Contents lists available at ScienceDirect

ELSEVIER



Neuropsychologia

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

Mean emotion from multiple facial expressions can be extracted with limited attention: Evidence from visual ERPs



Luyan Ji*, Valentina Rossi, Gilles Pourtois

Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium

ARTICLE INFO

Ensemble representation

Facial expressions

Keywords:

Emotion

Attention

N2pc

SPCN

Human obse

ABSTRACT

Human observers can readily extract the mean emotion from multiple faces shown briefly. However, it remains currently debated whether this ability depends on attention or not. To address this question, in this study, we recorded lateralized event-related brain potentials (i.e., N2pc and SPCN) to track covert shifts of spatial attention, while healthy adult participants discriminated the mean emotion of four faces shown in the periphery at an attended or unattended spatial location, using a cueing technique. As a control condition, they were asked to discriminate the emotional expression of a single face shown in the periphery. Analyses of saccade-free data showed that the mean emotion discrimination ability was above chance level but statistically undistinguishable between the attended and unattended location, suggesting that attention was not a pre-requisite for averaging. Interestingly, at the ERP level, covert shifts of spatial attention were captured by the N2pc and SPCN components. All together, these novel findings suggest that averaging multiple facial expressions shown in the periphery can operate with limited attention.

1. Introduction

Facial expressions carry important social and emotional information which can be used to guide and optimize communication between people. In many natural settings and environments, humans usually interact with multiple other individuals simultaneously. For example, in an auditorium, some of the students' faces may seem pleased while some others may display some signs of disapproval or concerns, leading in turn the instructor to experience a rather mixed feeling regarding the disposition of his/her audience. Growing evidence in experimental psychology suggests that human observers can rapidly and rather precisely extract mean emotion from mixed valences in multiple faces presented concurrently (e.g., Haberman and Whitney, 2007, 2009; Li et al., 2016). The perceived average emotion provides a rather accurate summary statistic of the complex scene or display composed of multiple emotional faces, usually referred to as ensemble representation (Alvarez, 2011; Whitney and Leib, 2018). As such, this averaging allows collapsing or combining multiple individual facial expressions into a coherent and integrated emotion percept that carries the information of mean intensity and/or valence of the scene. This remarkable ability of establishing ensemble representation is thought to provide an efficient way to cope with the bottlenecks in visual processing (Alvarez, 2011; Chong and Treisman, 2003; Whitney et al., 2014) and to reconcile the subjective impression of a rich visual world with the limited perceptual and attention capacities (Cohen et al., 2016). However, the mechanism underlying ensemble representation is still largely unclear, and discrepant results have been reported in the past.

Some earlier research argued that extracting mean information from multiple items is best conceived as a capacity-limited perceptual process (Attarha et al., 2014; Florey et al., 2016; Jacoby et al., 2013; Ji et al., submitted). In a recent study (Ji et al., submitted), we used the extended simultaneous-sequential paradigm (Scharff et al., 2011) to examine the processing capacity for extracting mean emotion from multiple facial expressions (with variations along the valence dimension; the faces being either happy or angry). The results showed that performance in the sequential condition (where two successive displays each containing 8 faces were shown) was better than that in the simultaneous condition (where the 16 individual faces were presented at the same time), which was consistent with the limited-capacity model assuming that all items in an ensemble could not be processed independently or without interference. In addition, a previous study using the attentional blink (AB) paradigm found a clear AB effect when estimating the average emotion from four faces (target 2) that followed, after a short lag, a first face whose gender had to be discriminated (target 1), indicating that average emotion processing suffers from the temporal limits of attention deployment (McNaire et al., 2017).

Although extracting the mean emotion from a set of individual facial expressions seems to obey to capacity limitations, it is not known

E-mail address: Luyan.Ji@Ugent.be (L. Ji).

https://doi.org/10.1016/j.neuropsychologia.2018.01.022

^{*} Corresponding author.

Received 7 September 2017; Received in revised form 4 December 2017; Accepted 16 January 2018 0028-3932/ © 2018 Elsevier Ltd. All rights reserved.

yet whether an ensemble representation for multiple facial expressions could be established with no attention or limited attention. Earlier psychophysical studies on low-level features showed that mean representation can be formed outside the focus of attention or with reduced spatial attention allocated to its actual content (Alvarez and Oliva, 2008, 2009; Demeyere et al., 2008; Leib et al., 2012; but cf. Huang, 2015; Jackson-Nielsen et al., 2017). The goal of the current study was twofold. (i) First, we examined whether the average emotion could be extracted with limited attention. For this purpose, we combined a spatial-cueing procedure (Posner, 1980; Mangun and Hillyard, 1991) with an average emotion task where we manipulated the average emotion of a face set by systematically varying across trials the ratio of positive and negative faces contained in the set. In short, in every trial, observers were required to judge the average emotion (either positive or negative) of a target set composed of four faces conveying a variable amount of happy and angry expressions and the target set was presented in a valid (75%) or invalid (25%) peripheral location, while distractor faces were presented in the opposite location. At the behavioral level, we compared performance between the valid and invalid condition. We reasoned that if performance in the invalid condition was above chance level (and/or similar to that in the valid condition), then this could be interpreted as evidence in favor of the possibility to extract the mean emotion from the set with limited attention. Conversely, if participants could not discriminate above chance level the mean emotion in the invalid condition, then this could be taken as evidence that the ensemble representation could not be established with limited attention. The presence of a significant validity effect (i.e., better performance in the valid compared to the invalid condition) would accord with a classical attention gating effect whereby a more accurate mean representation could be achieved when selective attention is allocated to this complex stimulus (Hillyard et al., 1998). (ii) Additionally, we also investigated whether averaging multiple facial expressions could be dissociated from recognizing a single emotional facial expression. To this aim, we compared behavioral performance of the same subjects between two tasks. Either participants had to discriminate the emotional expression of a single face (single emotion task) or they performed the average emotion task, as described here above.

Noteworthy, visual event-related brain potentials (ERPs) were recorded concurrently and used to gain insight into the time-course of establishing an ensemble representation for multiple facial expressions. More specifically, we capitalized on two well-known lateralized ERP components that are sensitive to spatial attention (and short term visual memory) manipulations, namely the N2pc (N2 posterior contralateral) and the SPCN (sustained posterior contralateral negativity).

The N2pc is a lateralized component, characterized by a larger negativity at posterior (occipito-temporal) electrodes sites contralateral versus ipsilateral to the attended location, and is usually observed approximately 200-300 ms after stimulus onset at lateral occipital leads (Luck and Hillyard, 1994). This ERP component is usually found in visual search tasks, and thought to reflect spatially selective attention to target stimuli (Eimer, 1996; Woodman and Luck, 1999). The N2pc was also observed previously during a spatial cueing task (Kiss et al., 2008; Woodman et al., 2009). For example, following an informative (100% valid) arrow cue in the center, an N2pc in response to targets was still reliably elicited (Kiss et al., 2008; Praamstra, 2006). More importantly, the N2pc also provides a neural index of individuation, or in other words, forming distinct representations of each individual item at the same time. In line with this assumption, the amplitude of N2pc usually increases with the number of to-be-selected items within the attended hemifield (Drew and Vogel, 2008; Ester et al., 2012; Mazza and Caramazza, 2011; Pagano and Mazza, 2012). For example, the more items need to be enumerated or to be tracked, the larger the amplitude of the N2pc, reaching a plateau usually at a set size over four items, indicating a limitation in simultaneously selecting or individuating these items (Drew and Vogel, 2008; Mazza et al., 2013). However, the relationship between the amplitude of the N2pc and set size is best

evidenced when the individuation of multiple stimuli is required. When observers had only to detect the presence of a specific color within the set for example, the N2pc amplitude was not influenced any more by the number of items therein (Mazza and Caramazza, 2011). The SPCN often follows the N2pc when more detailed processing and/or memory for the lateralized target is required (Brisson and Jolicœur, 2007; Jolicœur et al., 2008; Mazza et al., 2007). The SPCN (also called contralateral delay activity - CDA sometimes in the extant literature) has a strong link with visual short-term memory (Eimer and Kiss, 2010; Klaver et al., 1999; McCollough et al., 2007; Vogel and Machizawa, 2004). The amplitude of the SPCN usually increases as the number of items held in visual short-term memory increases, up to the observer's memory capacity (McCollough et al., 2007; Vogel and Machizawa, 2004).

Using these specific electrophysiological markers of attention selection (N2pc) and visual short-term memory (SPCN), we first hypothesized that the N2pc would be observed in the valid condition, indicating allocation of attention to the target face stimuli. In the invalid condition, we surmised that the N2pc would be strongly reduced, if not fully absent. With regard to our second research question, we conjectured that the N2pc could be larger in the average emotion task (four target faces) compared with the single emotion task (one target face), assuming that extracting the mean emotion from multiple facial expressions requires individuating/computing individual faces. On the other hand, if establishing a mean emotion representation does not require individuating but can be computed based on a so-called "total activation map" (Šetić et al., 2007), or alternatively participants use a subsampling strategy (selecting one face out of the four available in the average emotion task), then the amplitude of the N2pc should be similar for the two tasks. Although the current study was not designed a priori to explore visual short term memory, we nevertheless assessed whether a SPCN could be elicited following the N2pc in this experiment. As a matter of fact, it may be the case that the emotional information extracted from the target stimulus first needs to be shortly retained in a visual buffer ("short-term memory") after visual presentation for further processing and elaboration (e.g., discrimination), leading in turn to the generation of a SPCN. In this scenario, if the four facial expressions are collapsed or compressed into one summary statistic (i.e., mean) (Alvarez, 2011; Brady and Alvarez, 2011), then this memory load (or buffer information) would be comparable for the two tasks, leading thereby to a similar SPCN for them. On the other hand, if the individual representations are maintained along with the summary estimate, or the mean representation retained in visual short-term memory consisted of low compressed or less structured features (Baijal et al., 2013; Treisman, 2006), then a larger SPCN would be observed primarily in the average emotion task compared with the single emotion task.

2. Method

2.1. Participants

Thirty-six volunteers (age: M = 22.6 years, SD = 2.4; 24 females) from Ghent University participated in this study after giving written informed consent and were compensated €30. All participants reported to be right-handed and have normal or corrected-to-normal vision. The study protocol was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee.

2.2. Stimuli and design

Four male and four female face identities were selected from NimStim database (Tottenham et al., 2009). Each face identity shows happy, angry and neutral expressions, all with closed mouths. The hair, ears, neck and other external information were cropped by an oval frame. All images were converted to greyscale, and scaled to the same



Fig. 1. Procedure of the average and the single emotion task. Participants were required to judge the valence (positive or negative) of the average emotion from four target faces (Left, average emotion task) or the emotion of the single face in the target set (Right, single emotion task), present either in the validly cued (75%) or the invalidly cued location (25%). The next trial automatically began (randomly varying between) 1000–1200 ms after participant responded. The target emotion in this example was both positive in the average and the single tasks. The distractors (opposite side) were either four or one neutral face(s), respectively.

mean luminance and root-mean-square contrast (Bex and Makous, 2002).

In the average emotion task, the target set of four faces conveying a variable amount of happy and angry expressions was presented in either the left or the right visual field. The ratio of happy faces in the set was .25, .5 or .75. On the opposite side, there was another set containing four neutral faces, hence yielding bilateral stimulus presentations (Fig. 1).

In the single emotion task, there was either one happy or one angry face flanked by three scrambled faces in the target set (Fig. 1). The distractor set (shown in the opposite visual field, similarly to the average emotion condition) included one neutral face flanked by three scrambled faces. For scrambling, we used Adobe Photoshop (Adobe Systems Corporation, San Jose, CA) to crop and rearrange the key internal features (eyes, nose, mouth, and forehead) of the original happy and angry faces, thus the scrambled images were still face-like to some degree, and the low-level features were maintained while the emotional information could no longer be extracted from them. The emotion of the three scrambled faces in each set was randomly selected.

For both tasks, face identities in each set were randomly selected with two specific constraints: 1) an equal number of male and female faces were presented; 2) the same identity was never repeated in the pair of two sets.

The pair of one target and one distractor set was presented in the left and the right visual field 3.52° lateral to the fixation. The four faces (or one face with three scrambled ones) in each set were shown in a 2×2 invisible grid, on a homogenous black background (Fig. 1). The position of these four faces was randomly selected. In the single emotion task, the single intact face in the two sets was presented in mirror symmetry. A white fixation point was continuously present on the

center of the screen. The two inner faces (i.e., being closer to the fixation, $3.28^{\circ} \times 2.15^{\circ}$) subtended a visual angle of $2.38^{\circ} \times 2.38^{\circ}$. The two outer faces (6.35° \times 2.15°) were scaled to 3.42° \times 3.45° to compensate for differences in V1 cortical representation/magnification (Rousselet et al., 2005) following a standard formula¹ (Dougherty et al., 2003; Horton and Hoyt, 1991). The outline of the outer grid (3 pixels) was also always visible on the screen, to help the participants to attend to the part of the visual field where the faces were actually presented. At the onset of the two face sets (one target and one distractor), the outline was converted from white to either blue or green, indicating the target or the distractor respectively (this color/validity mapping was counterbalanced across participants), and these two colors remained when the masks appeared. The masks had the same size and appeared at the same location as the faces in the target and distractor set (Fig. 1C). We used MATLAB (Mathworks, Natick, MA) to divide neutral faces into a 10 $\,\times\,$ 12 matrix of randomly arranged squares and trimmed them with an oval frame against the black background. Before the presentation of the two face sets, there was a centrally located arrow cue (1.61° \times 1.28°), which had the same color as the outline of the target set.

¹ $M_{\text{linear}} = A/(E+e2)$, with *E* the eccentricity in degrees, *A* the cortical scaling factor in mm, and *e2* the eccentricity in degrees at which a stimulus subtends half the cortical distance that it subtends at the fovea. We used A = 29.2 mm and $e2 = 3.67^{\circ}$ based on a recent report of the cortical magnification factor in V1 (Dougherty et al., 2003), similarly to Rousselet et al. (2005). The face image at 3.28° stimulates a cortical surface area that is (6.35 + 3.67)/(3.28 + 3.67) = 1.44 times larger than the surface stimulated by the same image when it is presented at 6.35° .

2.3. Procedure

Participants seated at 75 cm in front of a 19" CRT screen (resolution of 1024 $\,\times\,$ 768 pixels, refresh rate 100 Hz) in a dimly lit cabin. To minimize head movements, a chinrest was used during both tasks. While fixating at the central cross, participants were required to discriminate (with peripheral vision) whether the (average) emotion of the target face set was positive or negative, by pressing one out of two predefined buttons with their right hand ("2" or "8" using a standard number keyboard; this mapping being counterbalanced across participants) as accurate and as fast as possible. A trial began with a fixation cross for 500 ms, followed by an arrow cue superimposed on the fixation cross for 200 ms, both of which were presented on the center of the screen. After an ISI randomly varying between 500 and 700 ms, the pair of two face sets was shown for 150 ms. The target face set mostly (p =.75) appeared at the spatial position indicated by the preceding arrow, and equally likely in the left or the right visual field. Then, the face sets were masked and the masks were presented for 1000 ms or terminated by response (Fig. 1). The next trial automatically began (randomly varying between) 1000-1200 ms after participant responded. Participants were informed about target validity and probability, and they were encouraged to rely on their first impression and not to think extensively.

The two tasks, either judging the average emotion from multiple faces (average emotion task) or identifying the emotion from the single face (single emotion task), were blocked, and the order of blocks was counterbalanced across participants. The ratio of happy faces (.25, .5, .75) in the average emotion task or the valence of one emotional face (happy, angry) in the single emotion task was randomized within blocks. Every trial had a unique face set to minimize the visual statistical regularity between trials. Participants performed 12 experimental blocks of 48 trials each (36 valid, 12 invalid) for the average task, and 6 blocks of 64 trials each (48 valid, 16 invalid) for the single task. Before starting the experiment, participants got acquainted with the two emotion judgment tasks with 20 practice trials each. Practice trials were excluded from all subsequent analyses. In order to encourage participants to focus on the center of the screen (fixation point) and only use peripheral vision to process face sets, the procedure also incorporated four catch trials each block, where a white dot (.60°) unexpectedly replaced the fixation cross. Participants were asked to press the spacebar with their left hand when they detected the dot. They did not need to judge the emotion when a dot appeared.

After the average and single emotion tasks, participants rated the valence and emotion intensity of each face previously presented. To this aim, one face appeared at a time in the center and had the same size as that in the previous task. Participants judged on a Visual Analogue Scale (VAS). The two anchors of the VAS for emotion valence and intensity were labeled Extremely positive and Extremely negative. Additionally, we also asked participants to rate their arousal level to each individual face. The two anchors of VAS for arousal were labeled Extremely calm and Extremely excited. The labels on the left and right side of both VASes were counterbalanced across participants. The main goal of these post-experiment ratings was to confirm that the happy, angry and neutral faces used in the main experiment were perceived as such and hence showed differences in terms of valence and arousal. Moreover, following the procedure adopted in our previous study (Ji et al., submitted), we directly used these post-experiment ratings to compute subject-specific mean emotions for the different face sets used in the main experiment and assessed if (objective) changes in the ratio of happy faces in these sets were related to the (subjective) estimates (see Supplementary materials section for details).

All the tasks were programmed and controlled using the E-Prime Version 2 software (Psychology Software Tools, Inc., 2001). The experiment lasted about 90 min.

2.4. Electrophysiological recording and preprocessing

The electroencephalographic (EEG) activity was continuously recorded from 64 active Ag/AgCl electrodes positioned according to the extended 10–20 system in an elastic cap (BioSemi ActiveTwo system) during the average and single emotion tasks. The EEG signals were referenced online to the CMS-DRL electrodes and sampled at 512 Hz. Additional bipolar electrodes were placed above and below the left eye, and the voltage difference between them was recorded as vertical electro-oculogram (EOG). The voltage difference between the electrodes at the left and the right outer canthus was recorded as horizontal EOG (HEOG).

The preprocessing was performed using EEGLAB (Delorme and Makeig, 2004) and ERPLAB (Lopez-Calderon and Luck, 2014). A highpass filter of .05 Hz and a low-pass filter of 80 Hz was firstly applied. Data were then referenced offline to the averaged reference after the noisy channels were interpolated by a spherical splines procedure. In order to exclude potential saccades to the peripheral face sets during the cue-target interval, EEG and EOG were first segmented into long epochs beginning 200 ms before cue onset and up to 2250 ms following it, baseline-corrected using the 200 ms pre-cue interval. Epochs containing horizontal eye movements (as identified using HEOG) within 900 ms after cue onset were automatically detected by the step-like artifact function (searching for step-shaped segments of data on the channel of HEOG, with window step set as 10 ms, moving windows full width as 400 ms, and voltage threshold as $15 \,\mu$ V), as implemented in ERPLAB.

A total of 41% of trials was marked to be excluded due to the steplike HEOG artifacts occurring during the cue-target interval. The number of trials detected was similar for all the experimental conditions (valid left, valid right, invalid left, invalid right). The large number of trials with HEOG during the cue-target interval indicated that for a large amount of trials, participants already moved their eyes following the cue, or had produced saccades towards the target location before its actual onset. Supplementary Fig. 3 shows representative examples of HEOG traces obtained for two subjects (one with no saccades and one with clear saccades) recorded during the cue-target interval. Out of 36 participants, we could actually retain only 19 who had more than half of the trials left without saccades (after analysis of the HEOG) and subsequent artifact rejections in both tasks, as well as no clear residual eye movements (less than 5 µV, deviated less than 0.3°, Lins et al., 1993). These 19 participants were used in the following EEG analyses. For this no-clear-saccade group, on average 23% of trials were rejected in the HEOG rejection procedure.² The original saccade-free long epochs were segmented into shorter epochs (-200 to 800 ms) time-locked to the onset of the target face sets. Independent component analysis (EEGLAB's runica algorithm) was adopted to remove components related to eye blinks. Epochs containing activity exceeding \pm 80 μ V in the scalp EEG electrodes were automatically rejected (on average 7% trials, a number which was not different between task and validity conditions). The artifact-free data were then baseline-corrected using the 200 ms pre-stimulus interval. We averaged only the correct trials (thus the Ratio0.5 condition in the average task was not included) separately for each type of task (average, single), validity (valid, invalid), and target location (left, right).³ Grand average ERPs were computed by averaging mean ERPs of these19 participants for each condition

 $^{^2}$ For this no-clear-saccade group, the number of trials rejected based on the HEOG rejection procedure was larger in the average task (25 \pm 10%, the Ratio0.5 condition was dropped to match the total number of trials with the single task) compared with the single task (21 \pm 10%), p=.016; however no significant difference was found between valid (22 \pm 10%) vs. invalid (23 \pm 10%) conditions, nor the interaction between task type and validity reached significance, ps>.34.

³ On average, there were 120 (SD = 27, Min = 79), 118 (SD = 25, Min = 81), 38 (SD = 9, Min = 24) and 36 (SD = 9, Min = 23) trials included in the ERP average of the average-valid, single-valid, average-invalid, and single-invalid condition, respectively. As expected, trial count in the valid condition (239 ± 48) was larger than in the invalid

separately. For the other 17 participants, a large amount of residual eye movements during the cue-target interval (more than 5 μ V) after HEOG rejection was observed and/or not enough trials were left after artifact rejection.⁴ The results obtained for this clear-saccade group (where trials with clear saccades only were eventually retained, based on the HEOG channel analysis) are presented in the Supplementary materials section for comparison purposes.

2.5. Data analyses

For the average and the single emotion tasks, the accuracy of catch trials was first calculated. The subsequent analyses were based on trials without catch trials. For the average task, the proportion of "positive" responses as a function of the ratio of happy faces in the sets was first computed to confirm that participants' average emotion judgments were sensitive to the variable amount of happy vs angry expressions contained in the sets. Next, the discrimination ability (d prime) was calculated. Importantly, the statistical analysis of behavioral data was aligned to the EEG data analysis: accuracy and reaction time (RT) for the saccade-free trials only were computed (hence trials contaminated by eve movements were first removed). To compute d prime scores, hits and false alarms were defined as follows. The positive face sets, containing three happy faces in the average emotion task or one happy face in the single emotion task were considered as target. The negative sets, which contained three angry faces (average emotion task) or one angry face (single emotion task), were considered as noise. Hits corresponded to judging the positive face sets as positive, and false alarms corresponded to judging the negative sets as positive. The sets composed of 50% happy (n = 2) and 50% angry (n = 2) faces in the average emotion task were not included in calculating d prime scores or RTs for correct responses. For each dependent variable separately (d prime and RT for correct trials), a two-way repeated-measures ANOVAs was carried out with the factors Task and Validity. As a control analysis, we also examined whether emotional faces shown close to fixation (two inner positions) had a larger impact on performance compared to faces shown further away from it (two outer faces; see Supplementary materials section).

For the face rating task, the actual positions participants clicked on the VASes were converted to data ranging from 0 to 100. After conversion, the larger the value, the stronger the emotion intensity or more aroused the participants judged the emotion of the face to be. For the no-clear saccade group, paired sample *t*-tests between angry and happy faces, or between emotional and neutral faces were conducted on emotion intensity and arousal scores using individual faces (as opposed to participants) as degrees of freedom in these analyses.

ERP mean amplitudes were computed at the lateral posterior electrodes PO7/PO8, during the 200–300 ms (N2pc) or 400–600 ms (SPCN) post-stimulus onset interval. The N2pc and SPCN were quantified by subtracting the mean amplitude recorded at the ipsilateral electrodes (relative to the location of the target faces) from that at the contralateral electrodes. Data were then submitted to repeated-measures ANOVAs with Task and Validity as factors. Greenhouse-Geisser correction was applied when assumptions of sphericity were violated. A Bonferroni correction was used when multiple comparisons were performed. We also examined the face-specific N170 component timelocked to target onset, as it was previously found to be modulated by the number of faces shown in the set (Puce et al., 2013). Details of these complementing analyses and results for the N170 are provided in the Supplementary materials section.

3. Results

3.1. Behavioral results

3.1.1. Catch trials

Accuracy was high in both the average and the single emotion tasks (average: M = .90, SD = .15; single: M = .91, SD = .10). Interestingly, when collapsing data across all the 36 participants, the accuracy for the catch trials negatively correlated with the overall amount of saccades, r = -.39, p = .019. It indicated that while these catch trials enforced central fixation to some degree, they did not however fully prevent the generation of saccades towards peripheral target locations.

3.1.2. Average and single emotion judgment

For the average task, the proportion of "positive" responses was sensitive to the ratio of happy faces contained in the sets, *F* (1.29, 23.15) = 42.70, p < .001, $\eta_p^2 = .70$, confirming that participants' judgments were influenced by this manipulation (Fig. 2). There was also a significant main effect of Validity in this analysis, *F* (1, 18) = 5.89, p = .026, $\eta_p^2 = .25$. As can be seen from Fig. 2, the proportion of "positive" responses was larger in the valid compared to the invalid condition. Tentatively, this effect could be explained by a response bias, namely being inclined to judge the face sets as positive in the valid condition. ⁵ However, the interaction between Ratio and Validity was not significant, *F* (1.97, 35.45) < 1, $\eta_p^2 = .04$, suggesting that in these two conditions (valid and invalid), participants' judgements were similarly influenced by the ratio of happy faces contained in the sets.

Table 1 shows the hit rate, the false alarm rate, and the corresponding d prime score for each condition separately. The discrimination ability (d prime) in the average and the single emotion task was generally not high (i.e. lower than one). However, in all four main conditions (average-valid, single-valid, average-invalid, and single-invalid), it was significantly above chance level (higher than zero, ps < .001, Fig. 3). The ANOVA on the d prime scores revealed no significant main effect of Task, F (1, 18) = 1.82, p = .19, $\eta_p^2 = .09$, however a marginally significant main effect of Validity, F(1, 18) =3.28, p = .087, $\eta_p^2 = .15$. The interaction between Task and Validity was also marginally significant, *F* (1, 18) = 3.29, *p* = .086, η_p^2 = .16. Simple effect analysis revealed that for the average task, there was no significant difference between the valid and the invalid condition, F(1, 1)18) < 1, η_p^2 = .01; while for the single task, the performance was significantly better in the valid compared with the invalid condition, F(1,18) = 6.23, p = .023, $\eta_p^2 = .26$. The performance in the valid condition of the average task did not differ significantly from that of the single task, *F* (1, 18) < 1, $\eta_p^2 = .02$; whereas for the invalid condition, the performance was slightly better in the average task relative to the single task, F(1, 18) = 3.03, p = .099, $\eta_p^2 = .14$. The ANOVA performed on the RTs revealed a significant main effect of Validity, F(1, 18) = 16.93, $p = .001, \eta_p^2 = .49$. The RTs were significantly shorter in the valid (M = 669.48, SD = 107.69) compared to the invalid condition (M =

⁽footnote continued)

condition (75 ± 15), *F*(1, 18) = 362.71, *p* < .001, η_p^2 = .95, while the number of trials in the average task (159 ± 34) did not differ significantly compared with the single task (155 ± 32), *F*(1, 18) < 1, *p* = .51, η_p^2 = .024. In addition, two-ways and the three-way interactions were all non-significant, *ps* > .087.

⁴ Eleven of them still had substantial residual eye movements (over 5 μ V) in at least one task after the HEOG rejection procedure (on average 61% trials were rejected; nine among them had less than 50% of trials left, and for the two other ones, less than half of the trials could be retained when more strict criteria were used to keep the residual deviance below 5 μ V); another six who had no clear deviance of eye movements after HEOG rejection had no more than half of the trials remaining after subsequent artifact rejection steps however (on average 51% trials were rejected due to HEOG artifacts) in at least one task.

⁵ The c score, an index of response bias (Stanislaw and Todorov, 1999), confirmed these observations. The c score in the valid condition of the average task (M = -.17, SD = .36) was smaller than zero, indicating a positive response bias, t (18) = -2.09, p = .051; while it did not differ significantly from zero in the invalid condition of the average task (M = .01, SD = .44), indicating no response bias, t (18) = .05, p = .96.



No-clear-saccade group

Fig. 2. Proportions of positive (happy) judgements shown as a function of the ratio of happy faces contained in the sets, as well as separately for the valid and invalid conditions. The error bar represents one standard error of mean.

Table 1

Summary of the hit rate, the false alarm rate, and the corresponding d-prime score in each condition.

	Average_valid	Average_invalid	Single_valid	Single_invalid
Hit	.67(.12)	.60(.18)	.58(.15)	.55(.10)
False alarm	.45(.17)	.40(.17)	.38(.14)	.41(.14)
d prime	.61(.41)	.58(.39)	.54(.35)	.35(.42)

Note. Mean (and standard deviation) is provided. The d prime score was significantly above chance level in all the four main conditions.

3.2. ERP results

As can be seen from Fig. 4A, in the valid condition, a reliable amplitude difference between the contralateral and ipsilateral waveforms in both the average and single emotion tasks started at around 200 ms after face set onset, corresponding to the N2pc. Later in time, these two waveforms converged and were then followed by another contralateral negativity emerging at around 400 ms post stimulus, corresponding to the SPCN. The topographic maps of the N2pc and SPCN in the valid condition were shown in Fig. 5. By comparison, no clear N2pc and SPCN were observed in the invalid condition, especially for the average emotion task. These two successive validity effects on the N2pc and the SPCN could also be revealed when computing the difference waves (see Fig. 4B), for which ERP at the ipsilateral electrodes were subtracted from that at the contralateral ones.

The presence of an N2pc was confirmed by one-sample *t*-tests against zero for the valid condition of both the average $(-1.22 \pm 1.64 \,\mu\text{V})$ and single emotion tasks $(-.63 \pm .88 \,\mu\text{V})$, t(18) = -3.25, p = .004, t(18) = -3.14, p = .006 (see Fig. 4). No reliable N2pc was elicited in the invalid condition of the two tasks (average: $-.01 \pm 2.02 \,\mu\text{V}$, single: $-.64 \pm 2.01 \,\mu\text{V}$), t(18) = -.02, p = .98, t(18) = -.1.38, p = .18. The ANOVA performed on the N2pc revealed no significant main effect of Task, F(1, 18) = 1.58, p = .23, $\eta_p^2 = .08$, or main effect of Validity, F(1, 18) < 1, $\eta_p^2 < .001$, but a significant interaction between Task and Validity, F(1, 18) = 6.59, p = .019, $\eta_p^2 = .27$. The validity effect on the N2pc was pronounced in the average emotion task, F(1, 18) = 4.55, p = .047, $\eta_p^2 = .20$, but very weak in the single emotion task, F(1, 18) < 1, $\eta_p^2 < .001$. Examining the N2pc for the valid and the invalid condition separately, we found that the N2pc



Fig. 3. Discrimination ability (d prime, Left) and reaction time for correct responses (correct RT, Right) shown separately for the two tasks and two levels of validity. The error bar represents one standard error of mean.

716.78, SD = 136.13). The main effect of Task, or the interaction between Task and Validity was not significant, ps > .19 (Fig. 3).

3.1.3. Face emotion ratings

Paired sample *t*-tests showed that the perceived emotion intensity of angry faces (M = 81.12, SD = 7.79) was overall stronger than that of happy faces (M = 73.51, SD = 5.52), t(7) = 2.61, p = .035. Angry faces (M = 53.07, SD = 3.70) were rated as equally aroused as happy faces (M = 54.03, SD = 7.34), t(7) = -.40, p = .70. Neutral faces were perceived as negative, since the comparison (based on a one-sample *t*-test) showed that their ratings were significantly lower than 50 (the smaller the value, the more negative the faces were perceived; M = 42.12, SD = 3.46), t(7) = -6.44, p < .001. In addition, neutral faces were rated as less aroused (M = 31.92, SD = 4.26) than both happy and angry faces, t(7) = -9.86, t(7) = -9.54, ps < .001.

was numerically larger in the average task than in the single task for the valid condition, F (1, 18) = 2.52, p = .13, η_p^2 = .12; while the comparison of N2pc between the two tasks showed the opposite pattern in the invalid condition, F (1, 18) = 1.35, p = .36, $\eta_p^2 = .07$, but this difference did not reach significance either. On the other hand, during this N2pc time window (200-300 ms after the onset of face sets), the mean amplitude was found to be bilaterally more negative in the average emotion task $(2.32 \pm 3.57 \,\mu\text{V})$ compared with the single emotion task (3.44 \pm 3.26 μ V) in both the valid and the invalid conditions, regardless of the electrodes being considered, either contralateral or ipsilateral to the location of target faces, F(1, 18) = 5.83, p = .027, η_p^2 = .25. We also analyzed the early N2pc in a narrower time window (i.e., 200-250 ms post-stimulus onset) to investigate whether spatial attention was first directed to the distractor locations (being opposite to the target location) following the invalid cue leading to a reversed N2pc. The early N2pc was numerically above zero in the



Fig. 4. (A) Grand-averaged ERPs for the correct trials in response to face sets at PO7/8 contralateral (solid lines) and ipsilateral (dashed lines) to the position of the target face(s). (B) Difference waveforms computed by subtracting ERP at PO7/8 ipsilateral to the target location from that at contralateral electrodes for the average and the single emotion tasks for the noclear-saccade group, shown separately for the valid cue (Left) and the invalid cue (Right) condition. The highlighted areas indicated the time-window of N2pc (200–300 ms) and SPCN (400–600 ms) after target onset.

invalid condition of the average task (.31 ± 1.84 µV, the mean amplitude in the ipsilateral electrode to the target was more negative than in the contralateral one), however not significantly so, t (18) = .73, p = .47. The early N2pc in the invalid condition of the single task ($-.76 \pm 2.63 \mu$ V) did not significantly differ from zero either, t (18) = -1.26, p = .23, confirming no reliable early N2pc elicited in the invalid condition might be imputed to the fact that this component could possibly be delayed in this condition compared to the valid condition. To test this hypothesis, we ran a control analysis and extracted the mean amplitude of the N2pc during the 250–350 ms interval post-target onset at the same electrode positions. However, no reliable N2pc (average: $-.32 \pm 2.36 \mu$ V; single: $-.65 \pm 1.61 \mu$ V) was detectable in this later time frame either, t (18) = -.59, p = .56, t (18) = -1.76, p = .10.

The ANOVA performed on the SPCN⁶ revealed no significant main effect of Task, *F* (1, 18) < 1, $\eta_p^2 = .05$, or main effect of Validity, *F* (1, 18) < 1, p = .40, $\eta_p^2 = .04$, nor significant interaction between the two factors, *F* (1, 18) = 2.42, p = .14, $\eta_p^2 = .12$. The SPCN was observed in the valid condition of both the average ($-.95 \pm 2.20 \,\mu$ V) and the single task ($-.88 \pm 1.36 \,\mu$ V), *t* (18) = -1.89, p = .075, *t* (18) = -2.83, p = .011, however they did not differ significantly between the two tasks, *F*

(1, 18) < 1, p = .90, $\eta_p^2 = .002$ (see Fig. 4). The SPCN was also present in the invalid condition of the single task ($-.89 \pm 1.88 \,\mu$ V), t (18) = -2.08, p = .052, but not of the average task ($-.05 \pm 2.79 \,\mu$ V), t (18) = -.08, p = .94. In addition, the mean amplitude in the 400–600 ms post-stimulus time window was generally more negative in the average emotion task ($2.26 \pm 2.92 \,\mu$ V) compared with the single emotion task ($3.58 \pm 3.26 \,\mu$ V), regardless of cue validity or contralaterality of the recording electrodes to the target location, F (1, 18) = 9.31, p = .007, $\eta_p^2 = .34$.

4. Discussion

The main aim of this study was to assess, using a Posner cueing paradigm, whether establishing an ensemble representation for multiple facial expressions is attention dependent or not. More specifically, we sought to investigate whether performance was still above chance level when the four individual faces to be averaged into a mean representation were presented outside the focus of attention (i.e., at an invalid location), suggesting in turn that this process could operate without the need to engage covert attention to some extent. Further, we also compared the processing of multiple faces (and computing the mean emotion thereof) to the processing of a single emotional face, while keeping low-level features and task demands as similar as possible between these two experimental conditions. At the ERP level, we recorded the N2pc and SPCN to explore possible differences in attention and short-term memory, respectively, as a function of task requirements (i.e., extracting the average emotion of four faces or discriminating the emotion from a single one). A number of important new results emerge

⁶ It might be argued that the mean amplitudes of the SPCN (as well as the preceding N2pc) could artificially be reduced due to the use of a .05 Hz high-pass filter during EEG data pre-processing. We re-analyzed the EEG data without any high-pass filter, but the SPCN (and N2pc) results remained unchanged.



Fig. 5. Topographic maps (back view) of the N2pc (200–300 ms) and SPCN (400–600 ms) components in the valid condition for the no-clear-saccade group, shown separately for the average emotion task (upper row) and the single emotion task (lower row).

from this study, as explained here after.

The first main result obtained in this study pertains to the observation, as evidenced based on a careful analysis of the HEOG channel, of an unexpectedly high number of cue-related saccades (towards the upcoming target location) in both tasks, despite the use of task demands requiring central fixation, as well as catch trials at fixation used to promote it. As the results obtained for the HEOG channel clearly showed, out of the 36 subjects included in the experiment, only 19 could eventually be retained in the subsequent analyses with saccade-free epochs/trials, enabling us to explore the averaging of multiple faces (vs. a single face) when covert attention was putatively used to carry out the two tasks. This first result, albeit unexpected, suggests indirectly that it was probably hard for participants in general to prevent gazing at the target location to carry out the emotion discrimination task, indicating thereby that the use of peripheral vision only was very challenging in the present case.

Interestingly, based on these "clean" data, we found that the performance of participants was reliably above chance level for the invalid condition in both tasks, suggesting that they could well discriminate the emotion (based either on the mean of the four faces shown, or a single face alone) with (very) limited attention engaged at the target location. Covert shifts of spatial attention in both tasks were evidenced at the behavioral level, with faster RTs for valid than invalid trials, yet regardless of task demands. They were also corroborated by the ERP results, showing that the N2pc and SPCN were clearly elicited in the valid condition, but not in the invalid one.

4.1. Averaging emotion from multiple faces with limited attention

To the best of our knowledge, our study is the first to show an above

chance level performance to discriminate the mean emotion of four faces shown at an invalid location in the visual field. The precision was not very high though (d prime scores remained below 1), indicating a rather coarse perception. Importantly, the performance in the invalid condition was statistically undistinguishable from the performance in the valid condition, suggesting a dissociation between covert shifts of spatial attention on the one hand (as evidenced by the N2pc) and the ability to compute the ensemble representation for multiple faces on the other. Hence, as these results suggest, the average emotion could be extracted with limited attention. We explicitly use "limited attention" here, as opposed to "no attention" for example, because with the spatial cueing paradigm used here and in the attention literature in experimental psychology (Posner, 1980, 2014), we cannot formally rule out the possibility that a certain amount of attention (e.g., diffuse/distributed attention, Chong and Treisman, 2005) actually spread to the invalid location, or some residual resources were used to process this specific location. Noteworthy, this conclusion does not necessarily contradict previous results (including from our group; see Ji et al., submitted) suggesting that extracting mean information actually requires attention (Huang, 2015; Jackson-Nielsen et al., 2017; McNair et al., 2017). The results obtained for the saccade-group (see Supplementary materials) also showed that directing overt attention, indicated by the presence of a clear saccade following the cue (as well as no N2pc or SPCN elicited in response to the valid target faces set), increased the averaging performance at the valid location. On the other hand, although attention can boost this complex perceptual process and increase accuracy or precision (at the behavioral level), when only limited attentional resources were available to process the visual input (e.g., because the stimulus presentation was short and the display contained many different faces; see simultaneous condition in Ji et al.,

submitted), behavioral performance was found to be still reliably influenced by the actual proportion of happy/angry faces included in the face set, suggesting that full or focused attention was not a pre-requisite for it.

The ability to extract the average emotion from multiple faces (and compute in turn a sort of affective gist) with limited attention, as our results here suggest, is deemed remarkable. One reason accounting for this phenomenon might be that similar to natural scene perception (Geisler, 2008; Peelen and Kastner, 2014), the face sets used as stimuli in our experiment always included some internal structure and carried statistical regularities (Alvarez, 2011). With different ratios of happy and angry faces, there was necessarily some redundant information in the set. As suggested by the co-activation model (Miller, 1982), neural signals from multiple redundant stimuli are summed up, which might enhance the robustness of the representation for them. Additionally, collapsing or averaging across noisy individual representations also contributes to obtain an estimate with relatively higher precision (Alvarez, 2011; Cohen et al., 2016).

At first sight, the lack of a reliable N2pc (or SPCN) in the invalid condition in our study contradicts earlier ERP results reported by Brisson and Jolicœur (2008), who found a N2pc and (delayed) SPCN in the invalid as well as the valid condition. A number of methodological differences between this earlier and the current study might explain this apparent discrepancy. These authors explored primarily effects of exogenous cueing (as opposed to endogenous here) and did not use masks after the target display, as we did here. Additionally, Brisson and Jolicœur (2008) had many more trials than we had in the invalid condition (although clear visual ERP components were generated for this condition in our study, see Fig. 4). Last but not least, the accuracy in their task (Brisson and Jolicœur, 2008) was on average very high (and close to ceiling), including in the invalid condition (94% of correct responses), while the averaging task used here appeared much more challenging for participants (60% of correct responses, which was above chance level), suggesting probably the involvement of different attention and perceptual processes between these two studies. More generally, the lack of a N2pc and SPCN in the invalid condition in our study might suggest that swift, unidirectional and covert shifts of attention towards the target location (re-orienting) were not carried out, possibly due to the short presentation of the face sets, and the use of a mask shown at their offset. Presumably, diffuse attention was perhaps used to perform the task in this condition. Moreover, unlike the valid condition where the two inner faces (relative to fixation) contributed more than the two outer faces to the averaging performance, no such differential effect of face position was found in the invalid condition (see Supplementary materials). This auxiliary result also suggests indirectly that different attention mechanisms were probably involved in the valid and invalid conditions. Interestingly, a similar dissociation between EEG brain activity and behavioral performance was reported recently by Trübutschek et al. (2017). These authors failed to evidence a sustained brain activity at the scalp level although accuracy for target detection was well above chance level in this condition (Trübutschek et al., 2017).

4.2. Dissociation between mean emotion and single emotion processing

The single face condition we used in this experiment actually shared similarities with a visual search task to some extent. In this condition, participants had to find the face among four items/objects and rapidly discriminate his/her emotional expression. In the valid condition, performance was similar for the single and average emotion tasks. However, for the invalid condition, discriminating emotion from a single face became worse than performance in the average emotion task. Hence, the lack of covert attention was clearly more detrimental to performance in the former compared to the latter task. This gain for the average emotion task could also be explained by the fact that multiple items have to be collapsed somehow, and noise reduction could take place, leading in turn to an advantage over the single face presentation ("visual search"), especially when the attentional resources are limited (Alvarez, 2011; Fischer and Whitney, 2011; Haberman and Whitney, 2009; Li et al., 2016).

At the ERP level, multiple faces (in the average emotion task) generally elicited increased neural activity compared with a single face (in the single emotion task), regardless of cue validity. Similar to Puce et al. (2013), we found that the N170 was (trend significantly) larger in the average task than in the single task (see Supplementary results). It was also evident in the later time window of N2pc (200-300 ms) and SPCN (400–600 ms). This amplification might be explained by the fact that compared with the single emotion task, there were obviously more (emotional) faces in the average emotion task, hence summation and/or a stronger emotion was elicited in this condition. Alternatively, when the average emotion task was required, it may be the case that the deviant emotional expression contained in the sets (i.e. the angry face in the Ratio .75 condition, or the happy face in the Ratio .25 condition) contributed to boosting early sensory processing and hence the amplitude of these early ERP components (Luck and Hillyard, 1994; Ritter et al., 1982).

Noteworthy, in the valid condition, we found that the mean amplitude of the N2pc in the average task was numerically larger than that in the single task. The N2pc was previously related to "individuation", and has been found to be modulated in amplitude by set size manipulations in the multiple objects tracking and enumeration tasks (Drew and Vogel, 2008; Ester et al., 2012; Mazza and Caramazza, 2011; Pagano and Mazza, 2012). These two tasks require early individuation, providing a coarse representation of the objects in the visual field, and then allowing the visual system to individuate each object as being separate from other ones. Using this framework, it could be argued therefore that the larger N2pc found in our study for the average emotion task compared to the singe one could reflect the fact that averaging emotion required individuating each facial expression, instead of using a "total activation map" (Šetić et al., 2007). This interpretation is consistent with a recent study showing that averaging face identities is not independent of processing individual identities (Neumann et al., in press).

Interestingly, the results obtained for the SPCN component in the valid condition supplemented the N2pc and suggest that short-term memory effects (as captured by this later ERP component, see Jolicœur et al., 2008) were balanced between the two tasks. This could tentatively be explained by the fact that the different emotional facial expressions extracted from the four faces shown concurrently were perhaps highly compressed into one summary statistic (i.e., mean), leaving one object or the mean information to be stored in short term memory and as such, being comparable to the single emotion condition. Alternatively, the four individual representations, even if computed, were perhaps rapidly lost and thus not encoded as such, or severely impoverished in visual short-term memory (Alvarez, 2001; Brady and Alvarez, 2011; McNair et al., 2016). Notwithstanding this possibility, Baijal et al. (2013) previously used a working memory (WM) paradigm and found that the amplitude of the CDA component (elicited 300-700 ms after target onset, and sharing similarities with the SPCN) was actually larger when a mean size of two circles versus two individual sizes had to be maintained in WM, suggesting that mean representation maintained in WM may not be compressed, but rather, be primarily feature-based and "under-structured". This apparent discrepancy between these earlier results and our new ERP findings could stem from a number of methodological factors, including the focus on low-level (such as size) versus high-level (such as facial expressions) visual properties, bearing in mind that establishing mean representations for high-level objects (such as faces) could very well be qualitatively different than for low-level features (Haberman et al., 2015; Haberman and Whitney, 2012). Hence, future studies using the same paradigm (and ERP methodology) but comparing the averaging of lowlevel vs. high-level features are needed to try to reconcile some of these

inconsistent results.

4.3. Limitations

Some methodological limitations warrant comment. As the analysis of eye movements (based on the HEOG channel) and behavioral results (see d prime scores) clearly showed, the emotion discrimination task used in this experiment (bilaterally stimulus presentation for target and distractor in the periphery) turned out to be quite difficult and challenging on average for participants, for the two tasks alike. As a matter of fact, at target onset, participants had first to discriminate each time color information in the periphery (in order to separate the target's location from the distractor's one), before either processing the single or averaging the four emotional faces at this specific (and presumably attended) location. We had to use this "second" cue (at target onset) as we used bilateral stimulus presentations, and distractors also included faces. Without this second cue, it would have been extremely difficult to separate the contribution of target from distractor to the behavioral or ERP results. The use of a double cueing technique (whereby the cue carried color and spatial location information concurrently) however may have actually hindered the use of rapid and covert shifts of spatial attention towards target location by participants across successive trials. Accordingly, we have to acknowledge that some of the ERP results reported in this study (e.g. N2pc) may also have been contaminated in part by this double cueing effect. To overcome this limitation, unilateral stimulus sets (combined with a shorter duration) could be used in future studies, as they would not require using such a double cueing (i.e., discrimination of target from distractor would not be required first; the target would always be shown alone in the peripheral visual field, either at an attended or unattended spatial location, as inferred from the preceding symbolic cue).

4.4. Conclusions

The current ERP study provides novel insights into the actual processes underlying the extraordinary human perceptual ability to rapidly extract the average emotion from a complex scene composed of multiple facial expressions shown concurrently. Strikingly, when attention was kept low and minimal (i.e., in the invalid condition for which no reliable N2pc was elicited and RTs were slower than in the valid condition), participants could still discriminate the mean emotion from the face set above chance level, suggesting that this process could well operate (albeit without a high precision or accuracy) under these impoverished conditions. Further, ERP results for the SPCN component show indirectly that the four individual faces were likely compressed and stored as one "single" object into visual short-term memory, suggesting in turn that averaging multiple faces likely operates by means of "contraction". This compression likely followed an earlier process where items individuation probably took place, at the N2pc level.

Acknowledgements

This work is supported by a China Scholarship Council (CSC) grant ([2014]3026) and a cofounding grant (BOFCHN2016000901) from Ghent University, both awarded to LJ.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2018. 01.022.

References

Alvarez, G.A., 2011. Representing multiple objects as an ensemble enhances visual cognition. Trends Cogn. Sci. 15 (3), 122–131. Alvarez, G.A., Oliva, A., 2008. The representation of simple ensemble visual features outside the focus of attention. Psychol. Sci. 19 (4), 392–398.

- Alvarez, G.A., Oliva, A., 2009. Spatial ensemble statistics are efficient codes that can be represented with reduced attention. Proc. Natl. Acad. Sci. 106 (18), 7345–7350.
- Attarha, M., Moore, C.M., Vecera, S.P., 2014. Summary statistics of size: fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. J. Exp. Psychol.: Human. Percept. Perform. 40 (4), 1440–1449.
- Baijal, S., Nakatani, C., van Leeuwen, C., Srinivasan, N., 2013. Processing statistics: an examination of focused and distributed attention using event related potentials. Vision. Res. 85, 20–25.
- Bex, P.J., Makous, W., 2002. Spatial frequency, phase, and the contrast of natural images. J. Opt. Soc. Am. A 19 (6), 1096–1106.

Brady, T.F., Alvarez, G.A., 2011. Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. Psychol. Sci. 22 (3), 384–392.

- Brisson, B., Jolicœur, P., 2007. A psychological refractory period in access to visual short-term memory and the deployment of visual–spatial attention: multitasking processing deficits revealed by event-related potentials. Psychophysiology 44 (2), 323–333.
- Brisson, B., Jolicœur, P., 2008. Express attentional re-engagement but delayed entry into consciousness following invalid spatial cues in visual search. PLoS One 3 (12), e3967.
- Chong, S.C., Treisman, A., 2003. Representation of statistical properties. Vision Res. 43, 393–404.
- Chong, S.C., Treisman, A., 2005. Attentional spread in the statistical processing of visual displays. Atten., Percept., Psychophys. 67 (1), 1–13.
- Cohen, M.A., Dennett, D.C., Kanwisher, N., 2016. What is the bandwidth of perceptual experience? Trends Cogn. Sci. 20 (5), 324–335.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.
- Demeyere, N., Rzeskiewicz, A., Humphreys, K.A., Humphreys, G.W., 2008. Automatic statistical processing of visual properties in simultanagnosia. Neuropsychologia 46 (11), 2861–2864.
- Dougherty, R.F., Koch, V.M., Brewer, A.A., Fischer, B., Modersitzki, J., Wandell, B.A., 2003. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J. Vision. 3 (10) (1-1).
- Drew, T., Vogel, E.K., 2008. Neural measures of individual differences in selecting and tracking multiple moving objects. J. Neurosci. 28 (16), 4183–4191.
- Eimer, M., 1996. The N2pc component as an indicator of attentional selectivity. Electroencephalogr. Clin. Neurophysiol. 99 (3), 225–234.
- Eimer, M., Kiss, M., 2010. An electrophysiological measure of access to representations in visual working memory. Psychophysiology 47 (1), 197–200.
- Ester, E.F., Drew, T., Klee, D., Vogel, E.K., Awh, E., 2012. Neural measures reveal a fixed item limit in subitizing. J. Neurosci. 32 (21), 7169–7177.
- Florey, J., Clifford, C.W., Dakin, S., Mareschal, I., 2016. Spatial limitations in averaging social cues. Sci. Rep. 6.
- Fischer, J., Whitney, D., 2011. Object-level visual information gets through the bottleneck of crowding. J. Neurophysiol. 106 (3), 1389–1398.
- Geisler, W.S., 2008. Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol. 59, 167–192.
- Haberman, J., Brady, T.F., Alvarez, G.A., 2015. Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. J. Exp. Psychol.: Gen. 144 (2), 432.
- Haberman, J., Whitney, D., 2007. Rapid extraction of mean emotion and gender from sets of faces. Curr. Biol. 17 (17), R751–R753.
- Haberman, J., Whitney, D., 2009. Seeing the mean: ensemble coding for sets of faces. J. Exp. Psychol.: Hum. Percept. Perform. 35 (3), 718.
- Haberman, J., Whitney, D., 2012. Ensemble perception: summarizing the scene and broadening the limits of visual processing. In: From Perception to Consciousness: Searching with Anne Treisman, pp. 339–349.
- Hillyard, S.A., Vogel, E.K., Luck, S.J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos. Trans. R. Soc. B: Biol. Sci. 353 (1373), 1257.
- Horton, J.C., Hoyt, W.F., 1991. The representation of the visual field in human striate cortex: a revision of the classic Holmes map. Arch. Ophthalmol. 109 (6), 816–824.
- Huang, L., 2015. Statistical properties demand as much attention as object features. PLoS One 10 (8), 1–9.
- Jackson-Nielsen, M., Cohen, M.A., Pitts, M.A., 2017. Perception of ensemble statistics requires attention. Conscious. Cogn. 48, 149–160.
- Jacoby, O., Kamke, M.R., Mattingley, J.B., 2013. Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. J. Exp. Psychol.: Hum. Percept. Perform. 39 (1), 233.
- Ji, L., Chen, W., Loeys, T., & Pourtois, G. (submitted). Ensemble representation for multiple facial expressions: Evidence for a capacity limited perceptual process.
- Jolicœur, P., Brisson, B., Robitaille, N., 2008. Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. Brain Res. 1215, 160–172.
- Klaver, P., Talsma, D., Wijers, A.A., Heinze, H.J., Mulder, G., 1999. An event-related brain potential correlate of visual short-term memory. NeuroReport 10 (10), 2001–2005.
- Kiss, M., Van Velzen, J., Eimer, M., 2008. The N2pc component and its links to attention shifts and spatially selective visual processing. Psychophysiology 45 (2), 240–249.
- Leib, A.Y., Landau, A.N., Baek, Y., Chong, S.C., Robertson, L., 2012. Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. Front. Hum. Neurosci. 6, 267.
- Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C.H., Fu, X., 2016. Processing of individual items during ensemble coding of facial expressions. Front. Psychol. 7.
- Lins, O.G., Picton, T.W., Berg, P., Scherg, M., 1993. Ocular artifacts in EEG and event-

related potentials I: scalp topography. Brain Topogr. 6 (1), 51-63.

Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. Front. Hum. Neurosci. 8, 213.

Luck, S.J., Hillyard, S.A., 1994. Spatial filtering during visual search: evidence from human electrophysiology. J. Exp. Psychol.: Hum. Percept. Perform. 20 (5), 1000.

- Mangun, G.R., Hillyard, S.A., 1991. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. J. Exp. Psychol.: Hum. Percept. Perform. 17 (4), 1057.
- Mazza, V., Caramazza, A., 2011. Temporal brain dynamics of multiple object processing: the flexibility of individuation. PLoS One 6 (2).
- Mazza, V., Pagano, S., Caramazza, A., 2013. Multiple object individuation and extract enumeration. J. Cogn. Neurosci. 25 (5), 697–705.
- Mazza, V., Turatto, M., Umiltà, C., Eimer, M., 2007. Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. Exp. Brain Res. 181 (3), 531–536.
- McCollough, A.W., Machizawa, M.G., Vogel, E.K., 2007. Electrophysiological measures of maintaining representations in visual working memory. Cortex 43 (1), 77–94.
- McNaire, N.A., Goodbourn, P.T., Shone, L.T., Harris, I.M., 2017. Summary statistics in the attentional blink. Atten., Percept., Psychophys. 79 (1), 100–116.
- Miller, J., 1982. Divided attention: evidence for coactivation with redundant signals. Cogn. Psychol. 14 (2), 247–279.
- Neumann, M., Ng, R., Rhodes, R., Palermo, R., 2018. Ensemble coding of face identity is not independent of the coding of individual identity. Q. J. Exp. Psychol. 1–27 (in press).
- Pagano, S., Mazza, V., 2012. Individuation of multiple targets during visual enumeration: new insights from electrophysiology. Neuropsychologia 50 (5), 754–761.
- Peelen, M.V., Kastner, S., 2014. Attention in the real world: toward understanding its neural basis. Trends Cogn. Sci. 18 (5), 242–250.
- Praamstra, P., 2006. Prior information of stimulus location: effects on ERP measures of visual selection and response selection. Brain Res. 1072 (1), 153–160.
- Posner, M.I., 1980. Orienting of attention. Q. J. Exp. Psychol. 32 (1), 3-25.
- Posner, M.I., 2014. Orienting of attention: then and now. Q. J. Exp. Psychol. 69 (10), 1864–1875.
- Puce, A., McNeely, M.E., Berrebi, M.E., Thompson, J.C., Hardee, J., Brefczynski-Lewis, J.,

2013. Multiple faces elicit augmented neural activity. Front. Hum. Neurosci. 7.

- Ritter, W., Simson, R., Vaughan, H.G., Macht, M., 1982. Manipulation of event-related potential manifestations of information processing stages. Science 218 (4575), 909–911.
- Rousselet, G.A., Husk, J.S., Bennett, P.J., Sekuler, A.B., 2005. Spatial scaling factors explain eccentricity effects on face ERPs. J. Vis. 5 (10) (1-1).
- Šetić, M., Švegar, D., Domijan, D., 2007. Modelling the statistical processing of visual information. Neurocomputing 70 (10), 1808–1812.
- Scharff, A., Palmer, J., Moore, C.M., 2011. Extending the simultaneous-sequential paradigm to measure perceptual capacity for features and words. J. Exp. Psychol.: Hum. Percept. Perform. 37 (3), 813.
- Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. Behav. Res. Methods, Instrum., Comput. 31 (1), 137–149.
- Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Nelson, C., 2009. The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res. 168 (3), 242–249.
- Treisman, A., 2006. How the deployment of attention determines what we see. Vis. Cogn. 14 (4–8), 411–443.
- Trübutschek, D., Marti, S., Ojeda, A., King, J.-R., Mi, Y., Tsodyks, M., Dehaene, S., 2017. A theory of working memory without consciousness or sustained activity. eLife 6, 1–29.
- Vogel, E.K., Machizawa, M.G., 2004. Neural activity predicts individual differences in visual working memory capacity. Nature 428 (6984), 748–751.
- Whitney, D., Haberman, J., Sweeny, T.D., 2014. From textures to crowds: multiple levels of summary statistical perception. In: Werner, J.S., Chalupa, L.M. (Eds.), The New Visual Neurosciences. MIT Press, pp. 695–710.

Whitney, D., Leib, A.Y., 2018. Ensemble perception. Annu. Rev. Psychol. 69 (12.1-12.25).

- Woodman, G.F., Arita, J.T., Luck, S.J., 2009. A cuing study of the N2pc component: an index of attentional deployment to objects rather than spatial locations. Brain Res. 10 (1297), 101–111.
- Woodman, G.F., Luck, S.J., 1999. Electrophysiological measurement of rapid shifts of attention during visual search. Nature 400 (6747), 867–869.
- Yang, J.W., Yoon, K.L., Chong, S.C., Oh, K.J., 2013. Accurate but pathological: social anxiety and ensemble coding of emotion. Cogn. Ther. Res. 37 (3), 572–578.