Visual periodicity reveals distinct attentional signatures for face and non-face categories

Genevieve L. Quek[1] & Adélaïde de Heering[2]

****

[1] The MARCS Institute for Brain, Behaviour and Development,
Western Sydney University, Sydney, Australia.

[2] Unité de Recherche en Neurosciences Cognitives (UNESCOG), Center for Research in Cognition & Neurosciences (CRCN), Université libre de Bruxelles (ULB), Brussels, Belgium.

Correspondence to:
Genevieve L. Quek
The MARCS Institute for Brain, Behaviour and Development
Western Sydney University
Westmead Innovation Quarter
160 Hawkesbury Rd, Westmead NSW 2145
E-mail: g.quek@westernsydney.edu.au

****

ACKNOWLEDGMENTS
The authors thank Daniela Ferreira Mesquita and Sofia Bonsignore for helping collecting the data and Fabienne Chetail for experimental programming.

COMPETING INTERESTS
The authors declare no competing interest.

AUTHORS' CONTRIBUTION
Both authors contributed to the experimental design, data analysis, and manuscript preparation. AdH oversaw participants testing.

ABSTRACT

Observers can selectively deploy attention to regions of space, moments in time, specific visual features, and individual objects. One level higher, we can also attend to particular high-level categories – for example, when keeping an eye out for dogs while jogging. Here we exploited visual periodicity to examine how the focus of category-based attention serves to differentially enhance and suppress selective neural processing of face and non-face categories. We combined electroencephalography (EEG) with an adaptation of the established Fast Periodic Visual Stimulation (FPVS) paradigm designed to capture selective neural responses for multiple visual categories contained in the same rapid stream of object images (faces and birds in Exp 1; houses and birds in Exp 2). Our key finding was that the pattern of attentional enhancement and suppression for face-selective processing is unique compared to non-face object categories. Specifically, where attending to non-face objects provides a strong boost their selective neural signals during the 300-500ms window, attentional enhancement of face-selective processing is both earlier and comparatively more modest. Moreover, only the selective neural response for faces appears to be actively suppressed when observers attend towards an alternate visual category. These results underscore the special status that human faces hold within the human visual system, and highlight the utility of visual periodicity as a powerful tool for indexing selective neural processing of multiple visual categories contained within the same visual sequence.

Keywords: faces, vision, selective attention, EEG, frequency tagging, object recognition

INTRODUCTION

With sensory input surging through the human visual system during our every waking moment, effective behaviour demands a mechanism for prioritizing relevant and important information within this immense data stream (Summerfield & Egner, 2009). Selective attention is the process by which the brain filters incoming information on the basis of both bottom-up (i.e., visual salience) and top-down factors (i.e., relevance to current goals) (Corbetta & Shulman, 2002; Johnston & Dark, 1986), prioritizing the selected input for subsequent higher processing (Allport, 1987). Where bottom-up attention is deployed rapidly and reflexively in response to salient visual events/features/stimuli (e.g., a peripheral flash) (H. J. Müller & Rabbitt, 1989; Posner, 1980), top-down attention is comparatively slower and requires more volitional effort on our part – hence the idiom "to pay attention" (Baluch & Itti, 2011). At the same time, prioritising what is most relevant to our real-world goals is often controlled by a dynamic interplay between bottom-up and top-down factors, wherein feature-based attentional sets (Maunsell & Treue, 2006) are subject to influence by expectations and knowledge about the world (Gayet & Peelen, 2022; Yeh & Peelen, 2022). Thus, if you lose your dog at the park, you might search for black, smallish objects, preferentially deploying this search template within the region of space where the target is most probable (e.g., on the grass, rather than in the lake).

In addition to individual visual features (Maunsell & Treue, 2006) and their conjunctions (Egeth, Virzi, & Garbart, 1984), visual selection can also be performed at the level of object categories (Addleman, Rajasingh, & Störmer, 2022; Battistoni, Kaiser, Hickey, & Peelen, 2018; Nako, Wu, Smith, & Eimer, 2014; Quek, Nemrodov, Rossion, & Liu-Shuang, 2018; Stein & Peelen, 2017; Störmer, Cohen, & Alvarez, 2019). One category that holds unique importance for humans is faces, which convey a great deal of meaningful information about age, sex, identity, emotion, and more, that guides effective behaviour in social environments. In particular, the human visual system appears remarkable adept at what has been referred to as generic face recognition (de Heering & Rossion, 2015; Jacques, Retter, & Rossion, 2016; Quek, Liu-Shuang, Goffaux, & Rossion, 2018; Quek, Rossion, & Liu-Shuang, 2021), that is, categorising a face *as a face.* Subtended by occipitotemporal regions (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Kanwisher, McDermott, & Chun, 1997; Le Grand, Mondloch, Maurer, & Brent, 2003; Liu, Harris, & Kanwisher, 2002), this ability develops remarkably quickly during infancy (de Heering & Rossion, 2015; Rekow et al., 2021) and arises based on very coarse visual input (Crouzet, Kirchner, & Thorpe, 2010; Crouzet & Thorpe, 2011; Quek, Liu-Shuang, et al., 2018; Quek et al., 2021). In fact, we are so good at detecting faces that we sometimes see a face when in fact, there is no face present, as in the case of face pareidolia (Wardle, Paranjape, Taubert, & Baker, 2022; Wardle, Taubert, Teichmann, & Baker, 2020).

Given its status as a uniquely robust brain function, it is somewhat surprising to find that the interplay between face processing and selective attention has historically been examined under sparse and simplified conditions (e.g., binary classification of a spatiotemporally isolated stimulus, e.g., face or house). Moreover, attention to faces has largely been operationalised as a unitary construct, in which faces function as either the actively attended or explicitly ignored category (Downing, Liu, & Kanwisher, 2001; Engell & McCarthy, 2010; Holmes, Vuilleumier, & Eimer, 2003; Vuilleumier, Armony, Driver, & Dolan, 2001; Wojciulik, Kanwisher, & Driver, 1998). Rapid visual presentation designs that exploit periodicity (de Heering & Rossion, 2015; Liu-Shuang, Norcia, & Rossion, 2014; Rossion, Torfs, Jacques, & Liu-Shuang, 2015) break with this standard, offering a way to quantify attentional effects for face-selective processing under high-competition conditions (i.e., distinguishing naturalistic images of faces amid a wide variety of other object categories). This approach recently revealed the degree to which face-selective signals can be respectively enhanced and suppressed (relative to a neutral baseline) by actively attending to faces as a category, or else towards another high-level category within the sequence (e.g., guitars, Quek, Nemrodov, et al., 2018).

At present, the relative contribution of enhancement and suppression components of category-based attention has only been documented using faces, leaving the attentional profile associated with selective visual processing of non-face categories comparatively unknown. On the one hand, perhaps when visual competition is high, task-relevance bestows a similar attentional benefit on selective visual

processing regardless of which specific high-level category is the focus of current task demands. On the other hand, as a high-priority stimulus within the human visual system, faces could well be expected to exhibit a unique attentional profile compared with other visual categories (Finkbeiner & Palermo, 2009; Lueschow et al., 2004; Reddy, Reddy, & Koch, 2006; Reddy, Wilken, & Koch, 2004).

To contrast how category-based attention modulates selective visual processing of face and non-face stimuli in the same group of observers, we adapted the fast periodic visual stimulation paradigm to index selective neural signals associated with multiple high-level visual categories contained within the same image sequence. We used an 'interlaced' periodicities approach inspired by both frequency-based techniques for targeting auditory (Jaeger, Bleichner, Bauer, Mirkovic, & Debener, 2018) and lower-level visual processing (Brummerloh, Gundlach, & Müller, 2019; M. M. Müller et al., 2006). Here, two critical categories of interest (faces and birds) were interlaced at distinct periodicities within a 6Hz sequence of various object images, effectively tagging each critical category with its own pre-specified stimulation frequency. We recorded 64 channel EEG while observers viewed these sequences, deriving separable neural indices of face- and bird-selective visual processing, each of which could be quantified with respect to the focus of category-based attention within that sequence. We manipulated category-based attention across blocks, evaluating the face- and bird-selective responses in the *attend-Faces* and *attend-Birds* conditions against those same signals elicited during a neutral-attention condition in which observers had to detect when the central fixation cross changed colour.

To anticipate our results, we observed distinct profiles of attentional enhancement and suppression for selective visual processing associated with individual high-level categories (i.e., faces/birds/houses). In Exp 1, we found that attending to faces enhanced their category-selective neural response to a much lesser extent than was evident for birds. Conversely, attentional suppression induced by attending to the alternate critical category was pronounced for face-selective processing, but non-existent for bird-selective processing. In Exp 2, we confirmed that this pattern across face and non-face categories did not arise as an artefact of the specific stimulation frequencies by substituting houses for faces. Temporal analyses revealed distinct time courses of attentional effects across the different categories; where enhancement of face-selective processing was most evident before 200 ms, the bird- and house-selective waveforms were not facilitated by attention until much later (i.e., 300-500ms).

METHODS

Participants

The study was carried out in accordance with the guidelines and regulations of the Research Ethics Board of the department of Psychology of the Université libre de Bruxelles (Belgium) (Reference - B406201734083). We tested two different groups of healthy adult volunteers with no reported psychiatric or neurological disorders. Twenty-one participants took part in Exp 1 in exchange for financial compensation, and 23 participants took part in Exp 2 in exchange for course credits. All were right-handed, reported having normal or corrected-to-normal vision, and gave written informed consent prior to the start of the experiment. Several participants' EEG data were discarded because of excessive noise (1 from Exp 1, 3 from Exp 2), with the final sample in each experiment consisting of 20 individuals (Exp 1 = 3 males; mean age 20±1.8 (SD) years; Exp 2 = 5 males; mean age 24±4 (SD) years).

Stimuli

We used a previously described set of 200 x 200-pixel colour images (de Heering & Rossion, 2015; Jacques et al., 2016; Rossion et al., 2015) that included 48 faces, 24 birds, 48 houses, and a further 176 images drawn from various other visual categories (e.g., chairs, guitars, flowers, animals…). The subject of each image remained unsegmented from its original natural background, such that the image set varied widely in terms of colour composition, viewpoint, lighting condition, etc. (see Figure 1A for examples). Each image subtended 6.86 x 6.86 degrees of visual angle when viewed at the distance of 50 cm from the computer screen (1920x1080 Lenovo monitor with a 60Hz refresh rate).

Apparatus

The experiments took place in a darkened, sound-attenuated room, where participants sat before a computer monitor (refresh rate 60Hz) that displayed a uniform grey background throughout the recording session. Image sequence presentation was programmed in Psychtoolbox (MATLAB, The MathWorks). We recorded scalp electrical activity using a 64-channel BioSemi ActiveTwo system (Amsterdam, Netherlands) with a sample rate of 1024Hz. During recording setup, we held individual electrode offsets below ±30 µV by softly abrading the participant's scalp with a blunt plastic needle and insulating the electrode tip with gel. The experimenter monitored the participant's eye movements throughout Exp 1 via 4 external electrodes, one placed at the outer canthi of each eye, and one placed immediately above and below the right orbit. An external camera achieved the same purpose in Exp 2 and provided the experimenter with online images of the participant's eye behaviour.

Design

Sequence Composition

Each sequence began with a "ready?" message. Once the experimenter confirmed that the EEG trace was artefact-free, they invited the participant to press the spacebar to launch the sequence. A black fixation cross appeared in the centre of the screen for a duration of 1 to 5 seconds, after which the participant experienced a rapid stream of various object images (e.g., vehicles, animals, trees, structures, etc.) appearing at a strict periodic rate of exactly 6 images per second. To achieve the 6Hz stimulation, we sinusoidally modulated the contrast of each image from 0-100-0% over a 167ms period (i.e., 10 screen refreshes). Within the image stream, two critical categories (birds/faces for Exp 1, birds/houses for Exp 2) were interlaced at distinct embedded periodicities: A bird appeared as every 4th image, and a face (Exp 1) or house (Exp 2) as every 5th image (see Figure 1B). We marked the onsets for object, bird, and face/house stimuli in the EEG trace using distinct numeric trigger values. As has been detailed elsewhere (de Heering & Rossion, 2015; Jacques et al., 2016; Quek & Rossion, 2017; Rossion et al., 2015), visual stimulation of this type reliably elicits neural responses at the exact frequencies of stimulation, with a general visual response arising at the image presentation rate (i.e., 6Hz), and category-selective responses at the embedded periodicities. Importantly for our purposes, since the critical categories here appear at distinct periodic intervals in the sequence (i.e., 1/4 and 1/5 image respectively), their differential neural signatures are expected to arise at distinct frequencies too (1.2Hz for faces/houses and 1.5Hz for birds). Where these frequencies overlapped (every 20 stimuli, see Figure 1B), the face (Exp 1) or house (Exp 2) was always presented, effectively skipping a bird instance. This served to equalize the total number of faces/houses and birds that appeared in each sequence: 48 occurrences of each critical stimulus type within each 40 second sequence. To encourage fixation and attention to the images, the fixation cross remained superimposed on the images throughout the entire sequence. To minimize blinks and artefacts elicited by the sudden appearance/disappearance of flickering stimuli, each 40-second sequence was preceded and followed by a transition period during which the maximum contrast of each image progressively ramped up or down (total sequence time = 43.33 seconds). These fade-in / fade-out periods were not included in analysis.

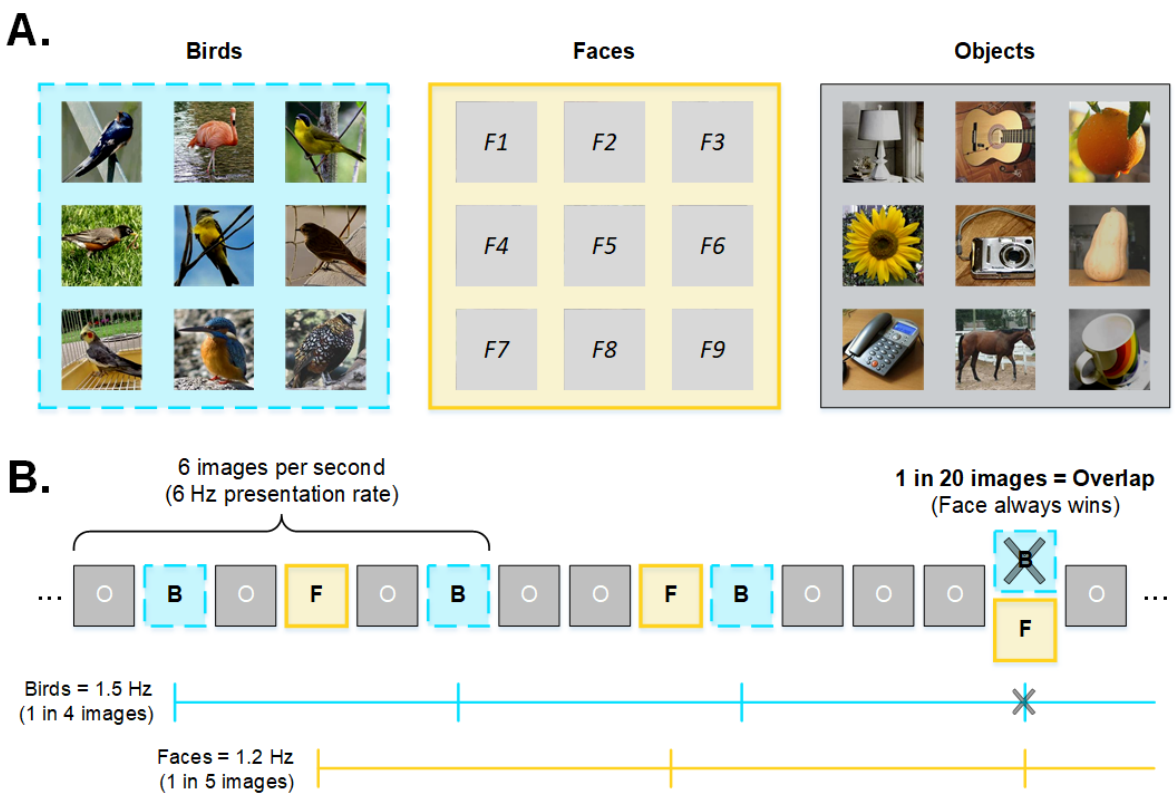

Figure 1. A) Example stimuli for the bird, face, and object categories in Exp 1. **B)** Schematic of the 'interlaced' rapid sequence. Within a 6Hz stream of object images, birds and faces appeared as every 4th and 5th image, respectively. These frequencies (1.2 and 1.5Hz) overlapped every 20 images; a face image always took precedence at overlap instances (i.e., effectively skipping a bird presentation). This allowed for an equal number (n=48) of bird and face instances in each 40 second sequence. Sequence structure in Exp 2 was identical, save that house images replaced the face images.

Attentional Manipulation

Participants in Exp 1 saw three blocks of 15 sequences that were strictly identical in visual content, varying only in terms of the task (order was counterbalanced across participants). In the *attend-Cross* block, participants had to press the spacebar whenever the overlying fixation cross changed colour from black to blue (200ms change duration, 8 random occurrences per sequence). This orthogonal change detection task is the paradigm standard (Jacques et al., 2016; Liu-Shuang et al., 2014; Quek, Nemrodov, et al., 2018; Quek & Rossion, 2017; Rossion et al., 2015) and was used as a baseline condition that encouraged a constant level of attention to the images themselves (which were irrelevant to the colour-change task). In the *attend-Birds* block, participants had to press the spacebar whenever a bird appeared in the image stream (i.e., 1 in 4 images), and in the *attend-Faces* block, whenever a face appeared (i.e., 1 in 5 images). Importantly, the sequence structure was identical across all conditions, with only the focus of participants' task-based attention changing across blocks. The structure of Exp 2 was near-identical, save that the *attend-Faces* block became an *attend-Houses* block, with images of houses replacing the faces to become the second critical category. In Exp 2, we also included a fourth block of sequences in which the participants' task was to attend to a target category which appeared nonperiodically throughout the sequence (guitars). The *attend-Guitars* block was always presented as the last block, with the first three task-blocks counterbalanced as in Exp 1. The total number of trials in Exp 1 and Exp 2 was thus 45 and 60 respectively. The 15 sequences in each condition took approximately 10 minutes to complete. Rest breaks were self-paced.

<u>EEG Analyses</u>

EEG Pre-processing

We analysed EEG data using Letswave6 (https://www.letswave.org/) running in MATLAB (MathWorks). After importing each participant's raw continuous recording, we applied a band-pass filter (0.1-100Hz) before downsampling the data to 250Hz for faster data handling and storage (linear interpolation). We then segmented 48-second epochs corresponding to each sequence, starting from 2 seconds before the onset of the sequence and lasting until 46 seconds. Noisy electrodes (identified by eye) were linearly interpolated using two immediately adjacent clean channels (no more than 4 channels interpolated per participant). We then re-referenced all scalp channels to a common average of the 64-scalp channels, before cropping each epoch to exactly 40 seconds, corresponding to the proper sequence (i.e., excluding the fade in/fade out period). This produced 15 epochs per task condition, per participant.

Frequency Domain Analyses

Response Significance & Harmonic Selection. To determine the harmonic range over which to quantify *i)* the common visual response and *ii)* the category-selective responses, we obtained each participant's grand average amplitude spectrum by averaging across all conditions and scalp channels and subjecting the resulting mean to Fast Fourier Transformation (FFT). We then computed a *z*-score for each frequency bin ($z = (x - \mu)/\sigma$), where x is the amplitude at a frequency bin of interest, μ is the mean amplitude of the 14 bins surrounding x (i.e., 7 bins either side, excluding the immediately adjacent bin), and σ is the standard deviation across that same 14 bin range. With our frequency resolution of 0.025 (i.e., 1/40 seconds), this 14-bin range ensured that noise estimates were based on a similar frequency range as in Quek et al. (2018) (i.e., 0.35Hz *vs.* 0.36Hz). We evaluated each participant's *z* scores on the harmonic frequencies of the image presentation rate (6Hz), bird presentation rate (1.2Hz), and face/house presentation rate (1.5Hz), imposing a significance criterion of $z > 3.1$ (i.e., $p < .001$, one-tailed).

Response Quantification. To quantify the common visual response (i.e., 6Hz & harmonics), we averaged the 15 epochs per condition for each participant and subjected these conditional means to FFT. From there we computed signal-to-noise (SNR) spectra by normalizing the raw amplitude on each frequency bin against the mean amplitude of the local noise range (i.e., 14 bins, as described above). Finally, we summarised the common visual response by averaging each electrode's SNR across the first 5 harmonics of the image presentation frequency (i.e., 6, 12, 18, 24, & 30Hz; see Results for details on why this range was chosen).

We quantified the bird- and face/house-selective signals for each individual participant in the same way, save that we performed an additional cropping step prior to FFT: First, we cropped the conditional means to 39.34 seconds (9834 bins), reflecting an integer number of cycles of the bird presentation frequency at 1.4999Hz. Separately, we cropped the conditional means to 39.17 seconds (9792 bins), reflecting an integer number of cycles of the face presentation frequency at 1.1999Hz. These different crop lengths ensured that the frequency resolutions of the subsequent FFT amplitude spectra were appropriate for the relevant signal of interest (i.e., with a bin centred exactly on the fundamental frequency under inspection). We generated SNR spectra as described above, before computing category-selective summary scores for each condition by averaging SNR on the fundamental frequency and its harmonics up to (but not including) the 6Hz image presentation rate (birds = 1.5, 3, & 4.5Hz; faces/houses = 1.2, 2.4, 3.6, & 4.8Hz; see Results for details on why this range was chosen). Conditional data for each signal type was then relabelled as either 'Attend Towards', 'Attend Away', or 'Baseline' to denote the relationship between the focus of selective attention and the signal type (e.g., For the face-selective signal, this corresponded to the blocks *attend-Faces, attend-Birds,* and *attend-Cross*).

As a final step, we followed Quek et al. (2018) by further computing indices of attentional enhancement and attentional suppression, separately for the bird-selective signal (enhancement = *attend-Birds* minus *attend-Cross*; suppression = *attend-Faces* minus *attend-Cross*) and face/house-selective signal (enhancement = *attend-Faces/Houses* minus *attend Cross*; suppression = *attend-Birds* minus *attend-Cross*). In Exp 2, we conducted this analysis using both the *attend-Cross* and *attend-Guitar* conditions as the baseline for comparison. Based on previous literature employing variants of the same experimental paradigm (Quek, Liu-Shuang, et al., 2018; Quek, Nemrodov, et al., 2018), we inspected

the summary scores and attentional indices in two occipitotemporal (OT) regions-of-interest (ROIs) comprising 5 electrodes each (*Left OT*: PO3, P7, PO7, P9, O1; *Right OT*: PO4, P8, PO8, P10, O2). Signals were compared across conditions using i) repeated measures ANOVAs, applying a Greenhouse-Geisser correction wherever the assumption of sphericity was violated, ii) paired *t*-tests with Bonferroni corrections, and iii) Bayesian paired *t*-tests. Attentional indices were evaluated against zero using one-sample *t*-tests (one tailed).

Time Domain Analyses

To inspect each participant's category-selective EEG responses in the time-domain, we subjected the preprocessed epochs to a 4th order Butterworth low-pass filter with a cutoff of 30Hz. We then removed the common visual response at 6Hz and harmonics using an FFT multi-notch filter (width = 3 bins) centred on each of the first 6 harmonics of the image presentation frequency (slope cutoff width = 0.0750Hz). We segmented an 833.33 ms window around each Face/House or Bird occurrence starting 167 ms before stimulus onset, discarding the final epoch of each stimulus type in each sequence since it overlapped the fade-out period (totalling 705 bird epochs and 525 face/house epochs). We then averaged sequences within each condition for each participant and performed a baseline correction by subtracting the mean amplitude in the -167ms to 0ms time-window from the entire epoch. Finally, we averaged the baseline-corrected conditional means across participants to produce group means for visualization. We assessed the presence of enhancement and suppression at each timepoint via paired Bayesian *t*-tests (two tailed) for the relevant condition contrasts (i.e., Baseline *vs.* Attend Towards, and Baseline *vs.* Attend Away). ROIs were the same as those used in the frequency-domain analyses.

RESULTS

EXPERIMENT 1

Frequency Domain Results

Response Significance & Harmonic Selection

We evaluated the presence of significant signal at the specific frequencies of interest by evaluating the *z*-score for each frequency bin (see Methods) against a criterion of $z > 3.1$ (i.e., $p < .001$, one-tailed). This analysis was aimed at determining the harmonic range which should contribute to response quantification (Talia L. Retter, Rossion, & Schiltz, 2021). To ensure impartiality, we inspected the significance of each harmonic based on each participant's grand average amplitude spectrum (i.e., collapsed across all scalp channels and all conditions).

As is standard in SS-EP designs (Jacques et al., 2016; Liu-Shuang et al., 2014; Quek, Nemrodov, et al., 2018; Quek & Rossion, 2017; Rossion, 2014), we observed a strong response at the image presentation frequency (see Figure 2), with all 20 participants exhibiting significant signal on the first four harmonics of 6Hz (i.e., 6, 12, 18, & 24Hz), and 15/20 participants exhibiting significant signal on the 5th harmonic (i.e., 30Hz). This 5-harmonic range formed the basis for our quantification of the common visual response (see below). In addition, we found clear evidence of category-selective processing for both birds and faces, with significant signal at their presentation frequencies and harmonics. For the face frequency (1.2Hz), all 20 participants exhibited significant signal on all harmonics up to 12Hz (i.e., harmonics 1-9 for 1.2Hz). For the bird frequency (1.5Hz), we observed strong signal on the first 3 harmonics (harmonics 1, 2, & 3 considered significant in 17/20, 16/20, and 18/20 participants respectively). There was less consistency across participants for bird harmonics above 6Hz, with harmonics 5, 6, 7 of 1.5Hz observed to be significant in 12/20, 10/20, and 10/20 participants respectively. Given that significant responses were most consistent for both category-selective signals under the 6Hz image presentation rate, we constrained the frequency range for category-selective quantification to the sub-6Hz range (i.e., Bird = 1.5, 3, & 4.5Hz; Face = 1.2, 2.4, 3.6, & 4.8Hz).

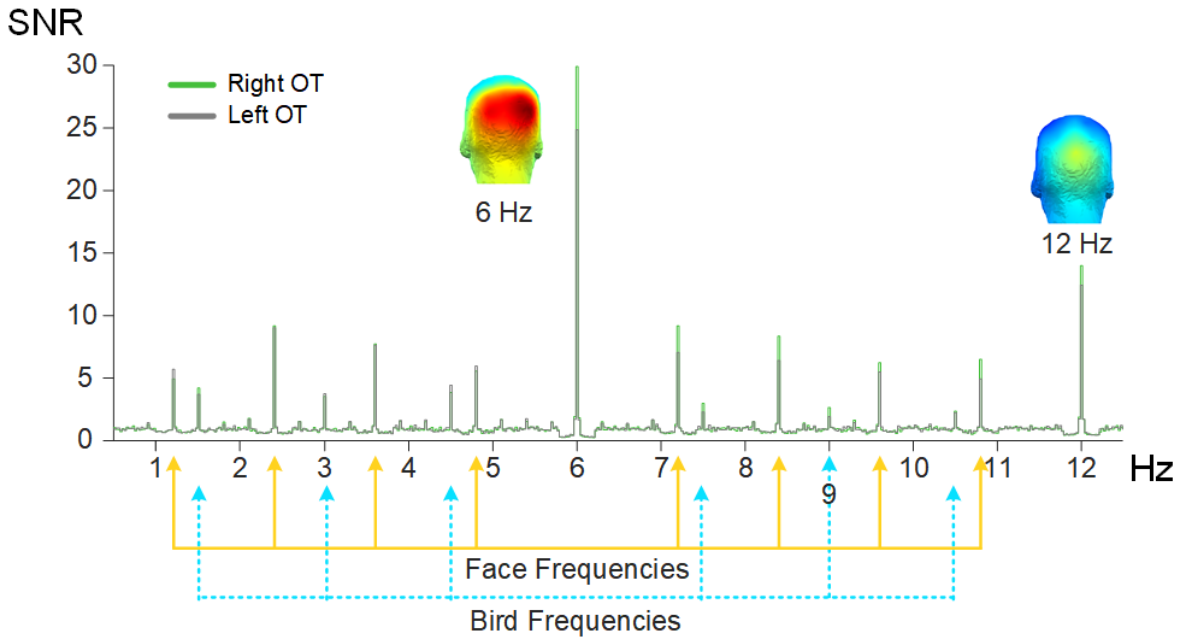

Figure 2. The group mean SNR spectrum for Exp 1, collapsed across conditions for the left and right occipitotemporal (OT) ROIs. Strong responses were evident at the image presentation frequency (i.e., 6Hz & harmonics), face-selective frequency (1.2Hz & harmonics, yellow solid arrows), and bird-selective frequency (1.5Hz & harmonics, blue dashed arrows).

Response Quantification

The **common visual response** arising at 6Hz and harmonics is understood to capture aspects of visual processing shared by all categories contained within the sequence (i.e., faces, birds, and the various other object categories such as vehicles, buildings, animals, etc.) (Jacques et al., 2016; Quek, Liu-Shuang, et al., 2018; Quek, Nemrodov, et al., 2018; Quek & Rossion, 2017; Rossion et al., 2015). We subjected the summarised values for the common visual response (see Methods) to a 2x3 repeated measures ANOVA with the factors *ROI* (left OT, right OT), and *Task* (attend-Cross, attend-Birds, attend-Faces). There was a significant main effect of *ROI*, $F(1,19) = 14.96$, $p = .001$, $\eta^2 = .174$, with larger SNR values for the right OT ($M = 8.52$, $SD = 1.82$) compared to the left OT ($M = 6.76$, $SD = 1.63$). Further inspection of the scalp topographies showed that the neural response at 6Hz was distributed most strongly over electrode PO8 (see scalpmap in Figure 2). Notably, however, *Task* had no significant influence on the magnitude of the common visual response, $F(2,38) = .246$, $p = .783$, $\eta^2 = .002$, nor was there a significant interaction between *ROI* and *Task*, $F(2,38) = 1.87$, $p = .168$, $\eta^2 = .005$.

Scalp topographies for the summarised **face-selective response** and **bird-selective response** in each attention condition are given in Figure 3A. We subjected the category-selective responses within our ROI to a 3-way repeated measures ANOVA with the factors *Signal Type* (face, bird), *Attention condition* (Baseline, Attend Towards, Attend Away), and *ROI* (left OT, right OT). We found a main effect of *Signal Type*, $F(1,19) = 60.75$, $p < .0001$, $\eta^2 = .338$, reflecting, as expected, overall stronger SNR values for the face-selective response ($M = 5.40$, $SD = 1.17$) than the bird-selective response ($M = 3.36$, $SD = 0.94$) (see Figure 3B). There was also a main effect of *Attention Condition*, $F(1.4,26.5) = 39.06$, $p < .000001$, $\eta^2 = .252$, as well as a significant interaction between *Signal Type* x *Attention Condition,* $F(1.34,25.51) = 6.30$, $p = .012$, $\eta^2 = .045$. The other main effects and interaction terms did not reach significance (all *p*-values > .05).

To clarify how the focus of task-based attention differentially influenced category-selective responses for faces and birds, we next examined indices of attentional enhancement (i.e., Attend Towards – Baseline) and suppression (i.e., Attend Away – Baseline). As difference scores calculated for each signal type, these indices were not influenced by the higher overall SNR for faces as compared to birds (see

Figure 3B). After confirming that each of these indices departed significantly from zero (all one-sample *t*-test *p* values < .05, Bonferroni-corrected), we examined whether either metric differed as a function of *Signal Type*. As it can be seen in Figure 3C, enhancement was significantly stronger for the bird signal (M = 2.44, SD = 1.79) compared to the face signal (M = 1.03, SD = 1.42), $F(1,19) = 8.49$, $p = .009$, $\eta^2 = .138$. In terms of suppression, directing attention to the other visual category served to suppress the face-selective signal (M = -0.64, SD = 0.91) to a greater extent than the bird-selective signal (M = 0.55, SD = 0.66), $F(1,19) = 25.48$, $p < .001$, $\eta^2 = .299$. In fact, suppression scores for the bird signal were in fact positive on average, suggesting that attending to faces within the image stream did not suppress the evoked response to birds at all, but in fact slightly facilitated this category-selective response as compared to the *attend-Cross* baseline condition.

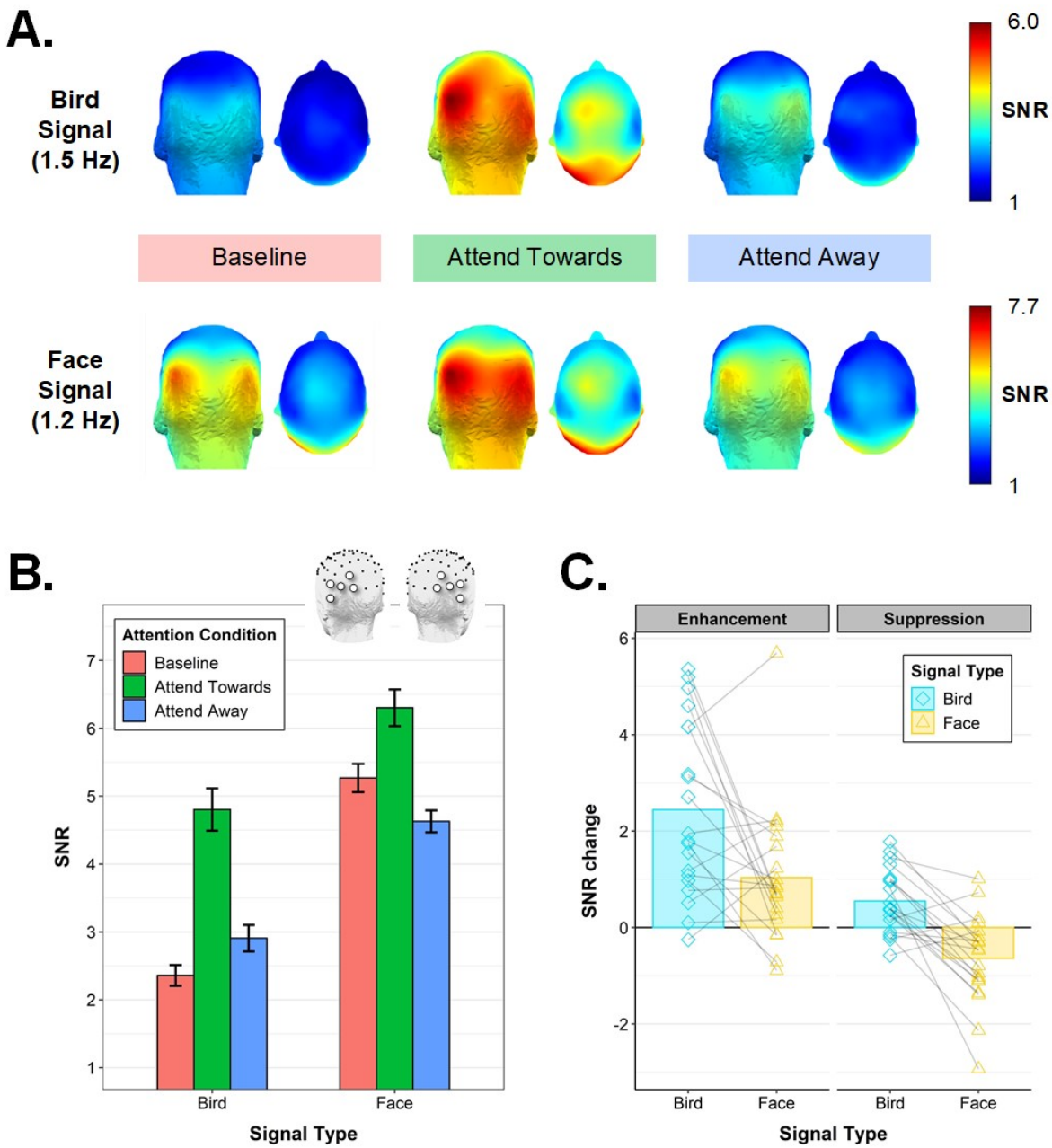

Figure 3. A) Conditional mean scalp topographies for the summarized bird-selective (top row) and face-selective response (bottom row) in Exp 1. Amplitude ranges are fixed across Attention Condition for each signal type. To see individual participant scalp topographies for each condition, see Figure S1A in Supplemental Material. **B)** Conditional mean bird-selective and face-selective responses averaged across the left and right OT ROIs (inset); error bars are within-subjects standard error. **C)** Indices of Enhancement (Attend Towards – Baseline) and Suppression (Attend Away – Baseline) for the bird-

selective and face-selective responses averaged across the left and right ROIs. Overlaid data points are individual participants. All attentional indices diverged significantly from zero (i.e., all *p* values < .05, Bonferroni corrected).

Time Domain Results

To understand how category-based attention modulated the bird-selective and face-selective signals over time, we inspected the stimulus-locked response to birds and faces in each condition (Figure 4) after filtering out the common visual response at 6Hz and harmonics (see Methods). The selective waveform for faces (Figure 4A) was highly consistent with previous reports using similar periodicity-based paradigms (Jacques et al., 2016; Quek & Rossion, 2017; T. L. Retter & Rossion, 2016; Rossion et al., 2015), evolving over some 600ms following stimulus onset[1]. The selective response elicited by periodic exemplars of birds amid various other natural object categories (Figure 4B) is a novel report that has not previously been characterized.

Bayesian paired *t*-tests contrasting amplitudes for the Attend Towards and Baseline conditions at each timepoint revealed markedly different patterns of attentional enhancement for face-selective and bird-selective processing: Although both signals showed a facilitation effect during the early stages of neural processing (with slightly later facilitation evident for birds as compared to faces), the bird-selective response was most strongly enhanced during a later time window (~300-500ms). Notably, this late facilitation was entirely absent for the face-selective response. Since we wondered how this timecourse might overlap with action preparation signals, we also inspected lateralised readiness potentials (LRPs) in each condition by subtracting electrode C4 amplitudes from C3 amplitudes, separately for the face and bird locked waveforms. We were specifically interested in whether LRPs would be strongly evident in the Attend Towards condition (which requires a behavioural response each time that attended category appears), however this was not the case (see Figure S5 in Supplemental Materials).

Conversely, contrasting the Attend Away and Baseline conditions revealed that attending to Birds significantly dampened the amplitude of the face-selective response from ~300ms until nearly 600ms post stimulus onset. This window of attentional suppression corresponds well with what Quek et al. (2018) observed for face-selective responses, although in our case the suppression effect endured for longer. Notably, the reverse effect was not evident: that is, attending to faces did not have a dampening effect on the bird-selective response, but in fact resulted in a response amplitude increase between 200-350ms.

[1] Our sinusoidal contrast manipulation means that the first frame for each image is displayed at 0% contrast, such that an image will be detectable minimally 16.67ms post stimulus onset (20% contrast, see Retter & Rossion, 2016, for comparison between sinewave and squarewave stimulation). This introduces some temporal displacement in the latencies of waveform components shown in the time domain.

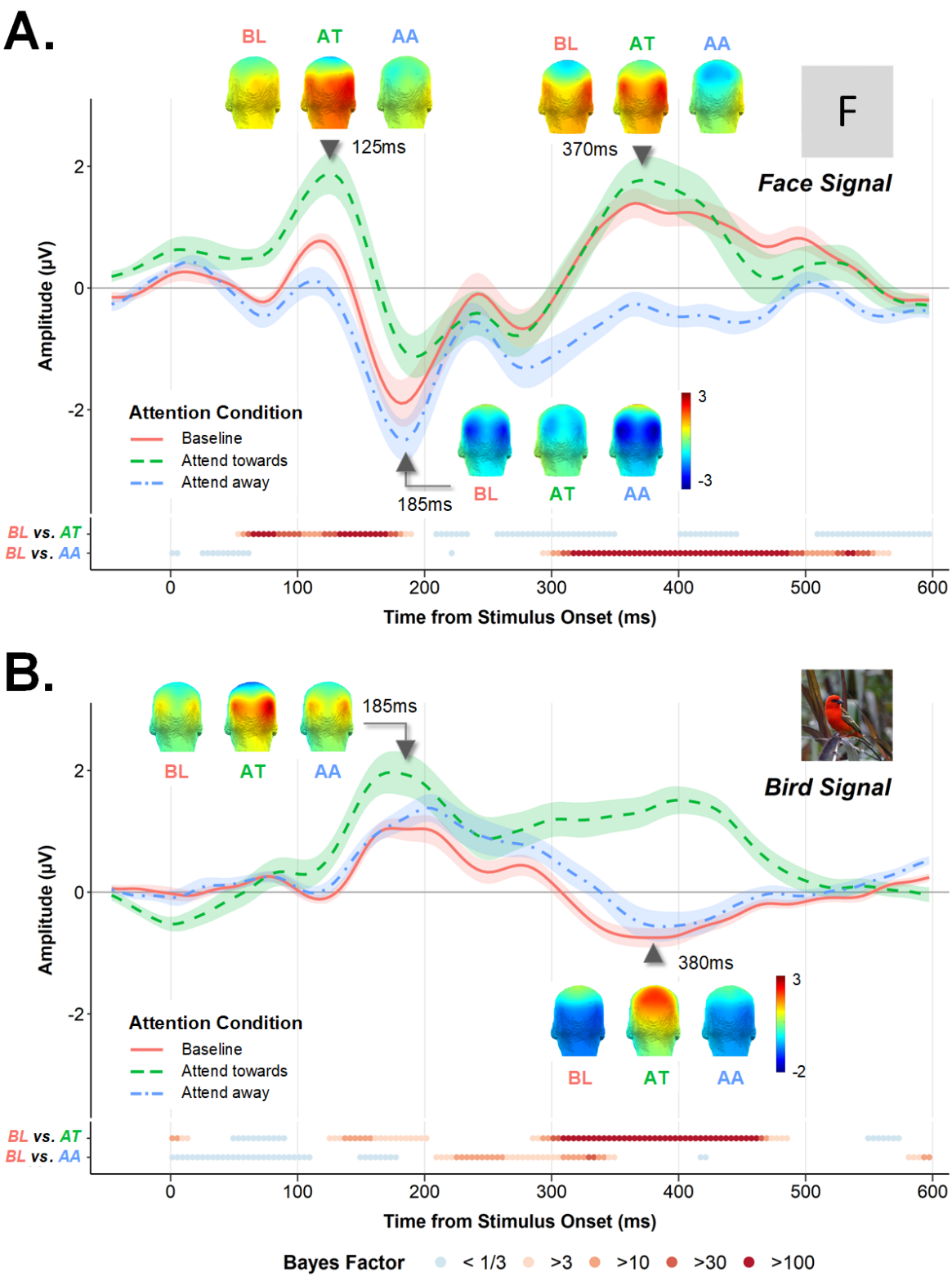

Figure 4. The conditional mean face-selective (A) and bird-selective (B) response from Exp 1, shown as a function of time from stimulus onset. Shaded regions are within-subjects standard error. Coloured points below each plot reflect Bayesian evidence at each timepoint for a difference between the Baseline (BL) & Attend Towards (AT) conditions and the Baseline (BL) & Attend Away (AA) conditions. Cool and warm colours denote evidence for the null and H_1, respectively. Group-averaged headplots are given for select timepoints; amplitude ranges are fixed within signal type (i.e., same colourbar for

all headplots for each signal). To see these category-selective waveforms separated by left and right occipitotemporal ROIs, see Figure S2 in Supplemental Material.

EXPERIMENT 2

Frequency Domain

To facilitate comparison across experiments, all response quantification parameters for Exp 2 were the same as those used in Exp 1. We began by replicating the structure of Exp 1 analyses by treating the *attend-Cross* condition as the relevant baseline. Here we subjected the **common visual response** scores to a 2x3 repeated measures ANOVA with the factors *ROI* (left OT, right OT), and *Task* (attend-Cross, attend-Houses, attend-Birds). We found a significant main effect of *ROI*, $F(1,19) = 7.27$, $p = .014$, $\eta^2 = .08$, with larger SNR values for the right ($M = 8.51$, $SD = 2.69$) than the left OT ($M = 7$, $SD = 1.66$). Contrary to Exp 1, *Task* had also a significant influence on the magnitude of the common visual response, $F(2,38) = 8.36$, $p < .001$, $\eta^2 = .045$, and this factor interacted significantly with *ROI*, $F(2,38) = 5.89$, $p = .006$, $\eta^2 = .009$. Follow-up ANOVAs within each ROI revealed that this modulation by *Task* did not manifest in the right ROI where the signal was maximal, $F(2,38) = 1.75$, $p = .188$, $\eta^2 = .011$, but was driven by responses in the left ROI, $F(2,38) = 17.94$, $p < .0001$, $\eta^2 = .148$. In this region, the common visual response was lower in the attend-Cross baseline condition than in either the attend-Bird or attend-House conditions (Bonferroni-corrected *p*-value < .0001 in both cases).

Mirroring Exp 1, summarised values for the **house-selective** and **bird-selective** responses were inspected via a 3-way repeated measures ANOVA with the factors *Signal Type* (house, bird), *Attention Condition* (Baseline, Attend Towards, Attend Away), and *ROI* (left OT, right OT). We found a main effect of *Signal Type*, $F(1,19) = 23.65$, $p = .0001$, $\eta^2 = .134$, reflecting stronger SNR for the house-selective response ($M = 4.59$, $SD = 1.65$) compared to the bird-selective response ($M = 3.16$, $SD = 0.76$) (see Figure 5B). There was also a main effect of *Attention Condition*, $F(1.15,21.92) = 42.713$, $p < .0001$, $\eta^2 = .247$ and, contrary to Exp 1, no interaction between *Signal Type* x *Attention Condition,* $F(1.09,20.64) = 0.724$, $p = .416$, $\eta^2 = .004$. The other main effects and interaction terms did not reach significance (all *p*-values > .05) except for a significant interaction between *ROI* x *Attention Condition,* $F(1.55,29.5) = 8.616$, $p < .01$, $\eta^2 = .015$.

We next inspected the indices of attentional enhancement (Attend Towards – Baseline) and suppression (Attend Away – Baseline), treating *attend-Cross* as the baseline condition. Since ROI was a significant modulator of the overall ANOVA, we retained this factor in these analyses. Except for the suppression score for the bird signal (see Figure 5C), each of the attentional indices departed significantly from zero (all one-sample *t*-test *p* values < .05, Bonferroni-corrected). For enhancement, a repeated measures ANOVA revealed no evidence of a difference in enhancement scores between the bird-selective and house-selective signals, $F(1,19) = .944$, $p = .343$, $\eta^2 = .016$ (see Figure 5C). Although not significantly different from one another, the pattern of enhancement across the two signal types was the reverse of what we observed in Exp 1, $F(1,38) = 7.25$, $p = 0.01$, $\eta^2 = .075$. Here, enhancement was numerically higher for the 1.2Hz signal (i.e., House: $M = 2.65$, $SD = 2.14$) than the 1.5Hz signal (i.e., Bird: $M = 2.14$, $SD = 1.47$). This is encouraging, since it suggests that the attentional differences observed for Birds and Faces in Exp 1 was not merely driven by the specific frequency of presentation assigned to each category. Lastly, there was a main effect of *ROI*, with enhancement being stronger overall in the left ($M = 2.88$, $SD = 1.87$) compared to the right *ROI* ($M = 1.90$, $SD = 1.21$), $F(1,19) = 10.53$, $p = .004$, $\eta^2 = .058$. Finally, no interaction between *ROI* and *Signal Type* was evident, $F(1,19) = 0.29$, $p = .599$, $\eta^2 = .001$.

Suppression scores for both signal types were slightly positive, suggesting that neither category-selective signal was actively attenuated by attending to the other critical category, but instead slightly facilitated relative to the attend-Cross baseline condition (right panel, Figure 5C). This somewhat paradoxical facilitation effect was greater for the house signal ($M = 0.66$, $SD = 0.54$) than for the bird signal ($M = 0.26$, $SD = 0.58$), $F(1,19) = 9.371$, $p = .006$, $\eta^2 = .069$, with only the former being different from zero, $t(19) = 1.23$, $p < .0005$. Notably, this pattern of numerically positive suppression for both signal types differed from what we observed in Exp 1, $F(1,38) = 34.82$, $p < .00001$, $\eta^2 = .260$, where we

found that attending to birds served to actively suppress face-selective processing, but attending to faces actually served to facilitate the bird-selective signal (Figure 3C). Finally, there was no main effect of *ROI* on suppression scores, $F(1,19) = .229$, $p = .637$, $\eta^2 = .003$, nor did *ROI* significantly interact with *Signal Type*, $F(1,19) = 486$, $p = .494$, $\eta^2 = .005$.

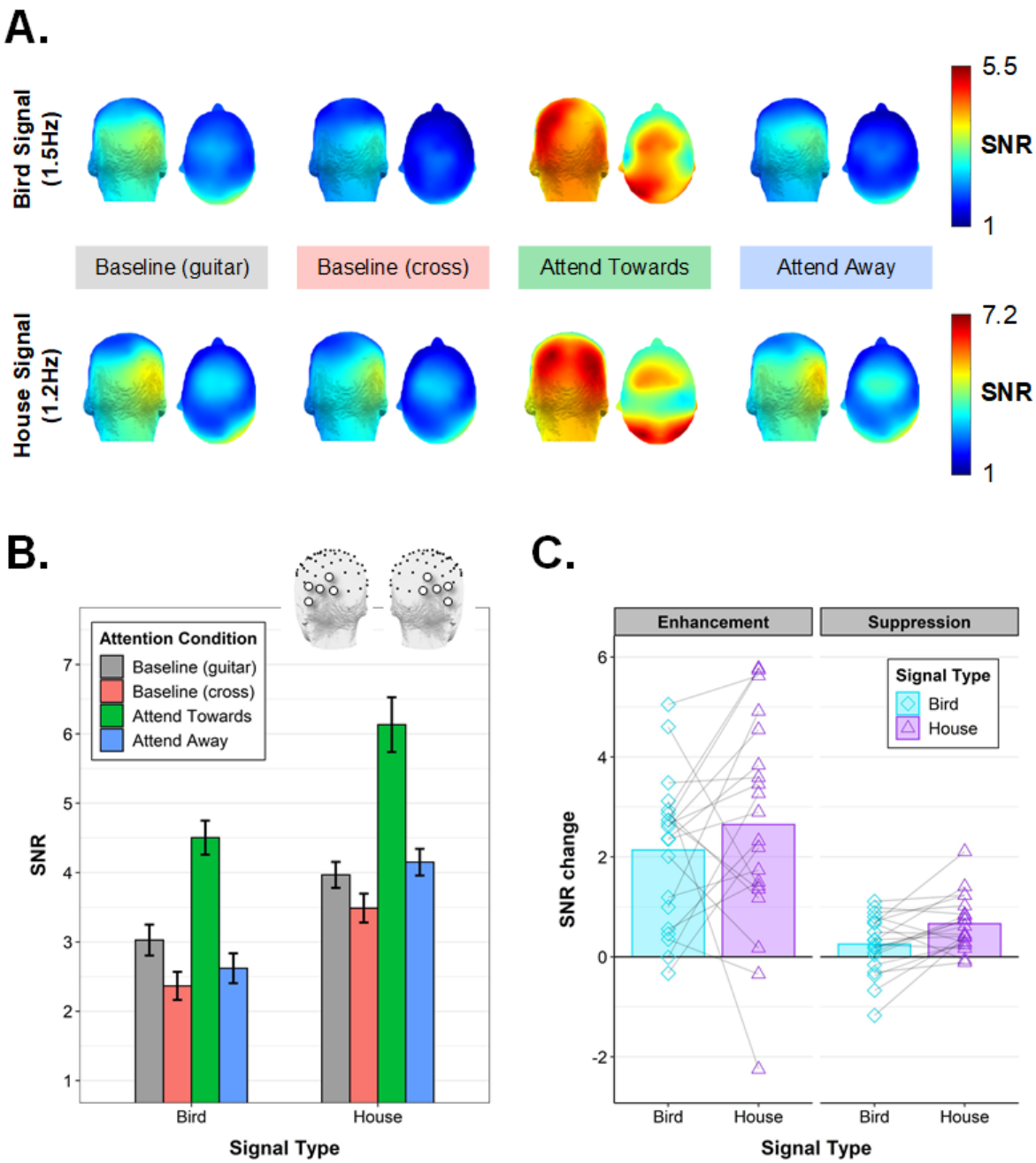

Figure 5. A) Conditional mean scalp topographies for the quantified bird-selective (top row) and house-selective response (bottom row) in Exp 2. To see individual participant scalp topographies for each condition, see Figure S1B in Supplemental Material. **B)** Conditional mean bird-selective and house-selective responses averaged across the left and right ROIs. Error bars are within-subjects standard error. **C)** Indices of enhancement (Attend Towards – Baseline, at left) and suppression (Attend Away – Baseline, at right) for the bird-selective (blue) and house-selective (purple) signals, where Baseline refers to the *attend-Cross* condition. Overlaid data points represent individual participants. All attentional indices except the suppression score for the bird-selective signal differed significantly from zero (*p* values < .05, Bonferroni corrected).

Next, we examined the pattern of results using the *attend-Guitars* condition as the neutral attentional baseline (grey bars in Figure 5B). We included this condition in Exp 2 to examine category-selective responses when attention was directed away from the critical categories, but still focused on the images themselves. We predicted this instruction would boost the bird- and house-selective signals relative to when attention was focused on the central fixation cross (red bars in Figure 5B). This was indeed the case; Figure 5B shows that the category-selective response was stronger in the *attend-Guitars* condition compared to the *attend-Cross* condition for both Birds, $t(19) = -5.75$, $p_{bonf} = .026$), and Houses, $t(19) = -2.75$, $p_{bonf} < .0001$. To examine whether the pattern of attentional modulation across the two signal types remained consistent when using this more conservative attentional baseline, we calculated new indices of enhancement and suppression using the *attend-Guitars* condition as the baseline and subjected these indices to the same analyses as reported above (see Figure S3 in Supplemental Materials). Enhancement was relatively unchanged, with one-sample t-tests confirming significant enhancement of the Bird signal, $t(19) = 4.31$, $p_{bonf} = .002$, $d = 0.963$, and House signal, $t(19) = 5.79$, $p_{bonf} < .0001$, $d = 1.30$. There was no significant difference in the magnitude of enhancement across Signal Type, $F(1,19) = 2.66$, $p = .12$, $\eta^2 = .035$. In contrast, suppression effects differed slightly as a function of which baseline was used: When calculated using *attend-Guitars*, suppression scores for the House signal were numerically positive (although not statistically different from zero, $t(19) = 1.02$, $p_{bonf} = 1$, d = 0.23, and suppression scores for the bird signal dropped significantly below zero, $t(19) = -3.84$, $p_{bonf} = .004$, d = 0.86. In other words, attending to houses served to significantly suppress the bird-selective signal when compared to a neutral attentional baseline in which observers had to monitor the images themselves, rather than just a focal part of the display (as in the *attend-Cross* condition).

Time Domain Results

The temporal dynamics of the bird-selective neural response were highly consistent with those observed in Exp 1 (Figure 6B), as was the pattern of attentional enhancement identified by contrasting the Attend Towards and Baseline conditions (where baseline = *attend-Cross*). Just as in Exp 1, explicitly attending to birds facilitated selective neural processing of this category in both an early time window (i.e., between 100-200ms) and, very markedly, during later time window between 320-420ms. Relative to the *attend-Cross* condition, attending to houses attenuated the amplitude of the bird-selective response for a brief period around 100ms post stimulus onset, but resulted in stronger negative amplitudes between 320-420ms. The latter effect differed from what we observed in Exp 1, where attending to the alternate category (Faces) resulted in stronger positive amplitudes for the bird-selective response over a slightly earlier window (~200-350ms, Figure 4B).

In contrast to the robust and replicable face- and bird-selective responses, the house-selective waveform was less well-formed. Under orthogonal task conditions (attend-Cross or attend-Guitars), houses evoked a small initial positivity around 70ms, followed by a prolonged negativity lasting until ~400ms. A qualitatively similar waveform has been documented previously for houses embedded amid other natural object categories, where the selective response to this category was notably less well-formed than those associated with faces and body parts (Jacques et al. 2016). Attending to houses resulted in a strong positivity during the early stages of the house-selective response, and again during a later time window between 300-400ms. This latter positive component strongly resembled the one that was evident in the bird-selective waveform when birds were the attended category. Notably, attending to an alternate category (either birds or guitars) did not serve to dampen house-selective response amplitudes at any timepoint, but rather resulted in stronger long-lasting negativity.

The comparison between the two baseline conditions was also of interest. When the task requires observers to monitor the image stream (but neither critical category is task-relevant, i.e., *attend-Guitars* condition), category-selective signals appear to be facilitated (i.e., larger absolute response amplitudes) relative to when the task requires monitoring of the central fixation cross. For birds, this manifests as a

stronger initial positivity and a stronger negativity later on. For Houses, this appears as a stronger initial negativity.

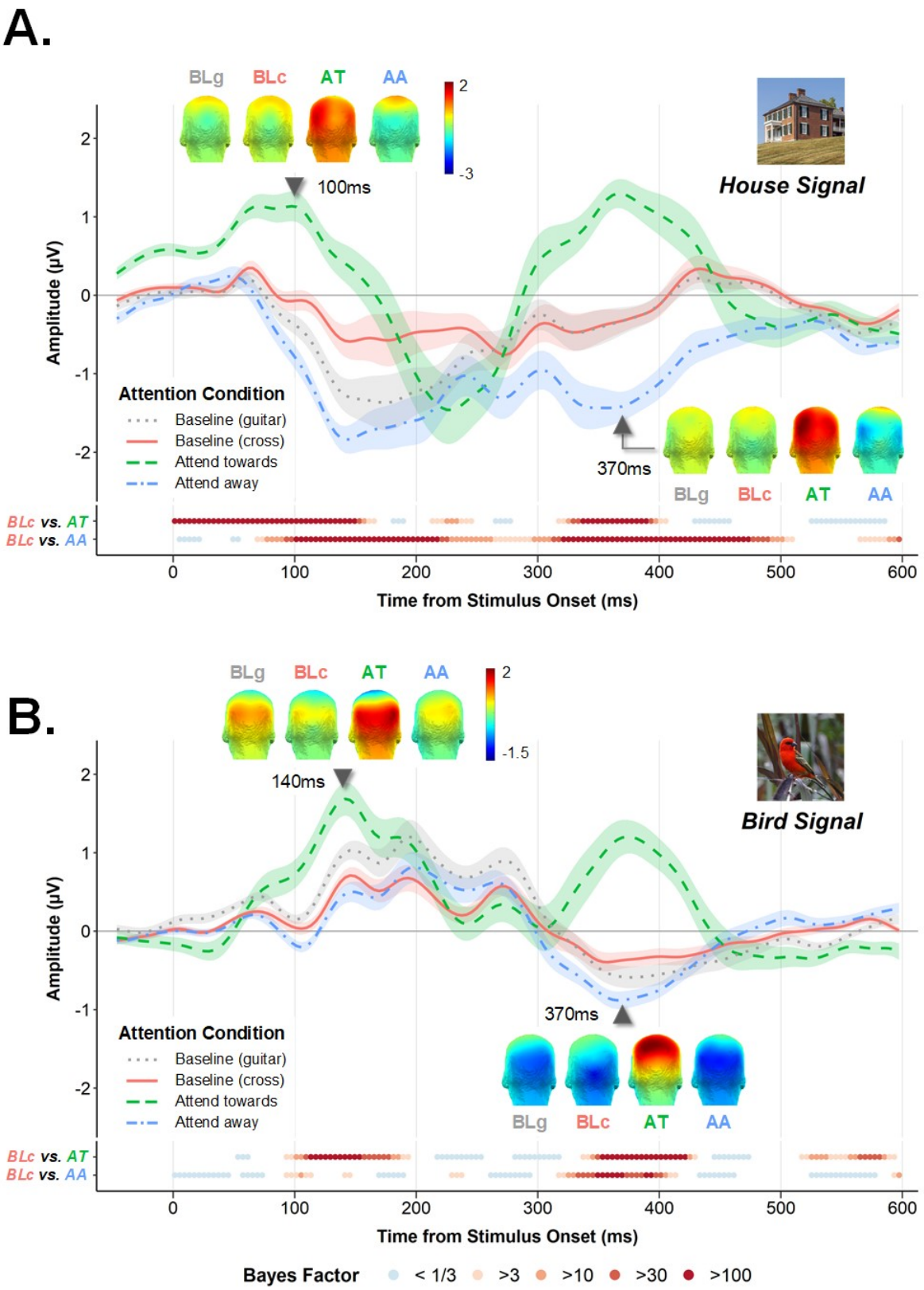

Figure 6. The conditional mean house-selective (A) and bird-selective (B) responses from Exp 2, shown as a function of time from stimulus onset. Shaded regions are within-subjects standard error. Coloured points below each plot reflect Bayesian evidence at each timepoint for a difference between the Baseline-cross (BLc) & Attend Towards (AT) conditions and the Baseline-cross (BLc) & Attend Away (AA)

conditions. Cool and warm colours denote evidence for the null and H_1, respectively. To see these category-selective waveforms separated by left and right occipitotemporal ROIs, see Figure S4 in Supplemental Material.

DISCUSSION

The current study is the first to exploit visual periodicity to dissociate concurrent selective neural signals for different high-level categories contained within the same rapid image stream. We leveraged this powerful 'interlaced' SS-EP design to capture the spatiotemporal dynamics of attention's influence on selective visual processing of faces, birds, and houses. Results reveal a qualitatively distinct attentional signature for face-selective visual processing compared to other high-level categories. Where attending to either birds or houses provides a strong boost to the selective neural signals associated with these categories, attentional facilitation of face-selective neural processing appears to be much more modest. Moreover, only the face-selective neural response was actively attenuated by attending to a competing visual category (i.e., no suppression of the bird- or house-selective response caused by attending to another category). These results underscore the special status that faces hold in the human visual system (Farah, 1996; Farah, Wilson, Drain, & Tanaka, 1998) and highlight the utility of frequency designs as a simple, yet powerful tool for indexing selective neural processing of multiple visual categories contained within the same visual sequence.

Distinct profiles of **attentional enhancement** for face and non-face stimuli

The data here establish that monitoring for task-relevant category exemplars amid a large and highly variable set of object images facilitates category-selective neural processing in distinct ways for face vs. non-face categories. Quantified in the frequency-domain, selective neural responses for birds increased by 177% in Exp 1 and 95% in Exp 2 when this category was the focus of the participants' task. In Exp 2, category-based attention enhanced house-selective processing by 91%. In contrast, attentional facilitation of face-selective visual processing was extremely modest (mean selective response increase of 26.4%, Exp 1). Inspection of the corresponding category-selective waveforms (i.e., time-domain analysis) revealed distinctions in the temporal unfolding of enhancement effects across the different high-level categories. Although category-based attention facilitated the early stages of selective neural processing for all three categories, the strongest enhancement of the bird- and house-selective waveforms arose much later (i.e., 300-500ms), with no evidence that category-based attention enhanced face-selective visual processing after 200ms.

Early facilitation of face-selective neural processing

Early enhancement of selective neural processing for task-relevant faces is consistent with previous reports of a stronger P1-face and a prolonged N1-face when observers actively attend to faces in a rapid image sequence compared to a overlaid fixation point (Quek, Nemrodov, et al., 2018). Early attentional enhancement for visual processing of task-relevant (vs. passively viewed) faces has also been found using a standard ERP approach (e.g., increased P1 amplitudes, earlier N1 latencies, Gazzaley et al., 2008; Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Zanto, Hennigan, Östberg, Clapp, & Gazzaley, 2010). These findings stand apart from studies that have operationalised selective attention as a simple binary (i.e., faces are either the *actively attended* or *actively ignored* category on each trial), where early attentional benefits (i.e., < 200ms post stimulus onset) tend to be evident only under high perceptual load conditions (Furey et al., 2006; Lueschow et al., 2004; Sreenivasan, Goldstein, Lustig, Rivas, & Jha, 2009). As such, it seems likely that the processing constraints conferred by the fast image presentation rate (6Hz) and highly variable stimulus set used here improved our chances of observing early enhancement of face-selective processing. Our observation that attending to faces amid a host of other natural categories facilitates the early stages of selective visual processing for this stimulus is also consistent with work linking P1 and N170 amplitudes to categorisation/recognition outcomes for faces (Liu et al., 2002). In one MEG study, Liu et al. showed that both M100 and M170 magnitudes are larger in response to degraded face stimuli that were successfully categorised as a faces (albeit only against

houses). Notably, only the M170 varied systematically as a function of successful identification (i.e., recognising the individual).

One interesting aspect of our data is that we did not find any evidence for differential attentional enhancement of face-selective signals at left and right occipitotemporal recording sites. Previous work has shown that attending to faces in a rapid visual sequence enhanced face selective processing much more evidently in the non-preferred hemisphere for faces (i.e., the left hemisphere for most participants, +43%) than in the preferred hemisphere (i.e., the right hemisphere for most participants, +3%) (Quek, Nemrodov, et al., 2018). In contrast, we found no distinction between the left and right occipitotemporal ROIs in terms of attentional enhancement, with category-based attention boosting SNR to a comparable extent in participants' preferred (+28%) and non-preferred (+25%) hemispheres. What accounts for this difference is unclear, as the two studies share the same image presentation rate and degree of target/distractor similarity (i.e., both use face targets amid object distractors at a rate of 6 images per second). One notable difference, however, concerns the participants' task: Where Quek et al. (2018) observers monitored the sequence for female faces (just 5 occurrences per sequence), observers here were instructed to press the spacebar with their (dominant) right hand every time a face appeared. If anything, however, additional motor responses in our *attend-Faces* block would be expected to drive a larger face-selective response in the (contralateral) left hemisphere, which is not the pattern we observe. As such, it seems unlikely that this task difference accounts for the discrepancy in hemispheric attentional differences. Another, more likely, explanation relates to the degree of right-lateralisation in the two participant samples: Where observers in Quek et al. (2018) exhibited predominantly right-lateralised face responses in the neutral attention condition (just 2/15 subjects showed the opposite pattern), our Exp 1 had a much higher proportion of left hemisphere dominant face responders in the neutral condition (10/20 = 50% of the sample). This is clearly reflected in the spatial topography of the group mean face-selective response, which is much more bilateral than what Quek et al. (2018) described. It is possible that this increased heterogeneity in the spatial distribution of face-selective processing masked hemispheric differences in attentional effects. Clearly, more work in this area is needed to determine whether attention truly does exert a differential influence in the left and right face perception networks.

Both early and late facilitation for bird- and house-selective neural processing

To the best of our knowledge, the data here are the first attempt to disentangle the timecourse over which category-based attention enhances and suppresses selective visual processing for a category other than faces. Over two participant samples, we found that attending to birds reliably enhanced the bird-selective neural signal during an early window between 100-200ms, and again between 300-500ms. A similar pattern of attentional enhancement was evident for house-selective neural processing in Exp 2, with even stronger early enhancement effects. These dynamics are broadly in keeping with the task-relevance effects for object representations that have been documented using neural decoding methods. In one EEG study, object representations elicited by a 2.5Hz image stream were more dissociable between 200 and 600ms when observers actively monitored the object stream for repeats (Grootswagers, Robinson, Shatek, & Carlson, 2021), compared to when they attended to a small stream of letters overlaid on the object images. Where these authors speculated that such long-lasting attentional effects might have arisen in part due to the memory-based nature of their 2-back task, the relatively late enhancement of visual processing for birds and houses here (i.e., 300-500ms) cannot be explained needing to hold object identities in mind. Instead, based on the latency and topographical distribution of the positive deflection that arises when these categories are actively attended, we interpret this relatively late enhancement as a P3 component related to attentional selection of target category exemplars (Hruby & Marsalek, 2003; Johnson, 1988; Luck & Yard, 1995). In particular, our data appear consistent with the P3b subcomponent that arises over parietal sites around 300-500ms, which is thought to reflect the cognitive evaluation and categorization of stimuli relevant to the task at hand.

Interestingly, the face-selective waveform also contained a similar positive deflection within the same temporal range, which has in the past been described as a P2-Face (Jacques et al., 2016; Retter & Rossion, 2016; Rossion et al., 2015; Quek et al., 2018). In our data, this component was evident in both

the attend-Faces and baseline conditions, but was obliterated when observers actively attended to the alternate category in Exp 1 (birds). This pattern is largely consistent with that observed by Quek et al. (2018), although in their data the late positive component was not entirely suppressed by attending to another category in the sequence. On one hand, this might suggest that the P2-Face reliably observed under orthogonal task conditions (akin to our baseline condition) could comprise a kind of automatic "targetisation" of faces driven by their high salience, which is eliminated when attentional selection is allocated to a different target category. On the other hand, however, the spatial topography of the P2-Face is not strictly the same as the more parietal positive scalp distribution typically associated with the P3b, suggesting that their mere coincidence in time should not be reason enough to believe they reflect the same underlying neural generators.

Given that in our paradigm, attending to a category also means responding behaviourally each time that category appears, it is relevant to consider the extent to which our enhancement effects reflect a motoric or premotor contribution. Importantly, our prespecified ROIs do not encompass the central electrodes that typically evidence action preparation signals (e.g., lateralised readiness potentials, or LRPs). Moreover, consequential LRPs should presumably be most evident for stimulus-locked waveforms in the 'attend-towards' condition, yet supplemental LRP analyses (Figures S5 & S6) show that in 3 out of 4 cases, this prediction was not met. As such, we interpret the enhancement effects here as more likely capturing an increase in selective *visual* processing for the attended category.

Attentional suppression: Exclusive to face-selective processing?

One contrast to which motoric elements cannot contribute in any way is attentional suppression, as neither the Baseline nor Attend Away condition is associated with a behavioural response at the category-selective frequency. A key finding of the current study is that only face-selective neural processing appeared to be readily suppressed by directing attention towards an alternate high-level category. In Exp 1, monitoring for instances of birds in the image sequence significantly dampened the face-selective signal relative to a neutral attention baseline, yet the reverse was not true. That is, attending to faces had no suppressive effect on the bird-selective signal, and in fact produced a somewhat paradoxical *increase* in the bird-selective response relative to baseline. We suspect this effect is tied to the nature of the participants' task in this condition: where monitoring the fixation cross demands a narrow, central focus of attention that is much smaller than the images, monitoring for faces necessarily requires attention to the image stream itself. In this way, attending to faces may have in fact created a spatial attention based facilitation of bird-selective processing, even though this category was not the focus of task-based attention in that condition. We tested this hypothesis in Exp 2 by including an additional baseline condition where participants monitored the sequence for (non-periodic) instances of a third object category (guitars), which enabled us to quantify both bird- and house-selective processing when the images themselves were attended, but neither category was task-relevant. A direct comparison of the two baseline conditions in Exp 2 confirmed that, as expected, attending to the image stream itself did indeed boost the magnitude of the bird- and house-selective responses, even though participants were not monitoring for instances of those specific categories. As such, we interpret the unexpected increase in category-selective processing observed in the Attend Away condition for birds (Exp 1 & Exp 2) and houses (Exp 2) as a spatial attention facilitation effect.

Faces were the notable exception to this pattern. The face-selective response in the neutral attentional baseline condition was already very strong, and was clearly suppressed by directing attention towards a non-face category (-12% in the left ROI, -8% in the right ROI) (Quek, Nemrodov, et al., 2018). Inspection in the time-domain suggested suppression of face-selective processing was largely evidenced after 300ms of visual processing time, enduring until at least 550ms post stimulus onset. Quek, Nemrodov et al. (2018) reported similar effects, however in their case suppression effects were even more pronounced (-23-25% in the two OT ROIs), with strong suppression observed between 200-300ms as well as during the later time window found here. In both studies, monitoring for an alternate high-level category seems to suppress the P2-Face – a face-selective component that is reliably observed using rapid periodic presentation designs (Quek, Nemrodov, et al., 2018; Quek & Rossion,

2017; Rossion et al., 2015). That the P2-Face is suppressed when category-based attention is directed away from faces would support the idea that this component may in part reflect automatic attentional selection of faces as they appear in the image sequence. When visual competition is high (i.e., dynamically changing visual input, fast presentation rate, naturalistic exemplars), perhaps neural architecture that typically prioritises face processing is recruited to serve current task demands (e.g., detect birds).

In contrast to faces, suppression of category-selective processing of non-face stimuli was a more complex picture. As noted above, when quantified against the *attend-Cross* neutral attention baseline, suppression was actually positive for non-face categories, likely reflecting a spatial attention benefit (see above). When we controlled for this spatial attention effect by quantifying suppression against the *attend-Guitars* neutral attention baseline (Exp 2), we found that attending to an alternative category did induce significant suppression of the bird-selective signal (-13.7%), but not the house-selective signal (+7.6%, not significantly different from zero). Further analysis in the time-domain provided more insight into how the temporal unfolding of house-selective processing varies with the focus of category-based attention. In effect, where explicitly attending to houses drove both early and late positivity in the house-selective response, attending away from houses (i.e., monitoring for either birds or guitars) seemed to enhance and prolong negativity in the house-selective signal. Given that birds and faces, and to a lesser extent, guitars, share curvilinear aspects that are missing from house images, we wonder if the degree of low- and mid-level featural overlap between the various target categories might contribute to these intriguing patterns. Future studies of attentional enhancement and suppression of high-level categories should aim to tease out the interplay between category or task-based attention and low-level featural overlap in distractor categories.

Visual competition in time

Visual periodicity paradigms are ideally suited for revealing the spatiotemporal dynamics of attention's influence on selective visual processing, since their rapid image presentation rate and highly variable naturalistic image set ensures that objects must compete for visual representation (Desimone & Duncan, 1995). Although visual competition is classically associated with spatial clutter (i.e., multiple items present in the same display) (Desimone & Duncan, 1995; Duncan, 1984; Kaiser, Stein, & Peelen, 2014), neural competition can also be induced when single items appear in close temporal proximity, as is the case in classic Rapid-Serial Visual Presentation (RSVP) paradigms (Keysers & Perrett, 2002; Zivony & Eimer, 2020). When images appear in rapid succession, neural representations of individual objects must necessarily overlap. As temporal overlap increases with faster image presentation rates (Grootswagers, Robinson, & Carlson, 2019), both representational strength (i.e., decodability) and category-selective response magnitudes reliably decrease (Quek et al., 2021; T. L. Retter, Jiang, Webster, & Rossion, 2020). In this way, the current findings suggest that category-based attention ameliorates visual competition induced by temporal clutter, similarly as other higher-order cognitive mechanisms have been shown to do (e.g., multi-object grouping, Quek & Peelen, 2020).

Along the same line, the data here suggest that faces as a category are comparatively robust to temporal competition effects, insofar as attentional enhancement for this category was rather modest compared to non-face categories. Moreover, category-selective signals for faces appear to be, at baseline, both larger and more complex than those evoked by non-face categories. In our case, the bird- and house-selective SNR reached 45% and 66% of the face-selective SNR respectively, and previous studies have observed face-selective signals 2-4 times larger than selective signals associated with houses or body parts (Jacques et al., 2016). This would imply that under orthogonal task conditions, faces are less affected by representational competition induced by neighbouring distractors than other high-level categories are.

Why is it that faces drive such a strong selective neural response amid object distractors, even when they are not relevant to current task demands? On one hand, if face-processing is highly automatized, perhaps this complex selective response is evoked even when faces are outside the focus of either spatial and/or task-based attention (Reddy & Kanwisher, 2007; Reddy et al., 2006; Reddy et al., 2004;

Vuilleumier, 2000). Alternatively, it could be that under orthogonal task conditions (e.g., detect colour changes of the fixation cross), faces that appear in the sequence are salient enough to *capture* attention, effectively inducing attentional slippage towards these socially relevant stimuli (Gluckman & Johnson, 2013; Lachter, Forster, & Ruthruff, 2004; Sato & Kawahara, 2015; Theeuwes & Van der Stigchel, 2006). Although the data here do not distinguish between these accounts, both explanations underscore the special status that faces hold in the human visual system, as a high-level category that drives a uniquely strong selective signal in the absence of overt attention.

To conclude, the current 'interlaced' periodicity-based design provides a simple, yet powerful way to cleanly dissociate selective neural responses associated with different high-level categories contained within the same rapid image stream. The approach offers a significant increase in experimental efficiency by tapping multiple selective neural signals contained within the same sequence, preserving both frequency- and time-domain information about these signals. The associated reduction in overall testing duration has a clear advantage in studying visual perception in populations with testing constraints, such as infants, specific patient groups, or individuals in residential care facilities.

REFERENCES

Addleman, D. A., Rajasingh, R., & Störmer, V. S. (2022). Attention to object categories: Selection history determines the breadth of attentional tuning during real-world object search.

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

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in neurosciences, 34*(4), 210-224.

Battistoni, E., Kaiser, D., Hickey, C., & Peelen, M. V. (2018). Spatial attention follows category-based attention during naturalistic visual search: evidence from MEG decoding. *bioRxiv*, 390807. doi:10.1101/390807

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience, 8*(6), 551-565.

Brummerloh, B., Gundlach, C., & Müller, M. M. (2019). Attentional Facilitation of Constituent Features of an Object Does Not Spread Automatically along Object-defining Cortical Boundaries. *Journal of Cognitive Neuroscience, 31*(2), 278-287. doi:10.1162/jocn_a_01350

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

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision, 10*(4), 16-16. doi:10.1167/10.4.16

Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology, 2*, 342. doi:10.3389/fpsyg.2011.00342

de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife, 4*, e06564.

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

Downing, P., Liu, J., & Kanwisher, N. (2001). Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia, 39*, 1329-1342.

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

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10*(1), 32-39. doi:10.1037//0096-1523.10.1.32

Engell, A. D., & McCarthy, G. (2010). Selective attention modulates face-specific induced gamma oscillations recorded from ventral occipitotemporal cortex. *Journal of Neuroscience, 30*(26), 8780-8786. doi:10.1523/JNEUROSCI.1575-10.2010

Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioural Brain Research, 76*(1), 181-189.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482-498. Retrieved from http://www.sciencedirect.com/science/article/pii/S0033295X02001688

Finkbeiner, M., & Palermo, R. (2009). The Role of Spatial Attention in Nonconscious Processing: A Comparison of Face and Nonface Stimuli. *Psychological Science, 20*(1), 42-51. doi:10.1111/j.1467-9280.2008.02256.x

Furey, M. L., Tanskanen, T., Beauchamp, M. S., Avikainen, S., Uutela, K., Hari, R., & Haxby, J. V. (2006). Dissociation of face-selective cortical responses by attention. *Proceedings of the National Academy of Sciences of the United States of America, 103*(4), 1065-1070.

Gayet, S., & Peelen, M. V. (2022). Preparatory attention incorporates contextual expectations. *Current Biology, 32*(3), 687-692. e686.

Gazzaley, A., Clapp, W., Kelley, J., McEvoy, K., Knight, R. T., & D'Esposito, M. (2008). Age-related top-down suppression deficit in the early stages of cortical visual memory processing. *Proceedings of the National Academy of Sciences of the United States of America, 105*(35), 13122-13126. doi:10.1073/pnas.0806074105

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature neuroscience, 8*(10), 1298-1300.

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology, 4*, 527. doi:10.3389/fpsyg.2013.00527

Grootswagers, T., Robinson, A., Shatek, S. M., & Carlson, T. (2021). The neural dynamics underlying prioritisation of task-relevant information. *Neurons, Behaviour, Data Analysis, and Theory, 5*(1).

Grootswagers, T., Robinson, A. K., & Carlson, T. (2019). The representational dynamics of visual objects in rapid serial visual processing streams. *Neuroimage, 188*, 668-679. doi:https://doi.org/10.1016/j.neuroimage.2018.12.046

Holmes, A., Vuilleumier, P., & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: evidence from event-related brain potentials. *Cognitive Brain Research, 16*(2), 174-184. doi:10.1016/s0926-6410(02)00268-9

Hruby, T., & Marsalek, P. (2003). Event-related potentials--the P3 wave. *Acta Neurobiologiae Experimentalis (Warsaw), 63*(1), 55-63.

Jacques, C., Retter, T. L., & Rossion, B. (2016). A single glance at a face generates larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. *Neuroimage, 137*, 21-33.

Jaeger, M., Bleichner, M. G., Bauer, A.-K. R., Mirkovic, B., & Debener, S. (2018). Did You Listen to the Beat? Auditory Steady-State Responses in the Human Electroencephalogram at 4 and 7 Hz Modulation Rates Reflect Selective Attention. *Brain Topography, 31*(5), 811-826. doi:10.1007/s10548-018-0637-8

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. *Advances in psychophysiology, 3*(April), 69-137.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual review of psychology, 37*(1), 43-75.

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 111*(30), 11217-11222.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience, 17*(11), 4302-4311. Retrieved from http://www.jneurosci.org/content/17/11/4302.abstract

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences, 6*(3), 120-125. doi:https://doi.org/10.1016/S1364-6613(00)01852-0

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review, 111*(4), 880-913. doi:10.1037/0033-295x.111.4.880

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience, 6*(10), 1108-1112. doi:10.1038/nn1121

Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia, 52*, 57-72. doi:10.1016/j.neuropsychologia.2013.10.022

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience, 5*(9), 910-916. doi:10.1038/nn909

Luck, S. J., & Yard, S. A. H. (1995). The role of attention in feature detection and conjunction discrimination: An electrophysiological analysis. *International Journal of Neuroscience, 80*(1-4), 281-297.

Lueschow, A., Sander, T., Boehm, S. G., Nolte, G., Trahms, L., & Curio, G. (2004). Looking for faces: Attention modulates early occipitotemporal object processing. *Psychophysiology, 41*(3), 350-360. doi:10.1111/j.1469-8986.2004.00159.x

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29*(6), 317-322. doi:https://doi.org/10.1016/j.tins.2006.04.001

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 315-330.

Müller, M. M., Andersen, S., Trujillo, N. J., Valdés-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences, 103*(38), 14250-14254. doi:10.1073/pnas.0606668103

Nako, R., Wu, R., Smith, T. J., & Eimer, M. (2014). Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans? *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1283-1288. doi:10.1037/a0036885

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32*(1), 3-25.

Quek, G. L., Liu-Shuang, J., Goffaux, V., & Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *Neuroimage, 176*, 465-476. doi:https://doi.org/10.1016/j.neuroimage.2018.04.034

Quek, G. L., Nemrodov, D., Rossion, B., & Liu-Shuang, J. (2018). Selective Attention to Faces in a Rapid Visual Stream: Hemispheric Differences in Enhancement and Suppression of Category-selective Neural Activity. *Journal of Cognitive Neuroscience, 30*(3), 393-410. doi:10.1162/jocn_a_01220

Quek, G. L., & Peelen, M. V. (2020). Contextual and Spatial Associations Between Objects Interactively Modulate Visual Processing. *Cerebral Cortex, 30*(12), 6391-6404. doi:10.1093/cercor/bhaa197

Quek, G. L., & Rossion, B. (2017). Category-selective human brain processes elicited in fast periodic visual stimulation streams are immune to temporal predictability. *Neuropsychologia, 104*, 182-200. doi:10.1016/j.neuropsychologia.2017.08.010

Quek, G. L., Rossion, B., & Liu-Shuang, J. (2021). Critical information thresholds underlying generic and familiar face categorisation at the same face encounter. *Neuroimage, 243*, 118481. doi:https://doi.org/10.1016/j.neuroimage.2021.118481

Reddy, L., & Kanwisher, N. (2007). Category Selectivity in the Ventral Visual Pathway Confers Robustness to Clutter and Diverted Attention. *Current Biology, 17*(23), 2067-2072. doi:https://doi.org/10.1016/j.cub.2007.10.043

Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Research, 46*, 2336-2343.

Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision, 4*(2), 106-117. doi:10.1167/4.2.4

Rekow, D., Baudouin, J. Y., Poncet, F., Damon, F., Durand, K., Schaal, B., . . . Leleu, A. (2021). Odor-driven face-like categorization in the human infant brain. *Proceedings of the National Academy of Sciences of the United States of America, 118*(21). doi:10.1073/pnas.2014979118

Retter, T. L., Jiang, F., Webster, M. A., & Rossion, B. (2020). All-or-none face categorization in the human brain. *Neuroimage, 213*, 116685. doi:https://doi.org/10.1016/j.neuroimage.2020.116685

Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia, 91*, 9-28. doi:10.1016/j.neuropsychologia.2016.07.028

Retter, T. L., Rossion, B., & Schiltz, C. (2021). Harmonic Amplitude Summation for Frequency-tagging Analysis. *Journal of Cognitive Neuroscience, 33*(11), 2372-2393. doi:10.1162/jocn_a_01763

Rossion, B. (2014). Understanding individual face discrimination by means of fast periodic visual stimulation. *Experimental brain research, 232*(6), 1599-1621.

Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision, 15*(1), 15.11.18. doi:10.1167/15.1.18

Sato, S., & Kawahara, J. I. (2015). Attentional capture by completely task-irrelevant faces. *Psychological Research, 79*(4), 523-533. doi:10.1007/s00426-014-0599-8

Sreenivasan, K. K., Goldstein, J. M., Lustig, A. G., Rivas, L. R., & Jha, A. P. (2009). Attention to faces modulates early face processing during low but not high face discriminability. *Attention, Perception, & Psychophysics, 71*(4), 837-846.

Stein, T., & Peelen, M. V. (2017). Object detection in natural scenes: Independent effects of spatial and category-based attention. *Attention, Perception, & Psychophysics, 79*(3), 738-752. doi:10.3758/s13414-017-1279-8

Störmer, V. S., Cohen, M. A., & Alvarez, G. A. (2019). Tuning attention to object categories: Spatially global effects of attention to faces in visual processing. *Journal of cognitive neuroscience, 31*(7), 937-947.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409.

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition, 13*(6), 657-665

Vuilleumier, P. (2000). Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia, 38*, 693-700.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron, 30*(3), 829-841. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627301003282

Wardle, S. G., Paranjape, S., Taubert, J., & Baker, C. I. (2022). Illusory faces are more likely to be perceived as male than female. *Proceedings of the National Academy of Sciences, 119*(5), e2117413119.

Wardle, S. G., Taubert, J., Teichmann, L., & Baker, C. I. (2020). Rapid and dynamic processing of face pareidolia in the human brain. *Nature Communications, 11*(1), 4518. doi:10.1038/s41467-020-18325-8

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI Study. *Journal of Neurophysiology, 79*(3), 1574-1578. Retrieved from http://jn.physiology.org/content/79/3/1574.abstract

Yeh, L.-C., & Peelen, M. V. (2022). The time course of categorical and perceptual similarity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 48*, 1069-1082. doi:10.1037/xhp0001034

Zanto, T. P., Hennigan, K., Östberg, M., Clapp, W. C., & Gazzaley, A. (2010). Predictive knowledge of stimulus relevance does not influence top-down suppression of irrelevant information in older adults. *Cortex; a journal devoted to the study of the nervous system and behavior, 46*(4), 564-574. doi:10.1016/j.cortex.2009.08.003

Zivony, A., & Eimer, M. (2020). Perceptual competition between targets and distractors determines working memory access and produces intrusion errors in rapid serial visual presentation (RSVP) tasks. *Journal of Experimental Psychology: Human Perception and Performance, 46*, 1490-1510. doi:10.1037/xhp0000871