This is the authors' version of the manuscript that appears in Neuroimage as

Ms. No.: NIMG-15-1876R2

Title: Electrophysiological correlates of Self-specific prediction errors in the human brain

Corresponding Author: Dr. Alejandra Sel

Authors: Rachel Harding; Manos Tsakiris, PhD Article Type: Regular Article

Title

Electrophysiological correlates of self-specific prediction errors in the human brain

Alejandra Sel*, Rachel Harding, Manos Tsakiris

Lab of Action & Body, Department of Psychology,

Royal Holloway University London,

Egham, Surrey, TW20 0EX, London, U.K.

Tel: +441784276551

*E-mail: alex.sel@rhul.ac.uk

Abstract

Recognising one's self, vs. others, is a key component of self-awareness, crucial for social interactions. Here we investigated whether processing self-face and self-body images can be explained by the brain's prediction of sensory events, based on regularities in the given context. We measured evoked cortical responses while participants observed alternating sequences of self-face or other-face images (experiment 1) and self-body or other-body images (experiment 2), which were embedded in an identity-irrelevant task. In experiment 1, the expected sequences were violated by deviant morphed images, which contained 33%, 66% or 100% of the self-face when the other's face was expected (and vice versa). In experiment 2, the anticipated sequences were violated by deviant images of the self when the other's image was expected (and vice versa), or by two deviant images composed of pictures of the self-face attached to the other's body, or the other's face attached to the self-body. This manipulation allowed control of the prediction error associated with the self or the other's image. Deviant self-images (but not deviant images of the other) elicited a visual mismatch response (vMMR) - a cortical index of violations of regularity. This was source localized to face and body related visual, sensorimotor and limbic areas and had amplitude proportional to the amount of deviance from the self-image. We provide novel evidence that self-processing can be described by the brain's prediction error system, which accounts for self-bias in visual processing. These findings are discussed in the light of recent predictive coding models of self-processing.

Keywords: self, face perception, body perception, prediction error, visual mismatch response

Introduction

Recognising and representing one's self, as distinct from others, is a fundamental component of human experience, essential for self-awareness and social cognition. However, the scientific accounts that have been proposed to explain how the brain processes self-related information remain controversial. Neuroimaging studies of self-face and self-body processing have shown activity in specific neural areas that differ from those involved in the processing of familiar and unfamiliar stimuli (Platek et al., 2006; Northoff et al., 2006; Frassinetti et al., 2008; Keyes et al., 2010). Such results support the hypothesis that there is brain specialisation for self-processing (e.g. Northoff et al., 2006). There are, however, significant inconsistencies in the reports of such studies. For example, some have shown that right prefrontal areas are particularly relevant for self-face processing (Platek et al., 2004), whereas others report activation of left frontal areas in recognition of one's own face (Gillihan et al., 2005). As a result we lack a unifying framework for self-processing in the brain, which can be embedded within wider theories of cortical function.

Current models of the self have proposed that self-related information has an overall advantage over the processing of non self-related information (Apps and Tsakiris, 2014). This accounts for the ability to respond rapidly to self-related stimuli, as opposed to stimuli related to others. Clinical, neuroimaging and behavioural studies have shown that presentation of one's own face leads to enhanced activation in visual and multimodal brain areas and is also associated with faster or more accurate performance when compared to processing of other faces (e.g. Northoff et al., 2006; Devue et al., 2007; Keyes et al., 2010). Moreover, self-related stimuli can influence the processing of subsequent information, as evidenced by studies showing self-biases in face recognition after the presentation of self-related primes (Platek et al., 2004; Pannese and Hirsch, 2011). Furthermore, studies on the effect of self-association have shown that neutral objects that have acquired personal significance, by

learned association with the self, lead to enhanced activity in brain regions linked to self-representation and behavioural self-biases (Sui et al., 2015).

In an attempt to account for the evidence of self-bias, recent theoretical models have proposed a unifying account of the self whereby self-related information takes the form of incoming sensory events that are compared and integrated with the mental representations of the self that have been formed from previous context-based sequential regularities (Fotopoulou, 2012; Limanowski and Blankenburg, 2013; Apps and Tsakiris, 2014; Moutoussis et al., 2014). These models fit within a general predictive coding framework of brain functioning, which proposes that sequential regularities are extracted from past sensory events, leading to the formation of predictions about the upcoming sensory events (Friston 2003). Importantly, these predictions allow us to automatically detect subtle unexpected changes in the environment and they thus play a central role in human cognition (e.g. Kimura, 2012; Stefanics, et al., 2012).

One way to test the validity of such theoretical predictive coding models of self-processing is to take advantage of the known properties of the electrophysiological signatures that reflect automatic change detection in vision, such as the visual mismatch negativity component (vMMN) (Stefanics, et al., 2014; Kimura et al., 2012; Winkler and Czigler, 2012). The vMMN is a counterpart of the auditory mismatch negativity (MMN - for reviews see Näätänen et al., 2007). It is also known as visual mismatch response (vMMR) and includes visual mismatch responses with both negative and positive polarity (Sulykos and Czigler, 2011; Stefanics et al., 2014). The vMMN is thought to be an electrophysiological correlate of the automatic prediction error responses that are generated when a current event is incongruent with events that are predicted on the basis of previous sequential regularities (Pazo-Alvarez et al., 2003; Czigler, 2007; Kimura, 2012). Typically, the vMMR is elicited by events with deviant visual features, such as changes in colour, orientation, or movement. And

there is evidence suggesting that its amplitude depends on the degree of visual mismatch between the expected and the current event (He et al, 2014). A vMMN has been also observed in response to changes in biological visual stimuli, such as changes in facial identity (Susac et al., 2004); changes in facial emotion (Astikainen and Hietanen, 2009; Kimura et al., 2011; Stefanics et al., 2012); and changes in hand laterality (Stefanics and Czigler, 2012), even when these changes were unrelated to the participant's task. An, as yet unanswered, question is whether the visual system can automatically detect changes in self-related information by extracting regularities embedded in the context of self- and other images, where, for example, a particular image may be expected, based on the previously extracted temporal regularities in the stimuli (e.g. Apps and Tsakiris, 2014).

Prediction error processing for biologically relevant visual stimuli has been assessed in a recent vMMN study (Kimura et al., 2012). Kimura et al. (2012) presented alternating sequences of happy and fearful faces in an identity-irrelevant task. Anticipated sequences of happy faces were violated by occasional fearful faces (and vice versa), giving rise to a vMMN. Importantly for the work presented here, the authors interpreted these vMMN responses as evidence of automatic prediction errors for emotional faces, based on the temporal context of the sequence. In other words, they propose a prediction-error account of the vMMN, which is distinct from other processes of memory mismatch (such as those involved in the sensory memory account of MMN) (Winkler, 2007, Stefanics et al., 2012, Kimura, 2012). Evidence for automatic detection of errors in self-related information has been tested further in two electrophysiological studies of self-voice processing (Graux et al., 2013; 2014). These studies used an oddball sequence consisting of frequent presentation of unknown voices, violated by infrequent familiar or unfamiliar voices. Graux et al. (2013, 2014) reported no difference in the MMN component between deviant self-voice stimuli and deviant familiar/ unfamiliar other voices. It remains unclear whether the visual processing of

self-images takes advantage of an automatic prediction error system. We predict that such a system would operate differently for self-related and other-related information and would give rise to self-biases in the processing of self-images. It might, however, be found that visual self-processing does not conform to the principles of automatic prediction error, as reported in Graux' studies of self-voice information.

In two linked vMMR experiments, we therefore set out to investigate whether automatic context-based predictions can account for the self-bias observed in the processing of self-face and self-body images. In experiment 1, participants were presented with alternating sequences of self-face and other face stimuli, in an identity-irrelevant task. These sequences were occasionally violated (by deviant images of the self-face, when other's face was expected and by the other's face when the self-face was expected). This allowed us to compare directly the prediction errors associated with the self and with the other's face. To control for familiarity effects in face processing, we alternated the presentation of the self-face with that of a familiar or unfamiliar person, in two separate sequences. In experiment 2, participants observed alternating sequences of whole-body images of the self and of another person, in an identity-irrelevant task. The expected sequence was occasionally violated by deviant whole-body images of the self when the other was expected (and likewise by the other when the self was expected). Finally, to explore whether the vMMR recorded in experiments 1 and 2 was associated with visual and multimodal cortical areas involved in self-facial and self-bodily processing, we examined the neural generators of the vMMR response by using standardized low-resolution brain electromagnetic tomography (s-LORETA).

According to previous vMMR studies of biologically relevant visual stimuli (Susac et al., 2004; Stefanics et al., 2012), we hypothesised that the alternating repetition of self and other images would build up an automatic prediction of a self-other sequential pattern. Therefore,

we hypothesised that the presentation of a deviant facial stimulus that differed from the predicted element in the series would lead to an automatic mismatch response (i.e. vMMR). Importantly, we predicted that deviant self-images would lead to a larger vMMR than deviant other-images. Previous evidence has demonstrated activity suppression in cortical areas when a self-related stimulus results in a predictable sensory input (Blakemore et al., 1998, 2000). However, we expected that when a self-related stimulus resulted in an unpredictable sensory input, activity would be enhanced. Furthermore, we expected that the magnitude of the vMMR to deviant self-images would be proportional to the amount of error in the deviant image, such that greater visual discrepancy between the expected and the current images would elicit a greater vMMR. Moreover, if the vMMR response indexes automatic prediction error for self-images, we expected that the vMMR would be source localized within visual, limbic and sensorimotor cortices, which are the brain areas that play a central role in the processing of facial and bodily information for the self (Northoff et al., 2006). We also hypothesised that the vMMR to deviant other-images would be of different magnitude and would have a separate neural source than that for the vMMR to deviant self-faces. Finally, if the vMMR to deviant self-images depends on the effect of familiarity, we expected that similar mismatch evoked responses would be observed to deviant self-faces and deviant familiar others' faces.

Study 1

Material and methods

Participants

Sixteen neurologically unimpaired paid participants (4 males, mean age 22.9 years; laterality quotient 87.81%) (Oldfield, 1971) were tested. Participants gave their informed consent, with approval by the Ethics Committee, Department of Psychology, Royal Holloway University of London.

169 *Stimuli and procedure*

170 Seven grey-scaled pictures of faces (250 x 343 pixels) were presented centrally on a black
171 background, using E-prime software (Psychology Software Tools). Facial stimuli comprised
172 the participant's face, the face of a gender-matched (familiar or unfamiliar) individual and 4
173 morphed faces that contained respectively 33% and 66% of participant's face and 66% and
174 33% of the familiar or of the unfamiliar gender-matched other person. Images were edited
175 with Photoshop software (Adobe Systems, San Jose, CA) and all the images were equalised
176 for luminosity, contrast, shadows, highlights, colour and image size. The photographs of the
177 participant's face were taken in a separate session, prior to the experimental session. Familiar
178 faces were those of famous people (e.g. Beyoncé, Angelina Jolie), selected on the basis of the
179 participants' ratings of their familiarity, which was assessed before the EEG recording.
180 Unfamiliar faces were images of two individuals, selected from our in-house database, that
181 had never been seen by the participants, prior to the experiment. In addition, seven target
182 faces were composed, by adding black sunglasses to the various facial stimuli. The stimulus
183 duration for presentation of each face was 250 ms and the stimulus onset asynchrony was 600
184 ms.

185 The stimuli were presented in two sequences, within which the presentation of the
186 participant's face was alternated with presentation of another face. In the 'familiar sequence',
187 the self-face was alternated with the face of a familiar other, while in the 'unfamiliar
188 sequence' the self-face was alternated with the face of an unfamiliar other person. The
189 percentage of regular trials of self-image and other images, presented in the sequence, was
190 81.4% i.e. in these trials there were no deviant images. In the remaining 18.6% of trials, the
191 alternating sequence was irregularly violated to create deviants that were associated either
192 with the self-face or with the face of another person. In half of these irregular trials,
193 participants were expecting to see a familiar/unfamiliar face but saw instead a face that

194 contained 33%, 66% or 100% of the self. This was the 'self-related deviant condition'. In the
195 remaining half of the irregular trials, participants were expecting to see their own face but
196 instead saw a face that contained 33%, 66% or 100% of the other face. This was the 'other-
197 related deviant condition' (Figure 1). Of the 18.6% irregular trials, one-third contained the
198 33% morph, one-third the 66% morph and one-third had 100% content of the unexpected
199 face. These three types were randomised across the deviant presentations. They are referred
200 to below as 33% error, 66% error and 100% error in expectation. Based on standard
201 paradigms used in vMMN studies (e.g. Kimura et al 2012), the deviant stimuli were designed
202 to randomly violate the regular order of the alternating sequences of facial stimuli, subject to
203 two limitations. Firstly, each sequence always started with 4 regular presentations of the
204 alternating stimuli, and secondly, two deviant stimuli were never presented sequentially.
205 On 9.3% of trials, target stimuli were presented within the sequence. Participants were
206 instructed to ignore all other stimulus attributes (e.g. their facial identity) and to respond as
207 quickly and accurately as possible by pressing a button when the target stimuli (any face
208 wearing black sunglasses) were presented. The presence of these targets ensured that
209 participants attended to the task. The experiment contained 2 experimental blocks (consisting
210 of one familiar other and one unfamiliar other sequence) separated by a break, with 2588
211 trials per block. Each block was made up of 240 deviant trials of each type (self and other),
212 including the 33% morph, the 66% morph and the 100% unexpected face. The order of the
213 blocks was pseudo-randomized and counterbalanced across participants. Participants were
214 seated in a dimly lit, sound attenuated and electrically shielded, chamber in front of a monitor
215 (Samsung SyncMaster 940N; size = 19 inches; resolution = 1280 X 1024) at a distance of 90
216 cm.
217 *Behavioural performance*

Participants' behavioural performance was measured in terms hit rate (%) and reaction time (ms). Because of the number of irregular target trials was quite low, the regular and the irregular targets were collapsed for purposes of analysis, resulting in 4 conditions ('self familiar', 'self unfamiliar', 'other familiar', 'other unfamiliar'). Responses that were made less than 250 ms after the trial onset were discarded. The measures were submitted to repeated-measures ANOVAs, with factors comprising identity of the facial image (self, other) and familiarity (familiar, unfamiliar)

EEG recording and data analysis

EEG was recorded with Ag-AgCl electrodes from 64 active scalp electrodes mounted on an elastic electrode cap, according to the International 10/20 system, using ActiveTwo system (AD-box) and Actiview software (BioSemi, Amsterdam, Netherlands; 512 Hz sampling rate; band-pass filter 0.16-100Hz (down 3 dB); 24 bit resolution). Electrodes were referenced to the Common Mode Sense and Driven Right Leg electrodes and rereferenced to the average reference off-line. Vertical and bipolar horizontal electrooculograms were recorded for artifact correction purposes. Off-line EEG analysis was performed using Vision Analyzer software (BrainProducts). The data were digitally low-pass-filtered at 30 Hz (12 dB/oct), and ocular correction was performed (Gratton et al., 1983). Epochs of 600 ms were extracted from the raw EEG data from 100 ms before the face onset to 500 ms after the face onset. Epochs were baseline corrected to the first 100 ms. Automatic artifact rejection was combined with visual inspection for all participants (±100 μV threshold; 0.15% mean percentage of data was rejected due to excessive amplitude) (see Supplemental table 1 for percentage of trials included in the analysis). Single-subject ERPs were calculated for each facial image (self, other), each expectancy (expected, 33%, 66%, 100% error) and the two sequences (familiar, unfamiliar) and were used to compute ERP grand averages across

subjects. The responses to the first four trials of the sequence were not included in the average.

To estimate the effects of violations in the regular sequence, single-subject averages of the irregular minus the regular difference waves (i.e. the difference between ERPs) were calculated in the following manner. For the self-related deviant conditions, the regularity violation effects were obtained by subtracting the ERPs elicited by regular self-face trials from (i) the ERPs elicited by 100% irregular self-face trials ('deviant self 100% error'); (ii) the ERPs elicited by 66% irregular self-face trials ('deviant self 66% error'); and (iii) the ERPs elicited by 33% irregular self-face trials ('deviant self 33% error'). For the other-related deviant conditions, the regularity violation effects were obtained by subtracting the ERPs elicited by regular other's face trials from (i) the ERPs elicited by 100% irregular other's face trials ('deviant other 100% error'); (ii) the ERPs elicited by 66% irregular other's face trials ('deviant other 66% error'): (iii) the ERPs elicited by 33% irregular other's face trials ('deviant other 33% error') (Supplementary table 3). This was done separately for both the familiar and the unfamiliar trials.

The differential activity in the ERPs was averaged across participants and then compared for the 33%, 66%, and 100% 'error' of the self and the other face, in the time windows 100-130 ms, 170-300 ms and 300-400 ms. These intervals were chosen in accordance with the latencies reported in previous vMMR studies (e.g. Kimura et al., 2012; Stefanics et al., 2012). In line with the standardised procedure (e.g. Gosling and Eimer, 2011; Stefanics et al., 2012), four regions of interests (ROIs) were defined on the basis of the difference potential maps, including those channels in which experimental effects could be predicted, based on previous vMMR literature (Astikainen and Hietanen, 2009; Stefanics and Czigler, 2012; Stefanics et al., 2012). There were two (right and left) posterior-temporal ROIs (P7/8, P9/10, PO7/8 and O1/2 electrodes of the 10/20 system); a central ROI (C1/2, Cz, FCz); and a frontal ROI (AFz,

Fz, F1/2) (Figure 2). Factors included in the analysis were: facial image (self, other); error level (33%, 66%, 100%); familiarity (familiar, unfamiliar); ROI (right/left posterior-temporal, central or frontal); and channel (4 levels). Mauchly's W was computed to check for violations of the sphericity assumption and Greenhouse–Geisser adjustments to the degrees of freedom was applied when needed. The p values were corrected for multiple comparisons using stepwise Bonferroni-Holm correction.

Current source density analysis

Standardized Low Resolution Brain Electromagnetic Tomography (s-LORETA) was used to estimate the brain generators associated with modulations of vMMN response. s-LORETA provides an approximate three-dimensional discrete solution to the inverse EEG problem. s-LORETA is used to computed statistical maps from EEG data to indicate the locations of the putative underlying source generators. These maps are derived by performing a location-wise inverse weighting of the results of a minimum norm least squares analysis, together with their estimated variances. s-LORETA performs source localization in 6239 cortical gray matter voxels, sized 5 mm^3. Localization inference is based on standardized values of the current density estimates. The solution space of s-LORETA is restricted to cortical and hippocampal and amygdala gray matter, defined via a reference brain from the Montreal Neurological Institute (MNI). The s-LORETA implementation incorporates a 3-shell spherical head model registered to a recognized anatomical brain atlas. MNI coordinates were translated to Talairach coordinates, by Talairach Daemon, according to the spatial association between anatomical brain landmarks and scalp position (Pascual-Marqui, 2002). Compared to dipole-based methods, s-LORETA has the advantage of estimating activity sources without any *a priori* assumptions regarding the number of sources, or their location. The sLORETA software package was used to perform the statistical analysis. The methodology used is non-parametric. It is based on estimating, via randomization, the empirical probability distribution

for the max-statistic (e.g. the maximum of a *t* statistic), under the null hypothesis. This methodology corrects for multiple testing (i.e. for the collection of tests performed for all voxels and time samples). Due to the non-parametric nature of the method, its validity need not rely on any assumption of Gaussianity (Nichols and Holmes, 2002). Source estimations were performed on single-subject vMMR, to determine the likely regions that significantly differ when observing deviant self-images or deviant other images at different levels of error (i.e. 33% *versus* 100%, 33% *versus* 66%, 66% *versus* 100% error). This analysis was undertaken in the time windows where deviant images significantly modulated the mean amplitude of difference ERPs.

Results

Behavioural performance

The hit rates for correctly identifying the target self-images (wearing sunglasses) were 89.13 % (SD = 9.54) and 88.79 % (SD = 7.07), in the familiar and the unfamiliar sequences respectively. The hit rates for the (target) images of others were 85.77 % (SD = 9.54) for the familiar other; and 90.44% (SD = 7.85) for the unfamiliar other. Repeated-measures ANOVA for hit rates revealed an interaction between facial image (self vs. other) and familiarity ($F(1,15) = 7.03$, $p = .018$). There were significant differences between the other's image when presented in the familiar sequence compared to when it was presented within the unfamiliar sequence, ($t(15) = -2.73$, $p = .015$). No differences were found when comparing self-images in the familiar compared with the unfamiliar sequence, ($t(15) = .13$, $p = .893$). The mean reaction times for the self-image targets were 435.36 ms (SD = 51.71) for familiar others; and 355.30 ms (SD = 42.73) for unfamiliar others. For images of others, the relevant mean was 436.86 ms (SD = 56.70) for familiar others; and 397.79 ms (SD = 42.73) for unfamiliar others. Repeated-measures ANOVA for RTs showed main effects of the facial (self vs. other) image

316 ($F_{(1,15)} = 97.20$, $p < .001$) and of familiarity ($F_{(1,15)} = 12.74$, $p = .003$). There was also an
317 interaction of facial image (self vs. other) with familiarity ($F_{(1,15)} = 60.63$, $p < .001$).
318 *vMMR to deviant self-faces*
319 Figure 2 shows grand-average ERPs elicited by deviant and non-deviant self and other faces,
320 at posterior-temporal, central and frontal ROIs. Both stimuli evoked the canonical P1,
321 N1/N170, P2 and P3 components. For deviant self-images, we found a positive deflection in
322 the difference ERPs at posterior-temporal sites, and a negative deflection in the difference
323 ERPs at central and frontal sites. By contrast, for deviant other faces we observed a deflection
324 in the difference ERPs at the right posterior-temporal sites and no clear changes at the left
325 posterior-temporal, the central or the frontal sites (with the exception of some amplitude
326 differences in the left posterior-temporal and central ROIs in the 33% error condition). We
327 performed repeated-measures ANOVA on the mean amplitudes of the difference ERPs in
328 right and left posterior-temporal ROIs and separately in central and frontal ROIs. Factors
329 comprised facial image (self, other), error level (33%, 66%, 100%), familiarity (familiar,
330 unfamiliar) and channel (4 levels).
331 In the 100-130 ms time window, the results showed a main effect of the level of error ($F_{(2,30)}$
332 $= 4.84$, $p = .015$); an error level X channel interaction (*F Greenhouse-Geisser* $_{(6,90)} = 6.32$, p
333 $= .001$); and a three way interaction of error level X channel X ROI ($F_{(18,270)} = 4.75$, $p <$
334 .001) interaction. Results also revealed a three way interaction of facial image X familiarity X
335 error level ($F_{(2,30)} = 4.05$, $p = .028$). However, follow-up analysis, with data collapsed over
336 channels in the four ROIs (conducted in self and other trials separately), show neither
337 significant main effects nor any significant interactions for the 'deviant self' condition (Facial
338 image = $F_{(1,15)} = 3.58$, $p = .078$; Error level = $F_{(2,30)} = 3.55$, $p = .041$; Facial image X Error
339 level = $F_{(2,30)} = 3.03$, $p = .063$). Likewise, there were neither main effects nor any significant

interactions for the 'deviant other' condition (Facial image = $F(1,15) = 5.95$, $p = .028$; Error level = $F(2,30) = 1.98$, $p = .156$; Facial image X Error level = $F(2,30) = 0.89$, $p = .418$). Analysis performed on the 100-130 ms time window also revealed a significant facial image X familiarity interaction ($F(1,15) = 7.75$, $p = .014$); and a facial image X familiarity X ROI interaction (*F Greenhouse-Geisser* $(3,45) = 4.66$, $p = .033$). Given that the factor familiarity did not interact with the factor error level, we described the details of the follow-up analysis with regard to the effects of familiarity in the supplementary material.

Analysis performed on the 170-300 ms time window revealed a main effect of facial image ($F(1,15) = 7.41$, $p = .016$); a facial image X channel interaction (*F Greenhouse-Geisser* $(3,45) = 5.32$, $p = .017$); a facial image X ROI interaction (*F Greenhouse-Geisser* $(3,45) = 8.56$, $p = .007$); and a facial image X channel X ROI interaction (*F Greenhouse-Geisser* $(9,135) = 3.83$, $p = .005$). Interestingly, we found a facial image X error level interaction ($F(2,30) = 6.34$, $p = .005$); and facial image X error level X ROI (*F Greenhouse-Geisser* $(6,90) = 13.89$, $p < .001$) interaction. This indicates a modulation of the vMMR that is dependent upon the amount of error in the deviant image (Figures 2 and 3). In view of the interactions involving facial image and error level but not familiarity or channel, we computed separate ANOVAs for the self and other trials on the averaged signal across channels at each ROI, collapsing across familiar and unfamiliar trials. Factors comprised error level (33%, 66%, 100%) and ROI (right/left posterior-temporal, central, frontal).

In the 'deviant self' condition, we found a main effect of error level ($F(2,30) = 7.74$, $p = .002$); and an error level X ROI interaction (*F Greenhouse-Geisser* $(6,90) = 14.50$, $p < .001$). Follow-up ANOVAs, with the factor 'error level' (33%, 66%, 100%), showed a main effect of error level in all four ROIs. In the posterior-temporal ROIs on the right ($F(2,30) = 29.25$, $p < .001$) and left ($F(2,30) = 10.10$, $p < .001$). In the central ROI ($F(2,30) = 11.16$, $p < .001$). In the frontal ($F(2,30) = 7.89$, $p = .002$) ROIs. We performed follow up *t* tests comparing the three levels of

deviance separately for each ROI. We found significant differences between 33% and 100% in all four ROIs. In the right posterior-temporal ROI (33% = 1.18 μV, 100% = 0.13 μV, $t(15)$ = -6.80, $p < .001$). In the left posterior-temporal ROI (33% = 1.04 μV, 100% = 0.16 μV, $t(15)$ = -3.80, $p = .002$). In the central ROI (33% = -0.82 μV, 100% = -0.20 μV, $t(15) = 3.84$, p = .002). In the frontal ROI (33% = -0.42 μV, 100% = .092 μV, $t(15) = 4.12$, $p = .001$).

When we then compared the 66% and 100% level of error we also found significant differences in all four of the four ROIs. In the right posterior-temporal ROI (66% = 0.76 μV, $t(15) = -5.46$, $p < .001$). In the left posterior-temporal ROI (66% = 0.63 μV, $t(15) = -3.11$, p = .007). In the central ROI (66% = -0.49 μV, $t(15) = 3.13$, $p = .007$). In the frontal ROI (66% = -0.30 μV, $t(15) = 2.91$, $p = .011$). Moreover, when we compared the 33% and 66% level of error and we found significant differences in the right posterior-temporal ROI ($t(15) = -2.95$, p = .009). However, there were not significant differences between 33% and 66% level of error in the left posterior-temporal ROI ($t(15) = -2.09$, $p = .054$), nor in the central ROI ($t(15) = 2.51$, $p = .024$), nor in the frontal ROI ($t(15) = 0.13$, $p = .412$). Interestingly, the 33% of error exhibited the largest vMMR when compared with the 66% or 100%.

When we then analysed the other-deviant condition, we did not find neither an effect of the factor error level ($F(2,30) = 0.45$, $p = .637$), nor interaction with ROI (*F Greenhouse-Geisser* $(6,90) = 3.74$, $p = .025$). This indicates that the 'deviant other' stimuli did not elicit a significant vMMR response when compared with 'deviant self-images'.

In addition, analysis of the 170-300 ms time window showed that the factor facial image (self vs. other) interacted with familiarity ($F(1,15) = 20.67$, $p < .001$); also with familiarity X ROI (*F Greenhouse-Geisser* $(3,45) = 11.86$, $p = .002$); and with familiarity X ROI X channel (*F Greenhouse-Geisser* $(9,135) = 3.13$, $p = .040$). Given the lack of interaction of the factor familiarity with the factor error level, the details of the follow-up analysis with regard to the effects of familiarity are described in the supplementary material.

Lastly, analysis performed in the 300-400 ms time window revealed a main effect of facial image ($F_{(1,15)} = 5.32$, $p = .036$) as well as the following interactions: facial image X ROI (*F Greenhouse-Geisser* $_{(3,45)} = 6.89$, $p = .010$); facial image X channel (*F Greenhouse-Geisser* $_{(3,45)} = 21.18$, $p < .001$); and facial image X ROI X channel (*F Greenhouse-Geisser* $_{(9,135)} = 7.77$, $p < .001$). Furthermore, facial image interacted with error level and ROI (*F Greenhouse-Geisser* $_{(6,90)} = 4.16$, $p = .021$). However, follow-up ANOVAs with factors comprising error level (33%, 66%, 100%) and ROI (right/left posterior-temporal, central, frontal), performed on the averaged signal across channels at each ROI, did not show a significant main effect of error level or an interaction with ROI for deviant self (error level = *F Greenhouse-Geisser* $_{(2,30)} = 1.27$, $p = .288$; error level X ROI = *F Greenhouse-Geisser* $_{(6,90)} = 2.25$, $p = .087$); or deviant other images (error level = $F_{(2,30)} = 1.20$, $p = .314$; error level X ROI = *F Greenhouse-Geisser* $_{(6,90)} = 2.73$, $p = .070$).

Moreover, analysis performed in the 300-400 ms time window revealed the following significant interactions: facial image X familiarity ($F_{(1,15)} = 23.78$, $p < .001$); facial image X familiarity X ROI (*F Greenhouse-Geisser* $_{(3,45)} = 7.17$, $p = .009$); facial image X familiarity X channel (*F Greenhouse-Geisser* $_{(3,45)} = 6.20$, $p = .017$); facial image X familiarity X ROI X channel (*F Greenhouse-Geisser* $_{(9,135)} = 4.32$, $p = .010$). Given the lack of interaction of the factor familiarity with the factor error level, the details of the follow-up analysis with regard to the effects of familiarity are described in the supplementary material.

Overall, these results show that the presentation of deviant self-faces in an alternating sequence leads to a vMMR in the 170-300 ms time window at posterior-temporal, central and frontal sites. This is demonstrated by the significant differences in vMMR amplitude between regular and deviant self-faces (Figures 2 and 3). The pattern of interaction shown in figure 3 illustrates that the amplitude of the vMMR to deviant self-images is proportional to the degree of error in the image, with 33% error level leading to the greatest vMMR. Conversely,

415 although visual inspection of the data might suggest that there was a vMMR in the 'deviant
416 other' condition, statistical analysis revealed that 'deviant other faces' did not significantly
417 differ from 'regular other faces' in the alternating sequence, in middle or late latencies.
418 Results also show that self-faces and other faces presented in the unfamiliar *versus* the
419 familiar sequence led to differential vMMR in the right posterior-temporal channels; and
420 likewise self-faces presented in the familiar sequence differed from those in the unfamiliar
421 sequence for the central ROI. Similarly, we observed an effect of familiarity in self and other
422 faces in the right posterior-temporal ROI in the 300-400 ms time window (Supplementary
423 material). In essence, our results show that the processing of deviant self-faces in an
424 alternating sequence significantly differs from similar processing of deviant other faces, thus
425 demonstrating a self-specific pattern of automatic prediction errors.
426 *Current source density analysis*
427 Source localization was performed on the time windows where error level significantly
428 modulated mean vMMR responses to deviant self-faces (170-300 ms). This identified a set of
429 regions whose peak activity was maximal for 33% *versus* 100%, for 66% *versus* 100%, and
430 for 33% versus 66% deviant self-faces (Figure 3). When comparing deviant self-faces with
431 error magnitude 33% *versus* deviant self-images with error magnitude 100%, maximum
432 differential activity was source localized within the parietal association cortex [Brodmann
433 area (BA) 39, 40], in the insula [BA 13] and superior temporal gyrus [BA 22], in the right
434 hemisphere. When contrasting deviant self-faces with error magnitude 66% *versus* deviant
435 self-images with error magnitude 100%, a cluster of sources was found in the left fusiform
436 gyrus [BA 18, 19, 37], in visual cortex, and in the inferior temporal gyrus [BA 20]. When
437 computing the difference between deviant self-faces with error magnitude 33% *versus*
438 deviant self-images with error magnitude 66%, a cluster of sources was found in the right

fusiform gyrus [BA 18, 19], parietal association cortex [BA 39], and the cingulate cortex [BA 31].

Second experiment

Current theories of the self have suggested that one's own face is considered among the most representative feature of the self. Therefore the majority of the studies on the self have focused on self-face processing (e.g. Northoff et al., 2006; Devue et al, 2007). Importantly, however, these studies tend to present self and other's faces in isolation (i.e. as disembodied heads) whereas the recognition of one's image in natural contexts requires the processing of one's own face integrated in one's body, in a holistic manner. A series of studies have investigated the processing of one's own body and body parts. They have shown that the recognition of the self-body differs from the recognition of others' bodies, as evidenced by enhanced cortical activity in visual and multisensory body areas (e.g. cingulate gyrus, insula). There is also better performance in response to self-body as opposed to others' body images (Devue et al., 2007; Frassinetti et al., 2008). However, evidence is lacking on how self-face and self-body are integrated and processed together as one self-image. Moreover, what the relative importance is of the face, in relation to the body, in the process of self-recognition is under-researched.

In experiment 2 we tested this issue by extending the paradigm used in experiment 1 by showing participants alternating sequences of images of the self-body and another's body. Specifically, we used the vMMR to investigate automatic context based predictions of the relative strength of self-face and self-body mental representations.

Material and methods

Participants

Eighteen neurologically unimpaired paid participants took part in experiment 2. One participant was excluded from the analysis because of inability to complete the task, resulting

on a total of 17 participants (5 males, mean age 21.58 years; laterality quotient 88.52% (Oldfield, 1971). Participants gave their informed consent, with approval by the Ethics Committee, Department of Psychology, Royal Holloway University of London.

Stimuli and procedure

Four grey-scaled whole-body images (238 x 575 pixels) were centrally presented on a black background, using E-prime software (Psychology Software Tools). Visual stimuli included the participant whole-body image; the image of another unfamiliar individual who was matched for age, gender and body size; and 2 composed images, one of which contained the participant's face superimposed on the other's body and the other containing the other's face superimposed on the participant's body. All individuals were dressed in standardised clothing comprising white, cropped vest and black shorts, excluding jewellery and any other clothing (Figure 4). Images were edited with Photoshop software (Adobe Systems, San Jose, CA) and were equalised for luminosity, contrast, shadows, highlights, colour and image size. Participants' photographs were taken in a separate session prior to the experimental session. Images of unfamiliar others were of four individuals (2 males, 2 females), selected from our in-house database, that had never been seen by the participants, prior to the experiment. Additionally, four target stimuli were constructed by adding either a black belt or a black sleeping mask to the whole-body images.

Similarly to experiment 1, stimuli were presented in a sequence (stimuli duration 250 ms, stimulus onset asynchrony 600 ms) where the participant's whole-body image was alternated with the whole-body image of another person. The percentage of regular trials of self and other's images presented in the sequence was 81.4%. The consistency of the alternating sequence was irregularly violated by deviant images that were associated either with the self-image (on half the irregular trials) or with the image of the other person. There were three types of deviant images that were randomly presented through the sequence (i.e. in 18.6% of

trials). One-third of the irregular trials included deviant whole-body images, which were images of the self when an image of the other was expected (or images of the other when an image of the self was expected). A further third of irregular trials were deviant face images, which contained the self-face superimposed on other's body when the whole-body image of the other was expected (or vice versa). The remaining third were deviant body images, which were images containing the other's face superimposed on the self-body when the whole body image of the other was expected (and vice versa). Target stimuli occasionally (p = .093) replaced the non-target stimuli. Participants were asked to ignore all other stimulus attributes (i.e. identity) and to press a button as accurately and quickly as possible whenever the target stimuli were presented. The experiment consisted of 2588 trials, including 240 deviant trials (80 deviant whole-body images, 80 deviant face images, 80 deviant body images).

Participants were seated in a dimly lit sound-attenuated and electrically shielded chamber in front of a monitor (Samsung SyncMaster 940N; size = 21 inches; resolution = 1280 X 1024) at a distance of 90 cm.

Behavioural performance

Participants' behavioural performance was measured using the same procedure described in experiment 1. The regular and irregular trials were collapsed, resulting in 2 conditions (self, other). Differences in hit rate (%) and reaction time (ms) between conditions were tested by a pairwise *t* test.

EEG recording and data analysis

EEG data recording and pre-processing were identical to Experiment 1. The same method and criteria were used for filtering, ocular correction and artifact rejection (the mean percentage of data rejected due to excessive amplitude was 0.14%) (see Supplementary table 2 for percentage of trials included in the analysis). In experiment 2 the single-subject ERPs

were computed for the following factors: image (self, other) and expectancy (expected, deviant whole-body image, deviant face image, deviant body image).

To estimate the effects of violations in the regular sequence, single-subject averages of irregular minus regular difference waves (difference ERPs) were calculated as follows. For the self-related deviant conditions, the regularity violation effects were obtained by subtracting ERPs elicited by regular self-image trials from: (i) the ERPs elicited by deviant self-whole-body image trials ('deviant self whole-body image'); (ii) the ERPs elicited by deviant self-face image trials ('deviant self face image'); and (iii) the ERPs elicited by deviant self-body image trials ('deviant self body image'). For the other related deviant conditions, the regularity violation effects were obtained by subtracting ERPs elicited by regular other image trials from: (i) the ERPs elicited by deviant other's whole-body image trials ('deviant other's whole-body image'); (ii) the ERPs elicited by deviant other's face image trials ('deviant other's face image'); and (iii) the ERPs elicited by deviant other's body image trials ('deviant other's body image') (Supplementary table 4).

The difference in activity was averaged across participants and contrasted for deviant whole-body image, deviant face image, and deviant body image, for the self and other, in the time windows 110-130 ms, 220-320 ms and 320-400 ms (e.g. Kimura et al., 2012; Stefanics et al., 2012). In line with standardised procedure (e.g. Gosling and Eimer, 2011; Stefanics et al., 2012), four regions of interests (ROIs) were defined on the basis of difference potential maps, including those channels in which experimental effects could be predicted, based on previous vMMR literature (Astikainen and Hietanen, 2009; Stefanics and Czigler, 2012; Stefanics et al., 2012). There were two (right and left) posterior-temporal ROIs (P7/8, P9/10, PO7/8 and O1/2 electrodes of the 10/20 system); a central ROI (C1/2, Cz, FCz); and a frontal ROI (AFz, Fz, F1/2) (Figure 4). Factors of the analysis comprised: image (self, other); error level (deviant whole-body image, deviant face image, deviant body image); and ROI (right/left

posterior-temporal, central, frontal) and channel (4 levels). Mauchly's W was computed to check for violations of the sphericity assumption and the Greenhouse-Geisser adjustments to the degrees of freedom were applied when needed. The p values were corrected for multiple comparisons using stepwise Bonferroni-Holm correction.

Current source density analysis

Current source density analysis was identical to Experiment 1, with the exception that the source estimations were performed within the time windows where deviant images (i.e. deviant whole-body image, deviant face image, deviant body image) significantly modulated mean amplitudes of difference ERPs (i.e. vMMR), independently for the self and the other conditions.

Results

Behavioural performance

The hit rate for the self-condition was 89.55% (SD = 12.17) and for the other condition 88.69 % (SD = 11.33). *t* tests showed no difference between the hit rates for the 'self' compared with the 'other' condition ($t(16) = -.71$, $p = .488$). The mean reaction time for the self-condition was 440.11 ms (SD = 74.59); and 438.90 ms (SD = 29.58) for the other condition. There were no significant differences in reaction time between the self and other images ($t(16) = -.14$, $p = .890$).

vMMR to deviant self-images

Figure 5 shows grand average ERPs elicited by deviant standard self and deviant other whole-body images at posterior-temporal, central and frontal ROIs. Both stimuli evoked the canonical P1, N1/ N170, P2 and P3 components. For deviant self-images, we found a positive deflection in the difference ERPs at posterior sites. By contrast, for deviant other faces, we observed a negative deflection in the difference ERPs at posterior sites. We performed separate repeated-measures ANOVAs on the mean amplitudes of difference ERPs in

posterior-temporal ROIs. Factors comprised image (self, other); error level (deviant whole-body image, deviant face image, deviant body image); and hemisphere (left ROI, right ROI). In the central ROI, the ANOVA was performed with factors comprising image (self, other) and error level (deviant whole-body image, deviant face image, deviant body image). Results of the ANOVA performed in the 220-320 ms time revealed a main effect of the factor image ($F_{(1,16)} = 57.40$, $p < .001$); and an interaction image X ROI (*F Greenhouse-Geisser* $_{(3,48)} = 23.58$, $p < .001$). There was also an interaction image X channel (*F Greenhouse-Geisser* $_{(3,48)} = 4.95$, $p = .019$); and a three-way interaction image X ROI X channel (*F Greenhouse-Geisser* $_{(9,144)} = 3.43$, $p = .012$). Moreover, results showed a three-way interaction of image X error level X ROI (*F Greenhouse-Geisser* $_{(5,96)} = 3.33$, $p = .029$). This suggests a modulation of the vMMR by the level of error in the deviant image (Figures 5 and 6). We computed separated ANOVAs for the self and the other condition, with factors comprising error level (deviant whole-body image, deviant face image, deviant body image) and ROI (right/left posterior-temporal, central, frontal).

In the 'deviant self' condition, there was a main effect of the factor 'error level' ($F_{(2,32)} = 6.07$, $p = .008$) as well as an interaction of error level X ROI (*F Greenhouse-Geisser* $_{(6,96)} = 5.00$, $p = .005$). We then computed four ANOVAs at each ROI separately, and we found a main effect of error level in the right ($F_{(2,32)} = 5.65$, $p = .008$) and left ($F_{(2,32)} = 5.96$, $p = .006$) posterior-temporal ROIs. In contrast, we did not find main effects at the central ROI ($F_{(2,32)} = 3.09$, $p = .059$), nor the frontal ROI ($F_{(2,32)} = 3.64$, $p = .037$). Interestingly, *t* tests contrasting the three levels of error, in the 220-320 ms time window, demonstrated significant differences between deviant whole-body image and deviant face image for the right posterior-temporal ROI (whole body = 1.42 μV, face = 0.39 μV, $t_{(16)} = 2.95$, $p = .009$); and for the left posterior-temporal ROI (whole body = 1.43 μV, face = 0.67 μV, $t_{(16)} = 2.93$, $p = .010$).

When we then compared deviant whole-body image and deviant body image, we found significant differences for the right posterior-temporal ROI (body = 0.28 μV, $t(16) = 3.02$, $p = .008$); and for the left posterior-temporal ROI (body = 0.48 μV, $t(16) = 2.85$, $p = .012$). However there were no significant differences between deviant face and deviant body image for the right posterior-temporal ROI ($t(16) = 0.320$, $p = .753$), or for the left posterior-temporal ROI ($t(16) = .69$, $p = .500$).

By contrast, in the 'deviant other' condition, there was no effect of error level ($F(2,32) = 0.65$, $p = .937$) nor any interaction with ROI (*F Greenhouse-Geisser* $(6,96) = .336$, $p = .764$).

In the 320-400 ms time window, results showed a main effect of image ($F(1,16) = 18.96$, $p < .001$); as well as an interaction of image X ROI (*F Greenhouse-Geisser* $(3,48) = 20.34$, $p < .001$); an interaction of image X channel ($F(3,48) = 4.23$, $p = .010$); and also an interaction of image X ROI X channel (*F Greenhouse-Geisser* $(9,144) = 2.46$, $p = .40$). This indicates significant differences in vMMR response between the self and the deviant other images, across sites. However, no main effect of error level, nor any interaction with image, was observed in the later time window. Moreover, there were no significant main effects, nor any interactions, with the factors 'image' or 'error level', for the 110-130 ms time window.

In summary, the results of the experiment 2 show that deviant self-images, presented in an alternating sequence, are associated with a vMMR in the mid-range latency (220-320 ms time window) of error processing, at posterior-temporal sites (Figures 5, 6). The amplitude of the vMMR to deviant self-images was proportional to the type of error in the deviant image, exhibiting greater amplitude to deviant whole-body images of the self. Similarly to experiment 1, the effect of 'deviant other' images did not significantly differ from the effect of 'regular other' images, in the alternating sequence, in middle or late latencies, even though visual inspection of vMMR might suggest a difference. Overall, the results of experiment 2 demonstrate that the processing of deviant self-images, as opposed to deviant other images,

when presented in the context of a temporal sequence, leads to a self-specific automatic prediction error response. This supports the results of the experiment 1.

Current source density analysis

Source localization was performed on the time window where the factor 'error level' significantly modulated mean vMMR responses to deviant self-images (i.e. the 220-320 ms time window). It was defined by that set of neural regions whose peak activity was maximal for (i) deviant self whole-body images *versus* deviant self-face images; (ii) deviant self whole-body images *versus* deviant self-body images; (iii) deviant self-face images *versus* deviant self-body images (Figure 6). When comparing deviant self whole-body images *versus* deviant self-face images, maximum differential activity was source localized within the left fusiform gyrus [BA 17, 18] and the cingulate cortex [BA 23, 30, 31], in the left hemisphere. The contrast between deviant self whole-body images *versus* deviant self-body images revealed source-localized activity within the insula [BA 13], the parietal association cortex [BA 40, 41], the postcentral gyrus [BA 2] and the cingulate cortex [BA 31], in the left hemisphere. When contrasting the difference between deviant self-face images *versus* deviant self-body images, a cluster of sources was found in the left precentral gyrus [BA 6] and postcentral gyrus [BA 1, 2, 3, 4], in sensory and motor areas.

Discussion

We investigated the neural signatures of automatic temporal context-based predictions of self-related or other-related visual stimuli in the brain. These stimuli were faces in Experiment 1, and were faces and bodies in Experiment 2. Our results showed that self-related stimuli that violated the regularities of sequences of self-other images, elicited a vMMR of positive polarity, while deviant images of others did not give rise to any vMMR. Moreover, the amplitude of the vMMR to deviant self-images was proportional to the degree of error in the image, so that images that differed most from the mental representations of

one's self led to the greater vMMR. This effect was source localized within visual and multimodal associative areas including frontal, cingulate and occipital cortices. Overall, our results provide novel evidence for an automatic detection of visual changes in self-related but not in other related visual stimuli. This is consistent with the theory that self-processing takes advantage of the brain's automatic prediction error system and accounts for self-bias in visual processing (Apps and Tsakiris, 2014).

Experiment 1 showed that deviant self-faces led to enhanced cortical responses when compared with regular (the expected) self-faces. By contrast, cortical responses to deviant other faces did not significantly differ from responses to regular other faces. As indicated by vMMR to deviant self-faces, these results suggest that, on the basis of a temporal sequential context, self-related but no other-related information is automatically predicted. Evidence for prediction errors in the processing of face-related changes comes from previous ERP studies showing vMMN responses to changes in both facial identity (Susac et al., 2004) and facial emotions (Astikainen and Hietanen, 2009; Kimura et al., 2012; Stefanics et al. 2012). Moreover, the processing of self-faces is associated with early changes in visual cortical signals, which supports the hypothesis that there are self-specific mechanisms in the human brain (e.g. Keyes et al., 2010; Gosling and Eimer, 2011). The results of experiment 1 complement and advance previous findings on the cortical mechanisms of self-face processing by showing self-specific automatic predictions to visual changes of self-faces but not to other faces (Northoff et al., 2006; Apps and Tsakiris, 2014). Additionally, we demonstrate that self-specific prediction errors cannot be attributed to the effects of familiarity (Platek et al, 2006) since the vMMR occurs irrespectively of whether the other's image is of a familiar or unfamiliar person. Furthermore, experiment 2 demonstrates that not only deviant self-faces but also deviant whole-body images of the self but not of others, are associated with prediction error responses as indexed by vMMR. These findings extend

previous studies on automatic processing of body parts (Stefanics et al., 2012) demonstrating neural self-specificity of facial and bodily information.

A possible explanation for the automatic prediction error responses for the deviant images of the self but not the other might be the existence of a self-specific mismatch detection mechanism in the brain, that allows us to detect self-related but no other related information. Alternatively, it is possible that self-specific prediction errors are associated with a rapid orienting response to highly salient self-stimuli (Folstein and Van Pettern, 2008). These hypotheses are difficult to tease apart within the context of this study. However, they might be mutually compatible and reinforcing, in the sense that the mismatch detection of deviant information tends to be amplified when mismatching images are a significant event, such as deviant self-images. Importantly, however, the results of the study show that both deviant self-face and self-body images result in a similar unique patterns of vMMR, whereby the cortical amplitude of response depends on the level of deviance in the image.

Converging evidence has shown that amplitude of the vMMN relies heavily on the differences between the current and the predicted image, generated on the basis of contextual regularities (Winkler and Czigler, 2012). Consistent with this, we observed that the vMMR amplitude was proportional to the magnitude of the deviance in the self-images. In experiment 1, there were two identical morphed deviant images, in both the self-related and the other related conditions (i.e. containing 33% and 66% error). Importantly, only in the self-related condition did these deviant images lead to mismatch responses. This suggests that the self-images processed as deviant images in the sequence only when participants were expecting an image of the other. More precisely, it seems that the deviant self-images not only violated the expectations regarding the regularities of the sequence but also violated the expectations regarding the mental representation of the self-face. Thus, the morphed deviant images that most differed from the cortical representation of the self (i.e. morphed images

688 with 33% error) led to the greatest mismatch responses. Generally, our results support the
689 self-related bias of deviant-related processing of faces. This effect is clearly observable in the
690 100% comparisons which are actually reverse control conditions. One should notice,
691 however, that the greater vMMR to the deviant morphed images as opposed to the 100%
692 error images can be associated not only to stimulus-related effect (i.e., the physical
693 differences between the self and the other facial features), but also to probability-related
694 effects (i.e., the lower overall probability of occurrence of the morph *vs.* the 100% images in
695 the sequence). These hypothesis are difficult to disentangle within the context of the current
696 study, and they could be explored in more detail in the future using an additional
697 equiprobable condition that would control for stimulus-related and refractory effects.
698 Experiment 2 demonstrated that deviant whole self-body images were associated with greater
699 mismatch responses, as compared with composite images where only the face (in deviant
700 self-face images), or the body (in deviant self-body image), did not match the expected
701 information. Overall, these results support the idea that the effects of deviance from
702 expectation are highly dependent on visual feature matching between the predicted and the
703 actual event.
704 According to models of self-processing, one's own face is among the most highly
705 representative features of the self (e.g. Northoff et al., 2006). In line with this idea, one would
706 have expected greater mismatch responses to deviant self-faces than deviant self-body
707 images. However, the results of the experiment 2 show no difference between deviant self-
708 face and self-body images. This suggests, in visual processing, that the strength of the mental
709 representation of one's own face is comparable to the strength of the representation of one's
710 self-body. This implies that both face and body are equally relevant in early stages of self-
711 processing. Our findings support and extend previous studies on self-biases in the processing

712 of whole-body images and body parts (Devue et al., 2007; Frassinetti et al., 2008) and
713 support the importance of holistic mechanisms in visual processing of oneself.
714 Interestingly, the results of the current study showed that both deviant self-face and whole-
715 body images are associated to enhanced positive responses, in comparison with regular self-
716 images. Although visual inspection of the data in the other-related condition seems to suggest
717 the presence of vMMR of negative polarity (i.e. vMMN, Pazo-Alvarez et al., 2003; Kimura,
718 2012), this effect was not significant. Our results thus contrast with several studies that have
719 indicated that cortical activity is dominantly negative over the posterior locations in the
720 vMMR latency range (e.g. Czigler, 2007). It should be noted, however, that in the visual
721 domain the cortical architecture of exogenous visual potentials is highly complex and
722 variable. In particular, the latency and polarity of the early visual ERPs rely on the spatial
723 orientation of their underlying dipolar sources which, in turn, depends on the folding
724 structure of the neural source area and its location relative to the recording electrodes (Di
725 Russo et al., 2002; Stefanics et al., 2014). Considering the cortical complexity of the visual
726 areas, the polarity reversal we found for the visual mismatch positivity response to deviant
727 self-images might suggest that self-related information is represented in anatomically
728 different cortical areas, within the extrastriate visual cortex, from information related to other
729 people (Northoff et al., 2006; Berlucchi and Aglioti, 2010). Alternatively, the positive
730 VMMR could be associated with the differences in VEPs between self and other images. Past
731 studies have shown that in comparison to other images, self-images lead to an enhancement
732 of the N250 (Keyes et al, 2010). Therefore the subtraction of self regular from the self
733 irregular could result in a positive mismatch response around the time window of the N250.
734 Although further investigations into the visual cortical properties of self-related automatic
735 predictions are required, the polarity reversal nature of the vMMR to deviant self-images that

736 was observed in the current study contributes to the idea of brain specialization for self-
737 processing and extends current knowledge about the nature of the visual mismatch response.
738 For both deviant self-face and whole-body images, the neural sources of the vMMR were
739 localized in the fusiform gyrus and in the superior/inferior temporal gyri. These areas have
740 previously been linked to the cortical generation of vMMR (Yucel et al., 2007; Kimura et al.,
741 2010, 2012). They are associated with the analysis of low-level facial features as well as high-
742 level facial information such as identity (Haxby et al., 2000). In addition, source-localised
743 activity associated with deviant self-images was observed in the insula and cingulate cortex.
744 The involvement of limbic areas in the processing of deviant visual information has been
745 observed in vMMN studies on emotional faces (Kimura et al., 2012), as well as other visual
746 stimuli (Huettel et al., 2002) and also as in the process of self-images (Berlucchi and Agioti,
747 2010). Moreover, the neural generators of vMMR responses to deviant self-images were
748 associated with activations within the parietal associative cortex (for self face and whole-body
749 images), as well as within the precentral and postcentral gyrus (for self whole-body). The
750 parietal associative cortex is responsible for the integration of visual and sensorimotor
751 information, with a fundamental role in face and body processing (Berlucchi and Agioti,
752 2010). Sensorimotor areas, by contrast, have been largely thought to index the embodiment of
753 other's expressions (Blakemore et al., 1998, 2000; Sel et al., 2014). Taken together, our
754 results are highly consistent with previous studies on automatic processing of facial and
755 bodily images and they provide further evidence of the engagement of visual, limbic and
756 sensorimotor areas in the self-specific prediction error processing.
757 The uniqueness of cortical responses during self-processing in other modalities, such as voice
758 processing, has been previously tested in two electrophysiological studies (Graux et al, 2013,
759 2014). These studies reported no MMN modulation when contrasting self-voice to other-
760 voice stimuli. The differences between self- and other-voice only appeared at later stages of

761 processing, for example, P3a latency (Graux et al, 2012; 2014). In contrast, the present study
762 demonstrates the existence of early and automatic detection of visual changes in self-
763 information, extending current knowledge on visual self-processing and suggesting a
764 sensitivity of the vMMR to identification of the self as a unique individual. Comparison
765 across current and former studies is difficult because of the various methodologies and
766 differing modalities of the stimuli. Whereas Graux et al. (2013, 2014) employed an oddball
767 paradigm related to activation of memory traces, we employed alternating sequences
768 associated with the processing of automatic temporal based context predictions (Kimura et al.,
769 2012; Stefanics et al., 2014). Such discrepancies between our current and other researchers'
770 previous studies might therefore be accounted for by different mechanisms of error
771 processing.
772 Recent vMMR studies have related the automatic prediction responses to the predictive
773 coding (PC) hypothesis (Kimura et al., 2012; Stefanics et al., 2014). This is a unifying theory
774 of cortical function that explain mismatch signals, among many other phenomena. According
775 to this view, the sensory input is compared with internal models, which are constantly
776 updated by compiling the statistical regularities of past inputs (Friston and Kiebel, 2009;
777 Apps and Tsakiris, 2013). The vMMR has been associated with the encoding of sensory input
778 (surprise or error) leading to adjustment of existing probabilistic mental models (Kimura
779 2012; Stefanics et al., 2014). Novel theoretical proposals have suggested that self-processing
780 is characterized by the principles of PC (Fotopoulou, 2012; Limanowski and Blankenburg,
781 2013; Apps and Tsakiris, 2014; Moutoussis et al., 2014). The key premise of these models is
782 that self-identification relies on hierarchical generative self-representations that arise from
783 multisensory information and are constantly updated through the prediction and integration of
784 unimodal sensory information (i.e. own face/ body) in multimodal areas. In the context of the
785 current study, the sensory events (self-images and other images) are contrasted with various

competing models (i.e. the mental representation of the self and others as driven by the sequential presentation of our stimuli). Our finding of a self-specific vMMR suggests that only deviant self-images and their associated prior mental representations compete with the sequential mental model, by taking the form of bottom-up error signals that are explained away by top-down processes in order to minimise the level of surprise. Furthermore, the activations of visual, limbic/ associative and frontal sensorimotor areas to deviant self-images fit nicely with studies on PC models (Kimura et al., 2012, Lieder et al., 2014). Several have reported activation in medial and superior frontal areas, related to the generation of rule structures and to error-awareness (Hester et al., 2005). Our results also accord well with PC models of the self, where the cingulate cortex has been proposed to house the generation, comparison and update of predictions of bodily information (Tsakiris et al, 2007; Seth, 2014, Sel, 2014). We therefore argue that our results can be accounted for within the PC brain hypothesis, thus providing empirical support for a PC model of self-processing in the human brain, such that self-related but not other related information result in modulation of the vMMR.

In conclusion, this study provides novel evidence for automatic prediction responses to visual changes in self-images but not other images, and supports the idea of self-specificity in the human brain. We designed two experiments that investigated the cortical processing of facial and bodily images of the self and other, showing that deviant self-images elicited a vMMR whose amplitude was proportional to the error magnitude. This vMMR response was source localized in visual, limbic/associative and sensorimotor areas, which are brain regions associated with facial and bodily processing. No such effects occurred when deviant other's faces were presented. Overall, our findings provide novel evidence to show that the processing of self-images takes advantage of the automatic prediction error system in the brain, leading to self-biases in self-related information.

811

812 **Acknowledgments**

813 This work was supported by the European Research Council Starting Investigator Grant
814 (ERC-2010-StG-262853) to MT and the Neuropsychoanalysis Foundation. The authors thank
815 Dr. Dagmar Muller for her advice in data analysis and interpretation, and Dr. Vivien Ainley
816 for her help in proof-reading the manuscript

817 **References**

818 Apps MAJ, Tsakiris M. 2013. Predictive codes of familiarity and context during the
819 perceptual learning of facial identities. Nat Commun. 4: 2698 (doi:10.1038/ncomms3698)
820 Apps MA, Tsakiris M. 2014. The free-energy self: a predictive coding account of self-
821 recognition. Neurosci Biobehav Rev. 41C: 85-97 (doi: 10.1016/j.neubiorev.2013.01.029)
822 Astikainen P, Hietanen JK. 2009. Event-related potentials to task-irrelevant changes in facial
823 expressions. Behav Brain Funct, 5:30-39 (doi:10.1186/1744-9081-5-30)
824 Blakemore SJ, Wolpert DM, Friston CD. 1998. Central cancellation of self-produced tickle
825 sensation. Nat Neurosci, 1:635-640 (doi: 10.1038/2870)
826 Blakemore SJ, Smith J, Steel R, Johnstone EC, Frith CD. 2000. The perception of self-
827 produced sensory stimuli in patients with auditory hallucinations and passivity experiences:
828 evidence for a breakdown in self-monitoring. Psycho Med. 30:1131-1139 (doi:
829 http://dx.doi.org/)
830 Berlucchi B, Aglioti SM. 2010.The body in the brain revisited. Exp Brain Res, 1:25-35 (doi:
831 10.1007/s00221-009-1970-7)
832 Devue C, Collette F, Balteau E, Degueldre C, Luxen A, Maquet P, Brédart S. 2007. Here I
833 am: The cortical correlates of visual self-recognition. Brain Res, 1143: 169-182
834 (doi:10.1016/j.brainres.2007.01.055)
835 Di Russo F, Martínez A, Sereno MI, Pitzalis,S, Hillyard SA.2002. Cortical sources of the
836 early components of visual evoked potential. Hum. Brain Map. 15:95-111
837 (doi:10.1002/hbm.10010)
838 Folstein JR, Van Pettern CV. 2008. Influence of cognitive control and mismatch on the N2
839 component of the ERP: A review. Psychophysiol, 45:152-170 (doi: 10.1111/j.1469-
840 8986.2007.00602.x)

841 Fotopoulou A. 2012. Towards psychodynamic neuroscience. In: Fotopoulou A, Conway M,
842 Pfaff D, editor. From the Couch to the Lab: Trends in Psychodynamic Neuroscience. New
843 York (NY): Oxford UP
844 Frassinetti F, Maini M, Benassi M, Avanzi S, Cantagallo A, Farnè A. 2008. Selective
845 impairment of self body-parts processing in right brain-damaged patients. Cortex, 46:322-328
846 (doi:10.1016/j.cortex.2009.03.015)
847 Friston K. 2003. Learning and inference in the brain. Neural Netw. 16:1325-1352 (doi:
848 10.1016/j.neunet.2003.06.005)
849 Friston K, Kiebel S. 2009. Predictive coding under the free-energy principle. Philos Trans R
850 Soc B Biol Sci. 364:1211-1221 (doi:10.1098/rstb.2008.0300)
851 Gillihan SJ, Farah MJ. 2005. Is self special? A critical review of evidence from experimental
852 psychology and cognitive neuroscience. Psycho Bull.131:76-97 (doi:10.1037/0033
853 2909.131.1.76)
854 Gratton G, Coles MG, Donchin E. 1983. A new method for off-line removal of ocular
855 artifact. Electroencephalogr Clin Neurophysiol. 55:468-484 (doi:10.1016/0013-
856 4694(83)90135-9)
857 Gosling A, Eimer M. 2011. An event-related brain potential study of explicit face
858 recognition. Neuropsychologia. 49:2736-2745 (doi:10.1016/j.neuropsychologia.2011.05.025)
859 Graux J, Gomot M, Roux S, Bonnet-Brilhaut F, Camus V, Bruneau N. 2013. My voice or
860 yours? An electrophysiological study. Brain Topogr. 26:72-82 (doi: 10.1007/s10548-012-
861 0233-2)
862 Graux J, Gomot M, Rouz S, Bonnet-Brilhault F, Bruneau N. 2014. Is my voice just a familiar
863 voice? An electrophysiological study. Soc Cogn Affec Neurosci, 10:101-105 (doi:10.1093/
864 scan/nsu031)

865 Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face
866 perception. Trends Cogn Sci. 4:223-233 (doi:10.1016/S1364-6613(00)01482-0)
867 He J, Hu Y, Pakarinen S, Li B, Zhou Z. 2014. Different effects of alcohol on automatic
868 detection of colour, location and time change: A mismatch negativity study. J
869 Psychopharmacol. 28:1109-1114 (doi:10.1177/0269881114548294)
870 Hester R, Foxe JJ, Molholm S, Shpaner M, Garavan H. 2005. Neural mechanisms involved in
871 error processing: A comparison of errors made with and without awareness. NeuroImage.
872 27:602-608 (doi:10.1016/j.neuroimage.2005.04.035)
873 Huettel SA, Mack PB, McCarthy G. 2002. Perceiving patterns in random series: dynamic
874 processing of sequence in prefrontal cortex. Nat Neurosci. 5:485-490 (doi: 10.1038/nn841)
875 Lieder F, Stephan KE, Daunizeau J, Garrido MI, Friston KJ. 2014. A neurocomputational
876 model of the mismatch negativity. Plos Compt Biology. 9(12): e10.1371 (doi:10.1371/
877 journal.pcbi.1003288)
878 Limanowski J, Blankenburg F. 2013. Minimal self-models and the free energy principle.
879 Front Hum Neurosci. 7:547 (doi: 10.3389/fnhum.2013.00547)
880 Keyes H, Brady N, Reilly RB, Foxe JJ. 2010. My face or yours? Event-related potential
881 correlates of self-face processing. Brain Cogn. 72: 224-254 (doi:10.1016/j.bandc.2009.09.
882 006)
883 Kimura M. 2012. Visual mismatch negativity and unintentional temporal-context-based
884 prediction in vision. Int J Psycho. 83:144-155 (doi: 10.1016/j.ijpsycho.2011.11.010)
885 Kimura M, Kondo H, Ohira H, Schoger E. 2012. Unintentional temporal context-based
886 prediction of emotional faces: an electrophysiological study. Cereb Cortex. 22:1774-1785
887 (doi: 10.1093/cercor/bhr244)

888 Moutoussis M, Fearon P, El-Deredy W, Dolan RJ, Friston KJ. 2014. Bayesian inferences
889 about the self (and others): A review. Consc Cognition. 25:67-76 (doi: 10.1016/j.concog.
890 2014.01.009)
891 Näätänen R, Paavilainen P, Rinne T, Alho K. 2007. The mismatch negativity (MMN) in basic
892 research of central auditory processing: a review. Clin. Neurophysiol. 118:2544–2590 (doi:
893 10.1016/j.clinph.2007.04.026)
894 Nichols TE, Holmes AP. 2002. Nonparametric Permutation Tests for Functional
895 Neuroimaging: A Primer with Examples. Hum Brain Map. 15:1-25 (doi: 10.1002/hbm.1058)
896 Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J. 2006. Self-
897 referential processing in our brain: A meta-analysis of imaging studies on the self.
898 NeuroImage. 31:440-457 (doi:10.1016/j.neuroimage.2005.12.002)
899 Pascual-Marqui RD. 2002. Standardized low-resolution brain electromagnetic tomography
900 (sLORETA): technical details. Methods Find Exp Clin Pharmacol. 24 [Suppl D]: 5-12
901 Platek SM, Loughead JW, Gur RC, Busch S, Ruparel K, Phend N, Panyavin IS, Langleben
902 DD. 2006. Neural substrates for functionally discriminating self-face from personally familiar
903 faces. Hum Brain Mapp. 27:91-98 (doi:10.1002/hbm.20168)
904 Pannese A, Hirsch J. 2010. Self-specific priming effect. Conscious Cogn. 19:962-968 (doi:
905 doi:10.1016/j.concog.2010.06.010)
906 Pazo-Alvarez P, Cadaveira F, Amenedo E. 2003. MMN in the visual modality: a review.
907 Biol. Psychol. 63:199–236 (doi:10.1016/S0301-0511(03)00049-8)
908 Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh Inventory.
909 Neuropsychologia. 9:97-113 (doi:10.1016/0028-3932(71)90067-4)
910 Sel A, Forster B, Calvo-Merino B. 2014. The emotional homunculus: ERP evidence for
911 independent somatosensory responses during facial emotional processing. J Neurosci.
912 34:3263-3267 (doi: 10.1523/jneurosci.0106-13.2014)

913 Sel A. 2014. Predictive codes of interoception, emotion, and the sel. Front Psychol. 5:189
914 (doi: 10.1016/j.tics.2013.09.007)
915 Seth AK. 2013. Interoceptive inference, emotion, and the embodied self. Trends Cogn Sci.
916 17:565-573 (doi:10.1016/j.tics.2013.09.007)
917 Sui J, Minghui L, Mevorach C, Humphreys GW. 2015. The salient self: the left intraparietal
918 sulcus responds to social as well as perceptual-salience after self-association. Cereb Cortex,
919 25:1060-1068 (doi: 10.1093/cercor/bht302)
920 Stefanics G, Csukly G, Komlósi S, Czobor P, Czigler I. 2012. Processing of unattended facial
921 emotions: a visual mismatch negativity study. NeuroImage 59:3042-3049 (doi:10.1016/
922 j.neuroimage.2011.10.041)
923 Stefanics G, Czigler I. 2012. Automatic prediction error response to hands with unexpected
924 laterality: an electrophysiological study. NeuroImage, 63:253-261 (doi:10.1016/j.neuroimage.
925 2012.06.068)
926 Stefanics G, Kremlácek J, Czigler I. 2014. Visual mismatch negativity: a predictive coding
927 view. Front Hum Neurosci, 8:666 (doi: 10.3389/fnhum.2014.00666)
928 Sulykos I, Czigler I. 2011. One plus one is less than two: Visual features elicit non-additive
929 mismatch-related brain activity. Brain Res, 1398:64-71 (doi:10.1016/j.brainres.2011.05.009)
930 Susac A, Ilmoniemi RJ, Pihko E, Supek S. 2004. Neurodynamic studies on emotional and
931 inverted faces in an oddball paradigm. Brain Topogr. 16:265-268 (doi: 10.1023/B:BRAT.
932 0000032863.39907.cb)
933 Tsakiris M, Hesse MD, Boy C, Haggard P, Fink GR. 2007. Neural Signatures of Body
934 Ownership: A Sensory Network for Bodily Self-Consciousness. Cereb Cortex. 17:2235-2244
935 (doi: 10.1093/cercor/bhl131)

936 Yucel G, McCarthy G, Belger A. 2007. fMRI reveals that involuntary visual deviance
937 processing is resource limited. NeuroImage. 34:1245-1252(doi:10.1016/j.neuroimage.
938 2006.08.050)
939 Winkler I. 2007. Interpreting the Mismatch Negativity. Psychophysiol. 21:147-163 (doi:
940 10.1027/0269-8803.21.3.147)
941 Winkler I, Czigler I. 2012. Evidence from auditory and visual event-related potential (ERP)
942 studies of deviance detection (MMN and vMMN) linking predictive coding theories and
943 perceptual object representations. Int J Psychophy. 83:132-143 (doi:10.1027/0269-
944 8803.21.34.147)
945
946

Captions to figures

Figure 1: Timeline of the experimental procedure of experiment 1. Timeline of the self-related and other related deviant conditions, where the regular alternation of stimuli (250 ms length) was irregularly violated with a face containing 33%, 66%, or 100% of the self-face or the other face, respectively (UO: Unfamiliar Other; FO: Familiar Other).

Figure 2: ERP responses to self and other face images (experiment 1). ERPs elicited by deviant and regular self-face and other's face containing 33%, 66%, or 100% of error, over posterior-temporal, central and frontal sites.

Figure 3: vMMR to deviant face images of the self and other (experiment 1). *A*, Grand average vMMR when observing deviant faces containing 33% (black), 66% (red), and 100% (blue) of the prediction error associated with the face, over posterior-temporal, central and frontal sites. Although the waveforms seem to suggest opposite vMMR for the 'deviant other' condition as opposed to the 'self-deviant' condition, this effect is not significant . *B*, Topographical maps showing vMMR to self-related and to other related deviant faces (interpolation by Spherical Splines, order of Splines = 4, maximum degree of Lengendre Polynomials = 10, precision = 1E-5). *C*, Pseudo-3D representation of s-LORETA statistical maps showing regions where maximal self-related versus other related deviant differential activity were source localized, at latency 170-300 ms (33% vs 100%, $t = 4.75$, $p = 0.0008$; 66% vs 100%, $t = 4.907$, $p = 0.0090$). *$p < 0.05$.

Figure 4: Timeline of the experimental procedure of experiment 2. Timeline of the self-related and other related deviant conditions. The regular alternation of stimuli (250 ms length) was irregularly violated with deviant whole-body (face and body), deviant face, or deviant body images of the self (resulting conditions were, respectively: deviant whole-body self, deviant self-face, deviant self-body), or with deviant whole-body, deviant face or deviant

971 body images of the other (resulting conditions were, respectively: deviant whole-body other;
972 deviant other-face; deviant other-body).

973 **Figure 5: ERP responses to self and other bodily images (experiment 2).** ERPs elicited by
974 deviant, regular self-images and other's images, over posterior-temporal, central and frontal
975 sites.

976 **Figure 6: vMMR to deviant bodily images of the self and other (experiment 2).** *A.* Grand
977 average vMMR when observing deviant whole-body (black), deviant face (red), and deviant
978 body (blue) images of the self or the other, over posterior-temporal central and frontal sites.
979 Although the waveforms seem to suggest opposite vMMR for the 'deviant other' condition as
980 opposed to the 'self-deviant' condition, this effect is not significant. *B.* Topographical maps
981 showing vMMR to self-related and to other-related deviant images (interpolation by
982 Spherical Splines, order of Splines = 4; maximum degree of Lengendre Polynomials = 10,
983 precision = 1E-5). *C.* Pseudo-3D representation of s-LORETA statistical maps, showing
984 regions where maximal self-related versus other-related deviant differential activity were
985 source localized, at latency 220-320 ms.

986
987
988
989
990
991
992
993
994

Figure 1

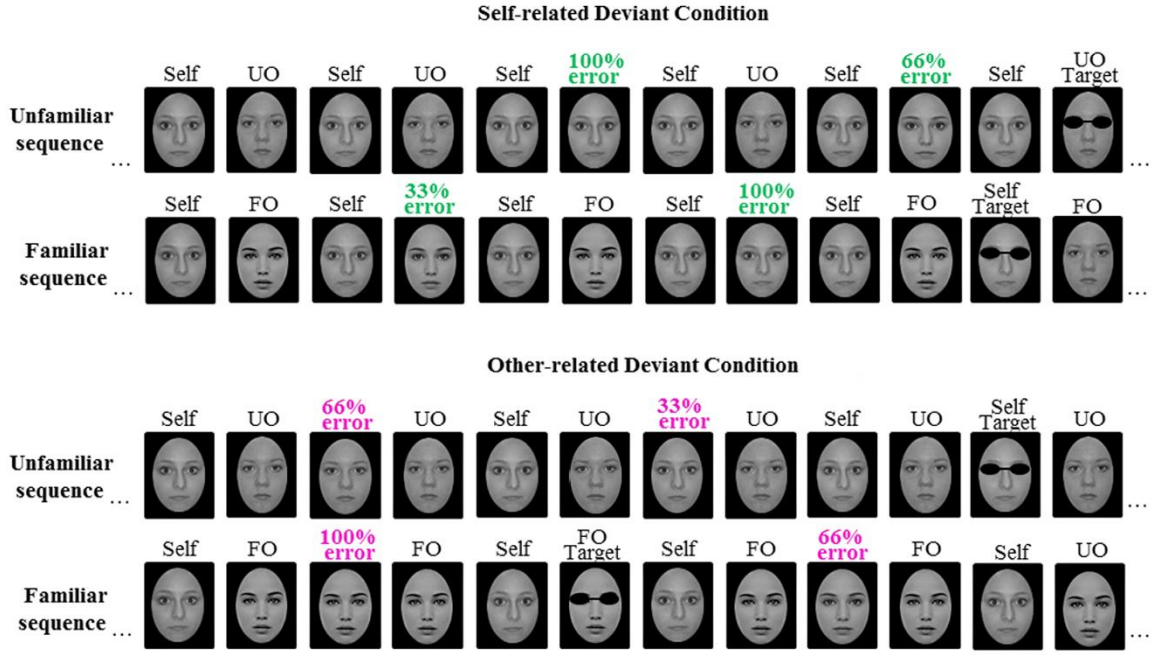

Figure 2

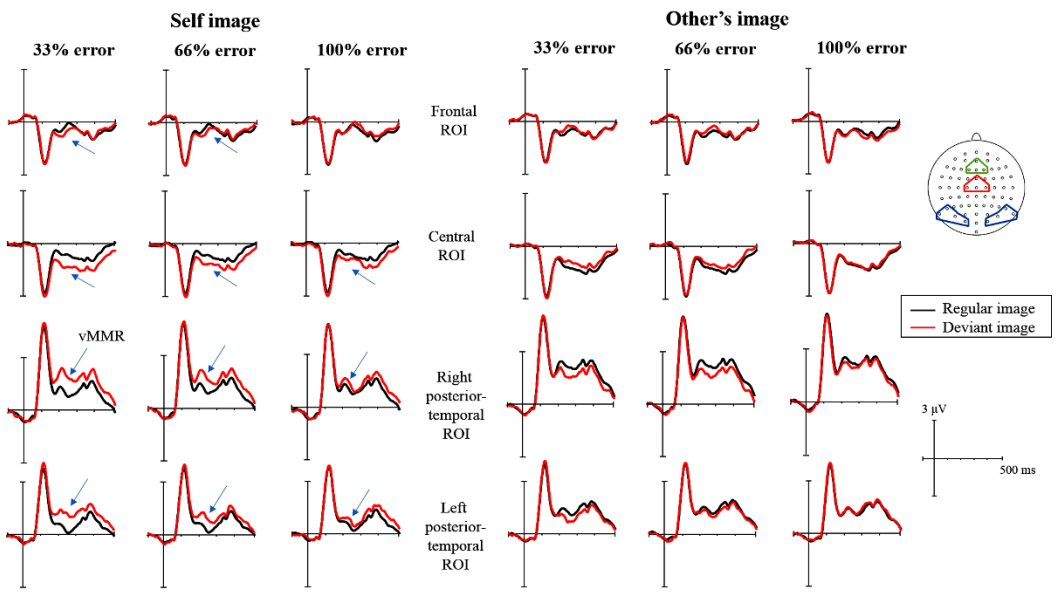

Figure 3

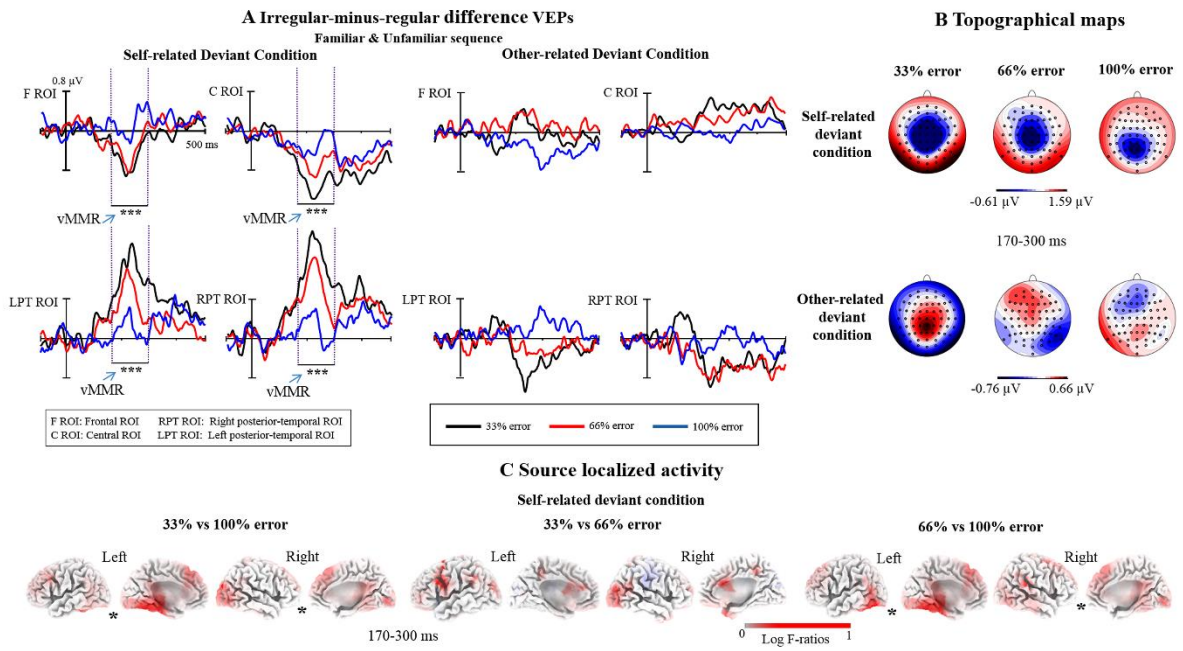

Figure 4

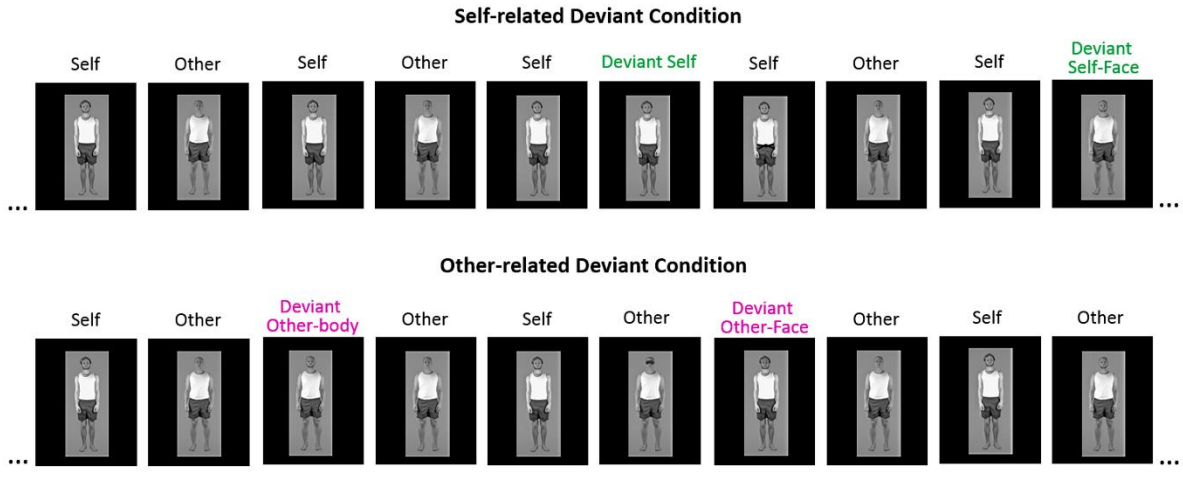

Figure 5

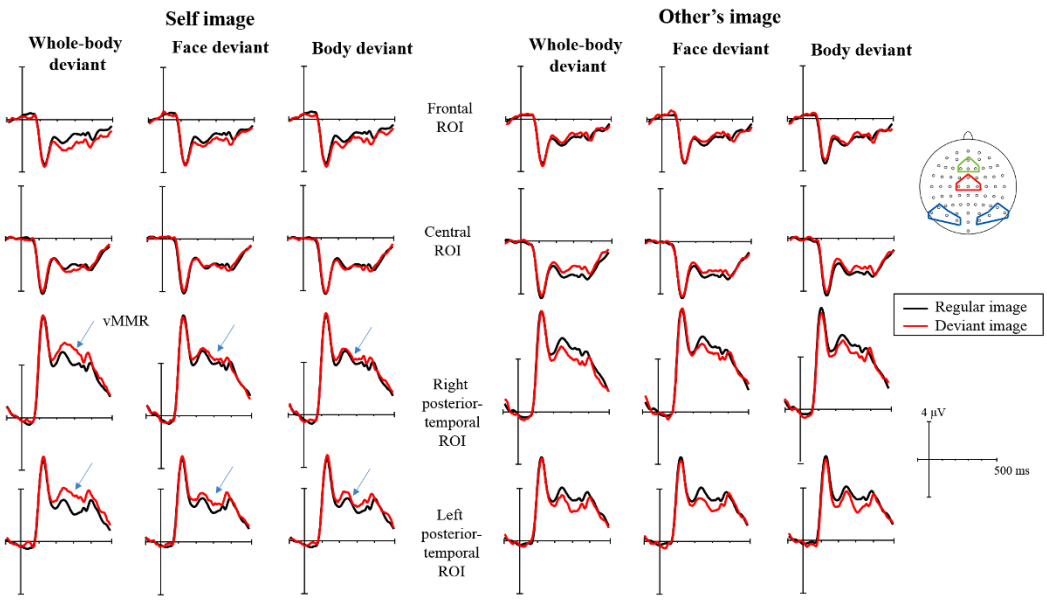

Figure 6