Brain representations of social knowledge during mental simulation

Daniel Alcalá-López[1], David Soto[1,2]

TH852

d.alcala@bcbl.eu

@danalclop

[1]Basque Center on Cognition, Brain and Language (BCBL), Donostia – San Sebastian, Spain
[2]Ikerbasque, Basque Foundation for Science, Bilbao, Spain

BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

Motivation

How does the brain represent conceptual meaning about the behavior of other individuals?

We have extensive evidence that we can use BOLD activity to classify conceptual knowledge using a variety of concrete concepts (e.g. animals or tools; JV Haxby 2012 Neuroimage).

Only recently there has been a similar interest in studying the brain representations of abstract concepts (M Ghio et al. 2016 Neuroimage; Y Wang et al. 2017 PNAS).

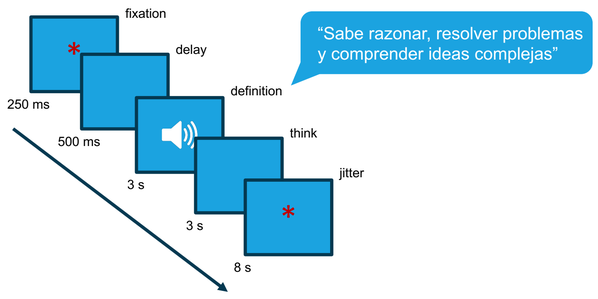
Abstract conceptual knowledge about others likely involves a range of underlying dimensions, yet it is unknown whether (and how) the brain maps different aspects of social information. We suggest and explore two candidate dimensions of interest for the study the representation of social information: **affect** and **likableness** (NH Anderson 1968 J Pers Soc Psychol).

The present study used MRI-based multivariate pattern analysis (MVPA) to investigate the brain representations of a specific type of abstract concepts: those we use to describe how others behave.

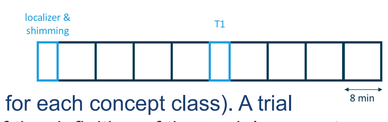
Methods

We asked participants (n = 30) to rate on a scale from 0 to 100 how *likable* and *affectionate* was each social concept before and after the scanning.

During the fMRI session participants listened to short **definitions of social concepts** and then mentally simulated another individual behaving the way it was described in the definition:



A scanning session was composed of 8 functional runs and a structural run halfway through the session.



Each functional run contained 36 trials (one trial per social concept; 9 concepts for each concept class). A trial consisted on the auditory presentation of the definition of the social concept for 3 seconds followed by another short period of 3 seconds to **mentally simulate the referred behavior**.

fMRI data was preprocessed as follows (i) removal of non-brain tissue using FSL's BET; (ii) volume realignment using MCFLIRT; (iii) gaussian kernel (FWHM = 3mm) for spatial smoothing; (iv) ICA-based automatic removal of motion artefacts; (v) temporal filtering (high-pass; cutoff = 60s); (vi) coalignment of each session to a reference volume (1st session).

After stacking, detrending, and z-scoring each ROI's BOLD data we used an **SVM-based linear classifier** to decode the brain representation of social concepts regarding: (i) their *likableness* (high vs. low), (ii) their *affect* (high vs. low), as well as (iii) in a multi-class classification problem that included both dimensions of interest. We used a PCA-based feature selection within each ROI and leave-one-out cross-validation (300 permutations).

Statistical significance of decoding performance at the group level was assessed using a one-sample t-test across subjects, corrected by the number of ROIs.

We focused on a set of **regions of interest** (ROIs) based on previous studies on semantic (JR Binder et al. 2009 Cereb Cortex) and social information processing (D Alcalá-López et al. 2017 Cereb Cortex):

+ 3 semantic regions (lateral temporal lobe, LTL; inferior frontal gyrus, IFG; and precuneus, Prec)
+ 3 social regions (insula, Ins; anterior cingulate cortex, ACC; posterior cingulate cortex, PCC)
+ 2 semantic & social regions (anterior temporal lobe, ATL; anterior prefrontal cortex, aPFC)
+ 1 control region (primary visual cortex, V1)

Take-home messages

Canonical semantic regions outperformed putative social regions when classifying distinct classes of conceptual knowledge about others. Particularly, the lateral temporal lobe (LTL) contains especially rich social information.

The anterior cingulate cortex (ACC) shows a clear preference to classify the *likableness* of a social concept. Can this be a consequence of the saliency of this conceptual domain?

Alcalá-López, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogeley, K., Mars, R. B., ... & Bzdok, D. (2017). Computing the social brain connectome across systems and states. *Cerebral cortex*, 28(7), 2207-2232. | **Anderson**, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of personality and social psychology*, 9(3), 272. | **Binder**, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767-2796. | **Haxby**, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage*, 62(2), 852-855. | **Ghio**, M., Vaghi, M. M. S., Perani, D., & Tettamanti, M. (2016). Decoding the neural representation of fine-grained conceptual categories. *Neuroimage*, 132, 93-103. | **Wang**, Y., Collins, J. A., Koski, J., Nugiel, T., Metoki, A., & Olson, I. R. (2017). Dynamic neural architecture for social knowledge retrieval. *Proceedings of the National Academy of Sciences*, 114(16), E3305-E3314.

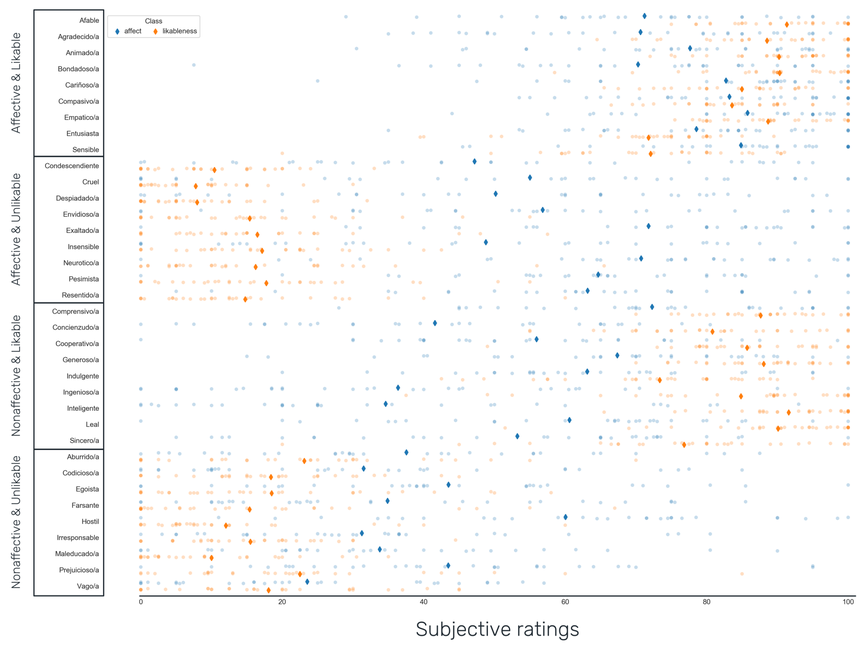
Decoding the brain representation of social concepts

ID	Concept	Pre	Post	Class
sub-001	Aburrido/a	20	10	likableness
sub-001	Afable	65	70	likableness
sub-001	Agradecido/a	81	72	likableness
sub-001	Animado/a	95	80	likableness
...	...	...	...	
sub-030	Resentido/a	50	70	affect
sub-030	Sensible	100	95	affect
sub-030	Sincero/a	50	50	affect
sub-030	Vago/a	10	0	affect

ICC_{affect} = 0'47 [0,36-0,60]

$ICC_{likableness}$ = 0'93 [0,89-0,96]

1 The intraclass correlation coefficient (ICC) shows that test-retest reliability of the subjective ratings was fair for *affect* ratings and excellent for *likableness* ratings.

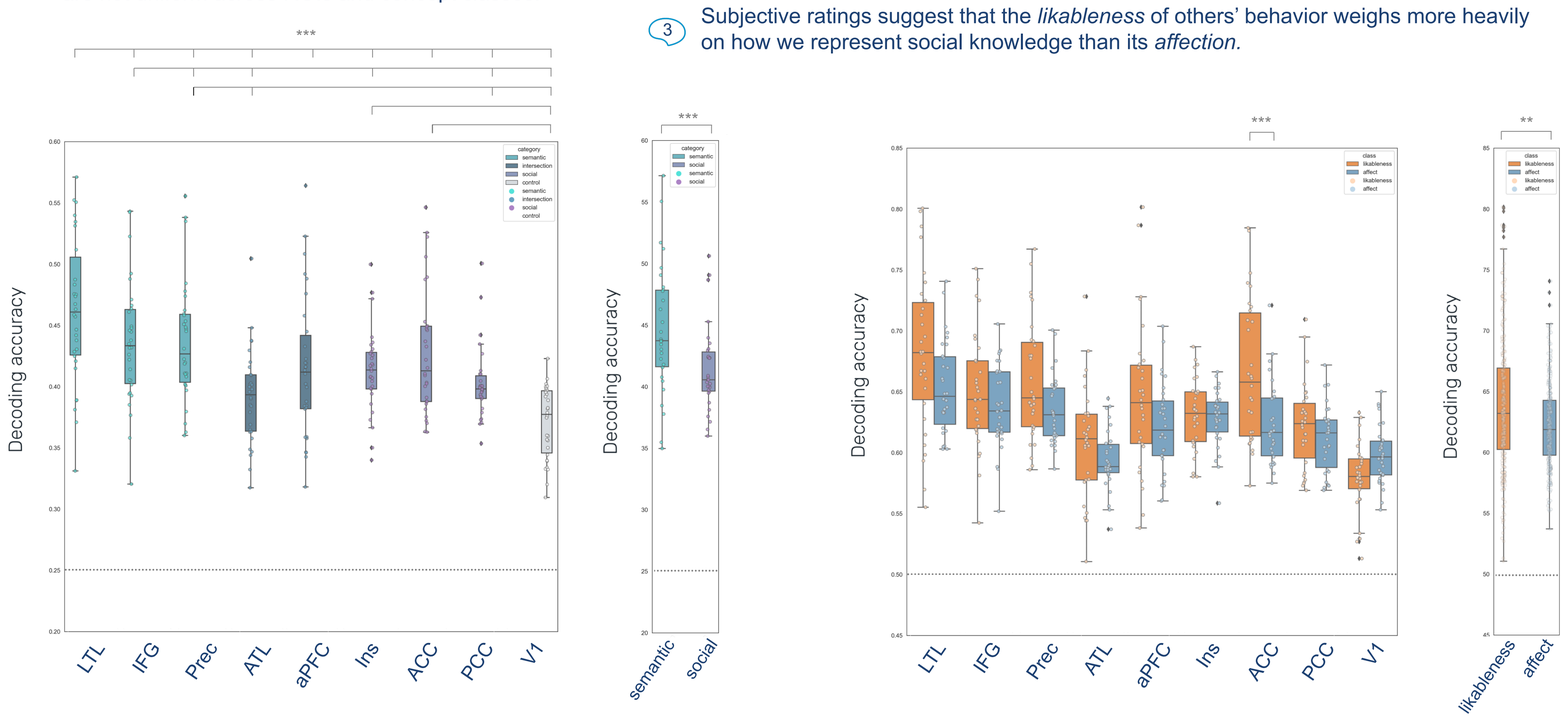


	Sum of Squares	df	Mean Square	F	p	η^2
ROI	0.289	4.495	0.064	39.395	< .001	0.576
CLASS	0.034	1.000	0.034	7.801	0.009	0.212
ROI * CLASS	0.037	5.699	0.006	7.640	< .001	0.209

2 Behavioral ratings show that participants could clearly identify highly likable as well as highly unlikable social concepts (*orange*), whereas they rated the affective component of social concepts (*blue*) in a less unambiguous fashion.

3 Subjective ratings suggest that the *likableness* of others' behavior weighs more heavily on how we represent social knowledge than its *affection.*

4 A repeated measures ANOVA with two factors (*ROI*, 9 levels; concept *CLASS*, 2 levels) shows that accuracies are not uniform across ROIs and concept classes.



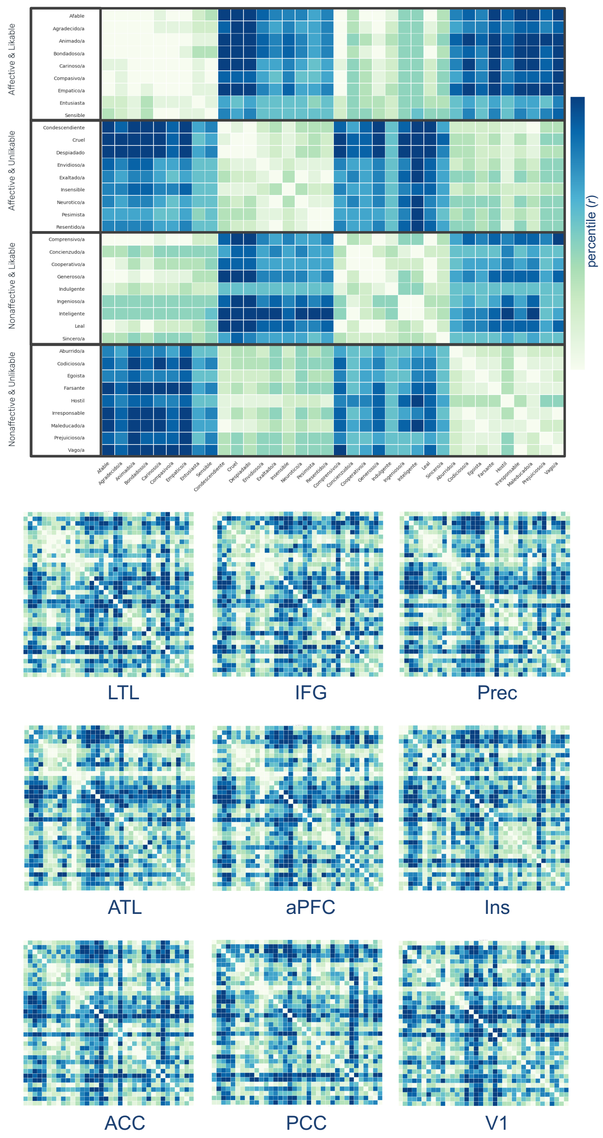
5 Decoding social concepts in a multi-class classification task shows clear regional differences.

6 *Semantic ROIs* show in average a statistically higher classification accuracy than *social ROIs*.

7 Decoding accuracies for the *likableness* classification task were significantly higher than for the *affect* classification task.

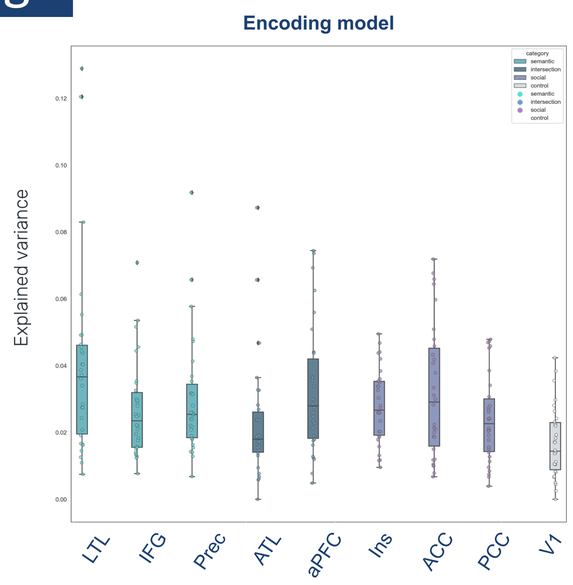
8 We found a significant interaction effect between *ROI* and concept class (*likableness* vs. *affect*) in the ACC.

Further exploration of subjective ratings



We used the subjective ratings to create a voxel-based **encoding model** of our social concepts.

The results from fitting the model to the voxels of each ROI are congruent with our previous decoding approach.



We then used the subjective ratings to create a 2D space that characterized the **dissimilarity** patterns between each pair of social concepts.

Looking at behavioral ratings we see that *likableness* seems to be the underlying feature driving (dis)similarity between social concepts.

We assessed the correspondence between the representational dissimilarity matrices of BOLD data for each ROI with behavior using Spearman's rank correlation.