

## Dyadic interaction processing in the posterior temporal cortex

Jon Walbrin<sup>\*</sup>, Kami Koldewyn

School of Psychology, Bangor University, Wales, UK



### ARTICLE INFO

#### Keywords:

Social interaction  
fMRI  
EBA  
pSTS  
LOTc

### ABSTRACT

Recent behavioural evidence shows that visual displays of two individuals interacting are not simply encoded as separate individuals, but as an interactive unit that is 'more than the sum of its parts'. Recent functional magnetic resonance imaging (fMRI) evidence shows the importance of the posterior superior temporal sulcus (pSTS) in processing human social interactions, and suggests that it may represent human-object interactions as qualitatively 'greater' than the average of their constituent parts. The current study aimed to investigate whether the pSTS or other posterior temporal lobe region(s): 1) Demonstrated evidence of a dyadic information effect - that is, qualitatively different responses to an interacting dyad than to averaged responses of the same two interactors, presented in isolation, and; 2) Significantly differentiated between different types of social interactions.

Multivoxel pattern analysis was performed in which a classifier was trained to differentiate between qualitatively different types of dyadic interactions. Above-chance classification of interactions was observed in 'interaction selective' pSTS-I and extrastriate body area (EBA), but not in other regions of interest (i.e. face-selective STS and mentalizing-selective temporo-parietal junction). A dyadic information effect was not observed in the pSTS-I, but instead was shown in the EBA; that is, classification of dyadic interactions did not fully generalise to averaged responses to the isolated interactors, indicating that dyadic representations in the EBA contain unique information that cannot be recovered from the interactors presented in isolation. These findings complement previous observations for congruent grouping of human bodies and objects in the broader lateral occipital temporal cortex area.

### 1. Introduction

Social interactions are ubiquitous, yet little research has investigated visual perceptual responses to these common social scenarios, relative to individual-person perception (Quadflieg and Koldewyn, 2017). Interestingly, recent behavioural evidence demonstrates that visual responses to two human individuals that are positioned to imply an interaction evoke different responses than when not positioned in this manner. These effects are demonstrated most strikingly by the findings of Papeo et al. (2017): In this study, subjects viewed pairs of briefly presented (30 ms) human bodies or control objects (i.e. chairs), that either faced *towards* or *away* from each other, in either upright or inverted orientation, and were instructed to respond to the stimulus category they saw (i.e. bodies or chairs). Greater recognition accuracy was shown for upright than inverted dyads when an interaction was implied by the two bodies facing towards each other, but crucially, not when facing away from each other. Similarly, visual search facilitation is shown for full body dyads that are positioned to *face towards* – rather than *away from* – each other (Vestner

et al., 2018), while facing direction effects are shown to modulate the evaluation of facial emotion of a target face (i.e. the perceived emotional expression of a target face is modulated by the emotion of a simultaneously presented non-target face, but only when positioned to *face towards* the target; Gray et al., 2017).

Together, these behavioural findings demonstrate that interacting individuals are not merely perceived as separate individuals, but as an *interactive dyad*. Indeed, similar *non-linear* neural responses have been observed recently – that is, *that responses to dyadic interaction stimuli are not the same as a linear combination of responses to the isolated elements of an interaction*. Specifically, Baldassano et al. (2017) demonstrated evidence of non-linear responses to human-object interaction stimuli in the posterior temporal cortex; the authors used a pattern classification approach to test whether responses to images of human-object interactions (e.g. a person pushing a shopping cart) are distinct from *the mean-averaged response* to the constituent parts of the interaction (i.e. the averaged response to an isolated human and isolated cart); it was found that voxel patterns for human-object interactions in the posterior superior temporal

<sup>\*</sup> Corresponding author.

E-mail address: [jon.walbrin@gmail.com](mailto:jon.walbrin@gmail.com) (J. Walbrin).

<https://doi.org/10.1016/j.neuroimage.2019.05.027>

Received 18 January 2019; Received in revised form 4 April 2019; Accepted 10 May 2019

Available online 14 May 2019

1053-8119/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

sulcus (pSTS) and lateral occipital cortex (LOC) were statistically distinct from the averaged patterns evoked by isolated ‘interaction parts’. These findings suggest that these regions are sensitive to *unique* interactive information that is accessed only through holistic processing of interactions, and not through part-wise analysis (i.e. processing of constituent ‘interaction parts’ in isolation).

Interestingly, this response in the pSTS complements previous findings that this region plays an important role in the visual processing of dynamic social interactions; for example, greater pSTS responses are shown for interacting point-light human dyads relative to two non-interacting figures, as well as for similar stimuli depicted by moving geometric shapes that do not contain body information (Isik et al., 2017; Walbrin et al., 2018). This region also differentiates between types of interactions performed by live-action human stimuli (Sinke et al., 2010), and is sensitive to ‘interactive’ motion cues such as the movement contingency between two interacting human figures (Georgescu et al., 2014), or the degree of correlated motion between interacting animate geometric shapes (Gao et al., 2009; Schultz, Friston, O’Doherty, Wolpert and Frith, 2005). These findings implicate the pSTS as a region that may be optimized for processing social interaction information.

The main aim of the present study was to determine whether pSTS encodes dynamic human interactions between two individuals in a non-linear fashion, using a similar approach to Baldassano et al. (2017). We herein adopt the phrase ‘*dyadic information effect*’ rather than ‘non-linear’ effect, to emphasize a *sensitivity to unique information that is only present in dyadic interactions and not the averaged responses evoked by each interactor, presented in isolation*. Specifically, we used support vector machine (SVM) classification to test whether voxel-pattern responses to dyadic stimuli in the pSTS were statistically differentiable from *averaged response patterns* of isolated interactors. Additionally, it was predicted that significantly

differentiable responses to *different types* of dyadic interaction would be observed in the pSTS, replicating previous findings (e.g. Isik et al., 2017; Walbrin et al., 2018). Responses were also tested in 3 other functionally localized regions of interest (ROIs) that are selective for social information that likely contributes to social interaction processing, and therefore might also plausibly show the hypothesized effects: Extrastriate body area (EBA), mentalizing-selective temporo-parietal junction (TPJ-M), and face-selective STS (STS-F).

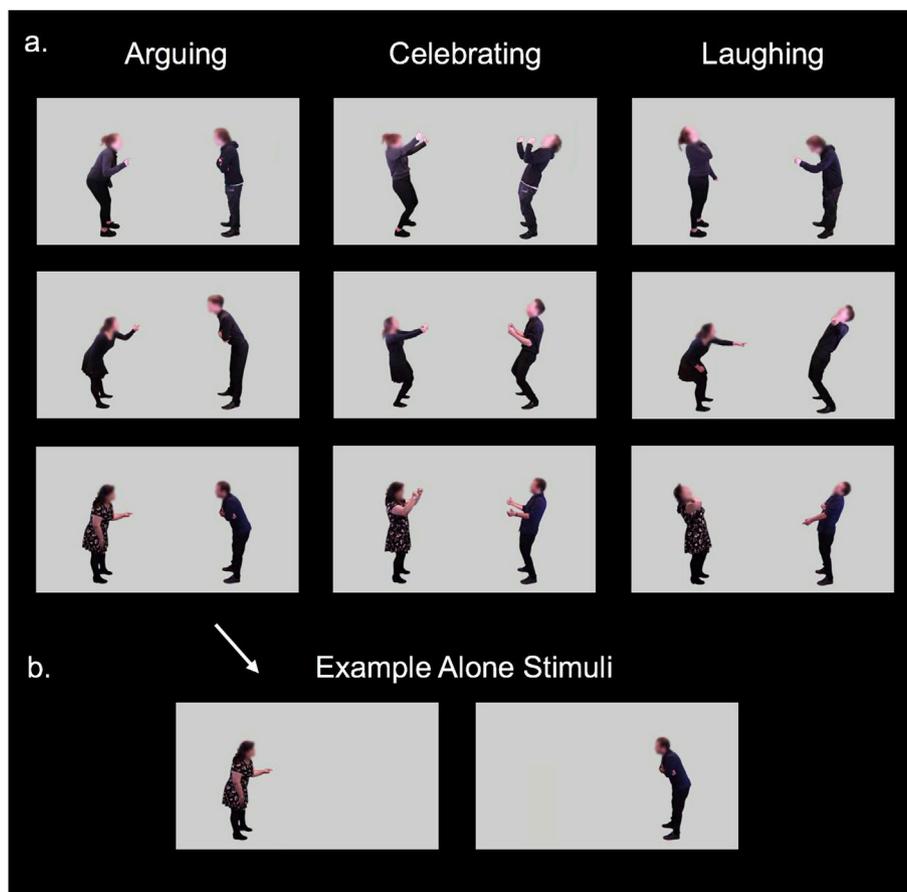
## 2. Material & methods

### 2.1. Participants

21 right-handed adults (mean age = 23.40 years; SD = 3.74; range = 18–35; 12 females) participated in the study. Participants gave informed consent and received monetary compensation for taking part. Ethical procedures were approved by the Bangor University psychology ethics board.

### 2.2. Stimuli

Stimuli consisted of 4 s (s) video clips that were taken from custom footage of paired actors engaging in semi-improvised interactions. Actors were instructed to improvise these scenarios while enacting scripted ‘*action-gestures*’; for example, for a given arguing scenario, one actor might be instructed to *point angrily* at the other person while the other *shook their fists* in frustration. Therefore, each interaction depicted two individuals performing a given pair of complementary action-gestures that they were encouraged to enact in a natural, authentic way (see supplementary materials A for example videos). An initial set of *dyad*



**Fig. 1.** a. Example video frames from the dyad versions of the three interaction scenarios. Each row represents one of three unique female-male interactor pairs. b. Two example alone stimuli (created from a given dyad stimulus).

stimuli were created (along with a separate set of *alone stimuli*, as described below; see Fig. 1 for examples of both dyad and alone stimuli). Dyad stimuli depicted two actors engaging in one of 3 *interactive scenarios*: *Arguing* (i.e. both actors engaging in an angry/frustrated confrontation), *celebrating* (i.e. both actors celebrating together, excitedly), and *laughing* (i.e. both actors were laughing together, or at each other). These specific scenarios were chosen for the ‘tonal consistency’ of actions performed by a given pair of interactors, such that the intentions, emotions, and valence information conveyed by both individuals in a given scenario were always similar (e.g. angry/frustrated) rather than contrasting (e.g. angry/sad). This ensured that successful classification of the different scenarios was not driven by systematic differences in intentional, emotional, or valence content *between* interactors. Therefore, these scenarios represented three interactive scenarios that were intended to be easily distinguishable.

Within each interaction scenario (e.g. arguing), 4 exemplar videos were created, each using a unique pair of action-gestures, such that each video showed the two individuals performing a complementary pair of action-gestures (e.g. while arguing, interactor A accusatorily points at interactor B who is shaking their hands in frustration). Importantly, no gestures were ‘reused’ in any of the other action-gesture pairings (i.e. a total of 8 action gestures were used across the 4 exemplar videos for each scenario). Similarly, 3 different female-male *interactor pairs* enacted these scenarios, yielding a total of 36 dyad stimuli: 3 interaction scenarios (arguing, celebrating, laughing)  $\times$  4 unique action-gesture pairings  $\times$  3 interactor pairs. The final stimuli were chosen from a wider set of stimuli based on the highest ‘interactive-ness’ and ‘naturalness’ ratings from a pilot study ( $N = 10$ ; see supplementary materials A).

For these stimuli, the average horizontal distance between actors was closely matched – the visual angle between the centre of each actor’s torso was approximately  $4.80^\circ$ , and actor height ranged between  $3.73$  and  $4.26^\circ$ . As dynamic facial information is known to activate the STS (e.g. Deen et al., 2015), the presence of facial information was controlled such that classification could not be attributed to different facial expressions. Accordingly, these stimuli did not contain high spatial frequency face information, but body information was preserved. To achieve this, a circle-shaped Gaussian blur mask was placed on each of the actors’ heads for each video frame. This preserved the overall shape of the head, preventing the potentially eerie appearance of headless interacting bodies.

To test neural responses to the same interactive information – but without specifically *dyadic information* (i.e. information available from two interactors presented simultaneously) – a separate set of 72 *alone stimuli* were created by removing either individual from each of the 36 dyad stimuli (see Fig. 1b for examples of two alone stimuli). It is important to note that although these stimuli depicted an isolated interactor by themselves, they still conveyed interactive information (e.g. communicative gesturing towards an implied interactor). Two horizontally-flipped variants of these 108 unique stimuli (36 dyad + 72 alone stimuli) resulted in a final set of 216 stimuli.

### 2.3. Design & procedure

A rapid event-related design was used, and each run was optimized using optseq2 (<http://surfer.nmr.mgh.harvard.edu/optseq>), based on differentiating 6 conditions (i.e. both dyad and alone variants of the arguing, celebrating, and laughing interaction scenarios), with an inter-stimulus interval range between 0 and 10s (along with 8s fixation at the beginning of each run, and 16 s at the end to capture most of the haemodynamic response). The 6 designs with the highest detection sensitivity were selected to determine event timings for runs.

Inside the scanner, participants viewed stimuli that were presented centrally on the screen within a  $9.17 \times 5.11^\circ$  rectangular space. 6 runs were completed, each lasted exactly 7 minutes and contained 8 stimuli for each dyad version and 16 stimuli for each alone version of each of the 3 scenarios, resulting in 72 experimental stimuli per run. Three important

stimulus ordering considerations are also noted here: Firstly, left and right horizontal presentations of each stimulus were balanced within the design, such that any resulting effects could not be attributed to low-level confounds in the horizontal position of interactors (i.e. left and right horizontally-flipped variants of the stimuli appeared equally often); secondly, that any given pair of alone stimuli (i.e. that originated from the same dyad stimulus) were always presented in the same run as each other so that classification of alone stimuli did not contain additional between-run variance that was not present for the dyad stimuli; thirdly, to minimize repetition effects (i.e. seeing the exact same action-gestures from a given dyad stimulus and the corresponding pair of alone stimuli), alone stimuli that appeared in any given run were always from dyad stimuli that were allocated to a different run.

In addition to the stimuli already described, nine additional catch stimuli were presented (three dyad stimuli, and six alone stimuli) but were not later analysed. These trials contained a ‘frame-freeze’ in which 12 consecutive video frames (duration = 500 ms) were randomly removed from the video and replaced with one repeated frame for that period, creating the impression of a momentary video pause. Participants were instructed to simply watch the videos and to give a button-press response whenever a frame-freeze was detected, and to refrain from making explicit judgements about the interactors.

### 2.4. Localizer tasks & ROI creation

Participants completed several localizer tasks in a separate scanning session, on a separate day (see supplementary materials B for full description of these tasks). Briefly explained, three different video tasks were used to localize brain regions that are sensitive to different types of social information: 1) A point-light figure social interaction task similar to that used previously (Isik et al., 2017; Walbrin et al., 2018) was used to localize interaction-selective pSTS (pSTS-I) regions of interest (ROI) with the interaction > scrambled interaction contrast (i.e. two intact human figures interacting vs. spatially scrambled versions of the same stimuli in which body and interactive information was disrupted). 2) A dynamic body and face localizer that was adapted from stimuli used previously (Pitcher et al., 2011) – this served to localize body-selective EBA and face-selective STS cortex (i.e. STS-F), with the bodies > objects, and faces > objects contrasts, respectively. 3) A free-viewing animated film (‘Partly Cloudy’; Pixar Animation Studios: <https://www.pixar.com/partly-cloudy>) identical to that used previously (Richardson et al., 2018) was used to localize mentalizing-selective TPJ-M with the mentalizing > pain contrast (i.e. mentalizing > pain time-points).

These tasks allowed for the localization of 4 bilateral subject-specific ROIs (i.e. pSTS-I, EBA, STS-F, & TPJ-M; see supplementary materials C for a visualization of these ROIs). These ROIs were created with a group-constrained definition procedure (e.g. Julian et al., 2012) as follows. For a given subject and contrast (e.g. interaction > scrambled interaction, for the pSTS-I), a 5 mm-radius ‘search sphere’ was created by running a whole-brain analysis for  $N-1$  group subjects (i.e. with the ‘current’ subject excluded) and centring the sphere at the peak voxel (i.e. highest  $t$ -value) in the designated region. This relatively small sphere was chosen to ensure subject’s ROIs did not deviate too far from a given designated anatomical region (e.g. pSTS). To determine the position of the final ROI, a whole-brain analysis for the current subject (for the same contrast) was run, and resulting activation was constrained to the search sphere. A 7 mm-radius sphere was then centred at the peak voxel in this search region; this ROI sphere size was chosen as an ideal compromise between capturing a relatively large number of voxels that would benefit classification performance (e.g. Coutanche et al., 2016), and ensuring minimal overlap between neighbouring STS ROIs.

All ROIs contained 179 voxels, with the exception of two subjects that had small regions of overlap between the right pSTS-I and right TPJ-M, and a further two subjects with similar overlap between the right pSTS-I and right STS-F. Across these four subjects, a mean overlap of 18 voxels (range: 12–24) was found. To ensure independence of ROI voxels

within each of these four subjects, overlapping voxels were removed and ROIs were recreated (respective final ROI sizes for these four subjects were: 167, 161, 161, 155 voxels; all other ROIs for these subjects contained 179 voxels).

### 2.5. MRI parameters, pre-processing, & GLM estimation

Scanning was performed with a Philips 3T scanner at Bangor University. Functional images were acquired with the following parameters: T2\*-weighted gradient-echo single-shot EPI pulse sequence; TR = 2000 ms, TE = 30 ms, flip angle = 83°, FOV(mm) = 240 × 240 × 108, acquisition matrix = 80 × 78 (reconstruction matrix = 80); 36 contiguous axial slices were acquired, with a reconstructed voxel size of 3 mm<sup>3</sup>. Four dummy scans were discarded prior to image acquisition for each run. Structural images were obtained with the following parameters: T1-weighted image acquisition using a gradient echo, multi-shot turbo field echo pulse sequence, with a five echo average; TR = 12 ms, average TE = 3.4 ms, in 1.7 ms steps, total acquisition time = 136s, FA = 8°, FOV = 240 × 240, acquisition matrix = 240 × 224 (reconstruction matrix = 240); 128 contiguous axial slices, acquired voxel size (mm) = 1.0 × 1.07 × 2.0 (reconstructed voxel size = 1 mm<sup>3</sup>).

Pre-processing was performed with SPM12 (fil.ion.ucl.ac.uk/spm/software/spm12). This entailed slice-timing correction, re-alignment (and re-slicing), co-registration, segmentation, normalization, and smoothing. All default parameters were used except for a 6 mm FWHM Gaussian smoothing kernel. General linear model (GLM) estimation was performed in SPM12 on participants' normalized images. For the main task, whole-brain beta maps were generated on a run-wise basis with events estimated as 6 *classification conditions* – both dyad and alone variants of the arguing, celebrating, and laughing stimuli. One further set of maps were created where each event was modelled separately, to allow for stimulus-wise analyses (see supplementary materials D).

### 2.6. SVM classification analyses

Leave-one-run-out linear support vector machine (SVM) classification was implemented with CoSMoMvPA (Oosterhof et al., 2016). Briefly explained, for a given subject, an SVM classifier was trained on ROI voxels (i.e. beta values) for the conditions of interest (e.g. dyad variants of the arguing, celebrating, and laughing conditions) in all but one run of data – with the 'left-out' run of data used to independently test classification performance on. This was iterated 6 times with each run serving as the left-out test run, and classification accuracy was averaged across iterations. These values were then entered into group level *t*-tests. All reported tests were significant at the corrected Bonferroni threshold ( $\alpha$ ) unless otherwise stated. A different threshold was calculated separately for each set of analyses (i.e. based on 8, 8, & 4 comparisons for dyad, alone, and cross-classification analyses, respectively), as stated in each sub-section in the results. All *t*-test *p*-values are one-tailed.

This approach was almost identical for both 'standard' classification (e.g. between the three dyad conditions, or between the three alone conditions) and *cross-classification* analyses except that the allocation of training and test conditions differed; that is, for cross-classification, the classifier was trained on the three dyad conditions, but *tested* on the three alone conditions. Significant cross-classification demonstrates that the patterns underlying the two sets of conditions are similar to each other, and therefore are largely driven by the same information. However, we reasoned that if a region showed significantly greater dyad classification than cross-classification (i.e. between dyad and alone conditions), this would indicate sensitivity to dyadic information that could not be 'recovered' from the individual interactors presented in isolation (i.e. averaged responses to alone stimuli). As explained previously (see section 2.3) several stimulus ordering constraints were imposed within each run, and importantly, alone stimuli from a given dyad stimulus were always presented in a different run to minimize repetition effects. Notably, this likely resulted in a *more conservative* estimation of the

dyadic information effect due to greater similarity between stimuli in test and train data splits for cross-classification, than for 'standard' classification (see supplementary materials E for further details).

## 3. Results

### 3.1. SVM classification analyses

For each of the 8 functionally localized ROIs, a series of analyses were performed in which a linear SVM classifier was trained and tested on different variants of the 3 interaction scenarios (i.e. arguing, celebrating, and laughing). One-sample *t*-tests were used to determine whether classification accuracy was above chance level (i.e. 100% / 3 categories = 33.3% chance accuracy; Bonferroni corrected  $\alpha = 0.006$ ).

Significant above-chance classification of the three interaction scenarios of dyad stimuli (see Fig. 2) was observed in the right pSTS-I (Classification accuracy (%):  $M = 41.39$ ,  $SD = 9.10$ ;  $t(19) = 3.96$ ,  $p < .001$ ) and both the right EBA ( $M = 49.38$ ,  $SD = 12.19$ ;  $t(17) = 5.59$ ,  $p < .001$ ) and left EBA ( $M = 50.88$ ,  $SD = 13.00$ ;  $t(18) = 5.88$ ,  $p < .001$ ), and at an uncorrected threshold in the left pSTS-I ( $M = 38.60$ ,  $SD = 10.55$ ;  $t(18) = 2.17$ ,  $p = .022$ ). None of the 4 other ROIs – bilateral STS-F and TPJ-M – showed above-chance classification of the dyad stimuli (all  $ps > .100$ ; see Fig. 3; see supplementary materials F for full statistics).

It is possible that significant classification of dyad stimuli in the bilateral pSTS-I and EBA does not completely rely on inherently dyadic information, and may also encode information conveyed by isolated individuals (e.g. interactive gestures directed towards an implied – but physically absent – interaction partner). To test if this was true, another classification analysis (Bonferroni corrected  $\alpha = 0.006$ ) was run to see if these regions could differentiate the three interaction scenarios for the alone stimuli (see Figs. 2 and 3). It is worth reiterating that the *same overall information* was present as in the dyad classification analysis (i.e. same scenarios, actors, & gestures). Above-chance classification was shown in right pSTS-I ( $M = 43.33$ ,  $SD = 12.57$ ;  $t(19) = 3.56$ ,  $p = .001$ ) but only marginally in left pSTS-I ( $M = 37.43$ ,  $SD = 12.81$ ;  $t(18) = 1.39$ ,  $p = .090$ ). Both right EBA ( $M = 46.30$ ,  $SD = 7.86$ ;  $t(17) = 7.00$ ,  $p < .001$ ), and left EBA ( $M = 46.49$ ,  $SD = 6.73$ ;  $t(18) = 8.52$ ,  $p < .001$ ) also showed significant classification. As for dyad classification, bilateral STS-F and TPJ-M ROIs did not show above-chance classification (all  $ps > .088$ ), and therefore, these regions were excluded from further analyses.

Together, these two classification analyses demonstrate interaction sensitive responses in the right pSTS-I and bilateral EBA regions, and to a marginal extent in the left pSTS-I; specifically, these regions were able to differentiate between the three different interaction scenarios, both when observing an intact dyad and when observing the same constituent interactors presented in isolation. However, although these regions are sensitive to both modes of presentation, this does not mean that the underlying information driving classification in both dyadic and alone scenarios is the same (e.g. information about the spatial-relations between interactors may contribute to classification of the dyad stimuli, but not the alone stimuli). Indeed, if voxel pattern classification in any region does not fully generalise from dyad stimuli to the alone stimuli, this would suggest that there is information encoded by these regions during dyadic interaction perception that cannot be recovered by the same information presented in the alone stimuli.

Next, a cross-classification analysis was implemented (Bonferroni corrected  $\alpha = 0.013$ ) whereby an SVM classifier was trained to discriminate responses to the three interaction scenarios with the dyad stimuli, but performance was tested on responses to the alone stimuli. Significant cross-classification was shown for all 4 ROIs (right pSTS-I:  $M = 41.39$ ,  $SD = 8.92$ ;  $t(19) = 4.04$ ,  $p < .001$ ; left pSTS-I:  $M = 40.64$ ,  $SD = 9.63$ ;  $t(18) = 3.31$ ,  $p = .002$ ; right EBA:  $M = 43.21$ ,  $SD = 7.75$ ;  $t(17) = 5.40$ ,  $p < .001$ ; left EBA:  $M = 46.20$ ,  $SD = 11.27$ ;  $t(18) = 4.97$ ,  $p < .001$ ), demonstrating that these regions encode similar information in both the dyad and alone stimuli.

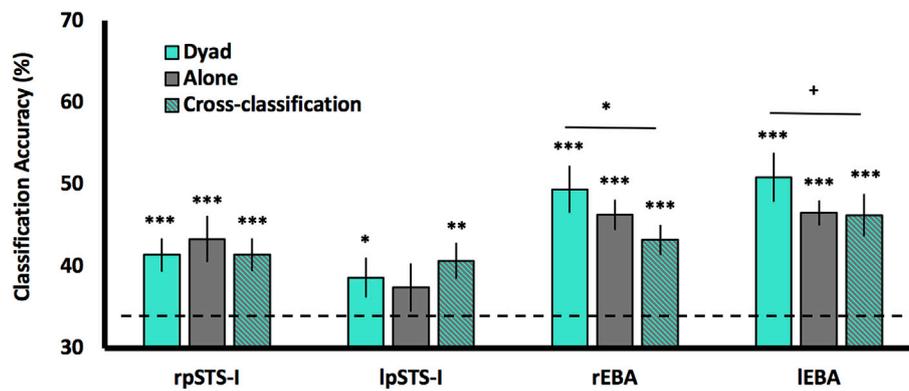


Fig. 2. A bar chart showing classification accuracy values for dyad, alone, and cross-classification analyses for bilateral pSTS-I and EBA ROIs. Dashed line represents chance-level accuracy (33.3%). \*\*\* =  $p \leq .001$ ; \*\* =  $p \leq .010$ ; \* =  $p \leq .05$ ; + =  $p = .073$ . Error bars are SEM.

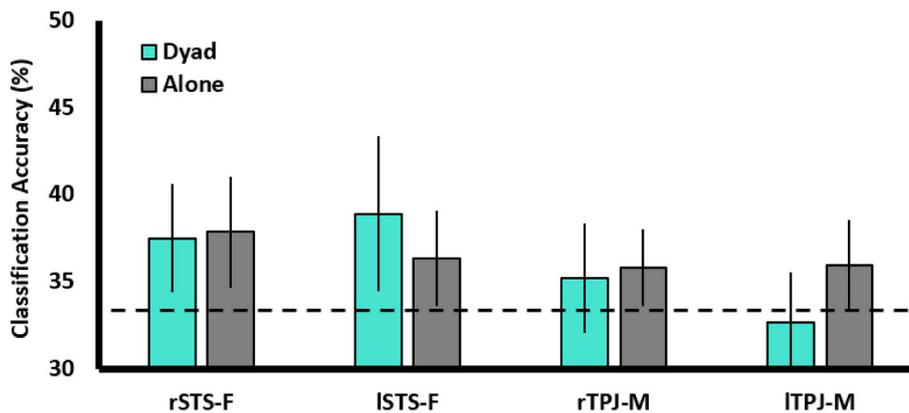


Fig. 3. A bar chart showing classification accuracy values for dyad and alone classification for bilateral STS-F and TPJ-M ROIs. Dashed line represents chance-level accuracy (33.3%). No results were significant. Error bars are SEM.

To test for the main hypothesis (i.e. a dyadic information effect) paired  $t$ -tests were then performed (Bonferroni corrected  $\alpha = 0.013$ ) between dyad classification accuracy scores and cross-classification accuracy scores. No difference was observed for either the right pSTS-I ( $t(19) = 0.00, p = .500$ ) or left pSTS-I ( $t(18) = -0.73, p = .763$ ), showing no dyadic information effect, indicating that the main hypothesis was not supported. However, significantly greater accuracy for dyad classification than cross-classification was shown in the right EBA at an uncorrected level ( $t(17) = 2.07, p = .027$ ). A similar, although weaker, marginal effect was also shown in the left EBA ( $t(18) = 1.52, p = .073$ ). Therefore, evidence suggestive of a dyadic information effect was shown in the bilateral EBA only.

To determine whether regions outside the functionally defined ROIs demonstrated a dyadic information effect, whole-brain searchlight analyses (Kriegeskorte et al., 2006) were performed (see supplementary materials G for a full description of searchlight methods and results). Peak classification accuracies (i.e. for dyad and alone classification separately, and also for cross-classification) were observed in the bilateral lateral occipito-temporal cortex (LOT) and pSTS, along with weaker responses in other areas. However, no dyadic information effects were observed in the LOT/EBA for this analysis (or in any other brain region), further demonstrating the subtle nature of the effect in the ROI analysis.

### 3.2. Reliability of the dyadic information effect in EBA

Due to the marginal nature of these results in the EBA, several follow up tests were performed to determine the reliability of this effect. First, Cohen's  $d$  effect-sizes were calculated for both the right and left EBA. A medium effect-size was found for the right EBA ( $d = 0.60$ ), and a small-

to-medium effect was shown in the left EBA ( $d = 0.38$ ).

To ensure that these effects were not spuriously driven by the 'direction' of cross-classification training and testing roles, cross-classification was performed again, but with the training and testing roles reversed. That is, the classifier was now trained on the *alone* stimuli and tested on the *dyad* stimuli. Both right EBA ( $M = 43.83, SD = 7.60; t(17) = 5.86, p < .001$ ) and left EBA ( $M = 46.20, SD = 10.32; t(17) = 5.43, p < .001$ ) showed significant cross-classification. Crucially, dyadic information effects were replicated; greater accuracy for dyad classification than cross-classification was again shown in the right EBA ( $t(17) = 2.03, p = .029; d = 0.55$ ) and marginally in the left EBA ( $t(18) = 1.41, p = .088; d = 0.40$ ).

One further test was performed to determine how reliable these effects were across different ROI sizes (i.e. in addition to the original 7 mm radius ROIs, 5, 6, 8, 9, 10, 11, & 12 mm radius ROIs were created). Consistent with the dyadic information effect in the original right EBA ROI, greater accuracy for dyad classification than cross-classification was shown across all ROI sizes, but was most pronounced in larger ROIs (i.e.  $ps < .05$  for 8, 9, 11, & 12 mm radii; see supplementary materials H). By contrast, in the left EBA, the dyadic information trend was only shown for smaller ROI sizes (i.e. 5 mm radius:  $p < .05$ ; 6 and 7 mm radii: marginal  $ps \leq .073$ ); indeed, these hemispheric differences appear to be consistent with larger regions of body selectivity in the right than left EBA as previously reported (Willems et al., 2009).

### 3.3. Results summary

In summary, although right pSTS-I – and marginally, left pSTS-I – differentiated between the three interaction scenarios, no evidence for specific *dyadic information* encoding was observed in these regions.

Instead, this effect was observed in the right EBA at an uncorrected threshold (the data for this analysis are available to download; see supplementary materials I). Follow-up analyses demonstrated that this effect was reliable and interpretable, and is further supported by similar (although weaker) effects in left EBA. Control analyses revealed that these effects are not accounted for by low-level differences in stimulus motion energy between conditions (see supplementary materials J). Additionally, exploratory representational similarity analyses were also performed to further characterize EBA responses to dyad and alone stimuli (see supplementary materials D).

## 4. Discussion

### 4.1. Overview of results

The present study aimed to determine whether the pSTS – or any other posterior temporal lobe region – showed sensitivity to *unique* dyadic information in visually observed interactive scenarios that is not present for isolated individual interactors. Two main findings were shown: 1) EBA – but not pSTS – showed evidence consistent with the encoding of unique dyadic information; 2) pSTS (and EBA) classified between three interaction scenarios (i.e. arguing, celebrating, & laughing) replicating similar differentiation of types of interactions between abstract moving shapes (Isik et al., 2017; Walbrin et al., 2018).

### 4.2. Interaction classification in the pSTS & EBA

Specifically, which type of information might drive differentiation of interaction scenarios in the pSTS and EBA? The pSTS plays an important role in biological motion perception (e.g. Deen et al., 2015; Grossman et al., 2000; Pelphrey et al., 2005), and is strongly responsive to movement contingencies between interacting figures (e.g. Georgescu et al., 2014), as well as dynamic cues that imply interactive behaviour between animate moving shapes (Schultz, Friston, O’Doherty, Wolpert and Frith, 2005; Gao et al., 2012). Similarly, the pSTS is also sensitive to the intentional contents of actions (Brass et al., 2007; Pelphrey et al., 2004; Saxe et al., 2004). It therefore seems plausible that classification in the pSTS is driven by differential intentional content between interaction scenarios that is extracted from different dynamic contingencies between interactors.

Additionally, the EBA also classified between interaction scenarios. A direct interpretation of this result is that body posture information contributes strongly to the differentiation of these three scenarios. EBA is shown to be sensitive to dynamic postural information (i.e. continuous sequences of body postures that form coherent actions) and is suggested to encode body-based actions (Downing et al., 2006). In the current study, distinctively different sequences of coherent body postures – or action-gestures – may have driven classification of interaction scenarios. Although distinct action-gestures were used within each interactive scenario, these tended to be relatively similar to each other (e.g. arguing gestures usually depicted short, sharp movements, while laughing gestures typically contained convulsive torso movements). Therefore, it seems possible that classification of interaction scenarios in the EBA was likely the result of similar action-gestures *within* each scenario, that were markedly different *across* the three scenarios.

### 4.3. No dyadic information effect in the pSTS

Despite the pSTS classifying interactive scenarios, the main prediction was not supported; no *dyadic information effect* was observed for the pSTS. This contrasts with the findings of Baldassano et al. (2017) that showed an analogous effect in the pSTS for static depictions of human-object (inter)actions compared to the averaged responses to isolated objects and humans. One possible explanation for this concerns STS sensitivity to implied biological motion in static images (Grossman and Blake, 2001; Peuskens et al., 2005); static human-object interactions might imply

greater biological motion or more effortful movement that is not ‘recoverable’ from isolated human and objects; for example, an image of a person pushing a cart implies greater movement than the same body pose and cart presented separately, by virtue of greater physical effort required to move the cart, along with the corresponding impression that the cart is moving. Additionally, pSTS sensitivity to causal contingencies (e.g. a billiard ball hitting another, causing a transfer in motion; Blake-more et al., 2001) suggests the strong influence of physical contact in human-object interactions that was not present in the isolated stimuli. By contrast, the current study used dynamic stimuli that contained biological motion information but no physical contact, and as such, the dyad and alone stimuli were closely matched for these two sources of information that might have driven responses to the stimuli used by Baldassano et al. (2017).

Although no dyadic information effect was found in the pSTS, it is important to note that interactive information was still conveyed in the alone stimuli (e.g. communicative gesturing to an unseen interactive partner was strongly implied). Therefore, successful classification of the alone stimuli does not necessarily reflect that pSTS responses are non-interactive. Indeed, in the context of the sorts of gestural interactions used in the current study, it is possible that classification of the alone and dyad stimuli relied on the same cues (i.e. communicative gestures). Similarly, the current data supports the possibility that representations of interactions in this region may encode the presence of two interactors in a linear fashion (i.e. dyad = the average of the two individuals). Alternatively, it is possible that the pSTS responses to both dyad and alone stimuli are driven by interactive gestures ‘directed’ at another individual, regardless of whether the other individual is present or not.

### 4.4. Dyadic information processing in the EBA

Although not observed for the pSTS, a dyadic information effect was shown for the right EBA and to a lesser extent, the left EBA. Although not predicted, this does fit with previous findings observed in the wider LOTC area. Specifically, Baldassano et al. (2017) observed differentiable responses to human-object interactions than averaged responses to humans and objects in object-selective LOTC (i.e. LOC – in close proximity to EBA); however, this trend did not quite reach significance in the EBA, likely due to weaker responses to object stimuli, suggesting that the currently observed EBA responses could be specific to human body information. Recent evidence also shows that object-selective LOTC is sensitive to ‘regular’ spatial configurations of objects that imply a congruent scene (e.g. different responses are shown for scenes that depict a sofa positioned in front of a television, rather than behind it; Kaiser and Peelen, 2018). Similarly, object-selective LOTC is sensitive to spatial configurations of objects that imply an action (e.g. a pitcher tilted towards an empty cup), relative to configurations that do not (Roberts and Humphreys, 2010).

Broadly, these findings might suggest a converging role for configural processing of *distinct* objects and people in the LOTC. In relation to the present findings, it is conceivable that LOTC – and here the EBA specifically – performs similar configural processing or grouping based on the action-, body-, and movement information conveyed by interactors. If true, to what extent does *dynamic* information contribute to this effect? In contrast to previous work investigating LOTC grouping responses for static stimuli (Baldassano et al., 2017; Kaiser and Peelen, 2018; Roberts and Humphreys, 2010), the current study used dynamic stimuli. Although the EBA is highly sensitive to static pose information, and may process body movements as a series of static ‘snapshots’ (Downing et al., 2006; Giese and Poggio, 2003) body (and face) responses are shown to generalise across static and dynamic depictions in broad regions of the posterior temporal cortex (O’toole et al., 2014). Similarly, representations in the LOTC generalise across dynamic and static depictions of actions and are invariant to other low-level features such as movement direction, or the specific hand used to perform an action (Hafri et al., 2017; Tucciarelli et al., 2015).

In line with these findings, it is likely that dyadic representations of (inter)actions in the EBA generalise across static-dynamic depictions. While dynamic information may not be necessary to encode such scenarios, it may, potentially, allow for more elaborate encoding of body-based actions than similar, static depictions. Additionally, other spatial cues (e.g. interpersonal distance, physical contact, and facing direction), and temporal cues (e.g. movement contingencies and correlated motion) may also contribute to dyadic encoding in the EBA, and further research may directly clarify which cues contribute most prominently.

It is also worth briefly considering the extent to which dyadic information processing is present for other types of interaction, for example, interactions depicted by moving geometric shapes that do not contain body information. These types of stimuli are known to drive responses in LOTC, ostensibly due to the presence of simple actions such as pushing and pulling movements (Walbrin et al., 2018). As mentioned previously, the wider LOTC area shows some sensitivity to spatial-temporal relations between interacting or scene entities, and therefore cortex in close proximity to (and overlapping with) EBA might plausibly encode dyadic information for these abstract scenarios.

The present stimuli consisted of interactions between individuals that did not involve physical contact, a potentially powerful interaction cue that is worthy of further investigation; indeed, stronger dyadic information effects might be predicted for contact-based interactions (e.g. two individuals shaking hands), by virtue of categorical differences in physical contact (i.e. presence of physical contact in dyadic interactions vs. absence of physical contact in ‘alone’ variants of these stimuli).

#### 4.5. Conclusion

In summary, the present results show that both EBA and pSTS differentiate between different types of social interactions. Crucially, representations of dyadic social interactions in the EBA are sensitive to information beyond that which is encoded by the simple average of two separate interactors presented in isolation. This so-called *dyadic information effect* suggests that the EBA is sensitive to unique interactive information that is present only when two individuals interact simultaneously. These findings complement previously observed sensitivity in the wider LOTC area to spatial configurations of objects or bodies that support the processing of holistic, congruent scenarios.

#### Author contributions

J.W & K.K: study design, data-collection, analysis, writing, and editing.

#### Conflicts of interest

None declared.

#### Funding

This work has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (ERC starting grant: Becoming Social).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.05.027>.

#### References

Baldassano, C., Beck, D.M., Fei-Fei, L., 2017. Human-object interactions are more than the sum of their parts. *Cerebr. Cortex* 27 (3), 2276–2288.  
 Blakemore, S.J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A.N., et al., 2001. How the brain perceives causality: an event-related fMRI study. *Neuroreport* 12 (17), 3741–3746.

Brass, M., Schmitt, R.M., Spengler, S., Gergely, G., 2007. Investigating action understanding: inferential processes versus action simulation. *Curr. Biol.* 17 (24), 2117–2121.  
 Coutanche, M.N., Solomon, S.H., Thompson-Schill, S.L., 2016. A meta-analysis of fMRI decoding: quantifying influences on human visual population codes. *Neuropsychologia* 82, 134–141.  
 Deen, B., Koldewyn, K., Kanwisher, N., Saxe, R., 2015. Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebr. Cortex* 25 (11), 4596–4609.  
 Downing, P.E., Peelen, M.V., Wiggett, A.J., Tew, B.D., 2006. The role of the extrastriate body area in action perception. *Soc. Neurosci.* 1 (1), 52–62.  
 Gao, T., Newman, G.E., Scholl, B.J., 2009. The psychophysics of chasing: a case study in the perception of animacy. *Cogn. Psychol.* 59 (2), 154–179.  
 Gao, T., Scholl, B.J., McCarthy, G., 2012. Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *J. Neurosci.* 32 (41), 14276–14280.  
 Georgescu, A.L., Kuzmanovic, B., Santos, N.S., Tepest, R., Bente, G., Tittgemeyer, M., Vogeley, K., 2014. Perceiving nonverbal behavior: neural correlates of processing movement fluency and contingency in dyadic interactions. *Hum. Brain Mapp.* 35 (4), 1362–1378.  
 Giese, M.A., Poggio, T., 2003. Cognitive neuroscience: neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* 4 (3), 179.  
 Gray, K.L., Barber, L., Murphy, J., Cook, R., 2017. Social interaction contexts bias the perceived expressions of interactants. *Emotion* 17 (4), 567.  
 Grossman, E.D., Blake, R., 2001. Brain activity evoked by inverted and imagined biological motion. *Vis. Res.* 41 (10–11), 1475–1482.  
 Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., Blake, R., 2000. Brain areas involved in perception of biological motion. *J. Cogn. Neurosci.* 12 (5), 711–720.  
 Hafri, A., Trueswell, J.C., Epstein, R.A., 2017. Neural representations of observed actions generalize across static and dynamic visual input. *J. Neurosci.* 37 (11), 3056–3071.  
 Isik, L., Koldewyn, K., Beeler, D., Kanwisher, N., 2017. Perceiving social interactions in the posterior superior temporal sulcus. *Proc. Natl. Acad. Sci. U.S.A.* 114 (43), E9145–E9152.  
 Julian, J.B., Fedorenko, E., Webster, J., Kanwisher, N., 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60 (4), 2357–2364.  
 Kaiser, D., Peelen, M.V., 2018. Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *Neuroimage* 169, 334–341.  
 Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U.S.A.* 103 (10), 3863–3868.  
 O’toole, A.J., Natu, V., An, X., Rice, A., Ryland, J., Phillips, P.J., 2014. The neural representation of faces and bodies in motion and at rest. *Neuroimage* 91, 1–11.  
 Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. CoSMoMVA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front. Neuroinf.* 10, 27.  
 Papeo, L., Stein, T., Soto-Faraco, S., 2017. The two-body inversion effect. *Psychol. Sci.* 28 (3), 369–379.  
 Pelphrey, K.A., Morris, J.P., Michelich, C.R., Allison, T., McCarthy, G., 2005. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cerebr. Cortex* 15 (12), 1866–1876.  
 Pelphrey, K.A., Morris, J.P., McCarthy, G., 2004. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J. Cogn. Neurosci.* 16 (10), 1706–1716.  
 Peuskens, H., Vanrie, J., Verfaillie, K., Orban, G.A., 2005. Specificity of regions processing biological motion. *Eur. J. Neurosci.* 21 (10), 2864–2875.  
 Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., Kanwisher, N., 2011. Differential selectivity for dynamic versus static information in face selective cortical regions. *Neuroimage* 56 (4), 2356–2363.  
 Quadflieg, S., Koldewyn, K., 2017. The neuroscience of people watching: how the human brain makes sense of other people’s encounters. *Ann. N. Y. Acad. Sci.* 1396 (1), 166–182.  
 Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., Saxe, R., 2018. Development of the social brain from age three to twelve years. *Nat. Commun.* 9 (1), 1027.  
 Roberts, K.L., Humphreys, G.W., 2010. Action relationships concatenate representations of separate objects in the ventral visual system. *Neuroimage* 52 (4), 1541–1548.  
 Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42 (11), 1435–1446.  
 Schultz, J., Friston, K.J., O’Doherty, J., Wolpert, D.M., Frith, C.D., 2005. Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron* 45 (4), 625–635.  
 Sinke, C.B., Sorger, B., Goebel, R., de Gelder, B., 2010. Tease or threat? Judging social interactions from bodily expressions. *Neuroimage* 49 (2), 1717–1727.  
 Tucciarelli, R., Turella, L., Oosterhof, N.N., Weisz, N., Lingnau, A., 2015. MEG multivariate analysis reveals early abstract action representations in the lateral occipitotemporal cortex. *J. Neurosci.* 35 (49), 16034–16045.  
 Vestner, T., Tipper, S.P., Hartley, T., Over, H., Rueschemeyer, S.A., 2018. Bound together: social binding leads to faster processing, spatial distortion and enhanced memory of interacting partners. *J. Exp. Psychol. Gen.* <https://psycnet.apa.org/record/2019-02038-001>.  
 Walbrin, J., Downing, P., Koldewyn, K., 2018. Neural responses to visually observed social interactions. *Neuropsychologia* 112, 31–39.  
 Willems, R.M., Peelen, M.V., Hagoort, P., 2009. Cerebral lateralization of face-selective and body-selective visual areas depends on handedness. *Cerebr. Cortex* 20 (7), 1719–1725.