitive functions and the recently accumulating findings from such naturalistic setups. More generally, our findings suggest that conceptualisations of ecological validity and naturalistic settings (Peelen et al. 2014; Shamay-Tsoory & Mendelsohn 2019; Vanderwal

et al. 2019; Eickhoff et al. 2020; Cantlon 2020) could benefit from addition to more traditionally invoked components (e.g. attention demands, context, socialness) also that of multisensory influences.

References

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and perceiving*, *23*(1), 3-38.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, *166*(3-4), 559-571.

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.

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., ... & Dikker, S. (2019). Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography study. *Journal of cognitive neuroscience*, *31*(3), 401-411.

Biasiucci, A., Franceschiello, B., & Murray, M. M. (2019). Electroencephalography. *Current Biology, 29(3),* R80-R85.

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational intelligence and neuroscience*, *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. *Journal of Neuroscience*, 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.

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibánez, A., ... & Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience*, *33*(27), 11194-11205.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*, *36*(1), 28-71.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Correa, Á., Lupiáñez, J., & Tudela, P. (2005). Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychonomic bulletin & review*, 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. *International Journal of Psychophysiology*, 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. *Frontiers in integrative neuroscience*, *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. *Frontiers in psychology*, 1, 177.

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research, 1242,* 136–50. https://doi.org/10.1016/J.BRAINRES.2008.03.071

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

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. *Electroencephalography and Clinical Neurophysiology, 99(3),* 225–234.

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

Eimer, M., Kiss, M., Press, C., & Sauter, D. (2009). The roles of feature-specific task set and bottom-up salience in attentional capture: An ERP study. *Journal of Experimental Psychology: Human Perception and Performance, 35(5),* 1316–1328.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5), 7-7.

Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & psychophysics*, *64*(5), 741-753.

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

Fort, A., Delpuech, C., Pernier, J., and Giard, M. H. (2002). Early auditory–visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research,* 14, 20–30. doi: 10.1016/S0926-6410(02)00058-7

Gaspelin, N., & Luck, S. J. (2019). Inhibition as a potential resolution to the attentional capture debate. *Current opinion in psychology*, *29*, 12-18.

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

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, 25(20), 5004-5012.

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

Girelli, M., & Luck, S. J. (1997). Are the same attentional mechanisms used to detect visual search targets defined by color, orientation, and motion? *Journal of Cognitive Neuroscience*, *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 and language*, *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.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28(2),* 240-244.

Hickey, C., Di Lollo, V., & McDonald, J. J. (2008). Target and distractor processing in visual search: Decomposition of the N2pc. *Visual Cognition*, *16*(1), 110-113.

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

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65-70.

Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N. Y., Vu, A. T., & Gallant, J. L. (2016). Decoding the semantic content of natural movies from human brain activity. *Frontiers in systems neuroscience*, *10*, 81.

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, *15*(3), 548-554.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerbral Cortex* 18, 1560–1574.

Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science*, *12*(5), 176-180.

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 topography*, *27*(1), 72-83.

Kuo, B. C., Nobre, A. C., Scerif, G., & Astle, D. E. (2016). Top–Down activation of spatiotopic sensory codes in perceptual and working memory search. *Journal of cognitive neuroscience*, *28*(7), 996-1009.

Lehmann D, Skrandies W (1980): Reference-free identification of components of checkerboard evoked multichannel potential fields. *Electroencephalography in Clinical Neurology, 48*, 609–621.

Lehmann, D., Ozaki, H., & Pal, I. (1987). EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and clinical neurophysiology*, *67*(3), 271-288.

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. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 509.

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. *Journal of Experimental Psychology: Human Perception and Performance, 20(5),* 1000–1014.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001-1010.

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.

Matusz, P. J., Wallace, M. T., & Murray, M. M. (2020). Multisensory contributions to object recognition and memory across the life span. In *Multisensory Perception* (pp. 135-154). Academic Press.

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019a). Are We Ready for Real-world Neuroscience?. *Journal of cognitive neuroscience*, *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. *Journal of cognitive neuroscience*, *31*(3), 412-430.

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. *European Journal of Neuroscience*, *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.

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. *Psychonomic bulletin & review*, *18*(5), 904.

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 EL, Coull JT, Nobre AC. (1999). Orienting atten- tion in the time domain: modulation of potentials. *Brain, 122,* 1507-18.

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 in Neurosciences*, *39*(8), 567-579.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain topography*, *20*(4), 249-264.

Murray, M.M., Foxe, J.J. & Wylie, G.R. (2005) The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage, 27,* 473–478.

Murray, M.M., Michel, C.M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S. & Schnider, A. (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage, 21,* 125–135.

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological science*, 13(5), 416-424.

Noonan, M. P., Crittenden, B. M., Jensen, O., & Stokes, M. G. (2018). Selective inhibition of distracting input. *Behavioural brain research*, *355*, 36-47.

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends in cognitive sciences*, *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. *Electroencephalography and clinical neurophysiology*, *66*(1), 75-81.

Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, *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. *European Journal of Neuroscience*, 31(10), 1772-1782.

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.

Retsa, C., Matusz, P. J., Schnupp, J. W., & Murray, M. M. (2018). What's what in auditory cortices?. *NeuroImage*, *176*, 29-40.

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *Journal of Neuroscience*, *38*(34), 7452-7461.

Rohenkohl, G., Gould, I. C., Pessoa, J., & Nobre, A. C. (2014). Combining spatial and temporal expectations to improve visual perception. *Journal of vision*, 14(4), 8-8.

Rowland, B. A., & Stein, B. E. (2007). Multisensory integration produces an initial response enhancement. *Frontiers in Integrative Neuroscience,* 1(4).

Sarmiento, B. R., Matusz, P. J., Sanabria, D., & Murray, M. M. (2016). Contextual factors multiplex to control multisensory processes. Human brain mapping, 37(1), 273-288.

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: an ecological approach to brain and behavior research. *Perspectives on Psychological Science*, *14*(5), 841-859.

Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, & Psychophysics*, *72*(6), 1455-1470.

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.

Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K., & Chait, M. (2017). Is predictability salient? A study of attentional capture by auditory

patterns. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714), 20160105.

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human perception and performance*, *38*(5), 1105.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, *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. *Journal of Cognitive Neuroscience,* 17, 1098–1114.

Ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences*, *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. *Experimental brain research*, 234(5), 1307-1323.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics, 50(2),* 184–193.

Thelen, A., Talsma, D., & Murray, M.M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition 138,* 148–160.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520-522.

Tovar, D. A., Murray, M. M., & Wallace, M. T. (2020). Selective enhancement of object representations through multisensory integration. *Journal of Neuroscience*. In press. DOI: https://doi.org/10.1523/JNEUROSCI.2139-19.2020

Tivadar, R. I., & Murray, M. M. (2019). A primer on electroencephalography and event-related potentials for organizational neuroscience. *Organizational Research Methods*, *22*(1), 69-94.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12(1),* 97–136.

Turoman, N., Tivadar, R. I., Retsa, C., Maillard, A. M., Scerif, G., & Matusz, P. J. (2020). The development of attentional control mechanisms in multisensory environments. *BioRxiv*. DOI**:** https://doi.org/10.1101/2020.06.23.166975

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. *Developmental neuropsychology*, *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 Learning Research,* 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.

Vanderwal, T., Eilbott, J.; Castellanos, F. X (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience, 36*, 100600. DOI: 10.1016/j.dcn.2018.10.004.

van Moorselaar, D., & Slagter, H. A. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, *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.

van Moorselaar, D., & Slagter, H. A. (2019). Learning what is irrelevant or relevant: Expectations facilitate distractor inhibition and target facilitation through distinct neural mechanisms. *Journal of Neuroscience*, *39*(35), 6953-6967.

Vaughan Jr, H. G. (1982). The neural origins of human event-related potentials. *Annals of the New York Academy of Sciences*, *388*(1), 125-138.

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. *Journal of Cognitive Neuroscience*, *27*(11), 2299-2307.

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience*, *27*(5), 1090-1096.

Acknowledgements and data sharing statement

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., and the National Centre of Competence in research project "SYNAPSY, The Synaptic Bases of Mental Disease" [project 51AU40_125759]). P.J.M. and M.M.M. are both supported by Fondation Asile des Aveugles. 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.

Author contributions

Nora Turoman: Investigation, Formal analysis, Data curation, Software, Visualisation, Writing - original draft.

Ruxandra I. Tivadar: Software, Writing - review & editing.

Chrysa Retsa: Software, Writing - review & editing.

Pawel J. Matusz: Conceptualization, Funding acquisition, Methodology, Resources, Formal analysis, Software, Supervision, Writing - review & editing.

Micah M. Murray: Funding acquisition, Methodology, Resources, Formal analysis, Software, Supervision, Writing - review & editing.

Conflict of interest statement

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.