

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

Correspondence to: Luyan Ji

Department of Experimental-Clinical and Health Psychology

Ghent University

Ghent 9000, Belgium

Email: Luyan.Ji@Ugent.be

1 Abstract

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

14

15 *Keywords:* ensemble representation; facial expressions; emotion; attention, N2pc, SPCN

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

3
4 Facial expressions carry important social and emotional information which can be used to
5 guide and optimize communication between people. In many natural settings and environments,
6 humans usually interact with multiple other individuals simultaneously. For example, in an
7 auditorium, some of the students' faces may seem pleased while some others may display some
8 signs of disapproval or concerns, leading in turn the instructor to experience a rather mixed
9 feeling regarding the disposition of his/her audience. Growing evidence in experimental
10 psychology suggests that human observers can rapidly and rather precisely extract mean emotion
11 from mixed valences in multiple faces presented concurrently (e.g., Haberman & Whitney, 2007,
12 2009; Li et al., 2016). The perceived average emotion provides a rather accurate summary
13 statistic of the complex scene or display composed of multiple emotional faces, usually referred
14 to as ensemble representation (Alvarez, 2011; Whitney & Leib, 2018). As such, this averaging
15 allows collapsing or combining multiple individual facial expressions into a coherent and
16 integrated emotion percept that carries the information of mean intensity and/or valence of the
17 scene. This remarkable ability of establishing ensemble representation is thought to provide an
18 efficient way to cope with the bottlenecks in visual processing (Alvarez, 2011; Chong &
19 Treisman, 2003; Whitney, Haberman, & Sweeny, 2014) and to reconcile the subjective
20 impression of a rich visual world with the limited perceptual and attention capacities (Cohen, &
21 Dennett, & Kanwisher, 2016). However, the mechanism underlying ensemble representation is
22 still largely unclear, and discrepant results have been reported in the past.

23 Some earlier research argued that extracting mean information from multiple items is best
24 conceived as a capacity-limited perceptual process (Attarha, Moore, & Vecera, 2014; Florey,

1 Clifford, Dakin, & Mareschal, 2016; Jacoby, Kamke, & Mattingley, 2013; Ji, Chen, Loeys, &
2 Pourtois, submitted). In a recent study (Ji et al., submitted), we used the extended simultaneous-
3 sequential paradigm (Scharff, Palmer, & Moore, 2011) to examine the processing capacity for
4 extracting mean emotion from multiple facial expressions (with variations along the valence
5 dimension; the faces being either happy or angry). The results showed that performance in the
6 sequential condition (where two successive displays each containing 8 faces were shown) was
7 better than that in the simultaneous condition (where the 16 individual faces were presented at the
8 same time), which was consistent with the limited-capacity model assuming that all items in an
9 ensemble could not be processed independently or without interference. In addition, a previous
10 study using the attentional blink (AB) paradigm found a clear AB effect when estimating the
11 average emotion from four faces (target 2) that followed, after a short lag, a first face whose
12 gender had to be discriminated (target 1), indicating that average emotion processing suffers from
13 the temporal limits of attention deployment (McNaire, Goodbourn, Shone, & Harris, 2016).

14 Although extracting the mean emotion from a set of individual facial expressions seems to
15 obey to capacity limitations, it is not known yet whether an ensemble representation for multiple
16 facial expressions could be established with no attention or limited attention. Earlier
17 psychophysical studies on low-level features showed that mean representation can be formed
18 outside the focus of attention or with reduced spatial attention allocated to its actual content
19 (Alveraz & Oliva, 2008, 2009; Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008; Leib,
20 Landau, Baek, Chong, & Robertson, 2012; but cf. Huang, 2015; Jackson-Nielsen, Cohen, & Pitts,
21 2017). The goal of the current study was twofold. (i) First, we examined whether the average
22 emotion could be extracted with limited attention. For this purpose, we combined a spatial-
23 cueing procedure (Posner, 1980; Mangun & Hillyard, 1991) with an average emotion task where
24 we manipulated the average emotion of a face set by systematically varying across trials the ratio

1 of positive and negative faces contained in the set. In short, in every trial, observers were required
2 to judge the average emotion (either positive or negative) of a target set composed of four faces
3 conveying a variable amount of happy and angry expressions and the target set was presented in a
4 valid (75%) or invalid (25%) peripheral location, while distractor faces were presented in the
5 opposite location. At the behavioral level, we compared performance between the valid and
6 invalid condition. We reasoned that if performance in the invalid condition was above chance
7 level (and/or similar to that in the valid condition), then this could be interpreted as evidence in
8 favor of the possibility to extract the mean emotion from the set with limited attention.
9 Conversely, if participants could not discriminate above chance level the mean emotion in the
10 invalid condition, then this could be taken as evidence that the ensemble representation could not
11 be established with limited attention. The presence of a significant validity effect (i.e., better
12 performance in the valid compared to the invalid condition) would accord with a classical
13 attention gating effect whereby a more accurate mean representation could be achieved when
14 selective attention is allocated to this complex stimulus (Hillyard, Vogel, & Luck, 1998). (ii)
15 Additionally, we also investigated whether averaging multiple facial expressions could be
16 dissociated from recognizing a single emotional facial expression. To this aim, we compared
17 behavioral performance of the same subjects between two tasks. Either participants had to
18 discriminate the emotional expression of a single face (single emotion task) or they performed the
19 average emotion task, as described here above.

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

1 negativity).

2 The N2pc is a lateralized component, characterized by a larger negativity at posterior
3 (occipito-temporal) electrodes sites contralateral versus ipsilateral to the attended location, and is
4 usually observed approximately 200-300 ms after stimulus onset at lateral occipital leads (Luck
5 & Hillyard, 1994). This ERP component is usually found in visual search tasks, and thought to
6 reflect spatially selective attention to target stimuli (Eimer, 1996; Woodman & Luck, 1999). The
7 N2pc was also observed previously during a spatial cueing task (Kiss, van Velzen, Eimer, 2008;
8 Woodman, Arita, & Luck, 2009). For example, following an informative (100% valid) arrow cue
9 in the center, an N2pc in response to targets was still reliably elicited (Kiss, et al., 2008;
10 Praamstra, 2006). More importantly, the N2pc also provides a neural index of individuation, or in
11 other words, forming distinct representations of each individual item at the same time. In line
12 with this assumption, the amplitude of N2pc usually increases with the number of to-be-selected
13 items within the attended hemifield (Drew & Vogel, 2008; Ester, Drew, Klee, Vogel, & Awh,
14 2012; Mazza & Caramazza, 2011; Pagano & Mazza, 2012). For example, the more items need to
15 be enumerated or to be tracked, the larger the amplitude of the N2pc, reaching a plateau usually
16 at a set size over four items, indicating a limitation in simultaneously selecting or individuating
17 these items (Drew & Vogel, 2008; Mazza, Pagano, Caramazza, 2013). However, the relationship
18 between the amplitude of the N2pc and set size is best evidenced when the individuation of
19 multiple stimuli is required. When observers had only to detect the presence of a specific color
20 within the set for example, the N2pc amplitude was not influenced any more by the number of
21 items therein (Mazza & Caramazza, 2011). The SPCN often follows the N2pc when more
22 detailed processing and/or memory for the lateralized target is required (Brisson & Jolicoeur,
23 2007; Jolicoeur, Brisson, Robitaille, 2008; Mazza, Turatto, Umiltà, & Eimer, 2007). The SPCN
24 (also called contralateral delay activity - CDA sometimes in the extant literature) has a strong link

1 with visual short-term memory (Eimer & Kiss, 2010; Klaver, Talsma, Wijers, Heinze, & Mulder,
2 1999; McCollough, Machizawa, & Vogel, 2007; Vogel & Machizawa, 2004). The amplitude of
3 the SPCN usually increases as the number of items held in visual short-term memory increases,
4 up to the observer's memory capacity (MaCollough et al., 2007; Vogel & Machizawa, 2004).

5 Using these specific electrophysiological markers of attention selection (N2pc) and visual
6 short-term memory (SPCN), we first hypothesized that the N2pc would be observed in the valid
7 condition, indicating allocation of attention to the target face stimuli. In the invalid condition, we
8 surmised that the N2pc would be strongly reduced, if not fully absent. With regard to our second
9 research question, we conjectured that the N2pc could be larger in the average emotion task (four
10 target faces) compared with the single emotion task (one target face), assuming that extracting the
11 mean emotion from multiple facial expressions requires individuating/computing individual
12 faces. On the other hand, if establishing a mean emotion representation does not require
13 individuating but can be computed based on a so-called "total activation map" (Šetić, Švegar, &
14 Domijan, 2006), or alternatively participants use a subsampling strategy (selecting one face out of
15 the four available in the average emotion task), then the amplitude of the N2pc should be similar
16 for the two tasks. Although the current study was not designed a priori to explore visual short
17 term memory, we nevertheless assessed whether a SPCN could be elicited following the N2pc in
18 this experiment. As a matter of fact, it may be the case that the emotional information extracted
19 from the target stimulus first needs to be shortly retained in a visual buffer ("short-term
20 memory") after visual presentation for further processing and elaboration (e.g., discrimination),
21 leading in turn to the generation of a SPCN. In this scenario, if the four facial expressions are
22 collapsed or compressed into one summary statistic (i.e., mean) (Alvarez, 2001; Brady &
23 Alvarez, 2011), then this memory load (or buffer information) would be comparable for the two
24 tasks, leading thereby to a similar SPCN for them. On the other hand, if the individual

1 representations are maintained along with the summary estimate, or the mean representation
2 retained in visual short-term memory consisted of low compressed or less structured features
3 (Baijal, Nakatani, van Leeuwen, & Srinivasan, 2013; Treisman, 2006), then a larger SPCN would
4 be observed primarily in the average emotion task compared with the single emotion task.

5

6

Method

7 **Participants**

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

13 **Stimuli and design**

14 Four male and four female face identities were selected from NimStim database
15 (Tottenham et al., 2009). Each face identity shows happy, angry and neutral expressions, all with
16 closed mouths. The hair, ears, neck and other external information were cropped by an oval
17 frame. All images were converted to greyscale, and scaled to the same mean luminance and root-
18 mean-square contrast (Bex & Makous, 2002).

19 In the average emotion task, the target set of four faces conveying a variable amount of
20 happy and angry expressions was presented in either the left or the right visual field. The ratio of
21 happy faces in the set was 0.25, 0.5 or 0.75. On the opposite side, there was another set
22 containing four neutral faces, hence yielding bilateral stimulus presentations (Figure 1).

23 In the single emotion task, there was either one happy or one angry face flanked by three
24 scrambled faces in the target set (Figure 1). The distractor set (shown in the opposite visual field,

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

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

10 The pair of one target and one distractor set was presented in the left and the right visual
11 field 3.52° lateral to the fixation. The four faces (or one face with three scrambled ones) in each
12 set were shown in a 2×2 invisible grid, on a homogenous black background (Figure 1). The
13 position of these four faces was randomly selected. In the single emotion task, the single intact
14 face in the two sets was presented in mirror symmetry. A white fixation point was continuously
15 present on the center of the screen. The two inner faces (i.e., being closer to the fixation, $3.28^\circ \times$
16 2.15°) subtended a visual angle of $2.38^\circ \times 2.38^\circ$. The two outer faces ($6.35^\circ \times 2.15^\circ$) were scaled
17 to $3.42^\circ \times 3.45^\circ$ to compensate for differences in V1 cortical representation/magnification
18 (Rousselet, Husk, Bennett, & Sekuler, 2005) following a standard formula¹ (Dougherty, Koch,
19 Brewer, Fischer, Modersitzki, & Wandell, 2003; Horton & Hoyt, 1991). The outline of the outer
20 grid (3 pixels) was also always visible on the screen, to help the participants to attend to the part

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 $A = 29.2\text{mm}$ 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° .

1 of the visual field where the faces were actually presented. At the onset of the two face sets (one
2 target and one distractor), the outline was converted from white to either blue or green, indicating
3 the target or the distractor respectively (this color/validity mapping was counterbalanced across
4 participants), and these two colors remained when the masks appeared. The masks had the same
5 size and appeared at the same location as the faces in the target and distractor set (Figure 1C). We
6 used MATLAB (Mathworks, Natick, MA) to divide neutral faces into a 10×12 matrix of
7 randomly arranged squares and trimmed them with an oval frame against the black background.
8 Before the presentation of the two face sets, there was a centrally located arrow cue ($1.61^\circ \times$
9 1.28°), which had the same color as the outline of the target set.

10 (insert Figure 1 about here, single-column)

11 **Procedure**

12 Participants seated at 75 cm in front of a 19" CRT screen (resolution of 1024×768 pixels,
13 refresh rate 100 Hz) in a dimly lit cabin. To minimize head movements, a chinrest was used
14 during both tasks. While fixating at the central cross, participants were required to discriminate
15 (with peripheral vision) whether the (average) emotion of the target face set was positive or
16 negative, by pressing one out of two pre-defined buttons with their right hand ("2" or "8" using a
17 standard number keyboard; this mapping being counterbalanced across participants) as accurate
18 and as fast as possible. A trial began with a fixation cross for 500 ms, followed by an arrow cue
19 superimposed on the fixation cross for 200 ms, both of which were presented on the center of the
20 screen. After an ISI randomly varying between 500-700 ms, the pair of two face sets was shown
21 for 150 ms. The target face set mostly ($p = .75$) appeared at the spatial position indicated by the
22 preceding arrow, and equally likely in the left or the right visual field. Then, the face sets were
23 masked and the masks were presented for 1000 ms or terminated by response (Figure 1). The
24 next trial automatically began (randomly varying between) 1000 ms-1200 ms after participant

1 responded. Participants were informed about target validity and probability, and they were
2 encouraged to rely on their first impression and not to think extensively.

3 The two tasks, either judging the average emotion from multiple faces (average emotion
4 task) or identifying the emotion from the single face (single emotion task), were blocked, and the
5 order of blocks was counterbalanced across participants. The ratio of happy faces (0.25, 0.5, 0.75)
6 in the average emotion task or the valence of one emotional face (happy, angry) in the single
7 emotion task was randomized within blocks. Every trial had a unique face set to minimize the
8 visual statistical regularity between trials. Participants performed 12 experimental blocks of 48
9 trials each (36 valid, 12 invalid) for the average task, and 6 blocks of 64 trials each (48 valid, 16
10 invalid) for the single task. Before starting the experiment, participants got acquainted with the
11 two emotion judgment tasks with 20 practice trials each. Practice trials were excluded from all
12 subsequent analyses. In order to encourage participants to focus on the center of the screen
13 (fixation point) and only use peripheral vision to process face sets, the procedure also
14 incorporated four catch trials each block, where a white dot (0.60°) unexpectedly replaced the
15 fixation cross. Participants were asked to press the spacebar with their left hand when they
16 detected the dot. They did not need to judge the emotion when a dot appeared.

17 After the average and single emotion tasks, participants rated the valence and emotion
18 intensity of each face previously presented. To this aim, one face appeared at a time in the center
19 and had the same size as that in the previous task. Participants judged on a Visual Analogue Scale
20 (VAS). The two anchors of the VAS for emotion valence and intensity were labeled *Extremely*
21 *positive* and *Extremely negative*. Additionally, we also asked participants to rate their arousal
22 level to each individual face. The two anchors of VAS for arousal were labeled *Extremely calm*
23 and *Extremely excited*. The labels on the left and right side of both VASes were counterbalanced
24 across participants. The main goal of these post-experiment ratings was to confirm that the happy,

1 angry and neutral faces used in the main experiment were perceived as such and hence showed
2 differences in terms of valence and arousal. Moreover, following the procedure adopted in our
3 previous study (Ji et al., submitted), we directly used these post-experiment ratings to compute
4 subject-specific mean emotions for the different face sets used in the main experiment and
5 assessed if (objective) changes in the ratio of happy faces in these sets were related to the
6 (subjective) estimates (see Supplementary materials section for details).

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

9 **Electrophysiological recording and preprocessing**

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

17 The preprocessing was performed using EEGLAB (Delorme & Makeig, 2004) and
18 ERPLAB (Lopez-Calderon & Luck, 2014). A high-pass filter of 0.05 Hz and a low-pass filter of
19 80 Hz was firstly applied. Data were then referenced offline to the averaged reference after the
20 noisy channels were interpolated by a spherical splines procedure. In order to exclude potential
21 saccades to the peripheral face sets during the cue-target interval, EEG and EOG were first
22 segmented into long epochs beginning 200 ms before cue onset and up to 2250 ms following it,
23 baseline-corrected using the 200 ms pre-cue interval. Epochs containing horizontal eye
24 movements (as identified using HEOG) within 900 ms after cue onset were automatically

1 detected by the step-like artifact function (searching for step-shaped segments of data on the
2 channel of HEOG, with window step set as 10 ms, moving windows full width as 400 ms, and
3 voltage threshold as 15 μ V), as implemented in ERPLAB.

4 A total of 41% of trials was marked to be excluded due to the step-like HEOG artifacts
5 occurring during the cue-target interval. The number of trials detected was similar for all the
6 experimental conditions (valid left, valid right, invalid left, invalid right). The large number of
7 trials with HEOG during the cue-target interval indicated that for a large amount of trials,
8 participants already moved their eyes following the cue, or had produced saccades towards the
9 target location before its actual onset. Supplementary Figure 3 shows representative examples of
10 HEOG traces obtained for two subjects (one with no saccades and one with clear saccades)
11 recorded during the cue-target interval. Out of 36 participants, we could actually retain only 19
12 who had more than half of the trials left without saccades (after analysis of the HEOG) and
13 subsequent artifact rejections in both tasks, as well as no clear residual eye movements (less than
14 5 μ V, deviated less than 0.3°, Lins, Picton, Berg, & Scherg, 1993). These 19 participants were
15 used in the following EEG analyses. For this no-clear-saccade group, on average 23% of trials
16 were rejected in the HEOG rejection procedure². The original saccade-free long epochs were
17 segmented into shorter epochs (-200 ms to 800 ms) time-locked to the onset of the target face
18 sets. Independent component analysis (EEGLAB's runica algorithm) was adopted to remove
19 components related to eye blinks. Epochs containing activity exceeding $\pm 80 \mu$ V in the scalp EEG
20 electrodes were automatically rejected (on average 7% trials, a number which was not different
21 between task and validity conditions). The artifact-free data were then baseline-corrected using

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, $ps > .34$.

1 the 200 ms pre-stimulus interval. We averaged only the correct trials (thus the Ratio0.5 condition
2 in the average task was not included) separately for each type of task (average, single), validity
3 (valid, invalid), and target location (left, right)³. Grand average ERPs were computed by
4 averaging mean ERPs of these 19 participants for each condition separately. For the other 17
5 participants, a large amount of residual eye movements during the cue-target interval (more than
6 5 μ V) after HEOG rejection was observed and/or not enough trials were left after artifact
7 rejection⁴. The results obtained for this clear-saccade group (where trials with clear saccades only
8 were eventually retained, based on the HEOG channel analysis) are presented in the
9 Supplementary materials section for comparison purposes.

10 **Data analyses**

11 For the average and the single emotion tasks, the accuracy of catch trials was first
12 calculated. The subsequent analyses were based on trials without catch trials. For the average
13 task, the proportion of “positive” responses as a function of the ratio of happy faces in the sets
14 was first computed to confirm that participants’ average emotion judgments were sensitive to the
15 variable amount of happy vs angry expressions contained in the sets. Next, the discrimination
16 ability (d prime) was calculated. Importantly, the statistical analysis of behavioral data was
17 aligned to the EEG data analysis: accuracy and reaction time (RT) for the saccade-free trials only

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 condition (75 ± 15), $F(1, 18) = 362.71$, $p < .001$, $\eta_p^2 = .95$, while the number of trials in the average task (159 ± 34) did not differ significantly compared with the single task (155 ± 32), $F(1, 18) < 1$, $p = .51$, $\eta_p^2 = .024$. 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.

1 were computed (hence trials contaminated by eye movements were first removed). To compute d'
2 prime scores, hits and false alarms were defined as follows. The positive face sets, containing
3 three happy faces in the average emotion task or one happy face in the single emotion task were
4 considered as target. The negative sets, which contained three angry faces (average emotion task)
5 or one angry face (single emotion task), were considered as noise. Hits corresponded to judging
6 the positive face sets as positive, and false alarms corresponded to judging the negative sets as
7 positive. The sets composed of 50% happy ($n=2$) and 50% angry ($n=2$) faces in the average
8 emotion task were not included in calculating d' prime scores or RTs for correct responses. For
9 each dependent variable separately (d' prime and RT for correct trials), a two-way repeated-
10 measures ANOVAs was carried out with the factors Task and Validity. As a control analysis, we
11 also examined whether emotional faces shown close to fixation (two inner positions) had a larger
12 impact on performance compared to faces shown further away from it (two outer faces; see
13 Supplementary materials section).

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

20 ERP mean amplitudes were computed at the lateral posterior electrodes PO7/PO8, during
21 the 200-300 ms (N2pc) or 400-600 ms (SPCN) post-stimulus onset interval. The N2pc and SPCN
22 were quantified by subtracting the mean amplitude recorded at the ipsilateral electrodes (relative
23 to the location of the target faces) from that at the contralateral electrodes. Data were then
24 submitted to repeated-measures ANOVAs with Task and Validity as factors. Greenhouse-Geisser

1 correction was applied when assumptions of sphericity were violated. A Bonferroni correction
2 was used when multiple comparisons were performed. We also examined the face-specific N170
3 component time-locked to target onset, as it was previously found to be modulated by the number
4 of faces shown in the set (Puce, McNeely, Berrebi, Thompson, Hardee, & Brefczynski-Lewis,
5 2013). Details of these complementing analyses and results for the N170 are provided in the
6 Supplementary materials section.

7 **Results**

8 **Behavioral results**

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

15 **Average and single emotion judgment.** For the average task, the proportion of
16 “positive” responses was sensitive to the ratio of happy faces contained in the sets, $F(1.29,$
17 $23.15) = 42.70$, $p < .001$, $\eta_p^2 = .70$, confirming that participants’ judgments were influenced by
18 this manipulation (Figure 2). There was also a significant main effect of Validity in this analysis,
19 $F(1, 18) = 5.89$, $p = .026$, $\eta_p^2 = .25$. As can be seen from Figure 2, the proportion of “positive”
20 responses was larger in the valid compared to the invalid condition. Tentatively, this effect could
21 be explained by a response bias, namely being inclined to judge the face sets as positive in the
22 valid condition of the average task (see also Yang, Yoon, Chong, & Oh, 2013 for a similar effect);

1 while no such response bias was present in the invalid condition⁵. However, the interaction
2 between Ratio and Validity was not significant, $F(1.97, 35.45) < 1$, $\eta_p^2 = .04$, suggesting that in
3 these two conditions (valid and invalid), participants' judgements were similarly influenced by
4 the ratio of happy faces contained in the sets.

5 (insert Figure 2 about here, single-column)

6 Table 1 shows the hit rate, the false alarm rate, and the corresponding d prime score for
7 each condition separately. The discrimination ability (d prime) in the average and the single
8 emotion task was generally not high (i.e. lower than one). However, in all four main conditions
9 (average-valid, single-valid, average-invalid, and single-invalid), it was significantly above
10 chance level (higher than zero, $ps < .001$, Figure 3). The ANOVA on the d prime scores revealed
11 no significant main effect of Task, $F(1, 18) = 1.82$, $p = .19$, $\eta_p^2 = .09$, however a marginally
12 significant main effect of Validity, $F(1, 18) = 3.28$, $p = .087$, $\eta_p^2 = .15$. The interaction between
13 Task and Validity was also marginally significant, $F(1, 18) = 3.29$, $p = .086$, $\eta_p^2 = .16$. Simple
14 effect analysis revealed that for the average task, there was no significant difference between the
15 valid and the invalid condition, $F(1, 18) < 1$, $\eta_p^2 = .01$; while for the single task, the performance
16 was significantly better in the valid compared with the invalid condition, $F(1, 18) = 6.23$, p
17 $= .023$, $\eta_p^2 = .26$. The performance in the valid condition of the average task did not differ
18 significantly from that of the single task, $F(1, 18) < 1$, $\eta_p^2 = .02$; whereas for the invalid
19 condition, the performance was slightly better in the average task relative to the single task, $F(1,$
20 $18) = 3.03$, $p = .099$, $\eta_p^2 = .14$. The ANOVA performed on the RTs revealed a significant main

⁵ The c score, an index of response bias (Stanislaw & 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, $t(18) = .05$, $p = .96$.

1 effect of Validity, $F(1, 18) = 16.93$, $p = .001$, $\eta_p^2 = .49$. The RTs were significantly shorter in the
2 valid ($M = 669.48$, $SD = 107.69$) compared to the invalid condition ($M = 716.78$, $SD = 136.13$).
3 The main effect of Task, or the interaction between Task and Validity was not significant, $ps > .19$
4 (Figure 3).

5 (insert Table 1 about here)

6 (insert Figure 3 about here, single-column)

7 **Face emotion ratings.** Paired sample t tests showed that the perceived emotion intensity
8 of angry faces ($M = 81.12$, $SD = 7.79$) was overall stronger than that of happy faces ($M = 73.51$,
9 $SD = 5.52$), $t(7) = 2.61$, $p = .035$. Angry faces ($M = 53.07$, $SD = 3.70$) were rated as equally
10 aroused as happy faces ($M = 54.03$, $SD = 7.34$), $t(7) = -.40$, $p = .70$. Neutral faces were perceived
11 as negative, since the comparison (based on a one-sample t test) showed that their ratings were
12 significantly lower than 50 (the smaller the value, the more negative the faces were perceived; M
13 $= 42.12$, $SD = 3.46$), $t(7) = -6.44$, $p < .001$. In addition, neutral faces were rated as less aroused
14 ($M = 31.92$, $SD = 4.26$) than both happy and angry faces, $t(7) = -9.86$, $t(7) = -9.54$, $ps < .001$.

15 **ERP results**

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

1 from that at the contralateral ones.

2 (insert Figure 4 about here, single-column)

3 (insert Figure 5 about here, single-column)

4 The presence of an N2pc was confirmed by one-sample t tests against zero for the valid
5 condition of both the average ($-1.22 \pm 1.64 \mu\text{V}$) and single emotion tasks ($-.63 \pm .88 \mu\text{V}$), $t(18) =$
6 $-3.25, p = .004, t(18) = -3.14, p = .006$ (see Figure 4). No reliable N2pc was elicited in the
7 invalid condition of the two tasks (average: $-.01 \pm 2.02 \mu\text{V}$, single: $-.64 \pm 2.01 \mu\text{V}$), $t(18) = -.02,$
8 $p = .98, t(18) = -1.38, p = .18$. The ANOVA performed on the N2pc revealed no significant main
9 effect of Task, $F(1, 18) = 1.58, p = .23, \eta_p^2 = .08$, or main effect of Validity, $F(1, 18) < 1, \eta_p^2$
10 $< .001$, but a significant interaction between Task and Validity, $F(1, 18) = 6.59, p = .019, \eta_p^2$
11 $= .27$. The validity effect on the N2pc was pronounced in the average emotion task, $F(1, 18) =$
12 $4.55, p = .047, \eta_p^2 = .20$, but very weak in the single emotion task, $F(1, 18) < 1, \eta_p^2 < .001$.
13 Examining the N2pc for the valid and the invalid condition separately, we found that the N2pc
14 was numerically larger in the average task than in the single task for the valid condition, $F(1, 18)$
15 $= 2.52, p = .13, \eta_p^2 = .12$; while the comparison of N2pc between the two tasks showed the
16 opposite pattern in the invalid condition, $F(1, 18) = 1.35, p = .36, \eta_p^2 = .07$, but this difference
17 did not reach significance either. On the other hand, during this N2pc time window (200-300 ms
18 after the onset of face sets), the mean amplitude was found to be bilaterally more negative in the
19 average emotion task ($2.32 \pm 3.57 \mu\text{V}$) compared with the single emotion task ($3.44 \pm 3.26 \mu\text{V}$)
20 in both the valid and the invalid conditions, regardless of the electrodes being considered, either
21 contralateral or ipsilateral to the location of target faces, $F(1, 18) = 5.83, p = .027, \eta_p^2 = .25$. We
22 also analyzed the early N2pc in a narrower time window (i.e., 200-250 ms post-stimulus onset) to
23 investigate whether spatial attention was first directed to the distractor locations (being opposite

1 to the target location) following the invalid cue leading to a reversed N2pc. The early N2pc was
2 numerically above zero in the invalid condition of the average task ($.31 \pm 1.84 \mu\text{V}$, the mean
3 amplitude in the ipsilateral electrode to the target was more negative than in the contralateral
4 one), however not significantly so, $t(18) = .73, p = .47$. The early N2pc in the invalid condition
5 of the single task ($-.76 \pm 2.63 \mu\text{V}$) did not significantly differ from zero either, $t(18) = -1.26, p$
6 $= .23$, confirming no reliable early N2pc elicited in the invalid condition of both tasks. The
7 absence of a N2pc in the invalid condition might be imputed to the fact that this component could
8 possibly be delayed in this condition compared to the valid condition. To test this hypothesis, we
9 ran a control analysis and extracted the mean amplitude of the N2pc during the 250-350 ms
10 interval post-target onset at the same electrode positions. However, no reliable N2pc (average:
11 $-.32 \pm 2.36 \mu\text{V}$; single: $-.65 \pm 1.61 \mu\text{V}$) was detectable in this later time frame either, $t(18) = -$
12 $0.59, p = .56, t(18) = -1.76, p = .10$.

13 The ANOVA performed on the SPCN⁶ revealed no significant main effect of Task, $F(1,$
14 $18) < 1, \eta_p^2 = .05$, or main effect of Validity, $F(1, 18) < 1, p = .40, \eta_p^2 = .04$, nor significant
15 interaction between the two factors, $F(1, 18) = 2.42, p = .14, \eta_p^2 = .12$. The SPCN was observed
16 in the valid condition of both the average ($-.95 \pm 2.20 \mu\text{V}$) and the single task ($-.88 \pm 1.36 \mu\text{V}$), t
17 $(18) = -1.89, p = .075, t(18) = -2.83, p = .011$, however they did not differ significantly between
18 the two tasks, $F(1, 18) < 1, p = .90, \eta_p^2 = .002$ (see Figure 4). The SPCN was also present in the
19 invalid condition of the single task ($-.89 \pm 1.88 \mu\text{V}$), $t(18) = -2.08, p = .052$, but not of the
20 average task ($-.05 \pm 2.79 \mu\text{V}$), $t(18) = -.08, p = .94$. In addition, the mean amplitude in the 400-
21 600 ms post-stimulus time window was generally more negative in the average emotion task

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 0.05 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.

1 (2.26 ± 2.92 μV) compared with the single emotion task (3.58 ± 3.26 μV), regardless of cue
2 validity or contralaterality of the recording electrodes to the target location, $F(1, 18) = 9.31$, p
3 = .007, $\eta_p^2 = .34$.

4 **Discussion**

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

18 The first main result obtained in this study pertains to the observation, as evidenced based
19 on a careful analysis of the HEOG channel, of an unexpectedly high number of cue-related
20 saccades (towards the upcoming target location) in both tasks, despite the use of task demands
21 requiring central fixation, as well as catch trials at fixation used to promote it. As the results
22 obtained for the HEOG channel clearly showed, out of the 36 subjects included in the
23 experiment, only 19 could eventually be retained in the subsequent analyses with saccade-free

1 epochs/trials, enabling us to explore the averaging of multiple faces (vs. a single face) when
2 covert attention was putatively used to carry out the two tasks. This first result, albeit unexpected,
3 suggests indirectly that it was probably hard for participants in general to prevent gazing at the
4 target location to carry out the emotion discrimination task, indicating thereby that the use of
5 peripheral vision only was very challenging in the present case.

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

14 **Averaging emotion from multiple faces with limited attention**

15 To the best of our knowledge, our study is the first to show an above chance level
16 performance to discriminate the mean emotion of four faces shown at an invalid location in the
17 visual field. The precision was not very high though (d' prime scores remained below 1),
18 indicating a rather coarse perception. Importantly, the performance in the invalid condition was
19 statistically undistinguishable from the performance in the valid condition, suggesting a
20 dissociation between covert shifts of spatial attention on the one hand (as evidenced by the N2pc)
21 and the ability to compute the ensemble representation for multiple faces on the other. Hence, as
22 these results suggest, the average emotion could be extracted with limited attention. We explicitly
23 use “limited attention” here, as opposed to “no attention” for example, because with the spatial
24 cueing paradigm used here and in the attention literature in experimental psychology (Posner,

1 1980, 2014), we cannot formally rule out the possibility that a certain amount of attention (e.g.,
2 diffuse/distributed attention, Chong & Treisman, 2005) actually spread to the invalid location, or
3 some residual resources were used to process this specific location. Noteworthy, this conclusion
4 does not necessarily contradict previous results (including from our group; see Ji et al.,
5 submitted) suggesting that extracting mean information actually requires attention (Huang, 2015;
6 Jackson-Neilsen et al., 2017; McNair et al., 2017). The results obtained for the saccade-group
7 (see Supplementary Materials) also showed that directing overt attention, indicated by the
8 presence of a clear saccade following the cue (as well as no N2pc or SPCN elicited in response to
9 the valid target faces set), increased the averaging performance at the valid location. On the other
10 hand, although attention can boost this complex perceptual process and increase accuracy or
11 precision (at the behavioral level), when only limited attentional resources were available to
12 process the visual input (e.g., because the stimulus presentation was short and the display
13 contained many different faces; see simultaneous condition in Ji et al., submitted), behavioral
14 performance was found to be still reliably influenced by the actual proportion of happy/angry
15 faces included in the face set, suggesting that full or focused attention was not a pre-requisite for
16 it.

17 The ability to extract the average emotion from multiple faces (and compute in turn a sort
18 of affective gist) with limited attention, as our results here suggest, is deemed remarkable. One
19 reason accounting for this phenomenon might be that similar to natural scene perception (Geisler,
20 2008; Peelen & Kastner, 2014), the face sets used as stimuli in our experiment always included
21 some internal structure and carried statistical regularities (Alvarez, 2011). With different ratios of
22 happy and angry faces, there was necessarily some redundant information in the set. As suggested
23 by the co-activation model (Miller, 1982), neural signals from multiple redundant stimuli are
24 summed up, which might enhance the robustness of the representation for them. Additionally,

1 collapsing or averaging across noisy individual representations also contributes to obtain an
2 estimate with relatively higher precision (Alvarez, 2011; Cohen et al., 2016).

3 At first sight, the lack of a reliable N2pc (or SPCN) in the invalid condition in our study
4 contradicts earlier ERP results reported by Brisson & Jolicoeur (2008), who found a N2pc and
5 (delayed) SPCN in the invalid as well as the valid condition. A number of methodological
6 differences between this earlier and the current study might explain this apparent discrepancy.
7 These authors explored primarily effects of exogenous cueing (as opposed to endogenous here)
8 and did not use masks after the target display, as we did here. Additionally, Brisson & Jolicoeur
9 (2008) had many more trials than we had in the invalid condition (although clear visual ERP
10 components were generated for this condition in our study, see Figure 4). Last but not least, the
11 accuracy in their task (Brisson & Jolicoeur, 2008) was on average very high (and close to ceiling),
12 including in the invalid condition (94% of correct responses), while the averaging task used here
13 appeared much more challenging for participants (60% of correct responses, which was above
14 chance level), suggesting probably the involvement of different attention and perceptual
15 processes between these two studies. More generally, the lack of a N2pc and SPCN in the invalid
16 condition in our study might suggest that swift, unidirectional and covert shifts of attention
17 towards the target location (re-orienting) were not carried out, possibly due to the short
18 presentation of the face sets, and the use of a mask shown at their offset. Presumably, diffuse
19 attention was perhaps used to perform the task in this condition. Moreover, unlike the valid
20 condition where the two inner faces (relative to fixation) contributed more than the two outer
21 faces to the averaging performance, no such differential effect of face position was found in the
22 invalid condition (see Supplementary materials). This auxiliary result also suggests indirectly that
23 different attention mechanisms were probably involved in the valid and invalid conditions.
24 Interestingly, a similar dissociation between EEG brain activity and behavioral performance was

1 reported recently by Trübtschek et al. (2017). These authors failed to evidence a sustained brain
2 activity at the scalp level although accuracy for target detection was well above chance level in
3 this condition (Trübtschek et al., 2017).

4 **Dissociation between mean emotion and single emotion processing**

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

16 At the ERP level, multiple faces (in the average emotion task) generally elicited increased
17 neural activity compared with a single face (in the single emotion task), regardless of cue validity.
18 Similar to Puce et al., (2013), we found that the N170 was (trend significantly) larger in the
19 average task than in the single task (see Supplementary results). It was also evident in the later
20 time window of N2pc (200-300 ms) and SPCN (400-600 ms). This amplification might be
21 explained by the fact that compared with the single emotion task, there were obviously more
22 (emotional) faces in the average emotion task, hence summation and/or a stronger emotion was
23 elicited in this condition. Alternatively, when the average emotion task was required, it may be
24 the case that the deviant emotional expression contained in the sets (i.e. the angry face in the

1 Ratio0.75 condition, or the happy face in the Ratio0.25 condition) contributed to boosting early
2 sensory processing and hence the amplitude of these early ERP components (Luck & Hillyard,
3 1994; Ritter, Simson, Vaughan, & Macht, 1982).

4 Noteworthy, in the valid condition, we found that the mean amplitude of the N2pc in the
5 average task was numerically larger than that in the single task. The N2pc was previously related
6 to “individuation”, and has been found to be modulated in amplitude by set size manipulations in
7 the multiple objects tracking and enumeration tasks (Drew & Vogel, 2008; Ester et al., 2012;
8 Mazza & Caramazza, 2011; Pagano & Mazza, 2012). These two tasks require early individuation,
9 providing a coarse representation of the objects in the visual field, and then allowing the visual
10 system to individuate each object as being separate from other ones. Using this framework, it
11 could be argued therefore that the larger N2pc found in our study for the average emotion task
12 compared to the single one could reflect the fact that averaging emotion required individuating
13 each facial expression, instead of using a “total activation map” (Šetić et al., 2006). This
14 interpretation is consistent with a recent study showing that averaging face identities is not
15 independent of processing individual identities (Neumann, Ng, Rhodes, & Palermo, 2017).

16 Interestingly, the results obtained for the SPCN component in the valid condition
17 supplemented the N2pc and suggest that short-term memory effects (as captured by this later
18 ERP component, see Jolicœur et al., 2008) were balanced between the two tasks. This could
19 tentatively be explained by the fact that the different emotional facial expressions extracted from
20 the four faces shown concurrently were perhaps highly compressed into one summary statistic
21 (i.e., mean), leaving one object or the mean information to be stored in short term memory and as
22 such, being comparable to the single emotion condition. Alternatively, the four individual
23 representations, even if computed, were perhaps rapidly lost and thus not encoded as such, or
24 severely impoverished in visual short-term memory (Alvarez, 2001; Brady & Alvarez, 2011;

1 McNair et al., 2016). Notwithstanding this possibility, Baijal et al. (2013) previously used a
2 working memory (WM) paradigm and found that the amplitude of the CDA component (elicited
3 300-700 ms after target onset, and sharing similarities with the SPCN) was actually larger when a
4 mean size of two circles versus two individual sizes had to be maintained in WM, suggesting that
5 mean representation maintained in WM may not be compressed, but rather, be primarily feature-
6 based and “under-structured”. This apparent discrepancy between these earlier results and our
7 new ERP findings could stem from a number of methodological factors, including the focus on
8 low-level (such as size) versus high-level (such as facial expressions) visual properties, bearing in
9 mind that establishing mean representations for high-level objects (such as faces) could very well
10 be qualitatively different than for low-level features (Haberman, Brady, & Alvarez, 2015;
11 Haberman & Whitney, 2012). Hence, future studies using the same paradigm (and ERP
12 methodology) but comparing the averaging of low-level vs. high-level features are needed to try
13 to reconcile some of these inconsistent results.

14 **Limitations**

15 Some methodological limitations warrant comment. As the analysis of eye movements
16 (based on the HEOG channel) and behavioral results (see d prime scores) clearly showed, the
17 emotion discrimination task used in this experiment (bilaterally stimulus presentation for target
18 and distractor in the periphery) turned out to be quite difficult and challenging on average for
19 participants, for the two tasks alike. As a matter of fact, at target onset, participants had first to
20 discriminate each time color information in the periphery (in order to separate the target’s
21 location from the distractor’s one), before either processing the single or averaging the four
22 emotional faces at this specific (and presumably attended) location. We had to use this “second”
23 cue (at target onset) as we used bilateral stimulus presentations, and distractors also included
24 faces. Without this second cue, it would have been extremely difficult to separate the contribution

1 of target from distractor to the behavioral or ERP results. The use of a double cueing technique
2 (whereby the cue carried color and spatial location information concurrently) however may have
3 actually hindered the use of rapid and covert shifts of spatial attention towards target location by
4 participants across successive trials. Accordingly, we have to acknowledge that some of the ERP
5 results reported in this study (e.g. N2pc) may also have been contaminated in part by this double
6 cueing effect. To overcome this limitation, unilateral stimulus sets (combined with a shorter
7 duration) could be used in future studies, as they would not require using such a double cueing
8 (i.e., discrimination of target from distractor would not be required first; the target would always
9 be shown alone in the peripheral visual field, either at an attended or unattended spatial location,
10 as inferred from the preceding symbolic cue).

11 **Conclusions**

12 The current ERP study provides novel insights into the actual processes underlying the
13 extraordinary human perceptual ability to rapidly extract the average emotion from a complex
14 scene composed of multiple facial expressions shown concurrently. Strikingly, when attention
15 was kept low and minimal (i.e., in the invalid condition for which no reliable N2pc was elicited
16 and RTs were slower than in the valid condition), participants could still discriminate the mean
17 emotion from the face set above chance level, suggesting that this process could well operate
18 (albeit without a high precision or accuracy) under these impoverished conditions. Further, ERP
19 results for the SPCN component show indirectly that the four individual faces were likely
20 compressed and stored as one “single” object into visual short-term memory, suggesting in turn
21 that averaging multiple faces likely operates by means of “contraction”. This compression likely
22 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 from Ghent University, both awarded to LJ.

References

- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122-131.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392-398.
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *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. *Journal of Experimental Psychology: Human Perception and Performance*, *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 Research*, *85*, 20-25.
- Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America 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. *Psychological Science*, *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 Research*, 43, 393–404.
- Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Attention, Perception, & Psychophysics*, 67(1), 1-13.
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience?. *Trends in Cognitive Sciences*, 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. *Journal of Neuroscience 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. *Journal of vision*, 3(10), 1-1.
- Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience*, 28(16), 4183–4191.
- Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, 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. *Journal of Neuroscience*, *32*(21), 7169–7177.
- Florey, J., Clifford, C. W., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports*, *6*.
- Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, *106*(3), 1389-1398.
- Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167-192.
- Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751-R753.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718.
- Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. *From perception to consciousness: Searching with Anne Treisman*, 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. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *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. *Archives of Ophthalmology*, *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. *Consciousness and Cognition*, *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. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 233.
- Jolicœur, P., Brisson, B., & Robitaille, N. (2008). Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. *Brain research*, *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. *Frontiers in Human Neuroscience*, *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. *Frontiers in Psychology*, *7*.

- Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain Topography*, 6(1), 51-63.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213.
- Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, 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. *Journal of Experimental Psychology: Human Perception and Performance*, 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. *Journal of Cognitive Neuroscience*, 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. *Experimental Brain Research*. 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.
- McNair, N. A., Goodbourn, P. T., Shone, L. T., & Harris, I. M. (2017). Summary statistics in the attentional blink. *Attention, Perception, & Psychophysics*, 79(1), 100-116.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive*

Psychology, 14(2), 247-279.

Neumann, M., Ng, R., Rhodes, R., & Palermo, R. (in press). Ensemble coding of face identity is not independent of the coding of individual identity. *Quarterly Journal of Experimental Psychology*, 1-27.

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 in Cognitive Sciences*, 18(5), 242-250.

Praamstra, P. (2006). Prior information of stimulus location: Effects on ERP measures of visual selection and response selection. *Brain Research*, 1072(1), 153-160.

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

Posner, M. I. (2014). Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, 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. *Frontiers in Human Neuroscience*, 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. *Journal of Vision*, 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. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 813.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *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 Research*, *168*(3), 242-249.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *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 J.S. Werner and L.M. Chalupa (Eds.), *The New Visual Neurosciences* (pp.695-710). MIT Press.
- Whitney, D., & Leib, A. Y. (2018). Ensemble perception. *Annual Review of Psychology*, *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 Research*, *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. *Cognitive Therapy and Research*, 37(3), 572–578.

Figure Legends

Figure 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 ms-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.

Figure 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.

Figure 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.

Figure 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 no-clear-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.

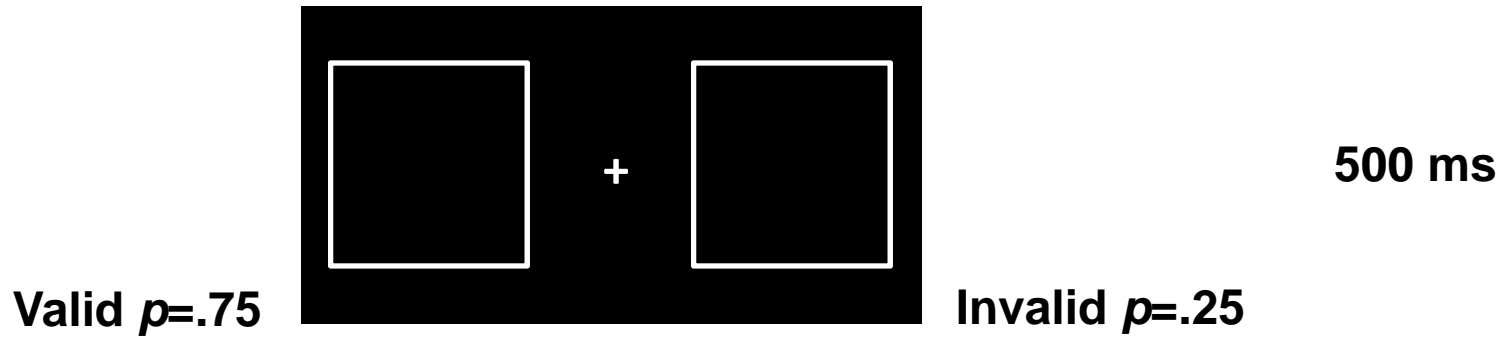
Figure 5. Topographic maps (back view) of the N2pc (200-300 ms) and SPCN (400-600 ms) components in the valid condition for the no-clear-saccade group, shown separately for the average emotion task (upper row) and the single emotion task (lower row).

Table 1

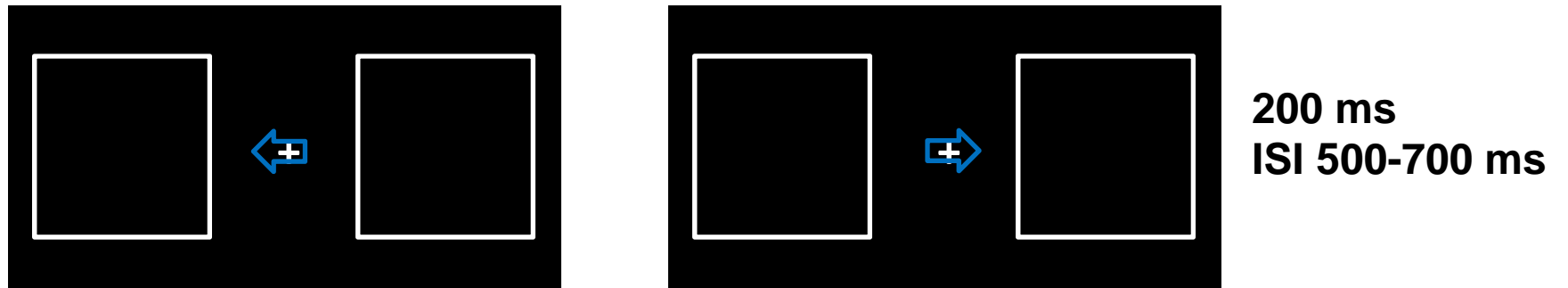
	Average valid	Average invalid	Single valid	Single invalid
Hit	.67(.12)	.60(.18)	.58(.15)	.55(.10)
False alarm	.45(.17)	.40(.17)	.38(.14)	.41(.14)
d prime	.61(.41)	.58(.39)	.54(.35)	.35(.42)

Note. The hit rate, the false alarm rate, and the corresponding d-prime score are presented for each condition separately. Mean (and standard deviation) is provided.

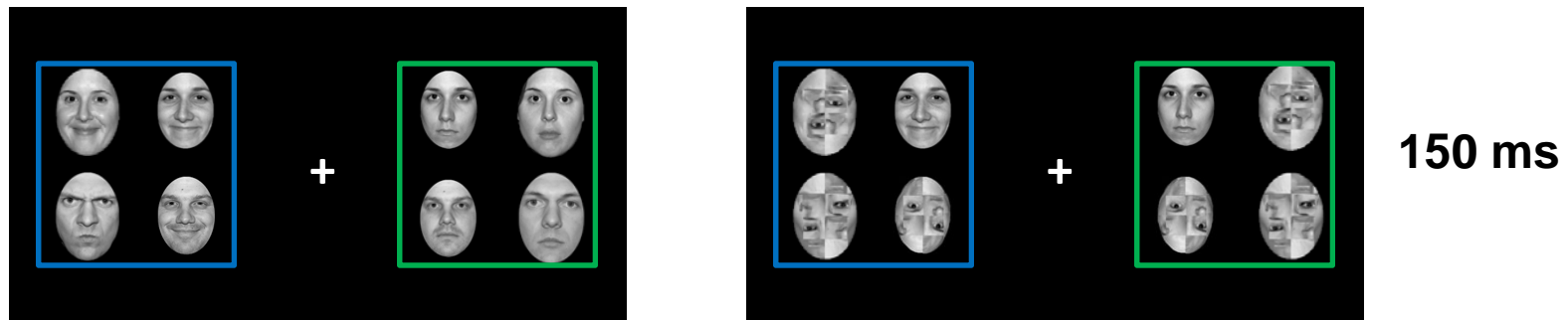
Fixation



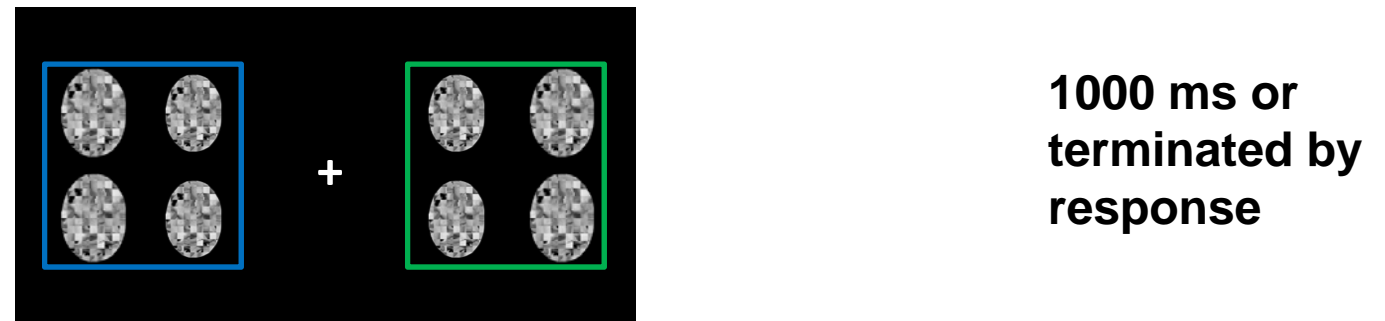
Cue



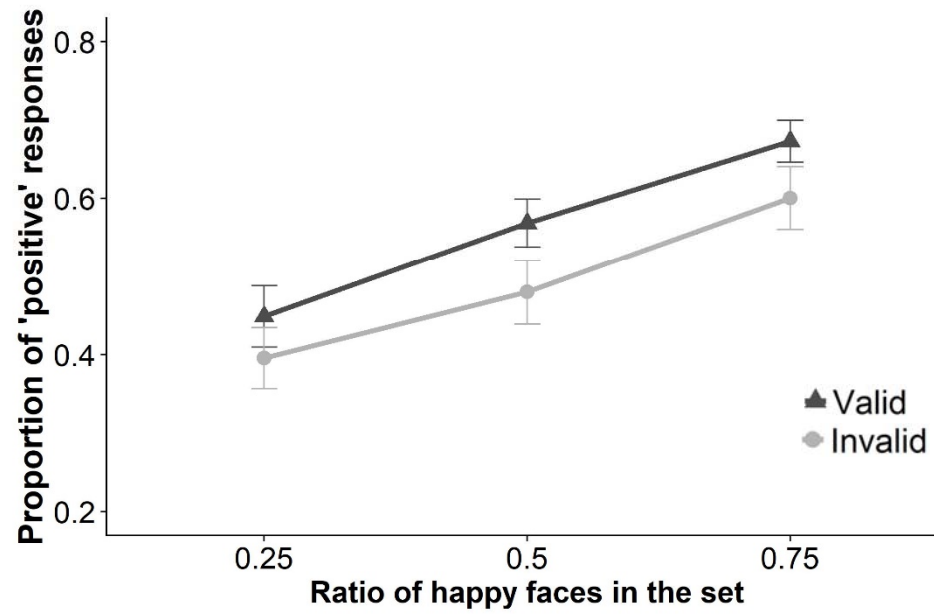
Target

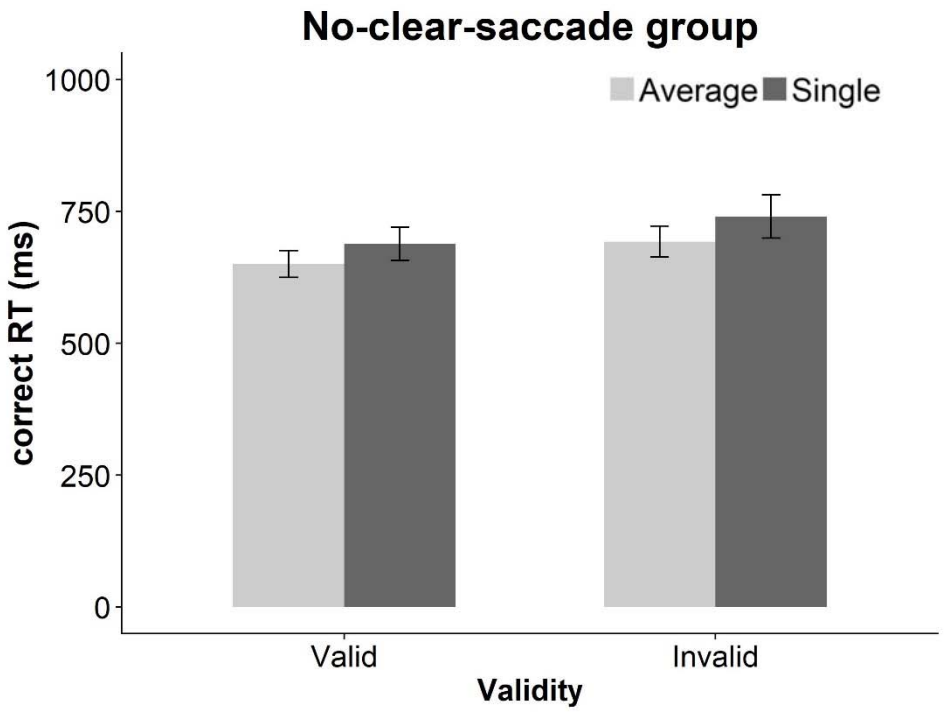
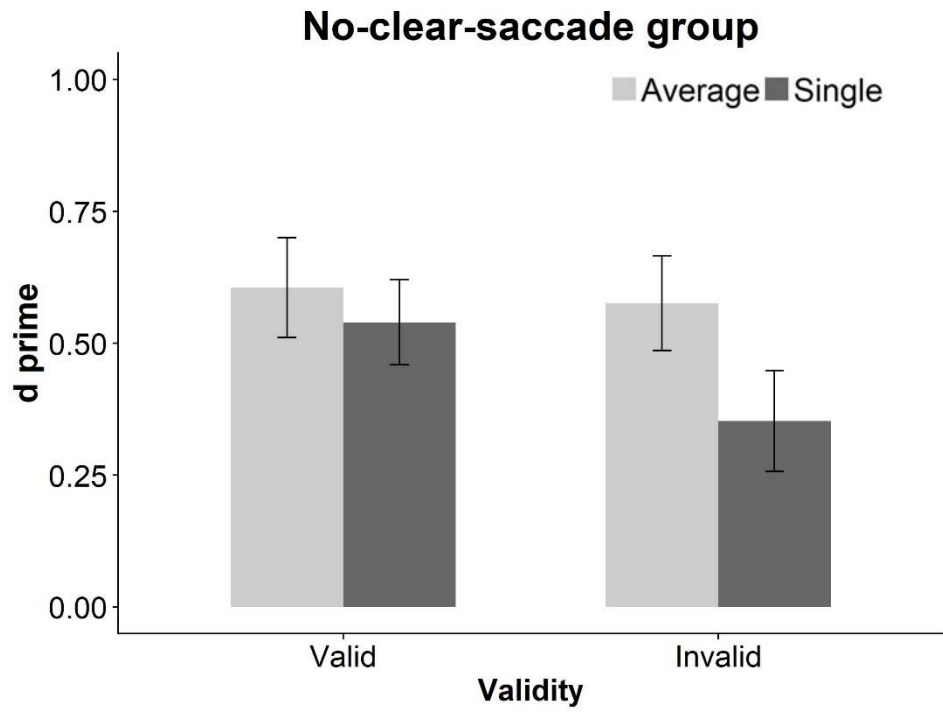


Mask

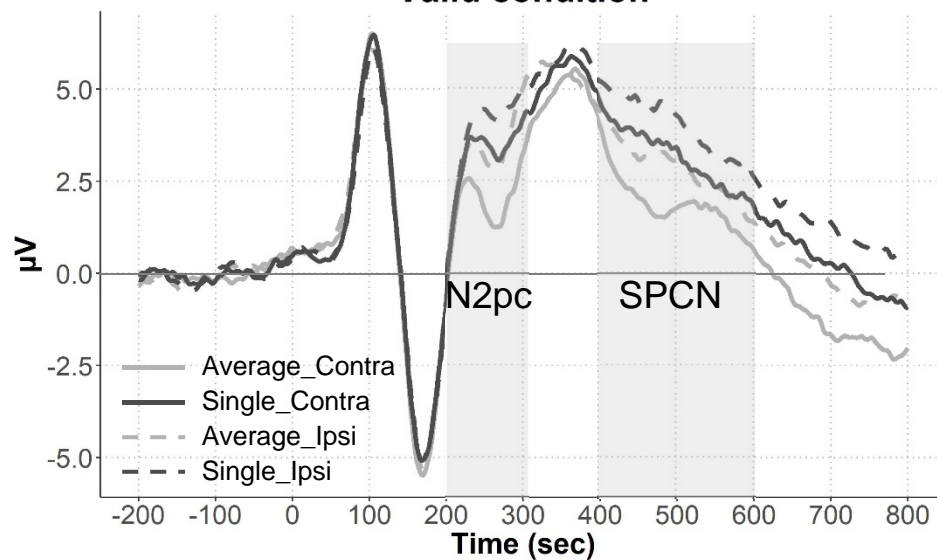


No-clear-saccade group

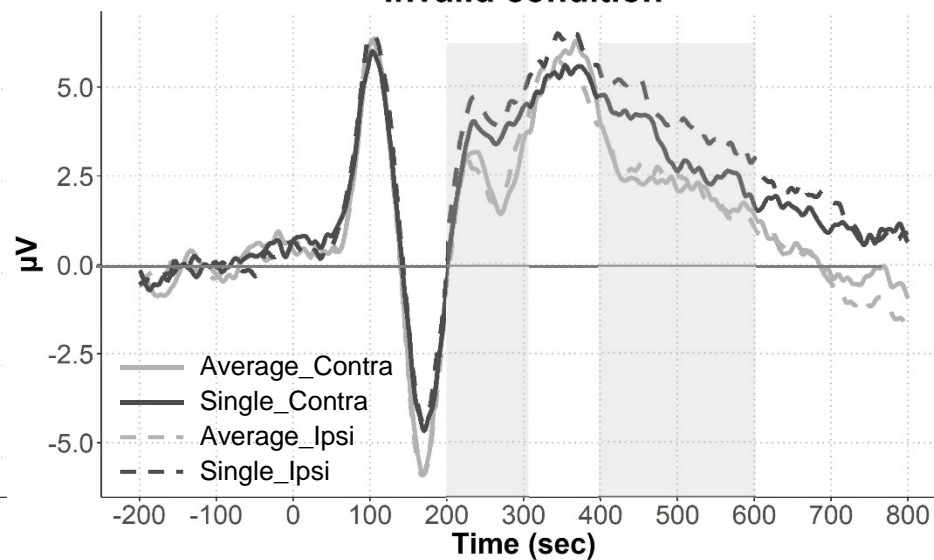




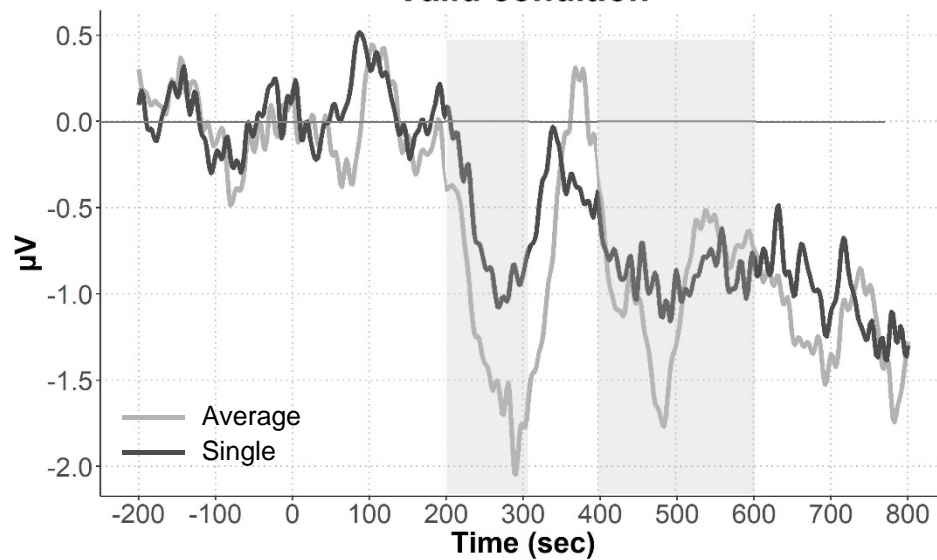
A) No-clear-saccade group
Valid condition



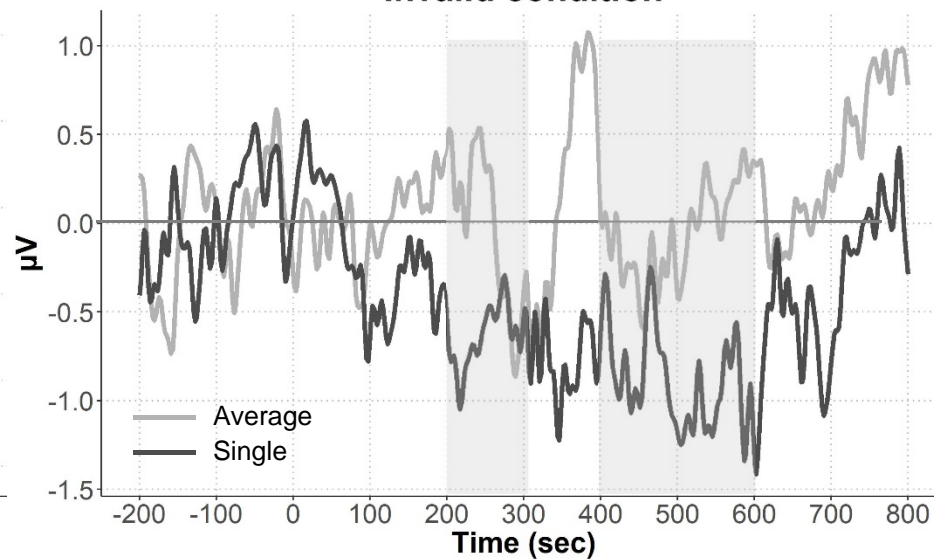
No-clear-saccade group
Invalid condition

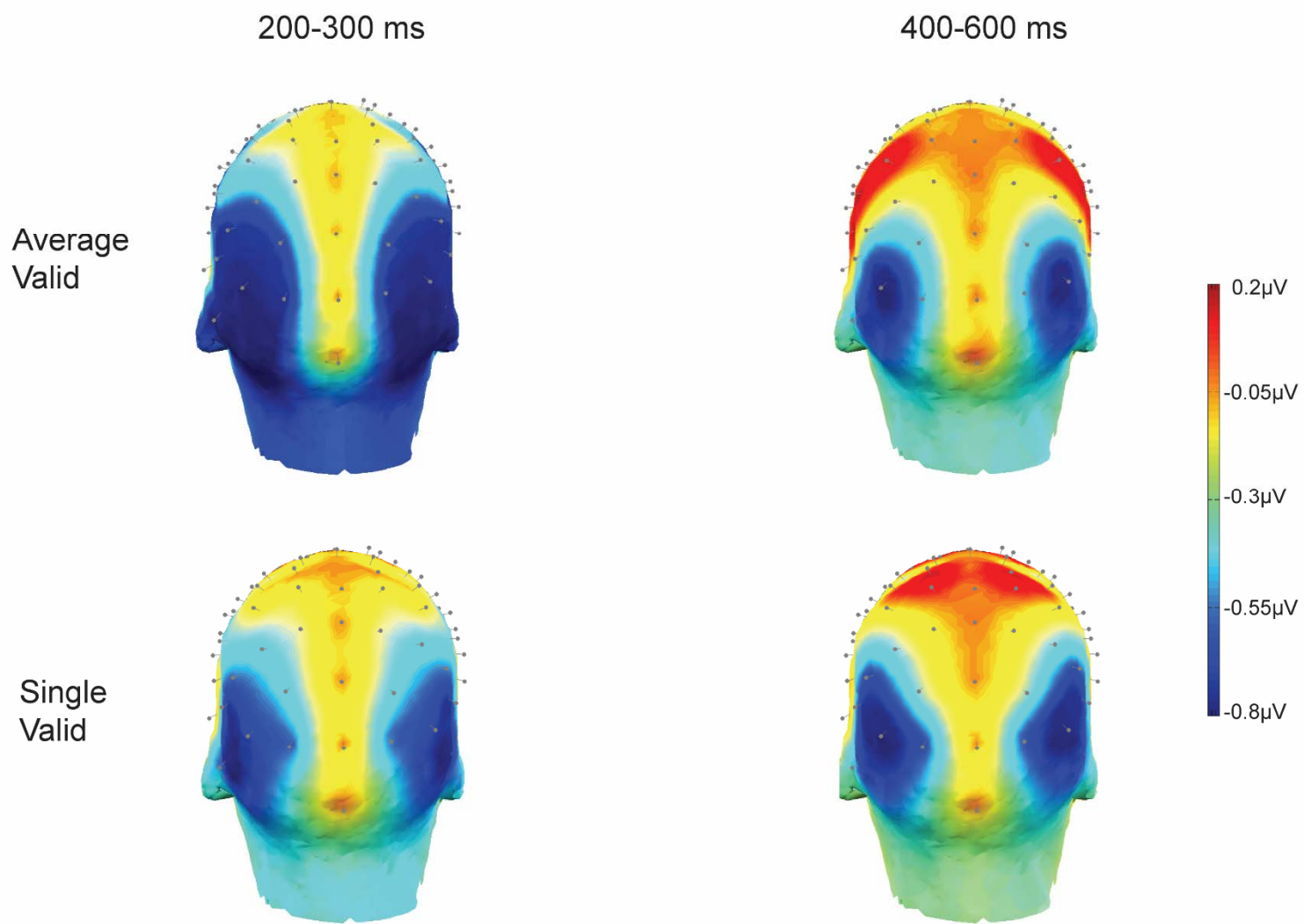


B) No-clear-saccade group
Valid condition



No-clear-saccade group
Invalid condition





Supplementary Materials

Method

Data analyses for the no-clear-saccade group

Mean emotion intensity of face sets. We computed the arithmetic mean emotion of the four faces in each set based on the subject-specific emotion intensity ratings obtained for these same faces (cf. post-experiment ratings). The larger the value, the more positive the computed mean emotion intensity was, while conversely, the smaller this value, the more negative the computed mean intensity was. A simple linear regression analysis was performed to assess if the computed mean intensity could be predicted by the ratio of happy faces in these sets.

Contribution of inner vs outer faces in the sets. We examined the effect of emotional face location (either inner/close to fixation or outer/further away from it) for the single and the average task separately. For the single task, the performance (*d prime*) was compared when the target face was present in the inner vs. the outer position. For the average task, the performance was compared when the single deviant face (i.e. the happy face in the Ratio0.25 condition, or the angry face in the Ratio0.75 condition) was presented in the inner vs. the outer position. The *d prime* scores were computed similarly as in the main analyses. Hits corresponded to judging the positive face sets as positive, and false alarms corresponded to judging the negative face sets as positive. In this control analysis, the sets composed of 50% happy and 50% angry faces were not included. For the two tasks separately, a two-way repeated-measures ANOVA was carried out with the factors Validity and (target or deviance) Location.

N170 component. The face-specific N170 component, measured from 150 to 190 ms post-stimulus onset (i.e., mean amplitude measurement) at lateral occipito-temporal electrodes (left cluster composed of P7, P9, PO7, and right cluster composed of P8, P10, PO8; same

methods as Puce et al., 2013) for the no-clear-saccade group, was submitted to a repeated-measures ANOVA with Task, Validity, Hemisphere, and target location as factors.

EEG preprocessing for the clear-saccade group

For the 17 participants of the clear-saccade group, we retained the trials with clear saccades only, based on the HEOG channel analysis. On average, 61% of trials were marked as contaminated by a horizontal saccade in the HEOG rejection procedure¹. The same artifact rejection procedure (removing the blink-related component and rejecting activity exceeding ± 80 μ V in the scalp EEG electrodes; on average 8% trials were rejected which did not differ across conditions) used for the no-clear-saccade group was applied to the clear-saccade group on their saccade-only epochs (-200 ms to 800 ms) time-locked to the onset of the target face sets. We also averaged only the correct trials separately for each type of task, validity and target location, and computed grand average ERPs of each condition for these 17 participants separately².

Data analyses for the clear-saccade group

The discrimination ability (*d prime*) and reaction time (RT) were computed for the clear-saccade group using the data aligned to the EEG data retained (i.e., only trials including clear eye movements were used; see here above). To further examine the role of saccade (and thus overt attention) on task performance, a three-way repeated-measure ANOVA was conducted on the *d prime* scores with Task (average, single) and Validity (valid, invalid) as two within-subject variables, and Group (no-clear-saccade, clear-saccade) as between-subjects variable. The N2pc and SPCN components were also computed similarly as for the no-clear-saccade group.

¹ The amount of saccades detected by the HEOG rejection procedure did not differ significantly between the average task ($63 \pm 23\%$) and the single task ($61 \pm 21\%$), $p = .36$. In addition, no significant difference was found between valid ($62 \pm 21\%$) vs. invalid ($61 \pm 23\%$) trials, nor the interaction between task type and validity reached significance, $ps > .30$.

² Trial count in the valid condition (209 ± 84) was larger than in the invalid condition (55 ± 19), $p < .001$, but the average task (133 ± 57) did not contain a significantly different number of trials compared with the single task (132 ± 55), $p = .91$.

Results

Behavioral results of the no-clear-saccade group

Mean emotion intensity of face sets. As can be seen from the supplementary Figure 1, the mean emotion intensity calculated based on individual intensity ratings (post-experiment rating) showed a large variance, but importantly, it was clearly sensitive to the ratio of happy versus angry faces contained in the sets, $F(1, 8241) = 24847, p < .001$, adjusted $R^2 = .75$, suggesting a reliable link between the (objective) ratio manipulation and the (subjective) subject-specific perception of these individual faces.

(insert Supplementary Figure 1 about here, single-column)

Contribution of inner vs. outer faces in the sets. In the single task, d prime scores were significantly above chance level (zero) in all conditions, except for the invalid-outer condition (i.e. the target face was presented in the outer position in the invalid condition). The ANOVA showed a significant main effect of Validity, $F(1, 18) = 7.25, p = .015, \eta_p^2 = .29$. There was also a significant main effect of (target) Location, $F(1, 18) = 17.46, p < .001, \eta_p^2 = .49$. When the target was in (one of the two) inner positions, the performance was better than when the target was in one of the two outer positions (Supplementary Figure 2). The interaction between Validity and Location was not significant, $F(1, 18) < 1, \eta_p^2 = .03$. In the average task, d prime scores were significantly above change level in all four conditions. The ANOVA revealed a significant interaction between (deviance) Location and Validity, $F(1, 18) = 12.46, p = .002, \eta_p^2 = .41$. Simple effects analyses showed that in the valid condition, performance was reliably influenced when the deviant face was presented in the inner compared to the outer position, $F(1, 18) = 21.77, p < .001, \eta_p^2 = .55$. Accordingly, the deviant face shown close to fixation had a larger

weight (and impact on the averaging process) than the same face shown further away from it, even though for this latter, we increased its size (and presumably salience, see Supplementary Figure 2). Interestingly, the different impact of inner vs. outer faces in the average task only occurred in the valid condition, but not in the invalid condition, providing an indirect manipulation check regarding the focus of attention (being on the target set of four faces for valid trials but presumably not for invalid trials).

(insert Supplementary Figure 2 about here, single-column)

N170 results of the no-clear-saccade group

There was a trend significant main effect of task on the N170 component, with larger amplitude in the average task ($-4.02 \pm 2.29 \mu\text{V}$) compared with the single task ($-3.44 \pm 2.78 \mu\text{V}$), $F(1, 18) = 3.15$, $p = .093$, $\eta_p^2 = .15$. None of the other main effects reached significance, nor was any interaction significant, $ps > .15$.

Behavioral results of the clear-saccade group

Average and single emotion judgment. The ANOVA on the d' prime scores showed no significant main effect of Task, $F(1, 16) = 2.74$, $p = .12$, $\eta_p^2 = .15$, or interaction between Task and Validity, $F(1, 16) = 2.10$, $p = .17$, $\eta_p^2 = .12$. However, there was a significant main effect of Validity, $F(1, 16) = 14.27$, $p = .002$, $\eta_p^2 = .47$. The performance in the invalid condition ($M = .28$, $SD = .45$) significantly dropped compared with that in the valid condition ($M = .77$, $SD = .32$) (Supplementary Figure 4). For RTs (correct responses), there was a significant main effect of Task, $F(1, 16) = 6.08$, $p = .025$, $\eta_p^2 = .28$. The average task ($M = 615.63$, $SD = 120.64$) was performed faster than the single task ($M = 661.10$, $SD = 134.49$). There was also a significant

main effect of Validity, $F(1, 16) = 32.12, p < .001, \eta_p^2 = .67$. The RTs were significantly shorter in the valid ($M = 600.73, SD = 114.82$) compared to the invalid condition ($M = 676.00, SD = 134.42$). The interaction between the two factors was not significant, $F(1, 16) < 1, \eta_p^2 = .67$ (Supplementary Figure 4).

When adding Group as an additional factor in the omnibus ANOVA, we found a significant three-way interaction on discrimination ability (i.e., d' prime scores), $F(1, 34) = 4.84, p = .035, \eta_p^2 = .13$. Further, the two-way interaction between Task and Group, $F(1, 34) = 4.26, p = .047, \eta_p^2 = .11$, and between Validity and Group, $F(1, 34) = 7.71, p = .009, \eta_p^2 = .19$ were also significant. The discrimination ability was numerically larger in the average task ($M = .59, SD = .37$) compared with the single task ($M = .45, SD = .35$) for the no-clear-saccade group, $p = .25$, while an opposite trend was found for the clear-saccade group (average: $M = .45, SD = .33$; single: $M = .60, SD = .35$), $p = .20$. In addition, compared with the no-clear-saccade group (valid: $M = .57, SD = .32$; invalid: $M = .46, SD = .29$), the clear-saccade group performed better in the valid condition, $p = .067$, but worse in the invalid condition, $p = .15$.

(insert Supplementary Figure 3 about here, single-column)

(insert Supplementary Figure 4 about here, single-column)

ERP results of the clear-saccade group

There was no reliable N2pc for either the valid ($-.38 \pm 1.58 \mu\text{V}$) or the invalid condition ($.64 \pm 3.35 \mu\text{V}$) in the average emotion task, $t(16) = -.99, p = .34, t(16) = .78, p = .45$. It was the same for the single emotion task (valid: $-.45 \pm 1.06 \mu\text{V}$; invalid: $.24 \pm 1.22 \mu\text{V}$), $t(16) = -1.77, p = .096, t(16) = .82, p = .43$ (Supplementary Figure 5). No reliable SPCN was elicited in the valid

($-0.47 \pm 1.85 \mu\text{V}$) or the invalid condition ($-0.40 \pm 3.78 \mu\text{V}$) of the average emotion task either, $t(16) = -1.06, p = .31, t(16) = -0.44, p = .67$. The SPCN was present in the valid ($-1.06 \pm 1.16 \mu\text{V}$) but not the invalid condition ($0.13 \pm 2.23 \mu\text{V}$) of the single task, $t(16) = -3.74, p = .002, t(16) = 2.45, p = .02$ (Supplementary Figure 5). These results therefore confirmed that no clear covert shifts of spatial attention were elicited in response to the target face sets onset in this group, because they had already overtly moved their eyes towards the expected target location in response to the cue (see Supplementary Figure 3).

(insert Supplementary Figure 5 about here, single-column)

Supplementary Figure Legends

Supplementary Figure 1. The computed mean intensities for the different sets (as established using the post-experiment rating) shown as a function of the ratio of happy faces contained in them. The larger the value, the more positive the computed mean emotion intensity was, while conversely, the smaller this value, the more negative the computed mean intensity was. A significant linear effect was found between these two variables.

Supplementary Figure 2. Discrimination ability (no-clear-saccade group) for the average emotion task (Left) and the single emotion task (Right) shown separately for the inner and outer conditions, as well as two levels of validity. The inner and outer locations correspond to the two possible positions relative to fixation where the target (single emotion task) or deviant face (average emotion task) was presented. The error bar represents one standard error of the mean.

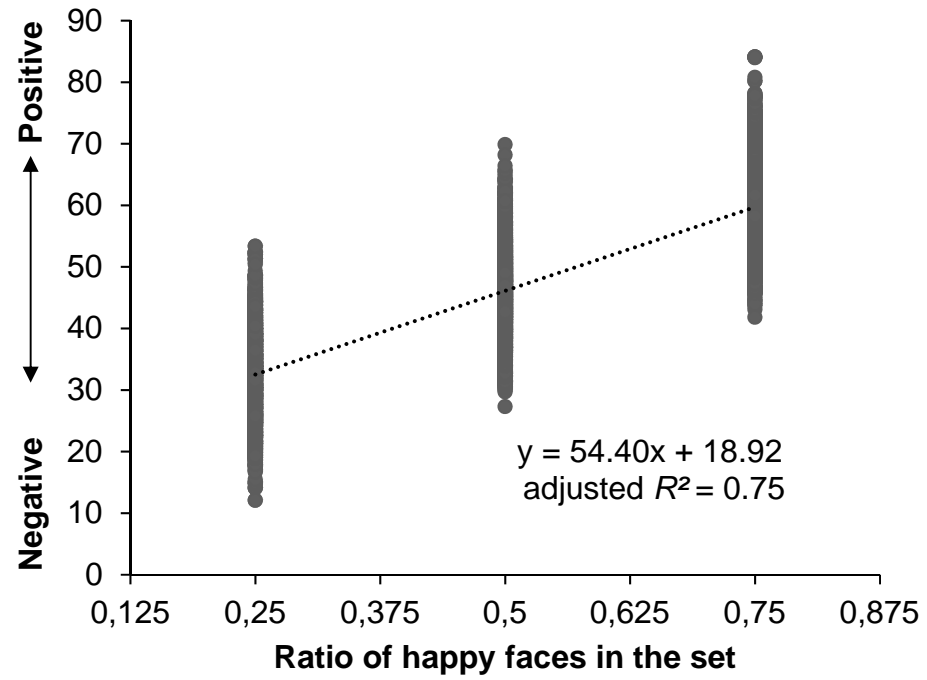
Supplementary Figure 3. Examples of horizontal eye movements during the cue-target interval recorded in the channel of HEOG for one representative participant with no clear saccade (Left) and one with clear saccades (Right). This subject had clear saccades even after running the HEOG rejection procedure based on a step-like artifact function, such as implemented in ERPLAB (see Method). The highlighted area indicated the time window of target onset (700-900 ms after cue).

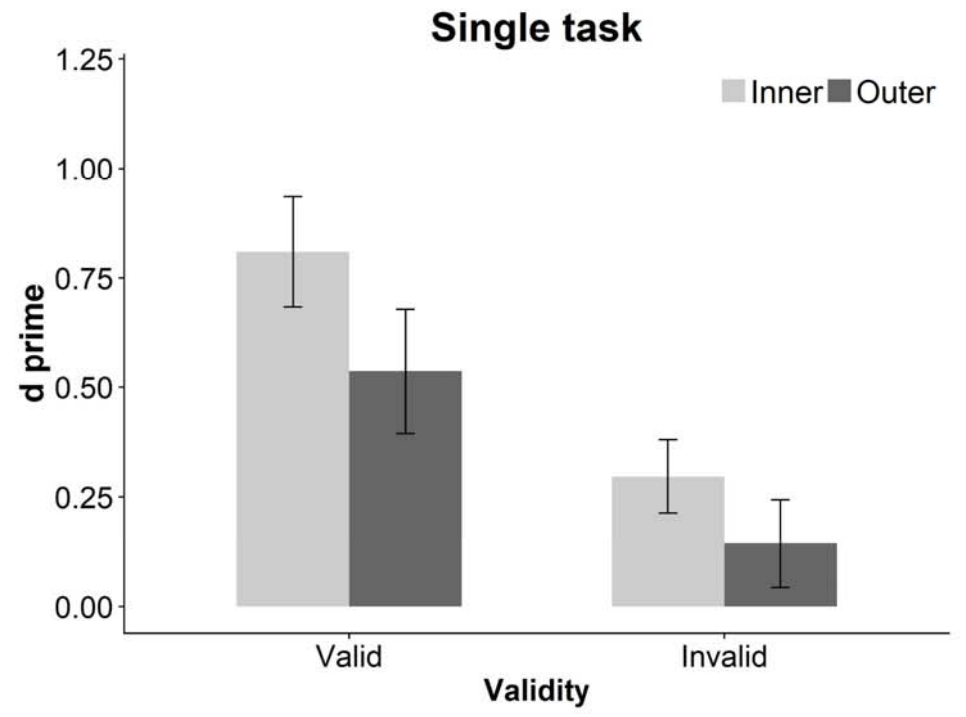
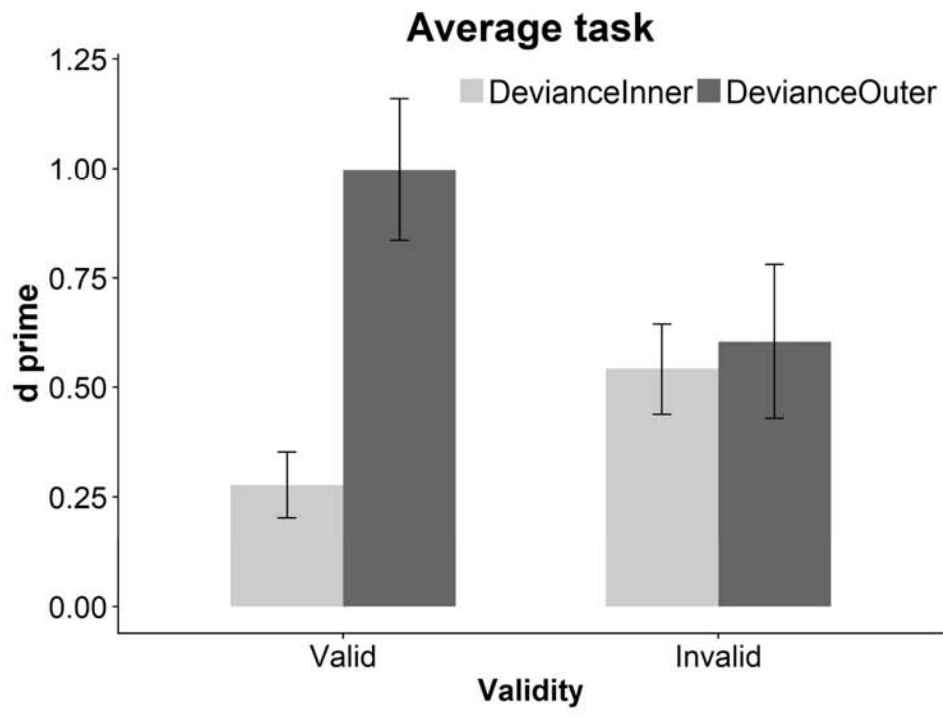
Supplementary Figure 4. Discrimination ability (d' prime, Left) and RT for correct responses (Right) shown separately for the two tasks and two levels of validity for the clear-saccade group. The error bar represents one standard error of mean.

Supplementary Figure 5. (A) Grand-averaged ERPs (clear saccade group) time-locked to the onset of the bilateral face sets at PO7/8, separately for the valid and invalid condition.

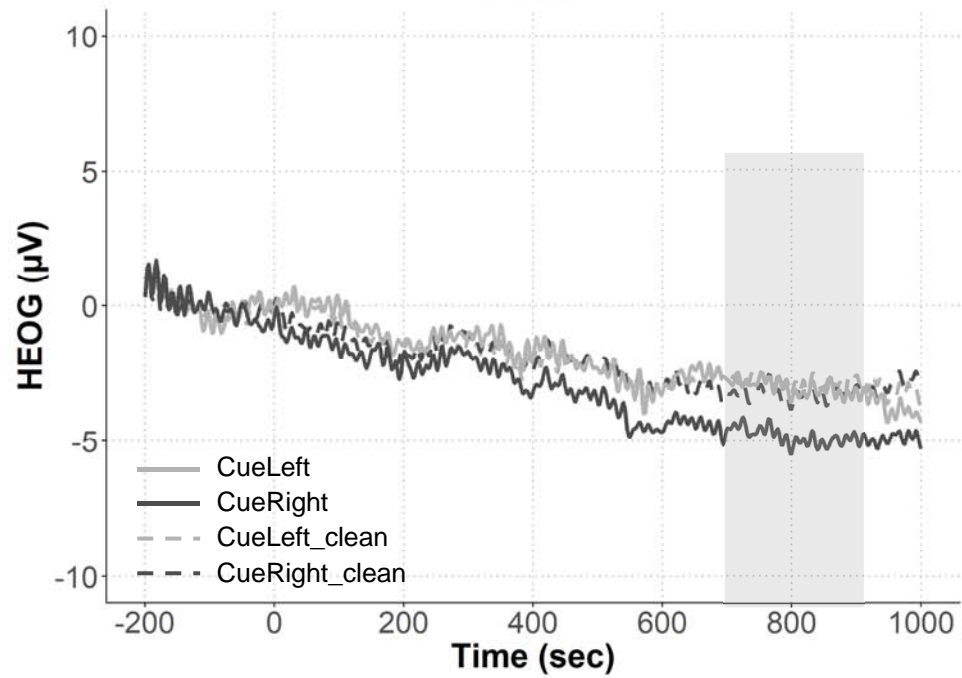
Contralateral (solid lines) and ipsilateral (dashed lines) ERP waveforms relative to the position of the target face(s) are presented. Note that these waveforms are computed for correct trials only.

(B) Difference waveforms (computed by subtracting ERPs at PO7/8 ipsilateral to the target location from that at contralateral electrodes) are presented separately for the valid (Left) and invalid condition (Right). The two highlighted areas indicate the time-windows used for N2pc (200-300 ms) and SPCN (400-600 ms) amplitude measurements (after target onset).





Sub1



Sub4

