OXFORD

An objective neural signature of rapid perspective taking

Alexy A. Beck, Bruno Rossion, and Dana Samson

Psychological Sciences Research Institute (IPSY), Institute of Neuroscience (IoNS), University of Louvain, 1348 Louvain-la-Neuve, Belgium

Correspondence should be addressed to Alexy A. Beck, Psychological Sciences Research Institute (IPSY), Institute of Neuroscience (IoNS), University of Louvain, 10 Place Cardinal Mercier, 1348 Louvain-la-Neuve, Belgium. E-mail: alexy.beck@uclouvain.be.

Abstract

The frequency-tagging approach has generally been confined to study low-level sensory processes and always found related activation over the occipital region. Here for the first time, we investigated with it, high-level socio-cognitive functions, i.e. the processing of what other people are looking at which is referred to as level 1 visual perspective taking (VPT). Sixteen participants were presented with visual scenes alternating at 2.5 Hz which were depicting a person and an object in a room, while recording electrophysiological brain activity. The person orientation and object position changed at every stimulus but the person in the room always faced the object, except on every fifth stimulus. We found responses in the electroencephalography (EEG) spectrum exactly at the frequency corresponding to the presentation of the scenes where the person could not see the object, i.e. 0.5 Hz. While the 2.5 Hz stimulation rate response focused on typical medial occipital sites, the specific 0.5 Hz response was found mainly over a centro-parietal region. Besides a robust group effect, these responses were significant and quantifiable for most individual participants. Overall, these observations reveal a clear measure of level 1-VPT representation, highlighting the potential of EEG frequency-tagging to capture high-level socio-cognitive functions in the brain.

Key words: EEG; fast periodic visual stimulation; frequency tagging; visual perspective taking; theory of mind

Introduction

Consider the visual scene displays shown in Figure 1A: which ones are the odd ones out? Answering this question requires considering not only the position of the person in the scene and the left/right location of the object in the scene, but the specific *relationship* between the two: in three of the visual scenes only, the person looks away from the object. The present study uses an original electrophysiological frequency-tagging approach to provide neural evidence for the rapid representation of this relationship between another person and what he or she is looking at, i.e. a basic visual perspective-taking (VPT) ability, without direct contamination by overt responses.

Considering another person's viewpoint plays a fundamental role in our everyday life social interactions. Amongst the various forms of perspective-taking we can engage in, processing what the other person is looking at is particularly important as it provides valuable information to make more complex social inferences, such as inferring what the other person likes, intends to do, or talks about, for example. This basic VPT ability is usually referred to as level 1 VPT [contrary to level 2 VPT which consists in establishing 'how' a stimulus is perceived by someone else (Flavell *et al.*, 1981)]. Level 1 VPT is one of the earliest developing form of explicit perspective taking in children (Sodian *et al.*, 2007) and it is a perspective-taking ability that higher primates such as chimpanzees might share with humans (Hare *et al.*, 2000).

Human adults easily represent what another person can see when explicitly instructed to do so. Under certain circumstances, they also do it involuntarily. For example, Samson *et al.* (2010) asked participants to count how many dots they could see in a visual scene. A human avatar similar to those depicted in
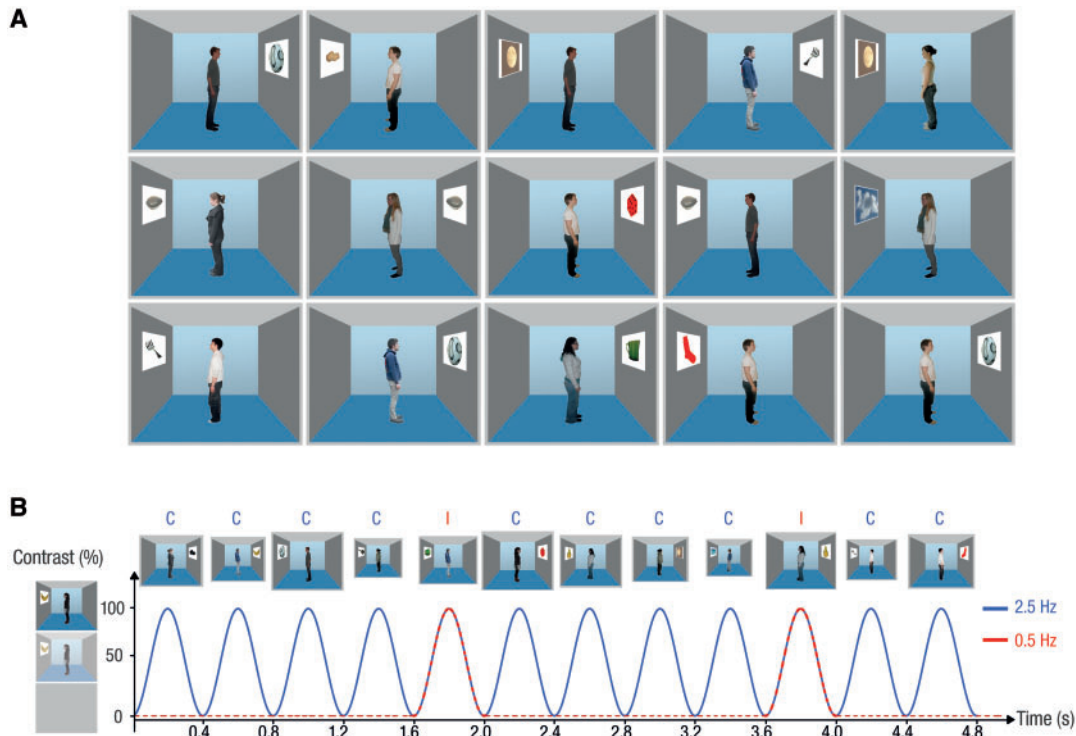
Fig. 1. Stimuli and paradigm. (A) Examples of stimuli used in the experiment. (B) Experimental paradigm (see also Video S1). Different stimuli appear at each stimulation cycle (no repetition within a stimulation sequence), with a randomly changed size to prevent pixel-based adaptation. Stimuli are presented by sinusoidal contrast modulation at the rate of 2.5 Hz (i.e. stimuli reached full contrast after 0.2 s). The embedded periodic response of interest was 0.5 Hz. Note that the side of the object location and the direction faced by the person could vary randomly at every cycle, and only the specific relationship between the two changed at 0.5 Hz. A full stimulation sequence lasted 94 s (84 s stimulation plus 5 s of fade-in and fade-out; see Supplementary Movie S1).

Figure 1 was present in the scene and could either see the same amount of dots visible to participants or could only see a subset of the dots. Participants were slowed down and more error prone in judging their own visual perspective when the avatar saw a different amount of dots than themselves (Baker *et al.*, 2015; Furlanetto *et al.*, 2016; Todd *et al.*, 2017). This observation indicates that participants involuntarily represented what the avatar was seeing. However, in this original paradigm, the signature of level 1 VPT is concomitant to the overt motor response. Isolating a neural marker of VPT that is independent from any motor response would not only be important for understanding better the information processing stages when making unconscious or conscious decisions about what others see, but may open new perspectives for measuring level 1 VPT in populations with immature or deficient decisional processes such as young children, individuals with cognitive deficits or non-human species without extensive training. The present study addresses this issue by developing a new marker of perspective taking. We measured level 1 VPT with scalp electrophysiology, taking advantage of a technique called 'frequency-tagging' or 'fast periodic visual stimulation' (FPVS), in which the process of interest is captured objectively (i.e. at the experimentally defined frequency) in the electroencephalography (EEG) frequency domain by Fourier Transform. Since its original demonstration by Adrian and Matthews (1934), this approach has essentially been confined to study low-level sensory processes ('Steady state visual evoked potentials', Regan, 1966), including their modulation by selective and spatial attention (e.g.Morgan *et al.*, 1996; Müller and Hubner, 2002; Keil *et al.*, 2005 Müller *et al.*, 2006). However, in recent years, changing high level visual properties at a periodic frequency rate and adapting this rate to the time-constant of the process of interest has shown that this highly

sensitive and objective approach can shed light on higher level brain functions, such as individual face perception (Rossion and Boremanse, 2011; Liu-Shuang *et al.*, 2014) or lexical processing (Lochy *et al.*, 2015). Here we extend this FPVS-EEG paradigm a step further to capture VPT processes in the human brain.

More specifically, we designed a paradigm based on Figure 1's display, in which we successively, rapidly (2.5 Hz) and continuously (i.e. for 84 s without interruption) presented visual scenes where the person faced an object. Every fifth stimulus, the person could not see the object. Participants were unaware of the ratio and periodic rate of the two types of visual scenes. We hypothesized to observe a differential EEG signal depending on whether the avatar could or could not see the object, i.e. a signal in the EEG spectrum corresponding exactly to the specific frequency at which the stimulus depicting a person who could not see the object was presented (i.e. 0.5–2.5 Hz/5). Furthermore, until now the scalp topography observed in studies using FVPS found activation on the occipital scalp region. However here, since previous studies highlighted the implication of parietal and frontal areas during VPT (e.g. McCleery *et al.*, 2011; Schurz *et al.*, 2015), we expected to find a more anterior neural response than the one usually found with FPVS. Such unique scalp activation would support the great interest of extending the use of the FPVS technique to investigate high-level cognitive functions.

Materials and methods

Participants

Sixteen healthy volunteers (female: 14, mean age: 21 ± 1.67, all right-handed) took part in the experiment in exchange of a

small honorarium. The number of participants was decided on the basis of the number of participants recruited in the original study by Samson *et al.* (2010) and on the basis of a previous EEG study investigating explicit VPT (McCleery *et al.*, 2011). All participants gave written informed consent prior to the experiment and the experiment received the approval from the ethics committee of the Psychological Sciences Research Institute. All experiments were performed in accordance with relevant guidelines and regulations.

Stimuli

The stimuli were created with Adobe Photoshop CS6. Sixteen photographs of the profile view of student volunteers (eight females) were taken, which all gave informed consent for publication of identifying images. The picture of the person was cropped and placed in the middle of a room, with the left, back, and right walls visible. The vanishing point of the room was localized at the level of the person's head. Half of the stimuli depicted the person looking at the left wall and the other half of the stimuli depicted the person looking at the right wall (Figure 1A).

Twenty objects with high familiarity and low visual complexity ratings were selected from the Moreno-Martínez and Montotro (2012) database and were displayed on either the left or right side wall. Two different sets of stimuli were made. For the first set, the objects were always placed in front of the person either on the right or the left wall (Figure 1A). For those stimuli, the person and the participant saw the object (i.e. *consistent perspective stimuli*). For the second set of stimuli, the objects were always placed on the wall behind the person. Here, the person and the participant had a different visual experience because only the participant could see the object (i.e. *inconsistent perspective stimuli*). Each person and each object were presented a similar amount of time. The color of two objects (the dice and the sock) was changed to red. These two objects were the only red objects (the other colors of the objects were yellow, green, black, gray, etc.).

Without any object and person displayed in the room, the luminance of the right and left walls was 127 cd/m^2, the luminance of the background wall was 199.55 cd/m^2 and the luminance of the floor was 117.37 cd/m^2. The size of the room was approximately 16.1° horizontal by 12.9° vertical, all pictures were displayed on a light gray background (198/255, 198/255 and 198/255), resulting in 641 × 480 pixel images (307 680 pixels). Note, however, that every displayed stimulus had a randomly varied image size (min: 227 683 pixels and max: 369 216 pixels) (see Supplementary Figure S1 for additional information about the size of the image).

Procedure

A stimulation sequence (or trial) consisted in a series of 235 pictures presented one after the other, at the relatively fast rate of 400 ms per picture (2.5 Hz; total trial duration: 94 s). We used slower frequency rate than generally used in previous studies because here, we investigated high-level cognitive function and assumed that stimuli need longer presentation to be processed. In the periodic stimulation condition, a trial consisted of a sequence composed of regular trains of four consistent perspective stimuli (C) followed by an inconsistent perspective stimulus (I) (i.e. CCCCICCCCICCCCI...). Hence, every fifth stimuli (i.e. 0.5 Hz), there was a change in perspective consistency. Critically, the side of the object (Right or Left, R or L)—and thus of the body orientation—was fully randomized, so that the inconsistent–consistent relationship concerned only the

relationship between the person and the object (object seen *vs* unseen) (e.g. LC-RC-RC-LC-*LI*-RC-RC-RC-LC-*RI*-RC, etc.) (Figure 1). In the nonperiodic stimulation condition, a trial consisted of a sequence composed of the exact same stimuli but shown in an entirely random order. In total, the same amount of consistent/inconsistent perspective stimuli were shown as in the periodic sequence (188 consistent perspective stimuli, 47 inconsistent perspective stimuli per trial). This control and nonperiodic sequence was used to ensure that any response of interest (0.5 Hz) was above noise level only due to the periodicity of the perspective change.

There were 8 different person's identity (4 females) and 10 different objects plus the 2 red objects (sock and dice) within one stimulation sequence. Furthermore, there were always the same amount of images with a male and a female person and the same amount of persons looking at the left or to the right wall equally spread across the two sexes.

During EEG recording, participants were seated in a light- and sound-attenuated room, at a viewing distance of 80 cm from an LED monitor (BenQ XL2420T) with a 1920 × 1080 resolution and a 120 Hz refresh rate. Stimuli were periodically presented through sinusoidal contrast modulation (Rossion and Boremanse, 2011) at a base frequency of 2.5 Hz (Figure 1).

Each stimulation sequence started with the onset of a fixation cross. After a random time period of 2 to 5 s, the stimulation gradually (i.e. linearly) appeared and reached full contrast after 5 s (contrast increase; fade-in 5 s). After 84 s, the scene gradually disappeared (contrast decrease; fade-out 5 s). The fade-in and fade-out periods (e.g. Liu-Shuang *et al.*, 2014) were used to prevent ocular artifacts and were not included in the analyses. In addition, to minimize low-level adaptation effects, every newly displayed stimulus had a randomly varied image size (min: 227 683 pixels and max: 369 216 pixels). The experiment started with one trial of periodic sequence and one trial of nonperiodic sequence (presented in counterbalanced order across participants) to practice the task before the EEG recording. Then, participants performed 16 test trials (8 periodic trials). Half of the participants started with a periodic trial and the other half started with a nonperiodic trial, and then we alternated trial by trial between these two conditions.

Participants received written instructions prior to the EEG experiment. They had to maintain fixation on the cross placed in the middle of the screen and had to respond by pressing the space bar as quickly and accuracy as possible. Their task was to detect the person viewing the red dice (12 targets per trial). They did not have to respond if the person saw any other object (including the red sock) or if the person could not see the red dice (because it was in his/her back). Thus, accurate performance in this task required detecting a specific visual experience of the person and the goal of this behavioral task was to ensure that participants maintained their attention to both stimuli throughout the trials. However, while participants had to monitor the other person's visual perspective to do the task, crucially, this task was not to respond whether the perspective was consistent or not. In both conditions (periodic and nonperiodic), stimuli depicting the person viewing the red dice were randomly inserted in the image sequences. Thus, correct behavioral responses (detecting the person viewing the red dice) were not periodic. Thereby, the EEG measure of interest was not concomitant with the motor response to the task instruction. A response was considered as correct if it was given within three standard deviations of the participant's mean RT.

EEG acquisition

EEG data were recorded using a 128 scalps channel at a sampling rate of 512 Hz with an ActiveTwoBiosemi electrode system (Biosemi, Amsterdam, Netherlands). Two additional electrodes, CMS (Common Mode Sense) and DRL (Driven Right Leg), were used as reference and ground. Each individual's electrode impedances were adjusted below $\pm 50\,\mu V$ prior to recording by injecting the electrode with a conductive gel (Signa). Eye-movements were recorded by four additional electro-oculograms around the subjects' eyes.

EEG preprocessing

Off-line analyses were performed with Letswave 5 (http://nocions.webnode.com/letswave) and Matlab 2012 (Math-Works, Natick, MA), similarly to previous FPVS studies that used this type of paradigm (e.g. Liu-Shuang *et al.*, 2014). Individual EEG data were first bandpass filtered between 0.1 and 120 Hz (zero phase shift Butterworth filter, order 4). Then, we reduced the sampling rate down to 256 Hz to save space and processing time. Noisy channels containing deflections larger than 200 μV were interpolated by the nearest neighboring electrodes. In addition, one subject had only two ocular electrodes instead of four due to a technical problem during the setting up. EEG data were subsequently segmented for each trial, 2 s prior to the start of the trial plus 2 s after the end (–2 s to 96 s). Afterwards, a common average reference computation was applied to all channels excluding ocular channels for each subject independently.

Frequency-domain analyses

Frequency-domain analyses were performed to quantify the magnitude of the response of the overall FPVS responses in each stimulation condition (periodic and nonperiodic). For that, each epoch was segmented a second time, right after the end of the fade-in to the end of the trial just before the beginning of the fade-out (21 504 time bins in total = 84 s). Then, the resulting segments were averaged separately for each participant and each periodic/nonperiodic trial. In addition, we also computed the grand average across participants and conditions. A fast Fourier transformation (FFT) was applied on the averaged segments to extract from the EEG signal, the amplitude spectra for each electrode. Thanks to the long-time windows (84 s), the frequency analysis yielded spectra with a high-frequency resolution (1/84 s, i.e. 0.0119 Hz) and allowed the unambiguous identification of the response at the exact frequencies of interest (i.e. 2.5 Hz for the base stimulation rate and 0.5 Hz for the perspective consistency discrimination response).

In order to correct for noise level for each participant's spectrum, at each frequency, we either subtracted (for baseline-corrected amplitude used for statistical analysis) or divided [for signal-to-noise ratio (SNR) used for the data visualization], the average voltage amplitude of the 10 surrounding bins (5 on each side, excluding the 2 immediately adjacent bins and the two extreme bins; e.g. Rossion *et al.*, 2012; Dzhelyova and Rossion, 2014). Here we used only 10 surrounding bins instead of 20 in previous studies since our periodic response of interest at 0.5 Hz is surrounded by frequency bins in the lowest part of the EEG spectrum, characterized by a high level of noise. Z-scores were also calculated to assess the significance of the response at specific frequencies (e.g. 0.5 Hz, 1 Hz, 1.5 Hz, etc.) by computing the difference between amplitude at the frequencies of interest and the mean amplitude of the 10 surrounding bins (excluding the 2 immediately adjacent bins and the two extremes bins) and dividing this difference by the standard deviation of the 10 surrounding bins (Rossion *et al.*, 2012). For the group analyses, individual baseline-corrected amplitude and SNR spectra were averaged separately for each periodic/nonperiodic trial per condition.

For further analysis, we determined a range of relevant harmonics (based on significant Z-scores) for each frequency based on the group-level data. Different regions-of-interest (ROIs) were also determined based on electrodes that showed a particularly large SNR on the average between periodic and nonperiodic trials.

Results

Behavioral data

All participants successfully performed the task (mean = 89% ± 3.54; false alarm = 14.08% ± 6.77; response time (RT) = 0.60 s ± 0.044). There was no significant difference of RT (periodic: 0.60 s ± 0.046; nonperiodic: 0.60 s ± 0.043; t (15) = –0.09, $p = 0.926$, Cohen's $d = –0.024$) and accuracy (periodic: 88.77% ± 4.80; nonperiodic: 89.17% ± 4.18; t (15) = –0.29, $p = 0.773$, Cohen's $d = –0.073$) between the periodic and nonperiodic trials, suggesting that the level of difficulty did not differ across the periodic and nonperiodic stimulation conditions.

EEG data

Discrimination between consistent and inconsistent perspectives. Inconsistent perspective stimuli correspond to images where the object was placed behind the person, so that the person could not see that object. In the periodic stimulation condition, these images were presented every fifth stimuli. Hence, neural discrimination between consistent and inconsistent perspectives should lead to a response at the exact periodic frequency of interest, which corresponds to 0.5 Hz (2.5 Hz/5) and to the harmonics in the EEG spectrum.

In order to determine the number of harmonics to take into account for further analysis, we pooled the 128 electrodes for each participant. We then computed the Z-score (computed as described above) for each harmonic on the grand averaged data across stimulation conditions (periodic and nonperiodic trials together). The highest significant harmonic was found at 3 Hz (sixth harmonic; threshold of significance placed at a Z-score of 1.65; P < 0.05, one-tailed, signal > noise).

We then cropped the grand average FFT spectrum, centered at the periodic frequency of interest (0.5, 1, 1.5, 2 and 3 Hz) with the 10 surrounded neighboring bins on each side. We excluded the fifth harmonic (2.5 Hz) because it corresponds to the base frequency rate. We summed the spectrum of the harmonics and their neighboring bins (Retter and Rossion, 2016), and computed the SNR (Figure 2A).

Based on the topographical map of the grand average of the combined (i.e. unbiased) periodic and nonperiodic conditions (Figure 2A), we selected two ROIs for further analysis. The main ROI was defined by 15 contiguous channels in the centro-parietal region (FCC1, FCC2, FCC1h, FCC2h, FCz, C1, C1h, C2, C2h, CCP1h, CCP2h, Cz, CCPz, CPz and CPPz), which were associated with the highest SNR on grand averaged data. Another ROI was defined by three channels in dorsolateral prefrontal region (AF8, F8 and FT8) associated with the highest SNR outside of the centro-parietal region.

As hypothesized, EEG spectra (Figure 2B) showed large and significant responses (based on the Z-score) at the frequency of interest and at its harmonics for each ROI, only in the periodic
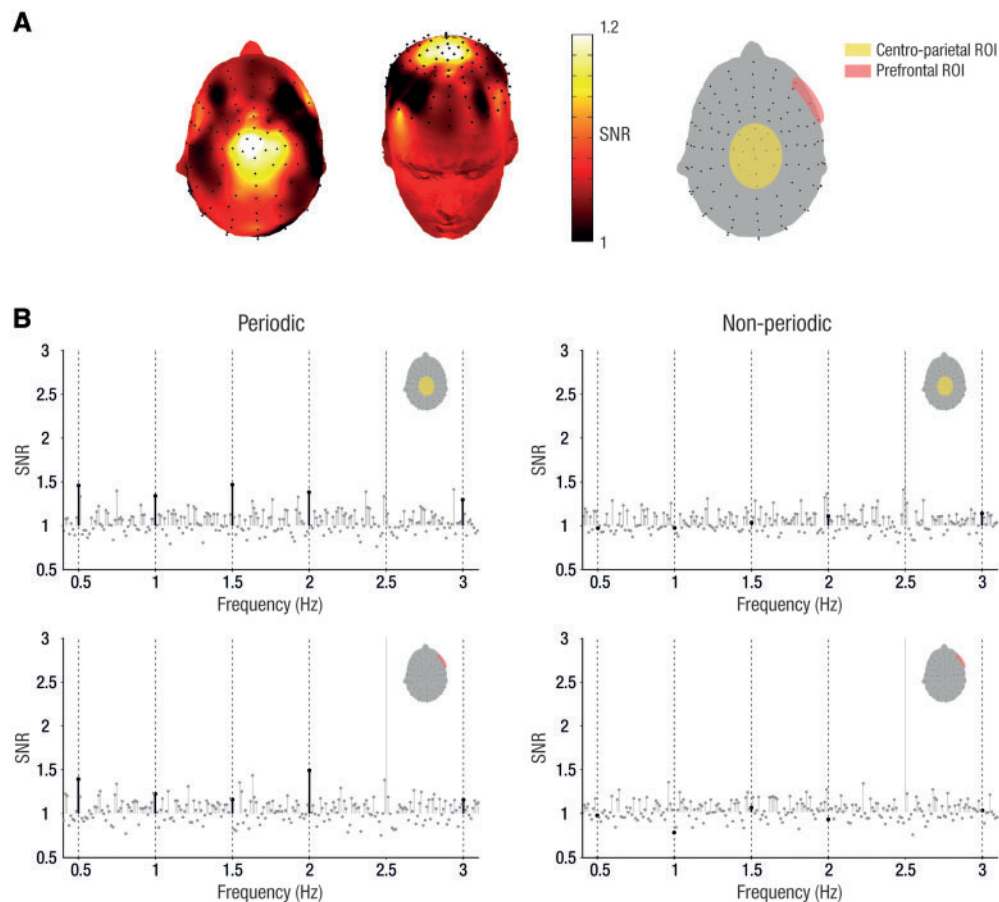
Fig. 2. Responses at the frequencies of interest. (A) Scalp topography (SNR) of the grand average of the periodic and nonperiodic trials of the sum of the 5th first harmonic (excluding 2.5 Hz) with the selected ROIs for further analysis. (B) SNR spectrum over the centro-parietal ROI and the frontal ROI of the periodic trials (left) and the nonperiodic trials (right).

condition (see Supplementary Table S1). Note that the results of the quantification analysis were independent of the number of harmonics included in the analysis (Supplementary Figure S2).

On one hand, we summed the spectrum of the harmonics and their neighboring bins and computed the SNR for the periodic and nonperiodic trials over the ROIs (see Supplementary Figure S3). The resulting Z-score were significant in the periodic trials only (one-tailed Z-score threshold set at 1.65, $P < 0.05$; frontal ROI: $Z = 7.59$ *vs* $Z = -2.54$ and centro-parietal ROI: $Z = 11.03$ *vs* $Z = 0.03$ for the periodic and nonperiodic trials, respectively).

On the other hand, we computed the baseline-corrected amplitude then summed the harmonics, and then we compared with one-sample t-test against 0 (i.e. noise level). t-Test showed a significant response only for the periodic condition (frontal ROI: t (15) = 5.35, $P = 0.001$, Cohen's $d = 1.34$ and t (15) = -1.63, $P = 0.125$, Cohen's $d = -0.41$, centro-parietal ROI: t (15) = 4.63, $P = 0.000$, Cohen's $d = 1.16$ and ROI: t (15) = 0.27, $P = 0.792$, Cohen's $d = 0.07$ for the periodic and nonperiodic trials, respectively). In summary, these results reveal an EEG signature of the detection of the change of perspective consistency.

Individual analysis. To evaluate the sensitivity of our paradigm, we tested whether each individual participant showed a discrimination response between the consistent and inconsistent perspective. To do so, for each participant and over each ROI, we cropped the FFT spectrum, centered at the periodic frequency of interest (0.5, 1, 1.5, 2 and 3 Hz) with the 10 surrounded neighboring

bins on each side. We summed the spectrum of the harmonics and their neighboring bins and computed the Z-scores and the SNR. The resulting Z-scores were significant in at least 1 of the 2 ROI for 14 participants of 16 (see Supplementary Table S2). Figure 3 shows an illustration of the individual results.

Base frequency analysis. As expected, we also found a response to the base frequency rate (i.e. the rate at which a new image was presented), merely reflecting the synchronization of the visual system to the visual stimulation. This response to the base frequency rate focused on medial occipital electrodes for both the periodic and nonperiodic trials (Supplementary Figure S4).

There was no significant difference in the baseline-corrected amplitude across the two stimulation conditions (periodic and nonperiodic; Supplementary Figure S4). Since frequency-tagged EEG responses are highly sensitive to spatial and selective attention (Morgan et al., 1996; Müller and Hillyard, 2000; Andersen et al., 2008; Walter et al., 2012; Müller et al., 2006), the absence of difference coupled with the equal behavioral performance suggest that participants paid equally attention to the stimuli in the two stimulation conditions.

Discussion

We found a neural signature at the exact frequency at which participants could witness the perspective of another person changing from being consistent with their perspective
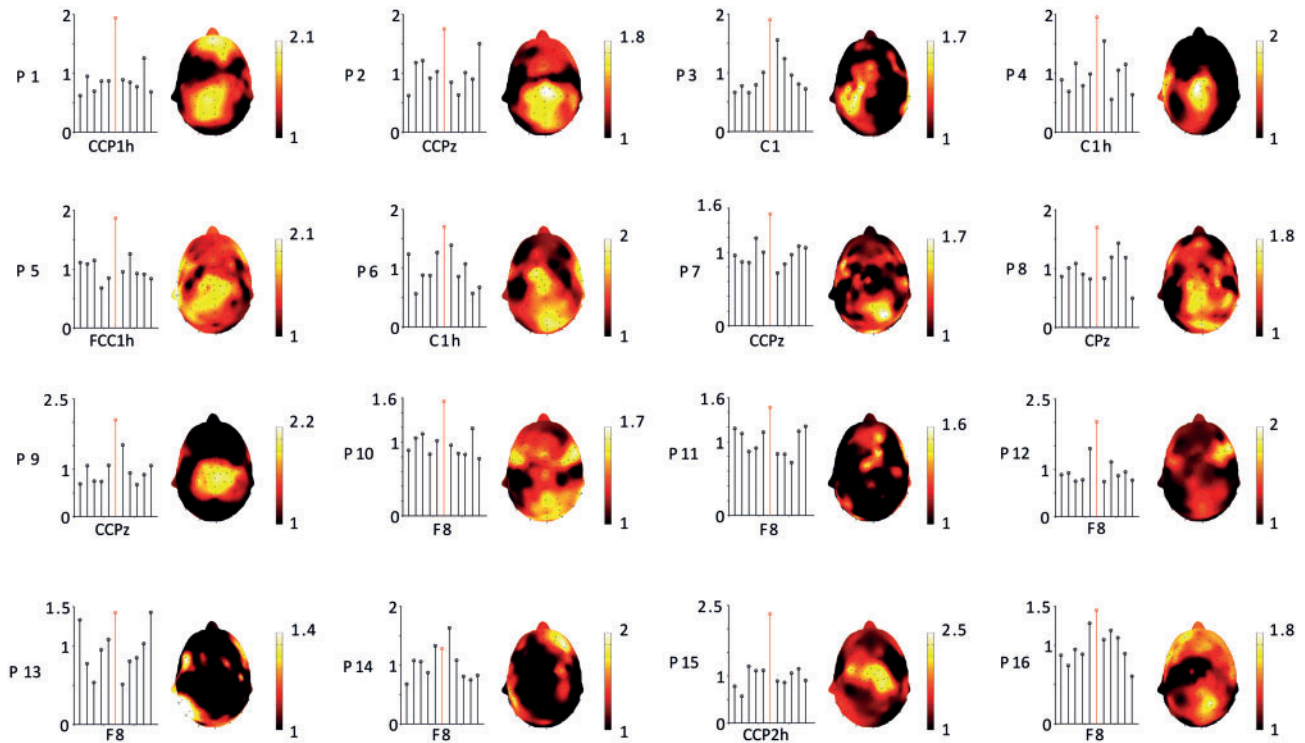
Fig. 3. Illustration of individual participants' SNR response to changes of perspective consistency over one electrode belonging to one ROI. The spectrum is centered at the response of the summed of the five significant harmonics, surrounded by the summed of the neighboring bins, which indicate noise levels (for visibility only the five neighboring bins on each side are shown).

(the other person saw the object on the wall that participants could also see) to becoming inconsistent (the other person could not see the object that participants could see). This EEG response was clear (i.e. well above noise level) and objective (i.e. occurring exactly at a frequency defined by the experimental design), and was significant in virtually all individual participants tested in the study. This process was measured at a relatively fast rate (i.e. every 2 s, with 400 ms/image) and, crucially, was not contaminated by a motor response. Since participants were explicitly asked to judge the avatar visual perspective throughout the experiment, the EEG measure reflects explicit perspective taking processes.

Besides a robust group effect, the responses were significant and easily quantifiable for most individual participants. The strength of the FPVS response may however vary according to factors (such as personality traits or clinical features) known to affect interindividual variability in perspective taking. This is an avenue for further investigation as FPVS may provide a more ecological and sensitive measure to capture such variability.

Virtually all EEG responses recorded so far in this type of fast periodic 'oddball' paradigm were localized over occipital and ventral occipito-temporal brain regions, reflecting the differential process of objects, faces, letters or words (e.g. Liu-Shuang *et al.*, 2014; Dzhelyova *et al.*, 2015; Lochy *et al.*, 2015). In contrast, the VPT neural response was found here over central and right prefrontal electrode sites. These scalp localizations are fully compatible with the known neurofunctional basis of VPT processing. The central topography may be due to processes generated in temporo-parietal brain areas, involved in the representation of the other person's perspective (Aichhorn *et al.*, 2006; Van Overwalle, 2009; McCleery *et al.*, 2011). Previous studies have also pointed to a role of the right prefrontal cortex in

inhibiting our own point of view to allow the selection of the other person's perspective when both perspectives are in conflict (Vogeley *et al.*, 2001; Samson *et al.*, 2005; McCleery *et al.*, 2011; Shibata and Inui, 2011; Hartwright *et al.*, 2015). However, a strong word of caution is required here as EEG alone is not suitable to make strong brain localization claims. Nevertheless, it is particularly interesting that the extension of the FPVS–EEG approach to higher-level cognitive processes reveals relatively anteriorly located responses on the scalp.

The current results do not inform about the nature of the information processed, and more specifically at which stage the information processed conforms to someone's mental state (e.g. the person can or cannot *see* the object). It is possible that at early stages of processing, only the spatial link between the other person and the object is computed and triggers a shift of attention to the object in front of the person. Future studies could address this issue with the FPVS–EEG approach by using obstacles in the other person's line of sight, similarly to what has been done in previous behavioral studies (Baker *et al.*, 2015; Cole *et al.*, 2016; Furlanetto *et al.*, 2016).

More generally, our study opens new ways to explore the neural basis of perspective taking in social cognition (Frith and Frith, 2007). For example, whether implicit and explicit perspective taking share the same type of processes is a highly debated question (Apperly and Butterfill, 2009; Heyes, 2014; Bardi *et al.*, 2016; Biervoye *et al.*, 2016; Grosse Wiesmann *et al.*, 2017). FPVS responses could be compared in the presence *vs* absence of explicit perspective taking instructions (without the need to change the nature of the stimuli) to provide further insights into this debate: will there be differences in terms of the strength of the EEG signal or is the topography on the scalp different? Furthermore, should future results show the presence

of a FPVS response even in the absence of explicit instructions, this could provide a powerful measure to investigate perspective taking in infants (see for example FPVS responses to faces in infants, de Heering and Rossion, 2015), nonhumans, perhaps even allow cross-ages or cross-species comparisons. Indeed, the strength of the technique relies in the fact that it only requires watching a train of images.

Acknowledgements

Funding

Author contributions

All authors contributed to the study concept and design. Testing, data collection and analysis were performed by A.B. All authors contributed to interpretation of the data. A.B. drafted the manuscript and D.S. and B.R. provided critical revisions. All authors approved the final version of the manuscript for submission.

Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

References

Adrian, E.D., Matthews, B.H.C. (1934). The interpretation of potential waves in the cortex. *The Journal of Physiology*, **81**(4), 440–71.

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, **30**(3), 1059–68.

Andersen, S.K., Hillyard, S.A., Müller, M.M. (2008). Attention facilitates multiple stimulus features in parallel in human visual cortex. *Current Biology*, **18**(13), 1006–9.

Apperly, I., Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, **116**(4), 953–70.

Baker, L.J., Levin, D.T., Saylor, M.M. (2015). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance*, **42**(4), 508–16.

Bardi, L., Desmet, C., Nijhof, A., Wiersema, J.R., Brass, M. (2016). Brain activation for spontaneous and explicit false belief tasks overlaps: new fMRI evidence on belief processing and violation of expectation. *Social Cognitive and Affective Neuroscience*, 2017 **12**(3), 391–400.

Biervoye, A., Dricot, L., Ivanoiu, A., & Samson, D. (2016). Impaired spontaneous belief inference following acquired damage to the left posterior temporoparietal junction. *Social Cognitive and Affective Neuroscience*, **11**(10), 1513–20.

Cole, G.G., Atkinson, M.A., Le, A.T.D., Smith, D.T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, **164**, 165–8.

de Heering, A., Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife*, **4**, 1–14.

Dzhelyova, M., Jacques, C., Rossion, B. (2015). At a single glance: fast periodic visual stimulation uncovers the spatio-temporal dynamics of brief facial expression changes in the human brain. *Cerebral Cortex*, **47**(1), 96–106.

Dzhelyova, M., Rossion, B. (2014). Supra-additive contribution of shape and surface information to individual face discrimination as revealed by fast periodic visual stimulation. *Journal of Vision*, **14**(14), 15.

Flavell, J.H., Everett, B. a., Croft, K., Flavell, E.R. (1981). Young children's knowledge about visual perception: further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, **17**(1), 99–103.

Frith, C.D., Frith, U. (2007). Social cognition in humans. *Current Biology*, **17**(16), 724–32.

Furlanetto, T., Becchio, C., Samson, D., Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, **42**(2), 158–63.

Grosse Wiesmann, C., Friederici, A.D., Singer, T., Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, **20**(5), 1–15.

Hare, B., Call, J., Agnetta, B., Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, **59**(4), 771–85.

Hartwright, C.E., Apperly, I., Hansen, P.C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, **67**, 183–92.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, **9**(2), 131–43.

Keil, A., Moratti, S., Sabatinelli, D., Bradley, M.M., Lang, P.J. (2005). Additive effects of emotional content and spatial selective attention on electrocortical facilitation. *Cerebral Cortex (New York, N.Y.: 1991)*, **15**(8), 1187–97.

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

Lochy, A., Belle, G.V., Rossion, B. (2015). A robust index of lexical representation in the left occipito-temporal cortex as evidenced by EEG responses to fast periodic visual stimulation. *Neuropsychologia*, **66**, 18–31.

McCleery, J.P., Surtees, A., Graham, K.A., Richards, J.E., Apperly, I. (2011). The neural and cognitive time course of theory of mind. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **31**(36), 12849–54.

Moreno-Martínez, F.J., Montoro, P.R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS One*, **7**(5), e37527.

Morgan, S.T., Hansen, J.C., Hillyard, S.A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(10), 4770–4.

Müller, M.M., Andersen, S., Trujillo, N.J., Valdes-Sosa, P., Malinowski, P., Hillyard, S.A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(38), 14250–4.

Müller, M.M., Hillyard, S.A. (2000). Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, **111**(9), 1544–52.

Müller, M.M., Hubner, R. (2002). Can the spotlight of attention be shaped like a doughnut? Evidence from steady-state visual evoked potentials. *Psychological Science*, **13**(2), 119–24.

Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, **20**(3), 238–48.

Retter, T.L., Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, **91**, 9–28.

Rossion, B., Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision*, **11**(2), 16.

Rossion, B., Prieto, E.A., Boremanse, A., Kuefner, D., Van Belle, G. (2012). A steady-state visual evoked potential approach to individual face perception: effect of inversion, contrast-reversal and temporal dynamics. *NeuroImage*, **63**(3), 1585–600.

Samson, D., Apperly, I.A., Kathirgamanathan, U., Humphreys, G.W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, **128**(5), 1102–11.

Samson, D., Apperly, I., Braithwaite, J.J., Andrews, B.J., Bodley Scott, S.E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, **36**(5), 1255–66.

Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: issues of spontaneity and domain-specificity. *NeuroImage*, **117**, 386–96.

Shibata, H., Inui, T. (2011). Brain activity associated with recognition of appropriate action selection based on allocentric perspectives. *Neuroscience Letters*, **491**(3), 187–91.

Sodian, B., Thoermer, C., Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science*, **10**(2), 199–204.

Todd, A.R., Cameron, C.D., Simpson, A.J. (2017). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, **159**, 97–101.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, **30**(3), 829–58.

Vogeley, K., Bussfeld, P., Newen, A., *et al.* (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage*, **14**(1), 170–81.

Walter, S., Quigley, C., Andersen, S.K., Müller, M.M. (2012). Effects of overt and covert attention on the steady-state visual evoked potential. *Neuroscience Letters*, **519**(1), 37–41.