# 1 Subjective understanding of actions and emotions involves the

2 interplay of the semantic and action observation networks in

# 3 the brain

- 4 Minye Zhan <sup>12</sup>, Rainer Goebel <sup>13</sup>, and Beatrice de Gelder <sup>14\*</sup>
- 5 <sup>1</sup>Faculty of Psychology and Neuroscience, Department of Cognitive Neurosciences, Maastricht University,
- 6 6229EV Maastricht, The Netherlands
- 7 <sup>2</sup> U992 (Cognitive neuroimaging unit), NeuroSpin, INSERM-CEA, 91191 Gif sur Yvette, France
- 8 <sup>3</sup> Department of Neuroimaging and Neuromodeling, Netherlands Institute for Neuroscience, Institute of
- 9 the Royal Netherlands Academy of Arts and Sciences, 1105 BA, Amsterdam, The Netherlands
- 10 <sup>4</sup> Department of Computer Science, University College London, WC1E 6BT, UK
- 11
- 12 \* Corresponding author and lead contact: b.degelder@maastrichtuniversity.nl

## 13 Summary

- How we subjectively generate an understanding of other people's bodily actions and emotions is
  not well understood. In this 7T fMRI study, we examined the representational geometry of
  bodily action- and emotion-understanding by mapping individual subjective reports with word
  embeddings, besides using conventional univariate/multivariate analyses with predefined
  categories. Dimensionality reduction revealed that the representations for perceived action and
  emotion were high dimensional, each correlated to but were not reducible to the predefined
  action and emotion categories. With searchlight representational similarity analysis, we found
- 21 the left middle superior temporal sulcus and left dorsal premotor cortex corresponded to the 22 subjective action and emotion representations. Furthermore using task-residual functional
- subjective action and emotion representations. Furthermore using task-residual functional
   connectivity and hierarchical clustering, we found that areas in the action observation network
- 24 and the semantic/default-mode network were functionally connected to these two seed regions
- 25 and showed similar representations. Our study provides direct evidence that both networks
- 26 were concurrently involved in subjective action and emotion understanding.
- 27 Keywords: body, action, emotion, subjective report, 7T fMRI, representational similarity analysis

28

# 29 Graphical abstract



30

### 31 Introduction

32 Imagine you are walking into a large conference hall and you see a person waving in your 33 direction from far away. Within hundreds of milliseconds, this triggers a cascade of implicit and 34 explicit brain processes: "Do I know this person? Is she waving towards me, perhaps to notify 35 me something? Since I don't know her and she seems to be guite happy, is she perhaps waving 36 towards someone close by who might be her acquaintance?" This is an example of the 37 multifaceted information about the action, emotion, intention and identity that human bodies 38 routinely convey, which we process effortlessly in daily life. Our brains extract and combine 39 those various dimensions of information, enabling us to reach a personal/subjective 40 understanding, and to act upon it.

41 These different dimensions of information have been the focus of specialized lines of inquiry.

42 Univariate studies of action observation have found a fronto-parietal network (action

43 observation network) including the intraparietal sulcus (IPS) and the dorsal and ventral

44 premotor areas (PMd, PMv)(Caspers et al., 2010; Grafton and Hamilton, 2007; Rizzolatti et al.,

45 2014). Studies focused on body forms found body/body parts-sensitive regions in the ventral-

46 lateral pathway, the extrastriate body area (EBA) and the fusiform body area (FBA) (Peelen and

47 Downing, 2007), while the posterior superior temporal sulcus (pSTS) is sensitive to the biological

48 motion of both faces and bodies (Allison et al., 2000). A study addressing the intention aspect of

49 action (de Lange et al., 2008) reported involvement of the inferior frontal gyrus (IFG) for

50 understanding unusual intentions, and involvement of the default mode network (DMN)

51 (Andrews-Hanna, 2012; Buckner et al., 2008) for attending to action intentions in general.

52 When the body posture and movements convey emotional information, the stimuli trigger

53 activation in areas associated with visual form, action and movement perception, in concert

54 with more emotion-specific areas including the IFG, insula and subcortical structures (Dricu and

55 Frühholz, 2016; de Gelder, 2006; de Gelder et al., 2004; Kober et al., 2008; Lindquist et al., 2012;

56 Molenberghs et al., 2012; Sinke et al., 2010). DMN areas including the temporo-parietal junction

57 (TPJ), precuneus, dorsal/ventral medial prefrontal cortex (dmPFC, vmPFC) (Amodio and Frith,

58 2006; Saxe et al., 2006), are involved in emotion and valence processing (Chikazoe et al., 2014;

59 Peelen et al., 2010; Skerry and Saxe, 2015), as revealed with multivariate methods such as

60 representational similarity analysis (RSA) (Kriegeskorte et al., 2008; Nili et al., 2014). However

61 univariate activation of DMN areas in action- and emotion-related studies had been scarce, the

62 DMN was considered as a separate system that does not interact with the action observation

63 network (Van Overwalle and Baetens, 2009).

Most previous fMRI studies presented participants with exemplars from a few predefined categories, and often used an explicit action or emotion recognition/categorization task, serving as a proxy for subjective action and emotion understanding. This approach also conforms to the traditional view of a few basic emotions (e.g. Ekman, 1999), assuming that (1) there exist a few

68 discrete categories of actions and emotions that participants are routinely able to recognize; (2)

69 there is high inter-individual similarity in recognizing and identifying these categories; (3) there

is a stable brain-behavior mapping between action and emotion categories and brain processesacross participants.

72 Under these assumptions, individual variability in perceiving the categories is typically treated as 73 noise. However, individual variability is prevalent in multiple functions from perception to 74 cognition, and some individual variability can be linked to the size or function of specific brain 75 areas (for reviews see Kanai and Rees, 2011; Seghier and Price, 2018). Specifically concerning 76 action and emotion perception, the first two assumptions of discrete categories and high inter-77 individual similarity may not hold either. Different observers may well reach a different 78 understanding of the action/emotion, depending on factors such as the expressiveness of each 79 expression (as could often be observed in validation studies for expression databases, e.g. 80 Goeleven et al., 2008; Langner et al., 2010), the personality of the viewer (e.g. Van den Stock et 81 al., 2015), and various contexts (for a review see Aviezer et al., 2008; for examples see Kret and 82 de Gelder, 2010; Righart and de Gelder, 2006). However, the third assumption may still hold. 83 Different individual behavior may still map to the same neural underpinning, as one fMRI study 84 found individualized and unique dissimilarity ratings for objects were still represented in the 85 brain activity patterns in participants' inferior temporal cortices (Charest et al., 2014).

86 The tendency to treat individual variability as noise was also partially related to the difficulty to

87 objectively quantify individual variability across participants, especially for the case of verbal

88 reports. With the advancement of deep neural networks and natural language processing

techniques, word embeddings (see Boleda, 2020 for a review) are now used to describe

90 semantic concepts in an objective and quantitative way, and recent studies started linking them

91 to brain activity (Hebart et al., 2019; Zhang et al., 2020).

92 Recently, researchers argued that studying subjective experiences is important for

93 understanding the high-level cognitions such as emotion, language, and music (Hartley and

94 Poeppel, 2020; LeDoux and Hofmann, 2018). The presence of subjective variability in emotional

95 perception and production is also acknowledged in the recent literature (for a review for facial

96 expressions, see Barrett et al., 2019; for a review for more general emotional stimuli, see Cowen

97 and Keltner, 2021). In response to these challenges, new studies utilized very large stimulus sets,

subjective reports, and the associated semantic space (Cowen and Keltner, 2017, 2020, 2021;

99 Cowen et al., 2019). Results indicated that for the subjective experiences elicited by facial, vocal,

100 musical emotional expressions, over 20 emotional categories could be observed in the

101 subjective reports for some modalities. There were no discrete category boundaries, and the

semantic space of these categories was also high-dimensional. The brain-behavior mappings of

103 such rich emotion representations seen in subjective reports were largely uncharted,

104 independent from the pre-defined basic categories.

105 In this high-resolution 7T fMRI study (1.2 mm isotropic resolution, voxel volume = 1.728 mm<sup>3</sup>)

106 with RSA (Kriegeskorte et al., 2008; Nili et al., 2014), our goal was to address both sides, the

107 subjective understandings of actions and emotions, and the researcher-defined action and

108 emotion categories. During functional runs, 10 participants passively viewed a large stimulus set

109 of 10 predefined bodily action categories (6 neutral, 4 emotional categories; each participant

- 110 viewed 40 stimuli). Immediately following the scanning session, participants provided free
- 111 reports of the subjectively perceived action and emotion for each stimulus.
- 112 We first examined the representational geometry of the subjective behavioral reports by
- 113 principal component analysis (PCA) and RSA, after mapping all subjective report entries into a
- 114 common vector space using Deconf word embeddings (Pilehvar and Collier, 2016). Then we
- examined the neural representations for predefined categories with RSA searchlight and RSA
- regression. There we estimated the body joint positions using the OpenPose library (Cao et al.,
- 117 2017), and computed low- and mid-level visual features. Lastly, we examined the relation
- 118 between subjective reports and brain activity. We searched for the subjective representations in
- 119 individual participant data by RSA searchlight, and examined their putative direct
- 120 upstream/downstream areas by task-residual functional connectivity and hierarchical clustering.
- 121 We found neural representations for perceived action and perceived emotion, and the two
- analyses with predefined categories and with the subjective reports converged as they both
- 123 indicated joint involvement of the action observation network and the DMN/semantic network.
- 124

# 125 **Results**

#### 126 **RDM construction**

- 127 In the scanning session, participants passively viewed 40 gray-scale images from one of two
- balanced stimuli sets (Fig. 1A. See Fig. S1 for the complete sets of stimuli). We used a slow
- event-related design, where each image (2.60 x 4.26 degrees) was presented for 500 ms,
- followed by an inter-stimulus interval of either 7.5, 9.5 or 11.5 s. Each image was presented 12
   times
- 131 times.
- 132 In the behavioral session directly after the fMRI scan, for each stimulus image, participants
- reported their subjective understanding of the action and emotion by typing a short description,
- and rated the implied motion and valence on a scale of 1 to 7 (See **Table S1** for the exact
- 135 questions, and examples of participants' free reports).
- 136 Individual reports were analyzed by mapping all response entries to the same high-dimensional
- 137 space using the pre-trained Deconf word embeddings (Pilehvar and Collier, 2016), which
- 138 combined the word2vec word embeddings and the WordNet database (see Methods for
- details). Specifically, we lemmatized all the verbs, nouns, adjectives, adverbs typed in by the
- 140 participants (e.g. fighting/fights  $\rightarrow$  fight), and selected the corresponding meaning for each word
- in WordNet 3.1 (<u>https://wordnet.princeton.edu/</u>). We then retrieved the corresponding 300-
- 142 dimensional vectors from Deconf embeddings, averaged the word vectors in each response
- 143 entry, and computed RDMs for perceived action and perceived emotion (cosine distance) for
- 144 each individual participant, see **Fig. 1Bab** and **1Cab**.

- 145 Implied motion and valence ratings were one-dimensional attributes provided by individual
- 146 participants, which were related respectively to the action and emotion aspects of each
- 147 stimulus. These RDMs were computed directly from the individual behavioral ratings (Euclidean
- 148 distance, Fig. 1Bcd, Ccd).
- 149 For predefined categories (10 action categories, non-emotional/emotional actions), we followed
- 150 the RSA and perceptual categorization literature and constructed model RDMs (Euclidean
- distance, Fig. 1B, first column), assuming that the stimuli representations were similar within
- 152 categories, but were different across categories between stimuli (e.g. Freedman et al., 2001).
- 153 We examined the correlations of the subjective report and the rating RDMs with the two
- 154 predefined RDMs. Throughout the whole study, correlations between RDMs were computed
- 155 with Spearman correlation, and submitted to a one-sample t test against 0 (two-tailed) at the
- 156 group level, after Fisher's Z transformation.

# Perceived categories largely corresponded to the predefined ones, but with considerable individual variability

- 159 All four subjective report and rating RDMs were significantly correlated to the predefined ones
- 160 (one-sample t test against 0; coefficient of variation, CV, showing inter-individual variability),
- 161 perceived action to predefined action RDMs: mean rho=0.324, p=8.55×10<sup>-8</sup>, CV=20.4%;
- perceived emotion to predefined emotion RDMs: mean rho=0.491, p=0.000164, CV=51.2%;
- rated implied motion to the predefined action RDMs: mean rho=0.166, p=0.000541, CV=60.4%;
- rated valence to the predefined emotion RDM: mean rho=0.496, p=5.91×10<sup>-7</sup>, CV=25.5%. This
- 165 indicates that the individual subjective understanding largely corresponded to the predefined
- 166 categories, especially for emotions, although there was considerable individual variability (Fig.
- 167 **1C, D**).
- 168 We then examined the individual variability, to assess whether it is suitable to use group-
- averaged RDMs to perform further analyses in individual participants. We first obtained a group-
- averaged RDM from all 10 individual RDMs, then computed its Spearman correlation to each
- 171 individual RDM (one-sample t test against 0; CV). When inter-individual consistency is high, the
- 172 group-averaged RDM should have high correlation to each individual RDM, and a low CV. For
- valence, the group-averaged RDM was very consistent with individual RDMs: mean rho=0.900,
- 174 p=6.91×10<sup>-8</sup>, CV=19.9%. However, the other 3 group-averaged behavioral RDMs were less
- 175 consistent with individual RDMs, again showing considerable individual variability: perceived
- action RDM: mean rho=0.517, p=4.74×10<sup>-6</sup>, CV=32.7%; perceived emotion RDM: mean
- 177 rho=0.565, p=6.02×10<sup>-5</sup>, CV=44.9%; rated implied motion RDM: mean rho=0.621, p=8.71×10<sup>-5</sup>,
- 178 CV=47.1%. With these moderate correlation coefficients, indicating less tight brain-behavior
- 179 links for individual participants, using group-averaged RDMs in subsequent analyses may be less
- 180 suitable to capture the individualized neural substrates. It was also the case for the pre-defined
- 181 categories.

- 182 On the other hand, analyses based on the individual subjective responses may better capture
- 183 the individual neural processes, because the neural activity and patterns generating these
- 184 individualized subjective reports (behavioral outputs) may still map to brain locations in a
- 185 consistent and meaningful way. If the neural processes found this way are robust and consistent
- 186 enough, individual-level replication can be achieved (Smith and Little, 2018). Adopting this logic,
- 187 our subsequent analyses looked at the mapping of subjective reports to brain activity, and
- accordingly focuses on individual-level data and results.

#### 189 The perceived action and emotion representations are high-

#### 190 dimensional

- 191 For investigating perceived action and perceived emotion, we examined the high dimensional
- space of all responses using principal component analysis (PCA), since all 400 subjective
- responses were in the same high-dimensional space, each represented by a 300-dimensionalvector.
- 195 The first few principal components (PCs) for the perceived action and perceived emotion

representations captured a large amount of variance (First 10 PCs: 86.17% and 89.63% variance;
 first 50 PCs: 97.76% and 99.18% variances, Fig. 1F, G, analysis performed on 1<sup>st</sup> to 50<sup>th</sup> PC for
 both representations).

- 199 The perceived action representation RDM showed considerable individual variability when 200 compared to the 10-action-category model (Fig. 1Ea, Ga). The first and biggest few components 201 did not show clear separation of action categories. Instead, individual variability was more 202 prominent (Fig. 1H), while it remains elusive what accounts for the majority of the variance in PC 1. Only the later and much smaller PCs were correlated more to the 10-action-category 203 model (7<sup>th</sup>, 9<sup>th</sup>, 10<sup>th</sup> PCs), the non-emotion/emotion model (7<sup>th</sup>, 22<sup>nd</sup>, 6<sup>th</sup> PCs), and the implied 204 motion ratings (6<sup>th</sup>, 10<sup>th</sup>, 18<sup>th</sup> PCs), and also showed relatively clear category separations (Fig. 205 206 1H, I). The presence of smaller PCs correlated to the action categories, emotions and implied motion ratings indicates that the perceived action representation is rich and high-dimensional. It 207 208 potentially supports extracting the relevant categorical and continuous information, but it is not
- 209 equivalent and not reducible to one of the predefined category structures.

210 The perceived emotion representation was lower-dimensional than the perceived action 211 representation (79 versus 174 independent PCs), and showed higher inter-individual consistency 212 (Fig. 1Eb, Gb, much lower correlation value to the individual subject model in Fig. 1H). The first 213 few PCs showed a moderate correlation to the non-emotion/emotion model and valence 214 ratings, and even to the 10-action-category model (Fig. 1H, J). Interestingly, visual inspection of 215 the RDMs showed that the perceived emotion representation contained the separation of non-216 emotion and emotion categories (Fig. 1Eb, PC1 and 2 in Fig. 1Gb), but also contained finer-217 grained separation for individual emotions. These were captured by the 3<sup>rd</sup>, 4<sup>th</sup> and 6<sup>th</sup> PCs, 218 corresponding to happy, fearful and sad (Fig. 1Gb, see the locations of the prominent white/red 219 bars across all conditions, showing similarity within the category and dissimilarity to the other 9 categories). In the 3<sup>rd</sup> and 4<sup>th</sup> PC, the angry emotion could further be separated. This indicates 220

7

- that the perceived emotion representation is also a high-dimensional one; brain areas
- 222 containing such representations could potentially support both the non-emotion/emotion
- 223 categorization, and categorization of each individual emotions.





Figure 1. The representational geometries of subjective reports were high-dimensional, and

showed considerable individual variability.

- A. Stimuli examples of 10 action categories, performed by one of 8 actors. Category
- abbreviations: CH: combing hair; DW: drinking water; OD: opening door; PH: phone; TR: putting
- 229 on trousers; NE: neutral standing still; FE: fearful; AN: angry; HA: happy; SA: sad.
- 230 B. RDMs for predefined categories, perceived categories, ratings, averaged across 10
- participants (40 by 40 matrices, correlation distances 1-r, all diagonals were 0, sorted by the 10
- action categories). The group-averaged RDMs showed correspondence to the predefined RDMs.
- a. perceived action; b. perceived emotion; c. rated implied motion; d. rated valence. In B, C, E,
- **G**: blue color: similar; red color: dissimilar. For all RDMs, the color ranges were scaled
- automatically by the minimal/maximal values within each RDM.
- 236 **C.** RDMs for individual participants showed considerable individual variability. RDM types **a** to **d**
- correspond to a to d in B. Each column corresponds to one participant. The left 5 participants
   viewed stimulus set A, the other 5 viewed set B.
- 239 **D.** Correlation coefficients for individual participants were not high. **a.** perceived action
- correlated to the 10- action-category model; **b.** perceived emotion correlated to the non-
- 241 emotion/emotion model; c. implied motion ratings correlated to the 10-action-category model;
- d. valence ratings correlated to the non-emotion/emotion model. Black bars: group average.
- **E** to J. PCA results of all 400 behavioral report items for perceived action and perceived emotion.
- The smaller PCs of the perceived action and the first PCs of the perceived emotion showed correspondence to the predefined RDMs.
- **E.** Same to **B**, but for all 400 behavioral report items across participants (400 by 400 matrices).
- 247 **F.** The first 10 principal components (PCs) explained a large amount of variance. Curves:
- 248 cumulative explained variance (%).
- G. RDMs for scores of the first 10 PCs (Euclidean distance), for perceived action (a) andperceived emotion (b).
- 251 H. RDMs the first 50 PCs, correlated to the predefined action, emotion models, behavioral rating
- 252 models, and to the individual subject model. The first PCs of the perceived action correlated
- 253 more to the individual subject model. Gray bars: non-significant correlation values (FDR q>0.05).
- **1.** Perceived action report items, plotted against the first two PCs, and against the two PCs with
- 255 highest correlation coefficients to each of the models.
- **J.** Perceived emotion report items, plotted against the first two PCs, and against the two PCs
- 257 with highest correlation coefficients to each of the models. Fewer dots were shown on the plots,
- 258 because some items have overlapping values.
- 259

### 260 Univariate results were consistent with the literature

- 261 For the fMRI data, as a sanity check, we first performed conventional group-level univariate
- analysis with data smoothing, for the 10 action categories (input for the participants), and
- 263 parametric modulations for implied motion and valence. The 7T data was very robust: the
- activation location and time course profiles at 7T were similar to 3T ones, but showed higher %-
- signal change (See Figure S2).

The 10-category one-way ANOVA showed differences across action categories in bilateral EBA, right FBA, right lateral occipital cortex (LOC) and right mIPS (**Fig. 2A**). Similarly, the parametric modulation of implied motion showed higher activity for actions rated with higher implied

- 269 motion, in bilateral EBA, right mIPS, left pIPS, right rostral cingulate zone (RCZ), where the
- former 3 overlapped with the ANOVA clusters (Fig. 2B).
- 271 When contrasting emotional and non-emotional categories (excluding the standing-still
- 272 condition), the left IFG and left middle temporal gyrus (MTG) showed higher activity for
- 273 emotional categories, while early visual areas (EVC) and right precentral gyrus showed higher
- activity for non-emotional categories (Fig. 2C). The activation map was very similar when
- including the standing-still condition. These univariate results were consistent with previous
- findings (Dricu and Frühholz, 2016; de Gelder et al., 2004; Kober et al., 2008; Molenberghs et al.,
- 277 2012; Sinke et al., 2010). However, for the parametric modulation of valence, we only observed
- 278 clusters modulated by positive valence, but not for ratings of negative valence (negative
- 279 modulation). These clusters were found outside the frontal lobe, in bilateral early visual cortices
- 280 (including bilateral V3a and right calcarine sulcus/cuneus), right posterior collateral sulcus, right
- 281 supramarginal gyrus, left precentral gyrus/central sulcus (**Fig. 2D**).
- 282 We cannot exclude that some of these may be false negatives due to the exacerbated inter-
- 283 individual variability at 7T, where activation clusters were small, and were constrained to the
- 284 gray matter even after spatial smoothing. This effect could be observed in the functional
- 285 localizer data: despite that we found robust EBA, FBA, FFA clusters in all individual participants
- 286 (data smoothed 3 mm FWHM, contrasts: bodies>faces, houses, tools, words; faces>bodies,
- 287 houses, tools, words), and the activation sites were reliable in one participant between 3T and
- 288 7T scans on different days (**Figure S2**), the across-participant overlapping of these clusters were
- low; the group-level GLM (data smoothed 6 mm FWHM) only showed an R EBA/pSTS cluster for
- bodies, L EVC and bilateral precuneus clusters for faces (Figure S3). Thus for whole-brain
- activation and searchlight analyses, we used the initial p threshold of 0.005 to balance betweenthe false positives and false negatives.
- 293



294

295 **Figure 2**. Univariate results were consistent with the literature. Data in volume space were

296 plotted on the group-aligned surface mesh after cortex-based alignment. Functional data were

- 297 smoothed at 6 mm FWHM, all maps were cluster-size thresholded with Monte-Carlo simulation,
- alpha=0.05, n simulations=5000. We used the initial p<0.005 to alleviate the false negatives due
- to high anatomical inter-individual variability, which was clearly observed in our functionallocalizer data.
- 301 **A**. ANOVA of 10 categories. Color bar p range: 0.005 to 0.0001.
- **B**. Parametric modulation of rated implied motion.
- 303 C. Emotional categories > non-emotional categories (excluding neutral standing still).
- **D.** Parametric modulation of rated valence. Color bar p range in **B, C, D**: 0.005 to 0.0005.
- Cluster size thresholds in A to D: 115, 55, 46, 56 functional voxels.

306

# 307 **Predefined non-emotion/emotion and 10-action-category**

### 308 representations in the brain

Next we followed the conventional predefined-model-driven approach, basing the models on
 the predefined action and non-emotion/emotion categories, and performed RSA searchlight

- 311 analysis for those RDMs. We constructed neural RDMs from univariate t maps of the 40
- 312 individual stimuli (no smoothing for functional data, Pearson's correlation distance), and
- 313 performed RSA searchlight (radius=5 voxels; Spearman correlation to the predefined category
- RDMs, one-sampled t test against 0 for the Fisher's Z-transformed rho values at the group-level,
- 315 cluster size thresholded at alpha of 0.05, Monte-Carlo simulation n=5000). The same group-level
- test and cluster thresholding scheme was used for all searchlight analysis throughout the study.
- For the predefined non-emotion/emotion categories, we found two areas positively correlated
- to the model RDM, in L central sulcus (adjacent to the cluster found in parametric modulation

for valence ratings in the univariate analysis, in Fig. 2D), and R PMd/FEF (Figure 3B). Another

320 four areas showed negative correlation to the RDM, in L precuneus, L caudal cingulate zone,

321 bilateral medial SFG, and L thalamus, indicating fine-grained processing for stimuli within the

322 non-emotion or emotion categories.

323 We found 29 brain areas positively correlated to the 10-action-category RDM (Figure 3A, S4,

**Table S2**, denoted "10-action-category areas" below), and no negatively-correlated areas. When

thresholding at p<0.001, eight of these areas were still present, including the R cuneus, R

326 EBA/hMT+, L central sulcus, L precuneus, R posterior cingulate gyrus and sulcus, bilateral medial

327 superior frontal gyrus, anterior RCZ. Some of these 10-action-category areas consistently

showed univariate activation in body or action perception, including R EBA/hMT+, L pIPS and L

329 RCZ; but several areas did not show above-baseline univariate activation at the group level,

including L central sulcus, R anterior SFS, R cerebellum, L vmPFC and bilateral medial SFG (Figure
 S5, Table S3).

332 We next examined the task-residual functional connectivity, which presumably reflects

333 information transfer, and displays certain levels of consistency with structural connectivity

334 (Bullmore and Sporns, 2009), and thus could potentially capture the direct

upstream/downstream areas for a seed area. Using these 29 areas as seed regions, we observed

336 fine-grained connectivity patterns for each seed region, showing hemispheric symmetry. These

337 areas were highly interconnected, and were also connected to the action-observation-related

338 areas (EBA, FBA, IPS, pSTS, SMA, PMd, PMv, M1, cerebellum), and DMN areas (dmPFC, vmPFC,

339 precuneus, TPJ, mSTS). Furthermore, most areas were heavily connected to bilateral caudate,

340 putamen and thalamus; and 14 out of 29 areas were connected to the hippocampus. See

341 supplementary **Figure S6** for a summary of their connectivity patterns.



342

Figure 3. Searchlight RSA showed multiple areas correlated to the predefined category models.
A. Results for the 10-action-category RDM, cluster size threshold=43 voxels. The labeled areas
were ones overlapped with the perceived action and emotion FC networks. B. Results for the
non-emotion/emotion RDM, cluster size threshold=35 voxels. The areas labeled were ones with
positive correlations to the RDM. Color bar p range in both A and B: 0.005 to 0.0005. Cluster size
thresholds: 43, 35 functional voxels. Initial p<0.005.</li>

349

# 350 The representations of low- or mid-level visual features and higher-

### 351 level attributes

352 We used the predefined 10 action categories to search for brain areas with corresponding

353 "category boundaries". However, these category boundaries in the brain may not only be

- 354 generated by higher-level abstractions between categories, but could also be generated by
- lower-level visual feature differences between the stimuli. For our stimuli set, since the actors
- received a fixed set of instructions for performing the actions (Stienen and de Gelder, 2011), the
- postures may be somewhat stereotypical (see all stimuli in **Figure S1**), where low- or mid-level

visual features alone could contribute to clear category boundaries, but co-vary with the
 abstract categories. Therefore, some of the brain areas found by the 10-action-category RDM
 may correspond to the low/mid-level visual features.

361 To examine this possibility in the 10-action-category areas found by searchlight, we performed 362 RSA regression, using the RDMs of higher-level stimuli attributes and low- or mid-level visual 363 features as predictors. The higher-level stimuli attributes included non-emotion/emotion, 364 implied motion, valence, and actor identity. The low level features included the raw pixel values 365 of each stimulus picture, and the body joint coordinates, extracted from each stimulus picture by the OpenPose library (Cao et al., 2019). The mid-level visual features were computed from 366 367 the body joint coordinates and based on suggestions from the literature: The body-part 368 orientations may signal the directions of social interactions; the hands/feet-to-head distances 369 may be related to the peripersonal space (Bufacchi et al., 2016) and the relaxation of the 370 arms/legs. Most of the low- mid-level RDMs indeed correlated to the 10-action-category RDM 371 (rho ranging from 0.08 to 0.36), apart from the raw pixel value and head orientation RDMs. See 372 Figure 4A, B. We acknowledge that these RDMs were only a very limited set from an infinitely 373 vast model space. All RDMs were put in the same linear model, so that the unique contribution 374 of each RDM could be examined.

375 We found that some of these stimulus features and attributes could partially explain the activity 376 in some of the 10-action-category brain areas. For the low-level features, the raw pixel values 377 representation could be found (with a positive beta estimate significantly bigger than 0) in R 378 cuneus and R CCZ, but showed a negative beta estimate in R SPOC; the raw joint coordinates 379 representation could be found in R EBA, but showed a negative beta estimate in R callosal 380 sulcus. For mid-level features, the head orientation representation could be found in R SPOC, 381 while showing a negative beta estimate in R dmPFC; the shoulder orientation representation 382 was found in L pIPS, and the shortest hand-head distance representation in R cerebellum. For 383 higher-level attributes, only the actor identity representation showed a positive beta estimate in 384 again the R SPOC, and a negative beta estimate in R cuneus.

We then performed whole-brain searchlight analysis for these attributes and features (Figure 4
C to G). For high-level features, we did not find areas positively correlated to the valence rating
RDMs, which may either correspond to a true negative, or a false negative due to again the
inter-individual anatomical variability. The implied motion rating RDMs showed positive clusters
in L FEF (adjacent but not overlapping with the L PMd cluster for perceived emotion) and R
SPOC. And interestingly, the actor identity RDM showed a positive cluster in L MTS, close to EBA
and FBA.

392 We found positive clusters for two mid-level visual features: head orientation and shortest

393 hand-to-head distance, but no positive clusters for any other low- or mid-level visual features,

despite their similarity to the two features with positive clusters (shoulder and waist orientation

395 RDMs were correlated to the head orientation RDMs: rho ranges from 0.317 to 0.412; the

396 shortest hand-to-head distance RDMs were correlated to the average hand-to-head distance

RDMs, rho=0.593, 0.646 for stimuli set A and B). This indicated that these two features may be

398 biologically meaningful ones for the brain.



399

Figure 4. Low/mid-level features (A and B), and whole-brain searchlight results corresponding to
 these features (C to G).

- 402 **A.** Skeleton in the left panel: coordinates of the 18 joints extracted by the OpenPose library;
- 403 white dots in the middle panel: joint coordinates after manual adjustment; right panel, magenta
- 404 lines: distances used to compute head, shoulder, waist orientations, normalized by the
- 405 corresponding distances in the neutral stand-still stimulus of the same actor (see middle panel);
- 406 green lines: hand/foot-to-head distance.
- 407 **B.** The low/mid-level features, showing RDMs for stimuli set A. The raw coordinates RDMs were

- 408 computed excluding joints on the ears and eyes, but was very similar to the ones computed
- 409 including these joints (RDM similarity rho=0.976, 0.98 for stimuli set A and B). Most of these
- 410 RDMs correlated to the 10-action-category RDM.
- 411 **C, D, E**. Searchlight results for higher-level attributes. The valence RDM did not show positively
- 412 correlated clusters. Cluster size thresholds: 38, 44, 31 functional voxels.
- 413 **F, G.** Searchlight results for mid-level features that showed clusters positively correlated to the
- 414 corresponding RDMs. Cluster size thresholds: 48, 37 functional voxels.
- 415 Color bar p range in **C** to **G**: 0.005 to 0.0005. Initial p<0.005. Clusters showing positive
- 416 correlations to the corresponding RDMs were labeled. In **D**, clusters showing negative
- 417 correlations (fine-grained processing with similar valence rating) were also labeled.

418

# 419**RSA searchlight analysis revealed representations for perceived action**

## 420 in left mSTS and for perceived emotion in left PMd

421 We next turned to the individual subjective reports, and searched for neural representations

- similar to the individualized perceived action and emotion RDMs. We found one positively
- 423 correlated cluster for each of the subjective report RDMs, in left mSTS for perceived action, and
- 424 in left PMd for perceived emotion, indicating that these two areas may be involved in subjective
- 425 understanding and reporting for action and emotion, respectively. The L PMd cluster is
- 426 consistent with the analysis based on predefined non-emotion/emotion representation, where
- the R PMd/FEF cluster was at a very similar location on the other hemisphere to the PMd clusterfound here.
- 429 All other clusters were negatively correlated to the subjective category RDMs, indicating fine-
- 430 grained processing for similarly perceived actions or emotions within those clusters. For
- 431 perceived action, the negative clusters include L RCZ and R supplementary motor area (SMA),
- 432 which were implicated in action observation in the literature (Caspers et al., 2010). For
- 433 perceived emotion, the negative clusters included R SMG/aIPS and R cerebellum for action
- 434 observation, R insula for emotion processing, and mentalizing areas of R precuneus, R vmPFC, R
- dmPFC/medial superior frontal gyrus (mSFG). See **Figure 5** and **TableS4** for the complete list of
- 436 clusters.
- 437 The L mSTS and L PMd clusters did not show reliable univariate activation for the 10 action
- 438 categories (one-sample t test of percent signal change against baseline, all p>0.052, except in L
- 439 PMd for the Anger category, p=0.031), thus they could not be found by conventional univariate
- 440 analysis. Neither was the 10-category univariate ANOVA significant in these two clusters,
- 441 p=0.309 and p=0.945 respectively.



442

443 Figure 5. Searchlight RSA revealed corresponding clusters to the perceived action model in L 444 mSTS (A), and to the perceived emotion model in L PMd (B). Color bar p range: 0.005 to 0.0005.

445 Cluster size thresholds in A and B: 39, 38 functional voxels. Initial p<0.005.

446

#### The putative direct upstream/downstream areas of L mSTS and L PMd 447

#### identified by task residual functional connectivity and hierarchical 448

#### clustering 449

The L mSTS and L PMd clusters were found by their representation similarity to individualized, 450 451 subjectively reported actions and emotions, which already showed considerable individual 452 variability. Thus, to further understand the involvement of these two areas in the current task, 453 we could no longer use the standard predefined-model comparison methodology, which 454 compares neural RDMs to a same model RDM, assumed by the researcher. Instead, we used a 455 data-driven approach based on two assumptions: (1) We could potentially capture the 456 information transfer stages of a certain area by functional connectivity; (2) Based on the idea 457 that the object representations were successively re-represented and "untangled" in the ventral 458 visual pathway (DiCarlo and Cox, 2007; DiCarlo et al., 2012), we assumed that this would be the 459 case for bodily action and emotion representations throughout the brain, that the information 460 (in our case the neural representations) should be gradually transformed across a chain or 461 network of brain areas, from the representation of the visual input, to the representation of the

462 subjective output, without sudden changes between two adjacent stages in the information-

- transfer chain. In this way, we linked the functional connectivity and representational analyses
- together (also see Ju and Bassett, 2020). To find the set of brain areas in the same information-
- transformation chain, we performed task-residual functional connectivity to the L mSTS seed
- and the L PMd seed, respectively. Using hierarchical clustering in the two resulting networks, we
- 467 could then find areas with most similar neural representations to the two seeds, which may
- 468 include their direct upstream/downstream areas.

#### 469 *Task-residual functional connectivity*

- 470 We regressed out the individual stimuli and head motion parameters from the functional data
- 471 (deconvolution, data smoothed at 3 mm FWHM), obtained task-residual time courses, and
- 472 performed functional connectivity (FC) analysis (Pearson's correlation), with L mSTS and L PMd
- 473 as seed regions. See Figure 6A, B, left panels. This analysis revealed two partially overlapping FC
- 474 networks for action/emotion understanding at the group level (**Figure 6D**), in action observation
- related areas including (bilateral if unspecified) IPS, M1, PMd, PMv, L IFG; and in DMN areas
- 476 including TPJ, mSTS/MTG, precuneus, posterior cingulate cortex, dmPFC/vmPFC, mSFG, parietal
- 477 occipital cortex (POC), retrosplenial cortex.
- 478 Interestingly, some overlapping areas in these FC networks are part of the semantic network,
- 479 including mSTS, IFG, TPJ/angular gyrus, dmPFC, vmPFC, posterior cingulate cortex, retrosplenial
- 480 cortex (Binder et al., 2009); and the L mSTS FC network corresponded to the semantic network
- 481 especially well. See **Figure 6D** and **E**. The seed cluster L mSTS itself is an important area in the
- 482 semantic network, which could be activated by written word stimuli (Binder et al., 2009).
- 483 Although word-specific clusters around left mSTS were found in individual participants in the
- 484 functional localizer data (words>other categories), neither the seed region nor the more
- extended FC cluster around the seed showed consistent group-level activation for words in the
- 486 functional localizer (one-sample t test against 0, L mSTS seed ROI: mean beta=0.589; t(9)=1.418;
- 487 p=0.190; L mSTS FC cluster: mean beta=0.369; t(9)=1.226, p=0.251), which could again be due to
- the inter-individual anatomical/functional variability. More interestingly, although both FC
- 489 networks overlapped at the L mSTS, the cluster for perceived emotion was much more posterior
- 490 than the one for the perceived action, which corresponded to previous findings that the STS is
- 491 an heterogeneous structure with several different functions (Hein and Knight, 2008).
- 492 In addition to the cortical clusters, the FC analysis revealed bilateral caudate and putamen 493 clusters in both FC networks (Figure S7, S8), which may be involved in categorization processing 494 (Seger, 2008; Seger and Miller, 2010). The left mSTS was further functionally connected to the R 495 hippocampus; the left PMd was further functionally connected to the bilateral thalamus, 496 pulvinar, cerebellum, the R septal nuclei, and the bilateral red nuclei (anatomical locations 497 clearly observable in the T2\*-weighted functional images, Figure S8). The involvement of the 498 hippocampus, the septal nuclei and the red nuclei in bodily action and emotion understanding 499 were not routinely observed in previous bodily-action perception studies with univariate 500 methods.
  - 19

501 The perceived action and emotion FC networks also overlapped with some of the clusters that

- negatively correlated to the perceived action/emotion RDMs in the searchlight analysis (**Figure**
- **3**). They include R mSFG in the perceived action FC network; R posterior cingulate gyrus/callosal
- sulcus, R EVC, R lingual gyrus/cerebellum in the perceived emotion FC network. This indicates
- that subjective category-like processing and fine-grained within-category processing could be
- 506 mediated within the same networks.
- 507 Furthermore, the perceived action FC network overlapped with the 10-action-category clusters
- 508 in the L central sulcus, R POC/retrosplenial cortex and bilateral callosal sulci; while clusters in the
- 509 perceived emotion FC network overlapped with the 10-action-category clusters in L precuneus, L
- 510 pIPS, L cingulate sulcus, R SFS, R posterior cingulate/POC/retrosplenial cortex, R caudate, R
- 511 cerebellum and bilateral medial SFG (labels in **Figure 3**), again showing consistency to the
- 512 analysis based on predefined categories.
- 513

#### 514 Hierarchical clustering

- 515 We then examined the possible direct upstream/downstream areas of the two seed regions,
- 516 which should show the most similar neural representations to the seeds. We performed
- 517 hierarchical clustering to all clusters within each FC network, on the group-averaged second-
- 518 level neural RDMs across areas (Spearman correlation distance, hierarchical cluster linkage
- arbitrarily thresholded at 0.8). Since both seed regions were encompassed by more extensive
- 520 clusters with the shortest hierarchical clustering distance to the seed (cluster in the perceived
- 521 action FC network: spanning L mSTS, MTG and ITS; cluster in the perceived emotion network: an
- 522 extensive one spanning L PMd, PMv and LR medial SMG), we plotted the extensive clusters
- 523 (denoted as "seed clusters" below) instead of the seed regions themselves, in the dendrograms
- 524 and second-level RDMs. See Figure 6A, B, Figure S9.
- 525 In the perceived action FC network, the FC cluster with the most similar representation (shortest
- 526 distance in the dendrogram) to the seed cluster was one spanning L p/mIPS, TPJ, angular gyrus.
- 527 The other areas under the same branch of the dendrogram were: the R mSTS contralateral to
- 528 the seed; action-perception related areas including bilateral PMv, MFG, L PMd, L IFG;
- 529 DMN/semantic areas including L precuneus, R TPJ, L posterior cingulate gyrus, LR dmPFC, LR and
- 530 medial SFG.
- 531 In the perceived emotion FC network, the FC clusters with the most similar representation to
- the seed cluster was one spanning R PMd and MFG, contralateral to the seed. The other FC
- clusters in the same branch were: L pSTS, MTG, R pSTS, action-perception related areas
- 534 including bilateral IPS, IPL, angular gyrus, L IFS, IFG; DMN/semantic areas including L TPJ,
- 535 bilateral precuneus, POC; also the lateral and anterior prefrontal cortices including bilateral
- 536 MFG, anterior MFG, L MFS, R SFS.
- 537 These clusters may be directly involved in understanding the bodily actions and emotions. In
- both FC networks, it is interesting to observe the strong involvement of both action-perception

539 related areas and DMN/semantic areas, showing symmetry across hemispheres. Especially for

- 540 the DMN/semantic areas, their direct involvement in bodily action/emotion understanding and
- 541 interactions with the action-perception related areas were not reported in previous studies. This
- 542 may be due to the observation that most of the areas in the two FC networks did not show an
- activation level different from baseline (one-sample t-test of beta values per action condition
- against baseline, FDR corrected across the 10 action categories per ROI). See **Figure 6A, B, C**.
- 545 Thus, they were not likely to be localized when using only univariate methods.
- 546 The clustering in individual participants was relatively consistent, with moderately correlated 547 second-level matrices across participants (averaged Spearman Rho, perceived action: 0.418, 548 SD=0.0473; perceived emotion: 0.374, SD=0.0578). See Figure 6F. Apart from the clustering 549 showing high similarity to the seed clusters in every participant (upper-left side of each RDM), 550 there appeared to be another prominent clustering of areas in each of the two FC networks of 551 individual participants (27-32<sup>nd</sup> areas in the perceived action FC RDM, 36-51<sup>st</sup> areas in the 552 perceived emotion FC RDM, both in the middle of the RDMs, see an enlarged version in Figure 553 S9), although less similar to the two seed clusters (red shade for FC clusters in Figure 6A, B, left 554 panels). The clustering in the perceived action FC network consisted of areas around the 555 bilateral central gyrus (lateral-central/central sulcus, post-central gyrus, bilateral SMA), most 556 showing above-baseline activation for 0 to 3 of the 10 action categories, apart from the L lateral 557 central sulcus showing activation for 8 categories. The clustering in the perceived emotion FC 558 network consisted of areas in the bilateral ventral-lateral visual pathway (EVC, lingual gyrus, 559 collateral sulcus, fusiform, EBA/hMT+, lingual gyrus, cuneus), where 8 areas showed above-560 baseline activation for all 10 action categories, another area showing activation for 8 categories, 561 the other 7 areas showing no activation for any of the categories. Despite the very strong 562 differences in their univariate activation levels, their pattern similarity were clustered in the 563 same branch of the dendrogram. Also, the second-level pattern-similarity clustering structure 564 across areas does not seem to be linked to the low-level visual features from the stimuli, but 565 seem to be a real organization feature, because the clustering structure could be seen in all 10 566 participants, despite the fact that each 5 participant saw a different stimuli set (set A and B).
- 567 For these two groups of clustered areas, their further dissimilarity distance to the seed regions 568 indicate, that these areas are at different processing stages than the final output stage of the 569 seed regions, and our two assumptions of the information-transfer-chains seems to be able to 570 discover meaningful and replicable functional organizations in the brain.

571



573 **Figure 6.** In the functional connectivity (FC) networks of the two seed regions overlapped with 574 the DMN/semantic networks.

575 **A** and **B**. Areas in the FC networks colorcoded by pattern similarity to the seeds, L mSTS (**A**), and

576 L PMd (B). Initial p<0.001, cluster size threshold in functional voxels: L mSTS network: 32; L PMd

- 577 network: 40. The seed ROIs were in yellow. Color bars: representational similarities to the FC
- 578 cluster containing the seed region. For each network, the hierarchical clustering dendrograms
- 579 computed on group-averaged second-level RDMs across areas (Spearman distance) were
- 580 plotted in the right-most panels. The areas with the most similar representations to the FC
- 581 cluster containing the seed region were colored in yellow in the dendrograms (linkage distance
- arbitrarily thresholded at 0.8, see a larger version in **Figure S9**), and labeled in the brain maps of
- the left panels. The name of the areas containing the seed regions were marked in blue. The
- areas with above-baseline univariate activation for more than 5 action categories were marked
  with a "+" sign.
- 586 **C**. Histograms of number of conditions with above-baseline activity, for clusters in the two
- 587 networks. Most areas did not show univariate activation for any of the 10 action
- 588 categories/conditions.
- 589 **D** and **E**. The two FC networks overlapped with the DMN/semantic networks.
- 590 **D.** The overlap of two FC networks.
- 591 **E**. The overlap of the two FC networks, only showing clusters with most similar representations
- 592 to the seed (the clusters labeled in **A** and **B**).
- 593 **F.** The group-averaged and individual second-level RDMs. The yellow and black bars for the
- 594 group-average RDM denote the same clustering of the dendrogram in A and B. See a larger
- 595 version in **Figure S9**.
- 596

### 597 **Discussion**

- In the current 7T fMRI study, we examined how participants perceived bodily actions, by
   analyzing participants' subjective understanding of the actions and emotions displayed, besides
- 600 the conventional analysis by the predefined action and emotion categories. Dimension
- 601 reduction (PCA) revealed that subjectively perceived action and emotion representations were
- high-dimensional and could not be reduced to the predefined category representations, despite
- 603 being correlated to them in smaller principle components. Some emotional categories were
- 604 reflected in the smaller principle components of the perceived emotion representation. Clusters
- 605 in L mSTS and L PMd had representations that were the most similar to the perceived action and
- 606 emotion representations. Areas located in the action-observation network and the semantic
- 607 network/DMN were functionally connected to these two clusters and also showed similar
- 608 multivariate patterns. This provided direct evidence for the involvement and interplay of both
- 609 networks in subjective understanding of bodily actions and emotions.
- For the predefined categories, the non-emotion/emotion representation corresponded wellwith the perceived emotion representation both behaviorally and neurally, showing a cluster in

- 612 R PMd/FEF, contralateral to the cluster for perceived emotion. For the predefined 10-action-
- 613 category representation though, the resulting clusters were more numerous, and likely
- 614 confounded by low/mid-level visual features that could co-vary with action/emotion categories.
- 615 However, all of the clusters were strongly connected to both the action-observation network
- and the semantic network/DMN, consistent with the perceived action results. For mid-level
- 617 visual features, we further found that the head orientation and the shortest hand-to-head
- distance were represented in the brain, indicating that these two features may be important for
- 619 the brain.

# The involvement of the DMN/semantic network in both action and emotionunderstanding

At first sight, it may seem that the involvement of the DMN/semantic network found in the 622 623 current study simply results from the use of word embeddings. We argue that this is not the case, and that our study reveals the individual brain underpinnings of action and emotion 624 625 understanding as involving both the action observation network and the DMN/semantic 626 network. Although the word embeddings indeed represent the semantic distances between the 627 words, these semantic distances reflect the similarities between different concepts. Our 628 perceived action and emotion representations were different, despite being computed from 629 word vectors from the same 300-dimensional space. Also, the perceived emotion cluster in L 630 PMd is not a key node area in the semantic but rather in the action observation network; the 631 two FC networks for perceived action and emotion did not completely overlap with each other

632 either.

633 Actually, the involvement of IFG (Caspers et al., 2010; Dricu and Frühholz, 2016; de Gelder et al., 634 2004; Molenberghs et al., 2012) and areas in the DMN/semantic network (Chikazoe et al., 2014; 635 Peelen et al., 2010; Skerry and Saxe, 2015) has been consistently found in action perception and 636 emotional expression perception, although in separate studies and in different contexts. For the 637 DMN and especially for the vmPFC area, their involvement in emotion perception was for 638 abstract representation of emotional stimuli found with multivariate RSA (Chikazoe et al., 2014; 639 Peelen et al., 2010; Skerry and Saxe, 2015), rather than with simple univariate contrasts. The 640 simultaneous involvement of the IFG and the DMN in action understanding was found in one 641 univariate study contrasting participant's attention on either the intention or means of the 642 performed actions (de Lange et al., 2008). In that study, the IFG activity was higher for actions 643 with unusual intentions compared to usual ones; the DMN was instead showing higher activity when participants were paying attention to the intentions rather than the means of the action. 644

Two recent studies provided more converging evidence to us. One RSA fMRI study about observed actions (Tucciarelli et al., 2019) used a large set of action categories, constructed a semantic similarity model of the meanings of the actions as well as other similarity models obtained from individual participants' behavioral sortings of similarities, and searched for corresponding RDMs in the brain. They found bilateral clusters in IFG/PMv, pIPS, and lateral occipitotemporal cortices (LOTC, close to EBA) corresponding to the semantic similarity model, although only the left LOTC cluster remained after RSA regression controlling the effects of

other models. That study controlled the inter-individual consistency of the stimuli perception
during the stimuli selection process, which was not done in the current study, and may explain
the discrepancy. Another study (Masson and Isik, 2021) examined the fMRI responses during
naturalistic movie watching using encoding models, and found that a social-affective model
(including features of an agent speaking, social interactions, theory of mind, perceived valence
and arousal) significantly explained the fMRI response in the left STS across two different
movies. This cluster fully overlapped with our L mSTS cluster.

In fact, the IFG and the areas in the DMN were all parts of the semantic network, thus their

660 involvements could also be studied in the future in the context of the semantic network, apart

661 from the context of "mentalizing" usually associated with the DMN. The importance of the

semantic network in emotion understanding is further supported by behavioral studies of

semantic dementia patients. In one such studies, three semantic dementia patients with left

temporal pole atrophy and impaired semantic knowledge were asked to sort faces into piles by

665 emotion. These patients were not able to distinguish between emotional faces with negative

valence, despite that their perception of positive/negative affect, the visual features for each

667 facial emotion and identity were intact (Lindquist et al., 2014).

668 Although we did not examine the temporal pole due to coverage limitation, our multivariate

669 RSA and functional connectivity results support the importance of the DMN/semantic network

670 in action and emotion understanding. These results are also consistent with the literature, that

the DMN did not show above-baseline activation, when passively observing actions (videos,

672 non-social, non-emotional) but without deliberating the goals/intentions (Van Overwalle and

Baetens, 2009). We further revealed that both the DMN/semantic network and the action

observation network were involved in the process of action and emotion understanding, that

they were consistently found in all 10 participants within the same functional connectivity

676 networks and showed similar multivariate patterns despite very different levels of univariate

activation. The multivariate methods worked in our study and in previous RSA studies, because

they took the multiple dimensions in the high-dimensional data into account, while the

679 univariate method considers only one or a few specific dimensions which associated with

680 specific contrasts (Haxby et al., 2011). Our study stresses the importance to further examine the

function of the DMN/semantic network in action and emotion understanding with multivariate

682 methods in future studies.

# 683 Understanding versus categorization

Previous experiments have mostly used explicit emotion and action categorization tasks with
 predefined categories, and implicitly used explicit categorization as a proxy to study subjective

686 understanding. However, categorization and understanding may involve different neural

substrates, as categorization involves some level of abstraction. Our study could not

688 differentiate between these two processes, because first, the RSA method could not disentangle

- 689 categorical boundaries driven by lower-level visual features, or concrete perceptual categories
- 690 (Hoemann et al., 2020; Mansouri et al., 2020) that were bound to individual exemplars, or
- abstract categories that generalize across exemplars. Second, what we obtained in the

- subjective reports were descriptions or labels of action and emotion for each individual
- 693 stimulus, rather than more abstract categorization to the group of stimuli. Thus, the resulting L
- 694 mSTS and L PMd clusters for perceived action and emotion may not represent the most abstract
- 695 level, and the more abstract categorization may happen in other areas functionally connected to
- 696 these two areas, and could perhaps utilize these category boundaries in computation. The
- 697 different levels of abstractions may also have driven the discrepancy between the analyses with
- subjective reports and predefined categories, although both analyses pointed to the strong
- 699 involvement of the caudate, which is in the executive loop of categorization learning tasks
- 700 (Seger, 2008; Seger and Miller, 2010). The areas found in our study could serve as target areas in
- future studies, to examine the level of abstraction.

#### 702 Brain areas for emotion understanding

Our analysis with the predefined non-emotion/emotion category model and with the perceived emotion model pointed to the PMd areas in the right and left hemispheres. This area may contain infromation exhibiting categorical boundaries between non-emotional/emotional stimuli, but again the level of abstraction could not be disentangled. Future studies could use emotional facial or voice stimuli, accompanied by subjective reports, to examine whether this area is specific to bodies and actions, or is more general for emotions in different modalities (Vaessen et al., 2019a).

- 710 We did not find evidence for coding of valence in frontal (higher-order) areas, either with
- 711 monotonic univariate activity modulation or with multivariate representations. However,
- valence coding may be bivalent, such that the vmPFC/mOFC activity monotonically increased for
- both positive and negative valence, as found in an RSA study (Chikazoe et al., 2014). This might
- be the case, as we found evidence that the vmPFC was showing a fine-grained pattern between
- similarly perceived emotions (Figure 5B). With only one category of positive emotion, we were
- not able to examine the possible bivalent activity.
- 717 In the analysis of the predefined action and emotion models, and in the univariate parametric
- 718 modulation of valence rating, we found adjacent/overlapping clusters in the L central sulcus
- 719 (Fig. 2D, Fig. 3). These clusters are also adjacent/overlapping to the two FC networks for
- perceived action and emotion (Fig. 6AB), but with different neural representations to the two
- seed regions for perceived action and emotion. This indicated that the primary sensorimotor
- areas were involved in action and emotion processing, although the representation there may
- be of lower-level features and may not be emotion-specific.

# Importance of using individualized subjective reports to study higher level cognition

- 726 With subjective reports, we found that the perceived action and emotion representations were
- 727 more high-dimensional and multi-faceted than the predefined category representations,
- consistent with the recent series of studies using subjective reports (Cowen and Keltner, 2017,
- 729 2020, 2021; Cowen et al., 2019). We also found neural representations corresponding to these
- rad subjective report representations in higher-level areas outside the ventral and dorsal pathways.

731 Apart from a few RSA studies which linked individualized behavioral data to the brain data 732 (Chikazoe et al., 2014; Stolier and Freeman, 2016; Tucciarelli et al., 2019), multiple previous RSA 733 studies either utilized only the predefined categories (Peelen et al., 2010), or the behavioral 734 ratings from an independent group (Bracci and Op de Beeck, 2016; Mur et al., 2013; Peelen et 735 al., 2014; Skerry and Saxe, 2015; Vaessen et al., 2019b), or the averaged behavioral responses of 736 the same participants scanned (Connolly et al., 2012; King et al., 2019). For high-level cognition 737 that has a more remote relation with the sensory stimuli and has more variability between 738 individual participants, the use of individualized behavioral data may be beneficial, as was 739 evidenced for studying object recognition (Charest et al., 2014), which was traditionally thought 740 to have considerable behavioral judgment similarities shared between participants. The multiple 741 pre-trained word embeddings are also relatively easy to use and enable objective analyses of 742 subjective reports.

#### 743 Advantages and limitations of the current 7T experiment

744 The use of high-resolution 7T fMRI in the current study has advantages and limitations. With 745 higher gray-white-matter contrast and higher temporal signal-to-noise ratio, the resulting data 746 were very robust (see the functional localizer data in Figure S2 and S3), the activation clusters in 747 individual participants were highly localized in the gray matter, and we observed small 748 subcortical clusters in multiple analyses at the group level, including periaqueductal gray, medial 749 geniculate nucleus, substantia nigra, red nucleus, and the septal nuclei. However, the current 7T 750 data also induced the possibility of false negatives, where the inter-individual 751 anatomical/functional variability was exacerbated and would no longer be compensated by 752 extensive smoothing, as observed again in the functional localizer data (Figure S3). Better 753 whole-brain group-level analysis schemes apart from the ROI analysis are needed to benefit 754 from both the high functional resolution and the large brain coverage, which potentially include 755 more fine-grained functional parcellations (Glasser et al., 2016; Yeo et al., 2011) and 756 hyperalignment (Haxby et al., 2011).

757

# 758 Acknowledgments

759 This study was supported by the European Research Council, under the European Union's

- 760 Seventh Frame-Work Programme (FP7/2007–2013)/ERC (grant numbers 295673 to BdG and
- 269853 to RG), by the ERC-Synergy program grant RELEVANCE (Grant agreement 856495 to
   BdG), and by the FPN-MBIC funding of Maastricht University to MZ and BdG. MZ was supported
- 763 by Fondation Bettencourt Schueller. We thank Federico de Martino for setting up the 7T
- 764 scanning sequences and Giancarlo Valente for providing a part of the searchlight code. We
- thank Maarten Vaessen for commenting on a previous version of the manuscript.

## 766 Author Contributions

- 767 Conceptualization: M.Z. and B.dG.; Methodology, Software, Validation, Formal Analysis,
- 768 Investigation, Data Curation, Writing Original Draft, Visualization: M.Z.; Resources: R.G.;
- 769 Writing Review & Editing: M.Z., R.G. and B.dG.; Project Administration, Funding Acquisition:
- 770 M.Z. and B.dG.; **Supervision**: B.dG.

# 771 **Declaration of Interests**

The authors declare no competing interests.

# 773 Data and code availability

- 774 The data and codes of this study are available at
- 775 <u>https://osf.io/cuh9v/?view\_only=efb12b7585ee4b6bbcfd7ca42c63b60d</u>, including: stimuli
- images with body joint estimations, subjective reports and corresponding word embeddings,
- 777 whole-brain statistical result maps, ROI raw data for the RSA analysis, structural images and
- 778 white-gray-matter-boundary segmentations.

# 779 Materials and Methods

#### 780 Participants

- 781 The data of 10 healthy right-handed participants recruited from the campus of Maastricht
- 782 University were included in the analyses (mean age =23.4, SD=1.955, 5 females.) Two more
- 783 participants took part in the study, but due to excessive head motion observed during the scan,
- their scanning sessions were either aborted, or data excluded from the analyses. All participants
- had normal or corrected-to-normal sight and had no history of psychiatric disorders. Participants
- 786 provided written consent before the study and received monetary reward afterwards. The
- 787 experiment was approved by the ethical committee of Maastricht University, and was carried
- out following the declaration of Helsinki. The experiment was conducted in English.

### 789 Data acquisition

- 790 The MR data were acquired in a 7T Magnetom full-body scanner (Siemens, Erlangen, Germany)
- in Scannexus, Maastricht University, with a Nova 1-transmitter/32-receiver head coil (Nova
- 792 Medical, Wilmington, USA). Dielectric pads were used for all participants except S10 (when the
- 793 head size didn't allow it), roughly covering bilateral occipito-temporal lobes. The stimuli were
- back-projected onto a screen behind the participants' head (Projector: Panasonic PT-EZ570EL,
- projected screen size 30 x 18 cm, resolution 1920 x 1200 pixels, refresh rate = 60 Hz, viewing
- distance ~99 cm, screen visual angle 17.23 x 10.38 degrees) and the participants viewed the
- rgr screen through a mirror fixed on the head coil. Participants came for two scanning sessions, a 2-
- hour session for the main experiment, and a 1-hour session for functional localizers and
- 799 anatomical scans.

800 Whole-brain anatomical data were collected for each participant with a resolution of 0.6 mm 801 isotropic (MPRAGE sequences, FOV=229 x 229 mm<sup>2</sup>, matrix size=384 x 384, flip angle=5. T1-802 weighted: TR=3100 ms, TE=2.52 ms; proton-density-weighted: TR=1440 ms, TE=2.52 ms). For 803 the functional runs, a 2D gradient-echo multi-band EPI sequence was used, with a resolution of 1.2 mm isotropic (multi-band acceleration factor=2 (Moeller et al., 2010), iPAT=3, FOV=172.8 x 804 805 172.8 mm<sup>2</sup>, matrix size=144 x 144, flip angle=75, number of slices=70, slice thickness=1.2 mm, no gap, ascending interleaved 2, TR=2000 ms, TE=21 ms, encoding direction Anterior to 806 807 posterior, reference scan mode: GRE, MB LeakBlock kernel: off, fat suppression enabled). In 808 each scanning session, a head scout was acquired for localization, then the B0 field map was 809 acquired and loaded in the console. The interactive shimming was performed before acquiring 810 the B1 field map. The system voltage was then computed according to the B1 map values (set to 811 a maximum of 190 V across all sessions), to have a 90 degree flipping angle at the white matter 812 beside the lateral ventricles, and the specific absorption rate (SAR) level for the longest 813 functional run (432 volumes) was controlled at below 75%. The slices were tilted in an angle that 814 covered most of the occipital lobe, parietal lobe and frontal lobe, while leaving out the anterior 815 temporal lobe, part of the motor cortex, and the orbitofrontal cortex. This was to ensure that 816 most of the important areas involved in body perception were covered, including EBA, fusiform 817 gyrus, IPS, IPL, PMd, PMv. The amygdala was not consistently covered given the relatively 818 limited coverage (covered in 6 out of 10 participants). Immediately before each functional run, a 819 5-volume run of the same setup but with posterior to anterior encoding direction was acquired 820 (Invert RO/PE polarity: on), for post-hoc top-up EPI distortion correction (See the fMRI data 821 preprocessing sub-section). We informed the participants the purposes of these distortion 822 correction runs, and instructed them not to move between the distortion correction run and the 823 actual functional run.

#### 824 Stimuli

- 825 The stimuli were gray-scale whole body images developed and validated in our lab (Stienen and
- de Gelder, 2011). They consisted of 8 actors, each posing 10 actions with or without emotional
- 827 content. The first 5 categories were neutral actions: combing hair (CH), drinking water (DW),
- opening door (OD), talking on the phone (PH), putting on trousers (TR); the 6<sup>th</sup> was neutral
- standing still (NE), and the last 4 were emotional expressions: fear (FE), anger (AN), happy (HA),
- sad (SA). The 80 postures were split into 2 balanced sets by randomly selecting images from 4
- actors for each category, resulting in 2 sets of 40 stimuli (4 images per category, 5 images per
- actor), which ensures that in each stimulus category the participants perceive as much
- variability in the posture and the identity as possible. Each participant saw one of the sets.
- The body stimuli were embedded a gray background (RGB value = 128, 128, 128), with all
- 835 internal facial information removed. They were sized to 400 x 600 pixels, and presented
- centrally on the screen (RGB value = 128, 128, 128). The whole-body shapes in the images
- overall spanned 309 x 492 pixels on the screen (visual angles=2.60 x 4.26 degrees).

#### 838 fMRI experiment design

839 The study used a slow event-related design. Stimuli were presented with Matlab (Version 840 R2012a, the MathWorks, Natick, USA) and Psychtoolbox 3.0.11 (Pelli, 1997). A white fixation 841 cross was present in the center of the screen throughout the experiment. The participants were 842 asked to always fixate on the fixation cross and take in the body posture as a whole. The 843 experiment consisted of 6 runs (14 min 18 s each, 429 volumes). Each run started with a fixation 844 period of 8 seconds, and then the whole set of 40 stimuli was presented to the participant twice 845 within each run. Each stimulus was presented for 500 ms, followed by an inter-stimulus interval of either 7.5, 9.5 or 11.5 s. The stimuli and the ISI were presented in a pseudorandomized order. 846 847 In addition, 4 catch trials were included in each run. Within each catch trial, a body posture was 848 randomly drawn from the stimulus set, while the fixation cross changed to either red or blue 849 (RGB color red= 195, 32, 30; blue= 10, 109, 195) during the stimulus presentation period. 850 Participants were asked to indicate the color by pressing the corresponding button of the button 851 box as soon as they saw the fixation change color. Two seconds were added to the inter-852 stimulus interval after each catch trial. Excluding the catch trials, each single stimulus image was 853 presented to the participant for 12 times throughout the main experiment. 9 participants 854 completed all 6 functional runs; 1 participant completed 5 runs.

#### 855 **Functional localizers**

856 A separate scanning session was devoted to acquiring functional localizer data and the structural 857 images. Stimuli were presented under passive viewing condition using Presentation software 858 (Version 16.0, Neurobehavioral Systems, Inc., Berkeley, USA). In the static localizer run (14 min 859 24 s, 432 volumes), after a 12 s fixation period, gray-scale stimuli of faces, houses, bodies, tools and words were presented in blocks of 12 s (12 stimuli per block, 800 ms stimuli presentation, 860 861 200 ms inter-stimulus interval), followed by resting periods of 12 s with the fixation cross on a 862 blank screen (RGB value=157,157,157). Each category block was presented 7 times, with a pseudorandomized presentation order for both the stimuli and the blocks. Facial stimuli were 863 864 front-view neutral faces from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998) 865 (24 identities, 12 males). The part below the neck (clothes, hair etc.) was removed from the face 866 images. Body stimuli were neutral still front-view bodies (de Gelder and Van den Stock, 2011) (20 identities, 10 males) from a different set than the one used in the main experiment, with the 867 868 facial information removed. House and tool images were obtained from the internet. The house 869 images consisted of 19 facades of houses with 2-to-3-storey height; the tool images consisted of 870 18 hand-held tools; words images consisted of high-frequency English words of 4-6 letters in Arial font. All the images were imbedded within a gray background (RGB value=157,157,157), 871 872 spanning a visual angle of 1.99 degrees (230 pixels).

The dynamic localizer run (5 min 36 s, 168 volumes) consisted of 1-s video clips of either facial or

bodily expressions, including neutral (coughing or clearing throat), angry, fear, happy (Kret et al.,

2011). The actors performed the actions against a green background, either wearing black

876 clothes for bodily expressions, or green clothes for facial expressions. Two exemplars were

selected for each expression category (in total 8 for faces and 8 for bodies). The actors in the

- 878 selected exemplars were all males. For the facial stimuli, two of the neutral expressions were
- 879 performed by the same identity. The facial and bodily expression clips were presented
- separately in blocks of 8 s, with the stimuli order pseudorandomized. The face and body blocks
- 881 were presented 10 times each, separated by an inter-block interval of 8 s, where a black screen
- and a white fixation cross was presented.

#### 883 Behavioral ratings

- 884 Right after the scanning session, participants completed a behavioral task outside the scanner. 885 Each of the 40 stimuli participant saw in the scanner was presented once using Psychopy (v1.83.04)(Peirce, 2007) on an LCD monitor (Acer VG248, resolution = 1920×1080, refresh rate = 886 887 60 Hz, whole-body size in the stimuli spanning the visual angle of 13.42 x 8.43 degrees). For each stimulus image, 6 questions were answered either with a 7-point scale (for implied motion and 888 889 valence respectively), or with an open answer with free typing (for the action the actor 890 performed, and the emotion). See Table S1 for details of the questions. We also recorded 891 whether the participants changed their perception during the scan (questions 2 and 3). Four 892 participants changed their perceptions (S4 and S9 for 2 stimuli, S6 for 7 stimuli, S10 for 12 893 stimuli). We used only the initial perception in the scanner (answer for question 1) for 894 subsequent analyses. The same stimulus image stayed on the screen for all 6 questions. 895 Participants answered the questions at their own paces (mean time spent=21.47 min, SD=6.98
- 896 min, range: 12.61 to 32.89 min).

#### 897 Data analysis

#### 898 Representational similarity analysis (RSA) for behavioral data

899 We mapped the subjective reports with Deconf word embeddings (Pilehvar and Collier, 2016), 900 which linked the Word2vec embeddings and the WordNet database. Word2vec embeddings 901 were trained on a very large corpus of text, able to capture various linguistic relationships 902 between words (Mikolov et al., 2013). After training, words with similar semantic meanings 903 were found to be situated closer to each other in the embedding vector space (cosine distance); 904 although for word2vec, different meanings of the same word were not disambiguated. On the 905 other hand, WordNet (Miller, 1995) is a curated lexicon database, where different meanings of 906 each single word were separated, and synonyms were grouped together; but it does not provide 907 a quantitative mapping for word similarities. Deconf embeddings mapped the individual word 908 meanings (senses) in WordNet into the 300-dimensional word2vec vector space, offering us 909 both a common high-dimensional semantic space trained by a large corpus, and the precise 910 separation of different meanings.

- 911 For behavioral free reports, we omitted the pronouns (e.g. he is, she is, his, her), as most of the
  912 times participants did not consistently type in the pronouns. The rest of the words/phrases
  913 (nouns, verbs, adjectives, adverbs) that had a corresponding entry in WordNet 3.1 were
  914 lemmatized (e.g. stretching→stretch). For words/phrases not found in WordNet 3.1, "how" was
  915 omitted; "something" was substituted by "thing"; "himself/herself/oneself" were substituted by
- 916 "self". The adverb "just" associated with verb phrases was omitted (e.g. just watching  $\rightarrow$  watch).

917 When describing the perceived action, "not sure" was substituted by "unsure"; when denoting

- that the participant had no idea about what the person was doing, "no idea" and "not sure"
- 919 were substituted by "not applicable". For words with multiple meanings ("senses" the term used
- 920 in WordNet), the corresponding sense was selected. When a noun denoting emotion has both
- 921 senses of <noun.feeling> and <noun.state>, the sense of <noun.feeling> was always selected.
- 922 There were very few cases that participants typed in "not". For one case the participant
- 923 denoting the emotion of the person in the stimuli "was not sure about something", it was
- substituted with the word "unsure". All the other cases were in responses for perceived acitons,
- about participant's own uncertain understanding for the stimuli. For these and similar cases we
- substituted the entry with "not applicable". For all response entries and word lists, see
- 927 <u>https://osf.io/cuh9v/?view\_only=efb12b7585ee4b6bbcfd7ca42c63b60d</u>.
- 928 For each free report entry after lemmatization, the 300-dimension word vector with the
- 929 corresponding sense number and sense key were selected from the Deconf pre-trained
- 930 embeddings. When an entry has multiple words, the vectors were averaged for this entry. The
- 931 RDM for each participant were then computed in cosine distance, as this metric was routinely
- used for computing word-embedding distances in the literature.
- The RDMs for behavioral ratings (implied motion, valence) were computed directly from theratings, in Euclidean distance.
- 935 For predefined RDMs, each category was binary-coded in vectors, with numbers of elements
- 936 corresponding to the total number of categories. E.g. for predefined actions, drinking water (2<sup>nd</sup>
- 937 category in 10) was coded as 0 1 0 0 0 0 0 0 0 0; for predefined non-emotion/emotion
- 938 categories, emotional ones were coded as 0 1. This coding assumes that each category was
- orthogonal to the others. The resulting RDMs were computed in Euclidean distance.
- 940 For RDM comparisons throughout the study, the Spearman's correlation was computed
- 941 between 40×40 RDMs, with the 780 values below the diagonal of the RDMs as inputs. The
- 942 resulting rho values were Fisher's Z transformed, submitted to a one-sample t test against 0
- 943 (two-tailed) at the group level. The group-averaged Z values were back transformed to rho
- 944 values (or 1 rho distance) and reported. To assess individual variability, we computed the
- 945 coefficient of variation (CV, sample standard deviation/sample mean, in %).
- 946 Since all 400 response entries for the perceived action and emotion representations were in the
- 947 common 300-dimensional word-vector space, we performed PCA on the 400 entries for each
- 948 representation (without centering, since the word vectors were all normalized). The PCA scores
- 949 for each principal component (PC) was computed into RDMs (Euclidean distance), and
- 950 correlated with 400×400 model RDMs. The model RDMs were created in the same way as the
- 40×40 RDMs, with the only difference that the 400 entries were sorted by action categories
- across all participants. Model RDMs included predefined action, predefined emotion, implied
- 953 motion, valence, individual subjects (entries of the same participant were coded as the same

954 vector). The PCA scores of all response entries were plotted in the two PCs that showed the

highest positive correlation to the model RDMs (Fig. 1I, J).

956

#### 957 fMRI data Preprocessing

The data were processed with BrainVoyager (version 20.0 and 20.2, Brain Innovation,
Maastricht, the Netherlands), MATLAB (version R2016a), and NeuroElf 1.0 toolbox implemented
in MATLAB (<u>http://neuroelf.net/</u>). Before any preprocessing, the functional data underwent in-

961 plane EPI distortion with the COPE plugin (v1.0) in BrainVoyager

962 (https://support.brainvoyager.com/brainvoyager/available-tools/86-available-plugins/62-epi-

963 <u>distortion-correction-cope-plugin</u>). The voxel-wise displacement was estimated between the

964 first volume of each functional run and the first volume of the preceding 5-volume correction

965 run (in reversed phase encoding direction). The in-plane voxel-wise displacement map was then

applied to all the volumes of the functional run. The first volume of each run was saved as a

separate file, and applied with distortion correction. This volume then served as the basis for
 subsequent motion correction and across-run alignments. After distortion correction, the

969 functional runs underwent slice scan time correction with the slice timetable information from

- 970 the scanner (interpolation: sinc), within-run 3D rigid motion correction with each run's first
- 971 volume as references (interpolation: trilinear for motion estimation, sinc for applying

972 transformations), and temporal high-pass filtering (GLM with Fourier basis set of 2 cycles,

973 including linear trend). The order of distortion correction and motion correction was assessed

974 on S06's sixth (last) main-task run, which has the biggest within-run rotation in all acquired runs

975 in the current experiment (about 3 degrees in the x axis). Between the two processing orders,

976 the resulting 429<sup>th</sup> volume showed negligible differences in the frontal lobe. Thus we believe the

order of distortion correction and motion correction was not critical. For the anatomical data,

978 the magnetic field inhomogeneity was corrected by dividing the T1 images with the PD images.

979 The anatomical data was then spatially normalized into the Talairach space.

980 To ensure that all functional runs of each participant (6 runs of the main experiment, 2

981 functional localizer runs, from two scanning sessions, distortion-corrected) were aligned well 982 with each other, we used a manual across-run alignment procedure, with careful visual

983 inspections and multiple iterations/checks. The first volume of the main-task run was aligned to

the anatomy, and then saved as an anatomical file in native space, keeping the T2\* weighted

985 contrast the same as the original functional run. This served as a "dummy" anatomical run for

986 the subsequent alignment procedure. Then the first volumes of all the functional runs (including

987 the first run itself) were aligned to this dummy anatomical run, and normalized into the

988 Talairach space (with the position information and transformation matrices for the T1 weighted

989 anatomical run). The first volumes of the 8 runs aligned in the Talairach space were then screen-

990 captured and saved as different layers in Photoshop (version CS6, Adobe, United States), toggled

991 on and off, and made into GIF animations to check alignment qualities across runs. Adjustments

992 were made for imperfect aligned runs, followed by the same checking procedure, until there

993 were not any big shifts/translations across runs at the whole-brain level. This procedure allowed

us to visually spot tiny alignment imperfections across runs with the same T2\* weighted
 modality. It is much easier than spotting misalignments across T2\* weighted and T1 weighted

100 ality. It is much easier than spotting misalignments across 12° weighted and 11 weighted

- 996 image modalities in the conventional alignment procedure, where the two images look very
- 997 different across modalities. The quality of the distortion correction was also checked when
- aligning the functional run 1 to the anatomy, where a good distortion correction resulted in a
- good shape correspondence between the T2\* weighted and T1 weighted data. In some cases,
  when the participants moved their head between the distortion correction run and the
- 1001 functional task run, despite our repetitive instructions, the distortion correction quality was
- 1002 affected. In those cases, we strived for aligning the occipital and temporal lobes.
- After alignment across runs, the functional data of the main experiment were then spatially smoothed with 6 mm FWHM for univariate analysis (comparable to the RSA searchlight radius), 3 mm FWHM for task residual functional analysis. The unsmoothed data were used for the representational similarity analysis, and univariate analysis of 10 categories in RSA searchlight clusters. The static functional localizer data was smoothed at 6 mm FWHM for group randomeffect GLM, 3 mm for individual participants, and not smoothed for univariate analysis in RSA searchlight clusters.
- 1010 All results in this study were computed in the volume space. For visualizing group-level results,
- 1011 cortex-based alignments (CBA) were performed to alleviate the high inter-individual anatomical
- 1012 variability. For each participant, the gray-white-matter boundary of the anatomical images in
- 1013 Talairach space underwent automatic segmentation, and careful manual correction slice-by-
- slice. Then a mesh with a high number of vertices was created for the white-gray-matter
- 1015 boundary of each hemisphere (number of vertices ranging from 370k to 434k), inflated into a
- sphere, corrected for distortions across vertices, and mapped to a high-resolution standard
   sphere with down sampling (number of vertices=163842). The curvature patterns of the original
- 1018 mesh were also smoothed and mapped to the standard sphere. The group-level cortex-based
- 1019 alignment was separately performed for the left and right hemispheres. For each hemisphere,
- the curvature patterns of all 10 participants' meshes were aligned into a dynamic group
- average. After aligning, the average left and right hemispheres of the whole group were created.
  This resulting group-average brain surface retained many anatomical landmarks, and was more
  detailed than the ones usually created with 3T data. However, note that it was a bit smaller than
- detailed than the ones usually created with 3T data. However, note that it was a bit smaller than
  the actual brain, and was only meant to provide visual landmarks to localize the clusters.

### 1025 Univariate analysis

- To compare the univariate activation to the literature, three different general linear models
  (GLMs, serial correlation correction: AR(2), data %-normalized before performing the GLM) were
  applied to the datasets smoothed 6 mm FWHM, with the following predictors: 1) 10 action
  categories; 2) the behavioral ratings of implied motion; 3) the behavioral ratings of valence. For
  and 3), the presentation of all stimuli were defined as a single main predictor, the ratings for
  40 stimuli were z-scored within each participant, and served as a parametric weight factor to
  each corresponding stimulus. The parametric modulation effects were subsequently computed
- 1033 for 2) and 3). For all the GLM models above, the time courses were %-transformed, the main

predictors were convolved with a two-gamma hemodynamic response function, and the 6
parameters of participant's head motion were z-scored and entered as confounding factors. The
catch trials were marked as a separate condition, and their parametric estimates (beta values)
were not used in subsequent analyses.

- 1038 The group random-effect GLM analysis was performed for each predictor set. For the 10-
- 1039 category predictor set, apart from the contrasts between action categories, a whole-brain
- 1040 ANOVA was performed, with a 10-level factor 'action categories'. Cluster size thresholds for all
- 1041 group-level maps in this study were estimated using Monte-Carlo simulation (alpha level=0.05,
- 1042 numbers of iterations=5000, initial p<0.005 for univariate and RSA searchlight analyses; initial
- 1043 p<0.001 for task-residual functional connectivity), with the BrainVoyager plugin Cluster-Level
- 1044 Statistical Threshold Estimator (<u>https://support.brainvoyager.com/brainvoyager/functional-</u>
- 1045 <u>analysis-statistics/46-tresholding-multiple-comparisons-problem/226-plugin-help-cluster-</u>
- thresholding), masked with the common functional data coverage across 10 participants. This
   mask was created from the averaged functional images across participants, covering 830178
- 1048 functional voxels.
- 1049 The GLM of the static functional localizer (without smoothing, and smoothed 3 mm FWHM) was
- 1050 performed for each individual participant. Since we do not assume that the bodies should only
- 1051 be processed in category-specific areas, the static functional localizer data were used as an
- 1052 indicator for inter-individual variability, and an indicator for the neural processes in some RSA1053 searchlight clusters.
- 1054 Representational similarity analysis for fMRI data
- 1055 For RSA analyses, the GLM model with 40 stimuli was fitted to each individual participants' unsmoothed data, resulting in 40 t-maps, one per stimulus. The neural RDMs across the 40 1056 1057 stimuli were computed with Pearson's correlation distance Searchlight spheres were 1058 constructed for each voxel, with a radius of 5 voxels (515 voxels within the sphere, 889.92 mm<sup>3</sup>), 1059 comparable to the value for univariate smoothing. Spheres containing more than 172 non-zero 1060 voxels were included in the analysis). The Spearman correlation values between the neural RDM 1061 and the model RDM was Fisher-Z transformed, and wrote back to the sphere center voxel. The 1062 group-level significance of the correlation was evaluated by a one-sample t-test against zero (2-1063 tailed).
- 1064 For the clusters found by the searchlight RSA and subsequent whole-brain analyses, the
- univariate percent signal changes for each of the 10 action categories were extracted (withoutsmoothing), and were compared to the baseline (one-sample t test against 0) in SPSS.
- 1067 **RSA regression with low/mid-level features**
- 1068 For each of the stimuli image, we extracted the body joint locations by the OpenPose library
- 1069 (Cao et al., 2017) (<u>https://github.com/CMU-Perceptual-Computing-Lab/openpose</u>). The output
- 1070 was the x, y values and a confidence score for the estimation, for 18 body joints (See **Figure 4A**).
- 1071 The joint locations were imported in Adobe Illustrator CS6, overlaid on the stimulus image, and

manually adjusted with visual estimation. The correspondence of the joints and their location
are as follows: joint 1 (head)-nose tip; joint 2 (neck)-the jugular notch of the sternum; joint 3 & 6
(shoulders)-humerus; joint 4 & 7 (elbows)-the connecting point of the upper/lower arms; joint 5
& 8 (hands)-the wrist end of the radial bone; joint 9 & 12 (waist)-the widest points of the femur;
joint 10 & 13 (knees)-on the patella; joint 11 & 14 (feet)-the ankle end of the tibia; joint 15 & 16
(eyes); joint 17 & 18 (ears). Since the eyes were not present on the stimuli images (facial

- 1078 features blurred), and the ears were often occluded by the hair or by the head, these joints were
- 1079 visually estimated with linear perspective in mind.
- We constructed low-level visual feature RDMs, for raw pixel values (as vectors), and raw
  coordinates of the 14 joints (excluding joints for ears and eyes), although the resulting RDMs
  were extremely similar to the ones including these 4 joints (RDM similarity rho=0.976, 0.98 for
  stimuli set A and B).
- The mid-level visual feature RDMs included head, shoulder, waist orientations, as well as hands/feet-to-head distances. The head orientation was individually computed by the distance between ears, normalized (divided) by the ear distance of the standing still condition, capped at 1 (1: facing the viewer; smaller than 1: facing left- or right-ward). The shortest hand-to-head distance may reflect whether the hand enters the head's peripersonal space of the actor; the
- 1089 average feet-to-head distance may reflect the relaxation of the torso and legs of that actor.
- 1090 For RSA regression in each of the 10-action-category areas (resulted from whole-brain 1091 searchlight analysis of the 10-action-category model RDM), the low/mid-level visual feature 1092 RDMs and the higher-level stimuli attribute RDMs (implied motion, valence, actor identity) were 1093 squared for each element (to satisfy the linearity for squared Euclidean distance values), z-1094 scored across elements, then put together in the same linear regression model as predictors. 1095 Each element of the neural RDM was also correspondingly squared. A set of beta values for 1096 these predictors were obtained for each participant, and their group-level significance was 1097 examined by one-sample t-test against zero (2-tailed), and FDR-corrected at q<0.05 (Storey,
- 1098 2002) across all predictors for each searchlight cluster.

#### 1099 Task residual functional connectivity analysis

1100 We smoothed the task data at 3 mm FWHM, regressed out the task-related activity by 1101 deconvolution analysis (5 stick predictors per stimulus, covering the evolvement of the BOLD 1102 shape per trial). For each seed region, the residual time course was extracted and averaged 1103 across voxels, and correlated with the residual time courses of all voxels in each run, resulting in 1104 one R map per run. The connectivity pattern across the runs were stable for all participants. The 1105 R maps were Fisher's Z-transformed, averaged across runs per participant, and put in a onesample t-test against 0 (2 tailed). The resulting group-level maps were thresholded at p<0.001 1106 1107 for cluster size correction.

- 1108 For the 10-action-category areas, the inter-area functional connectivity was also examined, by
- 1109 correlating the seed ROI time courses per run, then again with the same group-analysis
- 1110 procedure (Z-transform, averaging, t test), with FDR correction.

#### 1111 Hierarchical clustering of RDMs across functional connectivity clusters

- 1112 We computed neural RDMs for all group-level clusters obtained from the task-residual
- 1113 functional connectivity analysis (51 ROIs for the perceived action FC network, 76 ROIs for the
- 1114 perceived emotion FC network), and computed second-level RDMs (Spearman correlation
- distance, Figure 6F) from them within each participant. For the two FC networks, the second-
- 1116 level RDM were averaged (with Z-transform) across participants (Figure 6A, B, right panels), and
- served as inputs for hierarchical clustering (MATLAB function: linkage; method: average. Plot
- 1118 function: dendrogram; clustering threshold for the most similar areas to the seed region: 0.8).
- 1119 We also examined the univariate activation for 10 action categories, and plotted histograms.
- 1120
- 1121
- 1122

#### 1123 **References**

Amodio, D.M., and Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social
cognition. Nature Reviews Neuroscience *7*, 268–277.

- 1126 Andrews-Hanna, J.R. (2012). The brain's default network and its adaptive role in internal 1127 mentation. Neuroscientist *18*, 251–270.
- Aviezer, H., Hassin, R., Bentin, S., and Trope, Y. (2008). Putting facial expressions back in context.
  First Impressions 255–286.
- 1130 Barrett, L.F., Adolphs, R., Marsella, S., Martinez, A.M., and Pollak, S.D. (2019). Emotional
- 1131 Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements.
- 1132 Psychological Science in the Public Interest *20*, 1–68.
- Binder, J.R., Desai, R.H., Graves, W.W., and Conant, L.L. (2009). Where is the semantic system? A
  critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex *19*,
  2767–2796.
- Boleda, G. (2020). Distributional Semantics and Linguistic Theory. Annual Review of Linguistics 6,213–234.
- Bracci, S., and Op de Beeck, H. (2016). Dissociations and Associations between Shape and
  Category Representations in the Two Visual Pathways. J. Neurosci. *36*, 432–444.
- Buckner, R.L., Andrews-Hanna, J.R., and Schacter, D.L. (2008). The Brain's Default Network.
  Annals of the New York Academy of Sciences *1124*, 1–38.

- Bufacchi, R.J., Liang, M., Griffin, L.D., and Iannetti, G.D. (2016). A geometric model of defensive peripersonal space. Journal of Neurophysiology *115*, 218–225.
- Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience *10*, 186–198.
- 1146 Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime Multi-person 2D Pose Estimation1147 Using Part Affinity Fields. (IEEE), p.
- 1148 Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2019). OpenPose: Realtime Multi-1149 Person 2D Pose Estimation using Part Affinity Fields. ArXiv:1812.08008 [Cs].
- Caspers, S., Zilles, K., Laird, A.R., and Eickhoff, S.B. (2010). ALE meta-analysis of action
  observation and imitation in the human brain. Neuroimage *50*, 1148–1167.
- Charest, I., Kievit, R.A., Schmitz, T.W., Deca, D., and Kriegeskorte, N. (2014). Unique semantic
  space in the brain of each beholder predicts perceived similarity. PNAS *111*, 14565–14570.
- Chikazoe, J., Lee, D.H., Kriegeskorte, N., and Anderson, A.K. (2014). Population coding of affect
  across stimuli, modalities and individuals. Nat Neurosci *17*, 1114–1122.
- 1156 Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., and
  1157 Haxby, J.V. (2012). The Representation of Biological Classes in the Human Brain. Journal of
  1158 Neuroscience *32*, 2608–2618.
- 1159 Cowen, A.S., and Keltner, D. (2017). Self-report captures 27 distinct categories of emotion
  1160 bridged by continuous gradients. PNAS *114*, E7900–E7909.
- 1161 Cowen, A.S., and Keltner, D. (2020). What the face displays: Mapping 28 emotions conveyed by1162 naturalistic expression. American Psychologist *75*, 349–364.
- 1163 Cowen, A.S., and Keltner, D. (2021). Semantic Space Theory: A Computational Approach to1164 Emotion. Trends in Cognitive Sciences *25*, 124–136.
- 1165 Cowen, A.S., Elfenbein, H.A., Laukka, P., and Keltner, D. (2019). Mapping 24 emotions conveyed
  1166 by brief human vocalization. American Psychologist 74, 698–712.
- DiCarlo, J.J., and Cox, D.D. (2007). Untangling invariant object recognition. Trends in Cognitive
  Sciences *11*, 333–341.
- DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How Does the Brain Solve Visual ObjectRecognition? Neuron *73*, 415–434.
- 1171 Dricu, M., and Frühholz, S. (2016). Perceiving emotional expressions in others: Activation
- likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental
   perception of emotions. Neuroscience & Biobehavioral Reviews *71*, 810–828.
- 1174 Ekman, P. (1999). Basic emotions. Handbook of Cognition and Emotion *98*, 16.

- 1175 Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical Representation 1176 of Visual Stimuli in the Primate Prefrontal Cortex. Science *291*, 312–316.
- de Gelder, B. (2006). Towards the neurobiology of emotional body language. Nature ReviewsNeuroscience 7, 242–249.
- de Gelder, B., and Van den Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST).
  Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body
  Expression of Emotions. Front Psychol 2, 181–181.
- de Gelder, B., Snyder, J., Greve, D., Gerard, G., and Hadjikhani, N. (2004). Fear fosters flight: a
  mechanism for fear contagion when perceiving emotion expressed by a whole body. Proc Natl
  Acad Sci U S A *101*, 16701–16706.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K.,
  Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human
  cerebral cortex. Nature *536*, 171–178.
- 1188 Goeleven, E., De Raedt, R., Leyman, L., and Verschuere, B. (2008). The Karolinska Directed 1189 Emotional Faces: A validation study. Cognition & Emotion *22*, 1094–1118.
- Grafton, S.T., and Hamilton, A.F. de C. (2007). Evidence for a distributed hierarchy of action
  representation in the brain. Hum Mov Sci *26*, 590–616.
- Hartley, C.A., and Poeppel, D. (2020). Beyond the Stimulus: A Neurohumanities Approach toLanguage, Music, and Emotion. Neuron *108*, 597–599.
- Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke,
  M., and Ramadge, P.J. (2011). A Common, High-Dimensional Model of the Representational
  Space in Human Ventral Temporal Cortex. Neuron *72*, 404–416.
- Hebart, M.N., Dickter, A.H., Kidder, A., Kwok, W.Y., Corriveau, A., Wicklin, C.V., and Baker, C.I.
  (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object
  images. PLOS ONE *14*, e0223792.
- Hein, G., and Knight, R.T. (2008). Superior Temporal Sulcus—It's My Area: Or Is It? Journal of
  Cognitive Neuroscience 20, 2125–2136.
- Hoemann, K., Wu, R., LoBue, V., Oakes, L.M., Xu, F., and Barrett, L.F. (2020). Developing an
  Understanding of Emotion Categories: Lessons from Objects. Trends in Cognitive Sciences 24,
  39–51.
- Ju, H., and Bassett, D.S. (2020). Dynamic representations in networked neural systems. Nature
   Neuroscience 23, 908–917.
- Kanai, R., and Rees, G. (2011). The structural basis of inter-individual differences in human
  behaviour and cognition. Nature Reviews Neuroscience *12*, 231–242.

1209 King, M.L., Groen, I.I.A., Steel, A., Kravitz, D.J., and Baker, C.I. (2019). Similarity judgments and

- 1210 cortical visual responses reflect different properties of object and scene categories in naturalistic1211 images. NeuroImage *197*, 368–382.
- 1212 Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., and Wager, T.D. (2008).
- Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis ofneuroimaging studies. NeuroImage 42, 998–1031.
- 1215 Kret, M.E., and de Gelder, B. (2010). Social context influences recognition of bodily expressions.
  1216 Experimental Brain Research *203*, 169–180.
- 1217 Kret, M.E., Pichon, S., Grèzes, J., and de Gelder, B. (2011). Similarities and differences in
  1218 perceiving threat from dynamic faces and bodies. An fMRI study. NeuroImage *54*, 1755–1762.
- Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini,
  P.A. (2008). Matching categorical object representations in inferior temporal cortex of man and
  monkey. Neuron *60*, 1126–1141.
- de Lange, F.P., Spronk, M., Willems, R.M., Toni, I., and Bekkering, H. (2008). Complementary
  Systems for Understanding Action Intentions. Current Biology *18*, 454–457.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., and van Knippenberg, A.
  (2010). Presentation and validation of the Radboud Faces Database. Cognition & Emotion 24,
  1377–1388.
- LeDoux, J.E., and Hofmann, S.G. (2018). The subjective experience of emotion: a fearful view.Current Opinion in Behavioral Sciences *19*, 67–72.
- Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., and Barrett, L.F. (2012). The brain basis
  of emotion: a meta-analytic review. Behav Brain Sci *35*, 121–143.
- Lindquist, K.A., Gendron, M., Feldman Barrett, L., and Dickerson, B.C. (2014). Emotion
  perception, but not affect perception, is impaired with semantic memory loss. Emotion *14*, 375–
  387.
- 1234 Lundqvist, D., Flykt, A., and Öhman, A. (1998). Karolinska Directed Emotional Faces.
- 1235 Mansouri, F.A., Freedman, D.J., and Buckley, M.J. (2020). Emergence of abstract rules in the 1236 primate brain. Nature Reviews Neuroscience *21*, 595–610.
- Masson, H.L., and Isik, L. (2021). Functional selectivity for naturalistic social interaction
  perception in the human superior temporal sulcus. BioRxiv 2021.03.26.437258.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient Estimation of Word
  Representations in Vector Space. ArXiv:1301.3781 [Cs].
- Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., and Uğurbil, K. (2010).
  Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging

with application to high spatial and temporal whole-brain fMRI. Magn Reson Med *63*, 1144–1153.

Molenberghs, P., Cunnington, R., and Mattingley, J.B. (2012). Brain regions with mirror
properties: A meta-analysis of 125 human fMRI studies. Neuroscience & Biobehavioral Reviews
36, 341–349.

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P.A., and Kriegeskorte, N. (2013). Human
Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. Front.
Psychol. 4.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A
toolbox for representational similarity analysis. PLoS Comput Biol *10*, e1003553–e1003553.

Peelen, M.V., Atkinson, A.P., and Vuilleumier, P. (2010). Supramodal representations of
perceived emotions in the human brain. J Neurosci *30*, 10127–10134.

Peelen, M.V., He, C., Han, Z., Caramazza, A., and Bi, Y. (2014). Nonvisual and Visual Object Shape
Representations in Occipitotemporal Cortex: Evidence from Congenitally Blind and Sighted
Adults. Journal of Neuroscience *34*, 163–170.

1258 Peirce, J.W. (2007). PsychoPy--Psychophysics software in Python. J Neurosci Methods *162*, 8–13.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbersinto movies. Spatial Vision *10*, 437–442.

Pilehvar, M.T., and Collier, N. (2016). De-Conflated Semantic Representations. (Association forComputational Linguistics), p.

Righart, R., and de Gelder, B. (2006). Context Influences Early Perceptual Analysis of Faces—An
Electrophysiological Study. Cerebral Cortex *16*, 1249–1257.

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., and Rozzi, S. (2014). Cortical Mechanisms
Underlying the Organization of Goal-Directed Actions and Mirror Neuron-Based Action
Understanding. Physiological Reviews *94*, 655–706.

Saxe, R., Moran, J.M., Scholz, J., and Gabrieli, J. (2006). Overlapping and non-overlapping brain
regions for theory of mind and self reflection in individual subjects. Soc Cogn Affect Neurosci 1,
229–234.

Seger, C.A. (2008). How do the basal ganglia contribute to categorization? Their roles in
generalization, response selection, and learning via feedback. Neurosci Biobehav Rev 32, 265–
278.

Seger, C.A., and Miller, E.K. (2010). Category learning in the brain. Annu Rev Neurosci 33, 203–
219.

Seghier, M.L., and Price, C.J. (2018). Interpreting and Utilising Intersubject Variability in Brain
Function. Trends in Cognitive Sciences 22, 517–530.

- 1278 Sinke, C.B.A., Sorger, B., Goebel, R., and de Gelder, B. (2010). Tease or threat? Judging social 1279 interactions from bodily expressions. NeuroImage *49*, 1717–1727.
- Skerry, A.E., and Saxe, R. (2015). Neural Representations of Emotion Are Organized around
  Abstract Event Features. Current Biology *25*, 1945–1954.
- Smith, P.L., and Little, D.R. (2018). Small is beautiful: In defense of the small-N design. Psychon
  Bull Rev 25, 2083–2101.
- Stienen, B.M.C., and de Gelder, B. (2011). Fear detection and visual awareness in perceiving
  bodily expressions. Emotion *11*, 1182–1189.
- Stolier, R.M., and Freeman, J.B. (2016). Neural pattern similarity reveals the inherent
  intersection of social categories. Nature Neuroscience *19*, 795–797.
- Storey, J.D. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical
  Society: Series B (Statistical Methodology) *64*, 479–498.
- 1290 Tucciarelli, R., Wurm, M., Baccolo, E., and Lingnau, A. (2019). The representational space of 1291 observed actions. ELife *8*, e47686.
- Vaessen, M., Heijden, K.V. der, and Gelder, B. de (2019a). Decoding of emotion expression in the
  face, body and voice reveals sensory modality specific representations. BioRxiv 869578.
- Vaessen, M.J., Abassi, E., Mancini, M., Camurri, A., and de Gelder, B. (2019b). Computational
  Feature Analysis of Body Movements Reveals Hierarchical Brain Organization. Cerebral Cortex
  29, 3551–3560.
- 1297 Van den Stock, J., Hortensius, R., Sinke, C., Goebel, R., and de Gelder, B. (2015). Personality traits1298 predict brain activation and connectivity when witnessing a violent conflict. Scientific Reports 5.
- 1299 Van Overwalle, F., and Baetens, K. (2009). Understanding others' actions and goals by mirror1300 and mentalizing systems: A meta-analysis. NeuroImage *48*, 564–584.
- Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L.,
  Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral
  cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology *106*, 1125–
  1165.
- Zhang, Y., Han, K., Worth, R., and Liu, Z. (2020). Connecting concepts in the brain by mapping
  cortical representations of semantic relations. Nature Communications *11*, 1877.

1307