Rapid neural discrimination of communicative gestures

Elizabeth Redcay[1] **and Thomas A. Carlson**[2,3]

[1]Department of Psychology, University of Maryland, College Park, MD 20742, USA, [2]Perception in Action Research Centre & Department of Cognitive Sciences, Macquarie University, Sydney, NSW 2109, Australia, and [3]Centre for Cognition & Its Disorders, Macquarie University, Sydney, NSW 2109, Australia

Humans are biased toward social interaction. Behaviorally, this bias is evident in the rapid effects that self-relevant communicative signals have on attention and perceptual systems. The processing of communicative cues recruits a wide network of brain regions, including mentalizing systems. Relatively less work, however, has examined the timing of the processing of self-relevant communicative cues. In the present study, we used multivariate pattern analysis (decoding) approach to the analysis of magnetoencephalography (MEG) to study the processing dynamics of social-communicative actions. Twenty-four participants viewed images of a woman performing actions that varied on a continuum of communicative factors including self-relevance (to the participant) and emotional valence, while their brain activity was recorded using MEG. Controlling for low-level visual factors, we found early discrimination of emotional valence (70 ms) and self-relevant communicative signals (100 ms). These data offer neural support for the robust and rapid effects of self-relevant communicative cues on behavior.

INTRODUCTION

Social interactions play a fundamental role in social, cultural and language learning (e.g. Tomasello *et al.*, 2005; Meltzoff *et al.*, 2009; Mundy and Jarrold, 2010). Because of their fundamental role, communicative cues may receive preferential attention from early in infancy through adulthood (e.g. Csibra and Gergely, 2009; Frith, 2009; Senju and Johnson, 2009). Behavioral evidence suggests communicative cues are attended to rapidly and reflexively, and perceived interaction with others has robust effects on perceptual and cognitive processing (Teufel *et al.*, 2009; Laidlaw *et al.*, 2011).

In recent years, there has been increasing interest in the neural correlates of communicative cues. Functional MRI (fMRI) research suggests regions associated with the mentalizing system, particularly dorsal medial prefrontal cortex and posterior superior temporal sulcus, are engaged when a participant perceives a communicative signal [e.g. a direct gaze shift, a smile and wink or hearing one's own name, (Kampe *et al.*, 2003; Morris *et al.*, 2005; Schilbach *et al.*, 2006)] or when participants believe they are in an interaction with a person, as compared with a computer or video recording (Iacoboni *et al.*, 2004; Walter *et al.*, 2005; Redcay *et al.*, 2010; Schilbach *et al.*, 2013). Further, in addition to a 'mentalizing' system, studies suggest involvement of action observation and action production systems, including regions within premotor and parietal cortex when participants view a communicative, as compared with non-communicative, face or hand action (Schilbach *et al.*, 2008; Streltsova *et al.*, 2010; Conty *et al.*, 2012; Ciaramidaro and Becchio, 2013).

The recruitment of higher-order 'mentalizing' brain systems could suggest the time these cues are processed in the brain is relatively delayed, at least later than face recognition (i.e. 170 ms, Halgren *et al.*, 2000). Behaviorally, however, communicative cues have rapid and robust effects even on low-level visual perception. For example, when participants hear someone call their name (as compared with other names), they are more likely to perceive eye gaze that is slightly averted as direct gaze (Stoyanova *et al.*, 2010). Further, Posner-type gaze cueing paradigms suggest gaze cues elicit reflexive orienting in the direction of gaze even if the cues are irrelevant to the task (Frischen *et al.*, 2007). Finally, when participants believe an experimenter can see through opaque goggles, gaze adaptation effects are larger than when the participant believes the experimenter cannot see (though visual stimulation is identical from the perspective of the participant) (Teufel *et al.*, 2009). These studies suggest the processing of communicative cues is rapid and automatic, and influences perceptual and attention processes.

Taken together, behavioral and fMRI research suggests communicative cues may rapidly and reflexively recruit a network of brain regions associated with detecting these cues and adjusting perceptual and cognitive processes accordingly. However, this remains speculative because relatively few studies have investigated *when* the brain processes communicative cues. Methods with high temporal sensitivity, such as event-related potentials (ERP) or magnetoencephalography (MEG) are well suited to address this question.

ERP and MEG studies examining processing of social-emotional and social-communicative cues have predominantly examined the timing of neural responses to faces displaying an emotional expression or gaze cues, relative to a face-sensitive response consistently identified at 170 ms after presentation of a neutral face (Halgren *et al.*, 2000; Itier and Batty, 2009). Differences in brain activity to directional cues (i.e. averted *vs* direct gaze) occur at latencies later than 170 ms (Itier and Batty, 2009; Dumas *et al.*, 2013; Hasegawa *et al.*, 2013). Studies using emotional expressions, on the other hand, reveal earlier discrimination of fearful, happy and sad faces from neutral faces, over fronto-central and occipital regions between 80 and 130 ms (Halgren *et al.*, 2000; Eimer and Holmes, 2002; Liu *et al*, 2002; Pizzagalli *et al.*, 2002; Roxanne J Itier and Taylor, 2004; Pourtois *et al.*, 2004). Emotional expressions and gaze cues are social cues that reflect another person's mental states, but context determines whether they are to be interpreted as communicative, or deliberate social signals (cf. Frith, 2009). Certain hand gestures, on the other hand, always signal

communicative intent and can vary in the emotional content of the message (for example, "point", "ok sign" and "insult" hand gestures). ERP responses differentiate these signals as early as 100 ms (i.e. the P100) (Flaisch *et al.*, 2011; Flaisch and Schupp, 2013), reflecting early modulation based on the emotional content of the communicative cue.

While face and hand gestures are powerful social stimuli, real-world communication involves both face and body gestures. Examining faces or hands alone and out of context may not reveal the full brain dynamics supporting processing of communicative cues. Conty *et al.* (2012) examined the temporal response to face and body gestures varying on emotion (angry *vs* neutral), gesture (i.e. point *vs* no point) or self-directedness (i.e. direct *vs* averted). Participants reported significantly more feelings of self-involvement with each of these comparisons (e.g. point *vs* no point); thus providing evidence that multiple factors (e.g. expression, gaze and gestures) can contribute to the perception of communicative intent. Independent effects for each of these dimensions were seen as early as 120 ms, but integration across dimensions occurred later.

In sum, these data offer compelling preliminary evidence for early discrimination of communicative cues but are limited by several factors. First, these studies have taken a categorical experimental approach (e.g. emotional *vs* neutral or self-directed *vs* not), ignoring the subtleties between gestures. Second, these studies only used a single gesture to represent a category (e.g. "ok" represents positive emotional gestures), and thus could be confounded by other low-level features that differ between gestures. Finally, these studies have focused on ERP components rather than taking advantage of the full high temporal resolution of time series data.

In the present study, we overcome previous limitations in our understanding of *when* communicative cues are represented in the brain by taking a sliding window multivariate pattern analysis method approach to the analysis of MEG time series data, while participants view an actress making communicative hand and face gestures (Carlson *et al.*, 2011). Using behavioral ratings, we evaluated 20 gesture stimuli (Figure 1A) on two social-communicative factors: self-relevance and emotional valence. For each factor, the relationship among gestures is quantitatively represented in a dissimilarity matrix (DSM; Figure 1B–D). Next, using MEG decoding methods, we measured the neural discriminability (i.e. decodability) among the gesture stimuli on a moment-to-moment basis, which is represented as a set of time varying DSMs. Three factors, two social-communicative factors and one visual factor to control for low-level difference between the images, were then used as predictors for the brain's time varying representation of the stimuli. Correspondences between individual factor DSMs and an MEG DSM for specific time points is indicative that the brain is representing information about this factor at this time (c.f. Kriegeskorte *et al.*, 2008). Using these continuous measures of the communicative factors and this sensitive decoding approach, we precisely map the time course of the neural processing of communicative cues in the brain.

METHODS

Participants

Twenty-four volunteers (11 male, 13 female) with an average age of 21.2 years participated in the MEG experiment. Forty-nine additional volunteers participated in pilot behavioral rating experiments to select the stimuli for the MEG experiment. All participants had normal or corrected to normal vision. Prior to the experiment, informed written consent was obtained from each volunteer. MEG participants were paid $30 for their participation. Behavioral participants participated in exchange for course credit through the University of Maryland SONA system. The University of Maryland Institutional Review Board approved all experimental procedures.

Stimuli

Stimuli were still image frames selected from a set of 170 dynamic videos created in collaboration with the Institute for Disabilities Research and Training (idrt.com) for a separate project. These 170 videos of an actress making gestures that varied in the degree of communicative content (e.g. waving, folded arms, looking at her watch, etc; see Figure 1) were rated for communicative-relevant factors including self relevance, semantic content (or meaningfulness) and emotional valence using a 7-point rating scale with the three questions below:

(1) "How much did it feel like someone was communicating with you? (In other words, did it feel like she was conveying something to you with her hand movement)" with the scale ranging from "Not at all" to "Very much".
(2) "How easily could you understand what the person was communicating? (In other words, if we asked you to put into words what she was gesturing, would you be able to do that)" with the scale ranging from "Movements were meaningless" to "Easily".
(3) "Would you consider the gesture in this picture to be emotionally negative, neutral or positive?" with the scale ranging from "Very Negative" to "Very Positive".

Based on these ratings, 20 videos were selected qualitatively by the experimenters that represented a continuum of ratings across the three factors. Frames from these 20 videos were selected to create static images at the point at which the frame best captured the action in the video (Figure 1A).
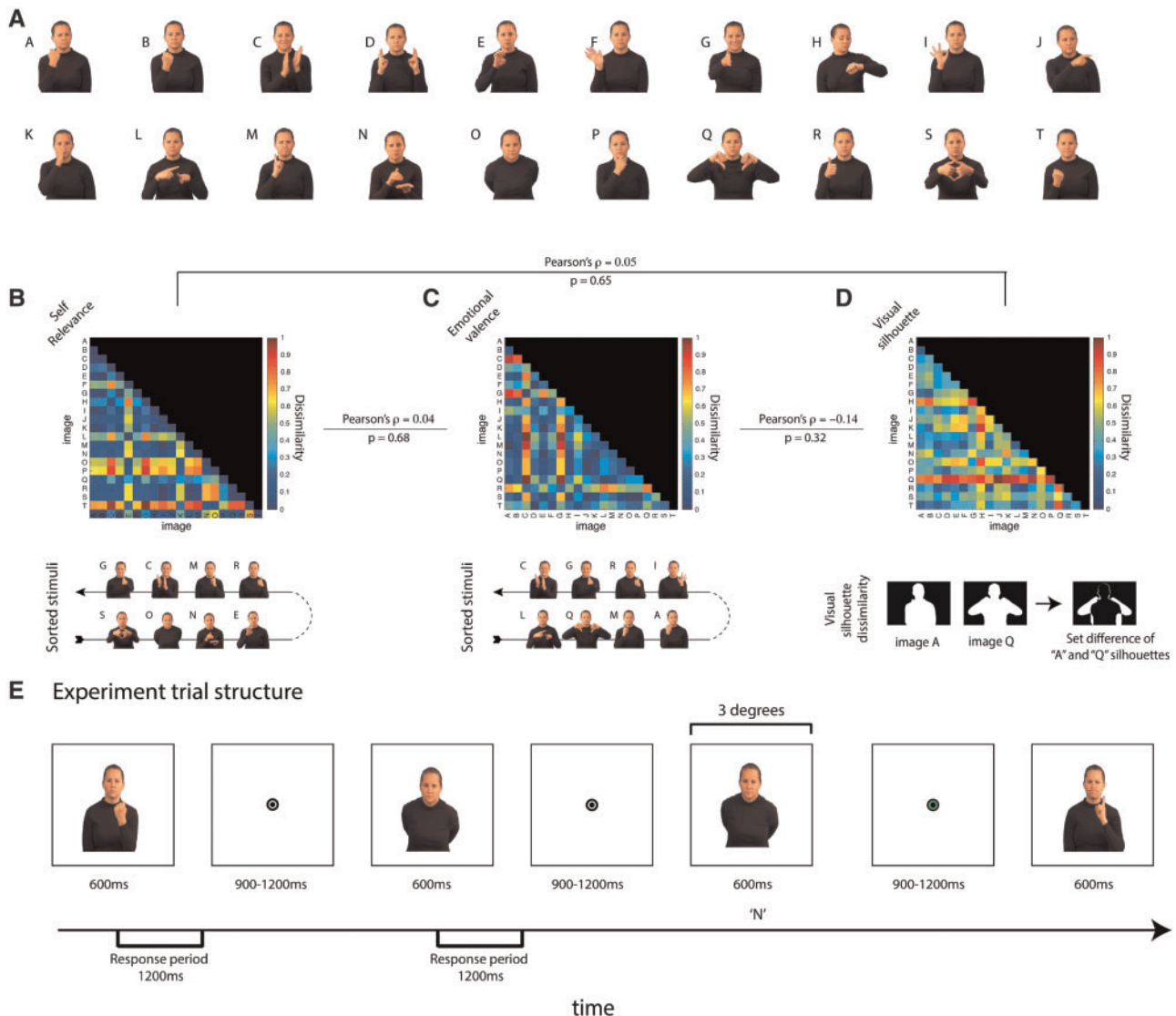
The social-communicative model

The model used in the study contains two social-communicative factors and a visual factor to account for differences in low-level visual feature differences between the stimuli.

The social-communicative factors

To construct the social-communicative factors, we collected independent ratings of the images. After the initial stimulus selection (described above), thirty-one new participants rated the content of each of the 20 images for self-relevance, semantic content and emotional valence (the questions listed above in Questions 1–3) as well as two additional questions relating to emotional valence listed below (Questions 4–5). The initial assessment of emotional valence (Question 3) used a scale with ratings ranging from negative to positive, reflecting a classic bipolar model of valence (e.g. Russell and Carroll, 1999). However, others (e.g. Norris *et al.*, 2010) have proposed that representations of positive and negative valence are distinct (the bivalent hypothesis) and thus unipolar scales within each emotion ranging from neutral to positive and neutral to negative may better capture these measures of emotional valence. Given these current theoretical debates, in this second round of ratings, we included two new questions (Questions 4–5 below) to capture unipolar valence effects.

(4) "How negative would you consider the gesture in this picture to be?" with the scale ranging from "Neutral" to "Very Negative".
(5) "How positive would you consider the gesture in this picture to be?" with the scale ranging from "Neutral" to "Very Positive".

The primary theoretically motivated factor in the model was self-relevance, with emotional valence as a secondary, comparison factor. We initially submitted the responses to all the questions to principle component analysis (PCA) with the aim that the social-communicative

Fig. 1 Social-communicative model and MEG experiment. (**A**) *Stimuli.* 20 images of an actress making gestures that vary in social-communicative content (**B**–**D**) *The communicative model.* A three-factor model including two social factors (**B**) *self-relevance* and (**C**) *emotional valence* and one *visual* factor (D). DSMs for each factor show the quantitative relationship between individual image pairs. Beneath the social-communicative factor, DSMs are the top four ranking and bottom four ranking images. Beneath the visual factor DSM, the computation of visual dissimilarity (the set difference of the two image silhouettes) is shown graphically. Lines connecting the DSM display the Pearson correlation between factor DSMs. (**E**) *Experimental trial structure.* Participants were shown blocks of images in pseudo random order with images repeating approximately once every 12 images in the MEG scanner. Each image was shown for 600 ms with a variable ISI between 900 and 1200 ms. The participant's task was to detect repetitions of any image. Participants were given feedback though changes in the color of the fixation bullseye (green for successful detection/red for false alarms and misses; no feedback was provided for correct rejections) immediately after the repetition occurred.

factors would be emergent. This analysis, however, produced confounded factors (e.g. a factor that ranged from "low self-relevance" to "high emotional negative content") and nonsensical factors that were not easily interpretable. To satisfy the theoretical motivation of the study, we opted to construct the two factors using the following method. The factor of central interest was self-relevance. The first question measured the perceived self-relevance of each gesture and the second rated the semantics, or meaningfulness. The answers to these two questions were highly correlated (Pearson's $\rho = 0.58$ $P < 0.01$), as it is difficult to have highly communicative gesture without meaning. The ratings from these two questions were combined into a single composite variable using PCA. The first component accounted for 94.4% of the variance. Figure 1B shows the factors' top four and bottom four images, which accord with intuitions of self-relevance. On one extreme, there are nonsense gestures, and on the other are gestures of interpersonal communication, e.g. greetings,

applause, disapproval. This component was taken as our metric of *self-relevance.* Questions 3–5 were also submitted to PCA. The first factor (accounting for 78.5% of the variance) captured bipolar emotional valence (i.e. from negative to positive). Figure 1C shows the top and bottom four images this factor. The second factor from PCA (accounting for 20.8% of the total variance, 96.7% of the remaining variance) captured a unipolar dimension of emotion, ranging from neutral to emotional (positive and negative). This factor correlated with both the self-relevance and bipolar emotional valence factors, introducing the potential problem of multicollinearity in the regression analysis (Dormann *et al.*, 2012). We therefore opted not to include this factor in the model. Of note, its inclusion in the analysis did not alter the pattern of findings (see Supplementary Figure S1).

For each of the two social-communicative factors, we constructed a DSM (Figure 1B–D). Each entry of a factor's DSM is the Euclidean distance between a pair of stimuli for that factor, and the complete

DSM describes the relationships between all the gestures. For example, for the self-relevance factor, the image of the women making a nonsense gesture (Image S) and the image of the woman extending her hand (Image G) are two extremes (see Figure 1B). Correspondingly in the self-relevance DSM, the value of the entry is high (colored red). In contrast, for emotional valence, the same two gestures are only moderately different, and thus in the emotional valence DSM, the value is more moderate (colored amber). Each factor's DSM describes a complex set of relationships between the stimuli—the representational geometry of the factor (Kriegeskorte and Kievit, 2013).

Visual factor

The use of naturalistic stimuli introduces the possibility that visual differences between the images might confound the findings. To remove this possibility, we modeled the differences between the images retinotopic projection using a silhouette model (Jaccard, 1901), which we have previously shown to account well for visually evoked decodable brain activity in MEG (Carlson *et al.*, 2011). Operationally defined, silhouette image dissimilarity is the image complement of the two image silhouettes in the comparison (Figure 1E). The visual model DSM was included in the analysis to remove the influence of low-level visual feature differences between images.

Experimental design

Figure 1D diagrammatically shows an example sequence of trials in the experiment. On the display, the images subtended $3°$ of visual angle, on average. Individual images were shown for 600 ms. Between images, there was a random interstimulus interval (ISI) that ranged from 900 to 1200 ms. The order of the images was pseudo-randomized within each block, such that images would repeat once every 12 images on average. This pseudo-randomization was introduced for the purposes of the repetition detection task (see below). Each block of trials was composed of a sequence of 240 images. In each block, each image was shown 12 times (20 gestures $\times$ 12 = 240 trials). Each participant performed eight blocks of trials for 96 trials ($12 \times 8 = 96$ trials) per gesture.

Experimental task/results

In the scanner, participants performed a repetition detection task to encourage them to attend to the images and maintain vigilance. Approximately once every 12 images, an individual image would repeat. Participants were instructed to report repetitions using a button press response. Feedback was given after each repetition in the form of changes in color in the fixation point. The fixation turned green if the subject correctly detected a repetition, and turned red if the subject either missed a repetition or made a false alarm response. No feedback was given for correct rejections. After each block, participants received a summary of their performance. The mean accuracy across participants was 92% correct (s.d. 4.6%). The average reaction time was 581 ms (s.d. 69 ms).

Display apparatus

Subjects viewed the stimuli on a translucent screen while lying supine in a magnetically shielded recording chamber. The stimuli were projected onto the screen located 30 cm above the participant. Experiments were run on a Dell PC desktop computer using MATLAB (Natick, MA).

MEG recordings and data preprocessing

Recordings were made using a 157 channel whole-head axial gradiometer MEG system (KIT, Kanazawa, Japan). The recordings were filtered online from 0.1 to 200 Hz using first-order Resistor-Capacitor (RC) filters and digitized at 1000 Hz. Time shifted PCA was used to denoise the data offline (de Cheveigne and Simon 2007).

Trials were epoched from 100 ms before to 600 ms after stimulus onset. Eye movement artifacts trials were removed automatically using an algorithm that detects deviations in the root mean square (RMS) amplitude over 30 selected eye-blink sensitive channels. The average rejection rate was 2.4% of trials with a standard deviation of 1.2% across participants. After eye movement artifact rejection, the time series data were resampled to 50 Hz [corrected for the latency offset introduced by the filter (see VanRullen, 2011) for the analysis].

Pattern classification analysis of MEG time series data

We used naïve Bayes implementation of linear discriminant analysis (Duda *et al.* 2001) to classify ('decode') the image that participants were viewing from the raw sensor data (see Carlson *et al.*, 2011, 2013 for detailed methods). To improve the signal to noise, trials were averaged in pseudo trials, which were averages of nine randomly selected trials (see Isik *et al.*, 2014). The set of 96 trials per gesture image (sometimes less after artifact rejection) was thus reduced to 10 trials per image though averaging. Generalization of the classifier was evaluated using k-fold cross-validation with a ratio of 9:1 training to test (i.e. nine pseudo trials used for training and one trial used to test the classifier). This procedure was repeated 50 times, each time with a new randomization. Performance of the classifier (reported as d-prime) is the average performance across the 50 iterations.

### Constructing time-resolved MEG DSMs

The procedure above was used to decode which image the observer was viewing from the MEG time series data on a time point by time point basis. For each time point, we constructed a DSM by attempting to decode the image of the gesture that participants were viewing for all possible pairwise comparisons of the images. Thus for each time point, we obtained a DSM, like those shown in Figure 1 that represents the decodability of all possible pairwise combinations of images. Across the time series, we computed a set of time varying DSMs for the analysis. Figure 2A show the decodability of the stimuli averaged over time.

Evaluation of time varying decoding performance

To evaluate significance for decoding on the time series data, we compared classification performance with chance performance (d-prime value = 0) using a nonparametric Wilcoxon signed-rank test for each time point. To determine the onset, we used a cluster threshold of two consecutive significant time points above chance [false discovery rate (FDR) < 0.05].

Evaluation of the communicative model

The time-resolved MEG DSMs were compared with the communicative model factors model using the representational similarity analysis framework (Kriegeskorte *et al.*, 2008). Each factor makes a prediction, which is represented in the factor's DSM, about the decodability of each pairwise comparison between images, which is represented in the MEG DSMs. Correspondences between individual factor DSMs and an MEG DSM for specific time points are indicative that the brain is representing information about this factor at this time (c.f. Kriegeskorte *et al.*, 2008). Each of the three model factors (self relevance, emotional valence and visual) was a predictor for an ordinary least-squares regression analysis on each MEG time point's DSM. The predictors were the entries from the lower left triangle of each factor's DSMs (note DSMs are symmetrical). The dependent variable for the regression analysis was the entries from the lower left triangle of the
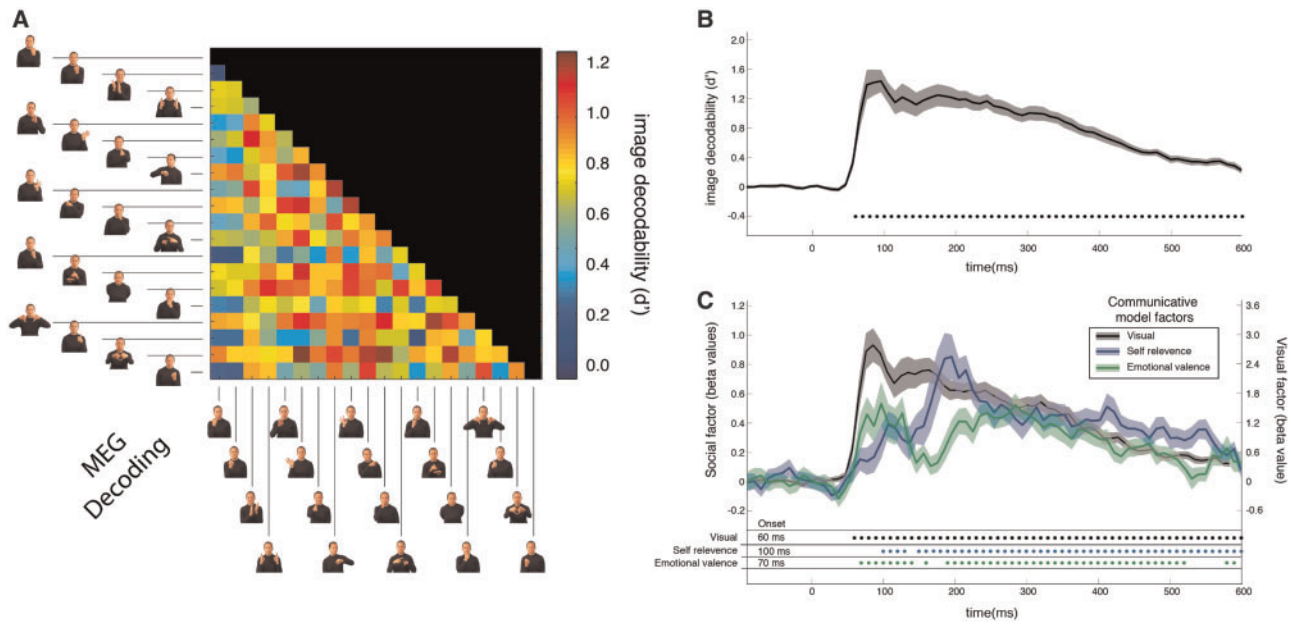
Fig. 2 MEG decoding and social communicativeness. (**A**) *Decodability of exemplar pairs*. DSM displaying the decodability of the exemplar pairing averaged over the time series. (**B**) *Average decodability across exemplar as a function of time.* Solid line is the average performance of classifier averaged across subjects. Shaded region is 1 SEM. Disks below the plot indicate above chance decoding performance, evaluated using a Wilcoxon signed-rank test with an FDR threshold of 0.05. (**C**) *The communicative model factor fits for decoding performance*. Plotted are beta weights for two communicative social factors and the visual factor from the regression analysis on time varying MEG decoding performance. The y-axis is the average beta weight across subjects. Note that the scale for the visual factor is shown on the left-hand side of the plot. The scale for the social-communicative factors is on the right. The solid line is the average beta weight across subjects. The shaded region is 1 SEM. Color-coded asterisks indicate a significant correlation, evaluated using a Wilcoxon signed-rank test with an FDR threshold of 0.05. Decoding onset determined using a cluster threshold of two consecutive significant time points above chance.

MEG DSM. We used a hierarchical analysis. The regression analysis was conducted on each subject's MEG data. Individual subject's beta weights from the regression analysis were then compared with the null hypothesis of a beta weight of 0 using nonparametric Wilcoxon signed-rank test to evaluate significance (FDR threshold 0.05). For evaluating onset latency, we used a cluster threshold of two consecutive significant time points below significance.

RESULTS

Decoding the stimuli

We first determined that we could decode the images participants were viewing from the MEG recordings. Figure 2A shows a DSM of decoding performance averaged across participants and across time (0–600 ms post-stimulus onset). Cool colors indicate poor decoding performance, and warm colors indicate good performance. While there is variation, indicating decoding performance depends on the comparison, most of the entries are warm colors indicating successfully decoding. Decoding performance as a function of time is shown in Figure 2B. Decoding first rises above chance 60 ms after stimulus onset, which accords with the estimated time for visual inputs from the retina to reach the cortex with single unit recordings, Electroencephalography (EEG), and our own findings using MEG decoding methods (Jeffreys and Axford, 1972; Di Russo *et al.*, 2002; Carlson *et al.*, 2013). Following the onset decoding remains above chance for the entire time the images are on the screen (600 ms).

Thus far, we have shown that it is possible to decode the images of the actress making gestures from the MEG time series data. We next examined how the different factors contribute to decodability.

The visual factor: robust visual representations

Visual differences between the stimuli will likely strongly affect decoding performance, which might confound the findings. To address this, we included a visual factor of the stimuli based on the images'

retinotopic silhouettes as a predictor (Figure 1D; see Methods). Figure 2C shows average beta weights from the regression analysis for the visual predictor as function of time. Retinotopic silhouette was an excellent predictor, as expected from earlier studies (Carlson *et al.*, 2011). From the time that decodability first rises above chance (60 ms) to the end of the time series, the visual factor was significant.

The social factors: rapid processing of communicative cues in the brain

The focus of the study was the social-communicative model factors, particularly self-relevance. Figure 2C shows the beta weights for both self-relevance and emotional valence factors as a function of time. While the communicative factor beta weights were smaller than the visual factors (note the different scales for visual and social factors), they also have explanatory power. Shortly after the onset of the stimulus, both factors are significant. Emotional valence first becomes significant at 70 ms, and self-relevance is first significant at 100 ms. And, like the visual model, both factors were significant for nearly the entire interval following the onset of the stimulus. Interestingly, emotional valence qualitatively appears to have two phases. Following the initial early onset and peak, emotional valence declines to chance levels around 160 ms and after this reemerges as a strong predictor.

DISCUSSION

Using a whole-brain multivariate pattern analysis approach that accounts for low-level visual differences, we mapped the time course of the processing of self-relevant social-communicative cues in the brain. Our findings provide neural support for behavioral data and theories suggesting early effects of self-relevant communicative signals on visual perception and cognition (Csibra and Gergely, 2009; Teufel *et al.*, 2009; Stoyanova *et al.*, 2010). We show that just 70 ms after stimulus onset, the brain begins to process the emotional valence of gestures and shortly thereafter (100 ms) represents whether or not an individual is

signaling self-relevant communication. While previous ERP and MEG studies of faces have demonstrated early discriminability between emotions, these data show processing of communicative cues earlier than previously reported using traditional analysis approaches (Itier and Batty, 2009; Flaisch et al., 2011).

The processing of social cues in the brain

Both classic and contemporary models of vision propose visual inputs are processed hierarchically from early retinotopic feature representations, to categories, and later to the extraction of semantic meaning (Marr, 1982; Biederman, 1987; Logothetis and Sheinberg, 1996; Riesenhuber and Poggio, 1999; DiCarlo and Cox, 2007). Physiological recordings indicate that this hierarchical processing takes place in a rapid feed forward sweep (for a recent review, see Vanrullen, 2007). Our data indicate that early within this rapid feed forward sweep, the brain determines the emotional valence and the communicative self-relevance of gestures. This early discrimination may reflect the salience of communicative cues in reflexively biasing the organism to relevant objects in the environment (Frischen et al., 2007; Stoyanova et al., 2010). While the onset of discrimination is seen early, later time windows also discriminated self-relevance and valence. This later time may represent a more evaluative phase of processing, consistent with studies of semantic processing (Habets et al., 2011; Kutas and Federmeier, 2011). In the current study, only communicative factors were included in the model. Thus, to determine where these factors fall in relation to other aspects of emotion and social perception, the relative processing time would need to be compared between these and other factors using the same methods. For example, a model could contain different actors and a factor that captured distance in face space (Valentine, 1991). Additionally, future models could contain stimuli reflecting a continuum from low to high arousal, as this dimension is thought to be relatively independent from valence (e.g. Feldman Barrett and Russell, 1999) and has been shown to produce rapid neural discrimination using categorical measures (e.g. Flaisch et al., 2011). These next steps would provide a better understanding of the time course of social and emotional processing, including which factors result in rapid top-down modulation of neural processing.

The advantage of naturalistic stimuli and continuous measures

We used stimuli that differed in several ways from previous studies. First, rather than using categorical measures of cues (e.g. communicative vs not), we used continuous measures of two factors (communicative self-relevance and emotional valence) that were based on independent behavioral ratings. Previous studies have used single gestures (e.g. insult vs ok) to represent positive and negative emotion (Flaisch et al., 2011; Flaisch and Schupp, 2013), but those effects could have been specific to the gesture rather than the category of gestures. However, like previous studies (Conty et al., 2012; Flaisch and Schupp, 2013), we also report early discrimination of emotional and communicative factors around 100 ms, suggesting the single, categorical gestures in the previous studies may have captured a similar dimension. Second, the gestures in the current study integrated emotional facial expressions and hand gestures, similar to how one would encounter communicative gestures in real-world settings. One previous study that examined face and body cues (e.g. an angry person pointing at you vs a neutral face without a point) found integration across cue categories at 200 ms but not earlier (Conty et al., 2012). The discrepancy between studies could be due to our inclusion of a wider range of more naturalistic gestures that integrate facial expression and body gesture.

CONCLUSION

The current study provides evidence for rapid discrimination of self-relevant communicative and emotional cues. This early neural sensitivity suggests these cues may receive preferential allocation of attention and act to modulate later stages of visual processing. Behaviorally, attention to social-communicative cues is critically important to social, cognitive and language learning from early in infancy and throughout life (Csibra and Gergely, 2009; Meltzoff et al., 2009). An important future direction will be to examine how moment-to-moment differences in neural discriminability to these cues are related to individual differences in social-communicative and social-cognitive abilities in adults and during development. These decoding methods also have promise to advance our understanding of autism spectrum disorder, a disorder characterized by atypical attention to social cues (Klin et al., 2003; Pierce et al., 2011).

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

Conflict of Interest

None declared.

REFERENCES

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. Psychological Review, 94(2), 115–47.

Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J. (2011). High temporal resolution decoding of object position and category. Journal of Vision, 11, 1–17.

Carlson, T., Tovar, D.A., Alink, A., Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. Journal of Vision, 13, 1–19.

Ciaramidaro, A., Becchio, C. (2013). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. Social Cognitive and Affective Neuroscience.

Conty, L., Dezecache, G., Hugueville, L., Grèzes, J. (2012). Early binding of gaze, gesture, and emotion: neural time course and correlates. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 32(13), 4531–9.

Csibra, G., Gergely, G. (2009). Natural pedagogy. Trends in Cognitive Sciences, 13(4), 148–53.

DiCarlo, J.J., Cox, D.D. (2007). Untangling invariant object recognition. Trends in Cognitive Sciences, 11(8), 333–41.

Di Russo, F., Martínez, A., Sereno, M.I., Pitzalis, S., Hillyard, S.A. (2002). Cortical sources of the early components of the visual evoked potential. Human Brain Mapping, 15, 95–111.

Dumas, T., Dubal, S., Attal, Y., et al. (2013). MEG evidence for dynamic amygdala modulations by gaze and facial emotions. PloS One, 8(9), e74145.

Eimer, M., Holmes, A. (2002). An ERP study on the time course of emotional face processing. Neuroreport, 13(4), 427–31.

Feldman Barrett, L., Russell, J.A. (1999). The Structure of current affect: controversies and emerging consensus. Current Directions in Psychological Science, 8(1), 10–14.

Flaisch, T., Häcker, F., Renner, B., Schupp, H.T. (2011). Emotion and the processing of symbolic gestures: an event-related brain potential study. Social Cognitive and Affective Neuroscience, 6(1), 109–18.

Flaisch, T., Schupp, H.T. (2013). Tracing the time course of emotion perception: the impact of stimulus physics and semantics on gesture processing. Social Cognitive and Affective Neuroscience, 8(7), 820–7.

Frischen, A., Bayliss, A.P., Tipper, S.P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. Psychological Bulletin, 133(4), 694–724.

Frith, C. (2009). Role of facial expressions in social interactions. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364(1535), 3453–8.

Habets, B., Kita, S., Shao, Z., Ozyurek, A., Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. Journal of Cognitive Neuroscience, 23(8), 1845–54.

Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. Cerebral Cortex, 10(1), 69–81.

Hasegawa, N., Kitamura, H., Murakami, H., et al. (2013). Neural activity in the posterior superior temporal region during eye contact perception correlates with autistic traits. Neuroscience Letters, 549, 45–50.

Iacoboni, M., Lieberman, M.D., Knowlton, B.J., et al. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. NeuroImage, 21(3), 1167–73.

Isik, L., Meyers, E.M., Leibo, J.Z., Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, *111*(1), 91–102.

Itier, R.J., Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neuroscience and Biobehavioral Reviews*, *33*(6), 843–63.

Itier, R.J., Taylor, M.J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex*, *14*(2), 132–142.

Kampe, K., Frith, C., Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(12), 5258–63.

Klin, A., Jones, W., Schultz, R., Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1430), 345–60.

Kriegeskorte, N., Kievit, R.A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–12.

Kriegeskorte, N., Mur, M., Ruff, D.A., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–41.

Kutas, M., Federmeier, K.D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–47.

Laidlaw, K.E.W., Foulsham, T., Kuhn, G., Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5548–53.

Liu, J., Harris, A., Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience*, *5*(9), 910–6.

Logothetis, N.K., Sheinberg, D.L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

Meltzoff, A.N., Kuhl, P.K., Movellan, J., Sejnowski, T.J. (2009). Foundations for a new science of learning. *Science (New York, N.Y.)*, *325*(5938), 284–8.

Morris, J.P., Pelphrey, K.A., McCarthy, G. (2005). Regional brain activation evoked when approaching a virtual human on a virtual walk. *Journal of Cognitive Neuroscience*, *17*(11), 1744–52.

Mundy, P., Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, *23*(8-9), 985–997.

Norris, C.J., Gollan, J., Berntson, G.G., Cacioppo, J.T. (2010). The current status of research on the structure of evaluative space. *Biological Psychology*, *84*(3), 422–36.

Pierce, K., Conant, D., Hazin, R., Stoner, R., Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, *68*(1), 101–9.

Pizzagalli, D.A., Lehmann, D., Hendrick, A.M., et al. (2002). Affective judgments of faces modulate early activity (approximately 160 ms) within the Fusiform Gyri. *NeuroImage*, *16*(3), 663–677.

Pourtois, G., Grandjean, D., Sander, D., Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex*, *14*(6), 619–33.

Redcay, E., Dodell-Feder, D., Pearrow, M.J., et al. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *NeuroImage*, *50*(4), 1639–47.

Riesenhuber, M., Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–25.

Russell, J.A., Carroll, J.M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*(1), 3–30.

Schilbach, L., Eickhoff, S.B., Mojzisch, A., Vogeley, K. (2008). What's in a smile? Neural correlates of facial embodiment during social interaction. *Social Neuroscience*, *3*(1), 37–50.

Schilbach, L., Timmermans, B., Reddy, V., et al. (2013). Toward a second-person neuroscience. *Behavior. Brain Sciences*, *36*(4), 441–62.

Schilbach, L., Wohlschlaeger, A.M., Kraemer, N.C., et al. (2006). Being with virtual others: neural correlates of social interaction. *Neuropsychologia*, *44*(5), 718–30.

Senju, A., Johnson, M.H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–34.

Stoyanova, R.S., Ewbank, M.P., Calder, A.J. (2010). "You talkin' to me?" Self-relevant auditory signals influence perception of gaze direction. *Psychological Science*, *21*(12), 1765–9.

Streltsova, A., Berchio, C., Gallese, V., Umilta', M.A. (2010). Time course and specificity of sensory-motor alpha modulation during the observation of hand motor acts and gestures: a high density EEG study. *Experimental Brain Research*, *205*(3), 363–73.

Teufel, C., Alexis, D.M., Todd, H., et al. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology*, *19*(15), 1274–7.

Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences*, *28*(5), 675–91; discussion 691–735.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204.

Vanrullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology*, *3*(1-2), 167–76.

Vanrullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 365.

Walter, H., Abler, B., Ciaramidaro, A., Erk, S. (2005). Motivating forces of human actions. Neuroimaging reward and social interaction. *Brain Research Bulletin*, *67*(5), 368–81.