Social Scene Perception in neurotypicals, bvFTD and AI models

Laura Iris Van Hove (laura.vanhove@kuleuven.be)
Department of Brain and Cognition, Leuven Brain Institute, Faculty of Psychology & Educational Sciences KU Leuven, 3000 Leuven, Belgium,
Neuropsychiatry, Department of Neurosciences, Leuven Brain Institute, KU Leuven, 3000 Leuven, Belgium

Elahe' Yargholi (elahe.yargholi@kuleuven.be)
Neuropsychiatry, Department of Neurosciences, Leuven Brain Institute, KU Leuven, 3000 Leuven, Belgium

Laurent Mertens (laurent.mertens@kuleuven.be)
KU Leuven, De Nayer Campus, Dept. of Computer Science, J.-P. De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium,
Leuven.AI - KU Leuven Institute for AI, 3000 Leuven, Belgium

Joost Vennekens (joost.vennekens@kuleuven.be)
KU Leuven, De Nayer Campus, Dept. of Computer Science, J.-P. De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium,
Leuven.AI - KU Leuven Institute for AI, 3000 Leuven, Belgium,
Vrije Universiteit Brussel, Brussels, Belgium,
Flanders Make@KU Leuven, 3000 Leuven, Belgium, [7]Geriatric Psychiatry, University Psychiatric Hospitals KU Leuven, 3000 Leuven, Belgium

Jan Van den Stock (jan.vandenstock@kuleuven.be)
Neuropsychiatry, Department of Neurosciences, Leuven Brain Institute, KU Leuven, 3000 Leuven, Belgium,
Geriatric Psychiatry, University Psychiatric Hospitals KU Leuven, 3000 Leuven, Belgium

Hans Op de Beeck (hans.opdebeeck@kuleuven.be)
Department of Brain and Cognition, Leuven Brain Institute, Faculty of Psychology & Educational Sciences KU Leuven, 3000 Leuven, Belgium

Social Scene Perception in neurotypicals, bvFTD and AI models

Abstract

Social-cognitive skills are required to successfully navigate the social world. Behavioural variant frontotemporal dementia is a neurodegenerative disorder characterised by a decline in social cognitive capabilities. One theoretical framework proposes that this decline is due to neurodegeneration affecting the way in which social contextual information is processed. Interestingly, preliminary evidence suggests that both bvFTD patients as well as simpler artificial network models display a contextual bias when interpreting a social scene in comparison to neurotypicals and multimodal AI models. We investigated this hypothesis by designing and running a novel social cognitive paradigm in neurotypicals, bvFTD patients and AI models. Our preliminary results in humans show initial support for the hypothesis that bvFTD patients are more influenced by contextual cues than neurotypicals. Further analyses will compare these behavioural findings with AI models.

Keywords: social perception; social cognition ; frontotemporal dementia ; EMONET, GPT-4

Introduction

Humans are highly social creatures. We navigate everyday life by perceiving, interpreting and responding to socially relevant information. We hereby rely on social cognition, which is an umbrella term for the central processes involved in these endeavours. Social cognition thus enables us to read people's emotions and intentions, experience empathy and succesfully interact with others and the world at large. Social cognitive difficulties have been associated with multiple conditions and disorders and have a profound impact on an individual's quality of life (e.g. Eddy & Rickards, 2022; Marafioti et al., 2024; Yogarajah & Mula, 2019).

Social Cognitive Decline

Social cognitive deficits are a key symptom of behavioural variant frontotemporal dementia (bvFTD). bvFTD is a neurodegenerative disorder which primarily affects the social functioning and personality of an individual (Rascovsky et al., 2011). It is the most prevalent form of frontotemporal lobar degeneration, which is one of the most common causes of early onset dementia (Grossman et al., 2023). Despite significant advances in our understanding of FTD over the past 30 years (Ahmed et al., 2021), the mechanisms driving the social cognitive decline in bvFTD remain unclear. The Social Context Network Model (SCNM), however, postulates that the neurodegeneration of the fronto-temporo-insular network in bvFTD dysregulates the integration and modulation of contextual information which leads to alterations in social cognition (Baez et al., 2016; Ibañez & Manes, 2012). Empirical testing of this framework, however, has been limited.

Measuring Social Cognition

Conventional measures of social cognitive capabilities are often stripped off contextual cues, challenging the testing of the SCNM, as well as the ecological validity of these measurements. Traditionally, the most studied components of social cognition are emotion recognition and mentalising (Fittipaldi et al., 2024). Emotion recognition is commonly measured through tasks in which an individual's attention is steered towards a face or part of a face in order to assign an emotion to highly sterile images. In order to move towards a more ecologically valid measurement of social cognition in which the importance of contextual cues and the complexity of everyday life is recognised, we designed a novel social cognitive paradigm.

Social Scene Perception, a uniquely human skill?

The novel task consists of the viewing and rating of highly complex social scenes. Participants are asked to rate the valence of the entire image, the valence of the context and the valence of a person within that scene. The people in the scene's affective valence can either be congruent or incongruent with the valence of the context (e.g. crying vs laughing at a funeral).

This task was run in neurotypicals and the results demonstrated consistent and reliable ratings (Yargholi et al., In prep.). When behavioural ratings of incongruent images were compared to the output of artifical intelligence (AI) models, however, only more complex multimodel models (e.g. GPT-4) showed some resemblance to the human ratings. Simpler models, like EMONET (Kragel et al., 2019), seemed to demonstrate a contextual bias. This is of particular interest considering a contextual bias has also been observed in bvFTD patients (Kumfor et al., 2018).

Research Question and Hypotheses

Our study investigates whether the proposed contextual bias in bvFTD can explain potential differences in the evaluation of complex social images in bvFTD patients versus neurotypicals. This study further aims to relate the responses of both groups to the ratings of AI models to explore how model performance relates to group-specific human performance. We hypothesise that people with bvFTD will be more heavily influenced by contextual versus facial affective cues as compared to neurotypicals. Furthermore, we expect the valence ratings of EMONET to show higher alignment to the valence ratings of bvFTD patients, whilst we hypothesise that GPT-4 ratings will more closely align to neurotypical responses.

Method

Participants

We aim to collect data from 30 bvFTD patients as well as age-matched controls. We have currently collected behavioural ratings in 107 neurotypical students (mean age = 19.09, SD = 2.45) and 5 bvFTD patients (mean age = 70, SD = 9.06).

Materials and Procedure

The task consisted of the viewing and rating of 4 congruent and 8 incongruent images. After each image was displayed for a 5s duration, participants were shown the image again and asked to rate the valence of the entire image and the valence of the context. In the final round, participants were shown the same images but with one person outlined and were asked to rate the valence of this person. The valence of the person highlighted in the incongruent conditions was consistently opposite to the valence of the context. Ratings were made using a 5-point Likert scale from very negative (-2) to very positive (+2).

Figure 1: An example of an incongruent image from our dataset. The left image enables the rating of the valence of the entire image and the valence of the context, whilst the right image is used to rate the valence of the outlined person in the picture.

Data Analysis

To test our main hypothesis regarding a contextual bias in our patient population, we examined the difference scores between the three valence ratings. In particular, we calculated the absolute difference between image valence – context valence and people valence – context valence. We used a mixed effect model to account for the dependency in the data (multiple ratings per participant) and the unequal sample sizes. Considering our dependent variable is bounded and discrete (from none to 4 point differences), we applied an ordinal logistic mixed effect model. We hereby used the CLMM function from the ordinal package in R. Considering our model required enough observations of each of the point differences to produce stable estimates, which our small patient sample did not meet, we collapsed our dependent variable into 3 categories (0, 1 and >1 absolute difference in ratings). Correlation analyses with the AI models are yet to be run but will be presented at the conference.

Preliminary Results

Summary statistics suggested that the absolute difference in valence ratings of the entire image – the valence rating of the context would only significantly differ when images were incongruent (Figure 2).

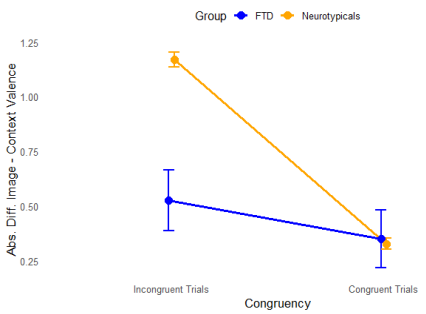

Figure 2: Interaction plot displaying the mean absolute difference between image valence and context valence for each group. Error bars represent +- 1 SE.

This group × congruency interaction was quantified by running a likelihood ratio test which confirmed a significant interaction (LRT: $\chi^2 = 8.33$, p = 0.004).

We then ran an ordinal logistic mixed effect model separately on the congruent and incongruent images. Each model consisted of the collapsed absolute difference ratings as the DV, group as fixed effect and random intercepts for both participant and image. For congruent images, no significant group effect in the difference scores were found ($p = 0.95$), however, for incongruent images, a significant effect of group emerged (OR = 6.64, 95% CI: [2.19–20.1], $p < 0.001$). These results imply that it is 6.64 times more likely that FTD patients scored the valence of the image more similarly to the valence of the context than neurotypicals.

We ran the same analyses on the absolute difference in the valence of the depicted person – the valence of the context. A significant group x congruency interaction in these scores emerged as well (LRT: $\chi^2 = 4.81$, p = 0.028). We again found no significant group effect when running our mixed model on the congruent condition only ($p = 0.72$). In the incongruent condition, the data yielded a significant effect of group (OR = 4.49, 95% CI: [1.61–12.53], $p = 0.004$). These results suggest that it is thus 4.49 x more likely that our patient sample rated the people and context valence more similarly in the incongruent condition than neurotypicals.

Discussion

Our preliminary results suggest that neurotypicals and bvFTD patients rate the valence of incongruent images in a distinct manner. On closer inspection, we found that bvFTD patients displayed a context bias when rating the valence of both the entire image as well as the people valence. Our initial results hereby provide support for our first hypothesis and are thus in line with the results of Kumfor et al. (2018) as well as the social context network model. However, considering the small patient sample size, as well as the fact that our groups were not age-matched (see limitations), more data is needed to reliably confirm the effects.

Next steps and Current Limitations

In order to test our second hypothesis concerning how the valence ratings of AI models compare to our groups of interest, we will compare image valence ratings from EMONET as well as the image, scene and people valence from GPT-4 with the valence ratings of both human participant groups. The valence ratings will be extracted from EMONET, which predicts image valence from its last fully connected layer, whilst the participants' task instructions will be used as prompts for GPT-4.

Data collection will furthermore continue. One major limitation of our current data is that our groups are not age-matched and that any effect we currently see is confounded by age. We therefore plan to also include a second neurotypical group age-matched to our patient sample.

References

Ahmed, R. M., Hodges, J. R., & Piguet, O. (2021). Behavioural Variant Frontotemporal Dementia: Recent Advances in the Diagnosis and Understanding of the Disorder. In B. Ghetti, E. Buratti, B. Boeve, & R. Rademakers (Eds.), *Frontotemporal Dementias* (Vol. 1281, pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-51140-1_1

Baez, S., García, A. M., & Ibanez, A. (2016). The Social Context Network Model in Psychiatric and Neurological Diseases. In M. Wöhr & S. Krach (Eds.), *Social Behavior from Rodents to Humans* (Vol. 30, pp. 379–396). Springer International Publishing. https://doi.org/10.1007/7854_2016_443

Eddy, C. M., & Rickards, H. (2022). Social cognition and quality of life in Huntington's disease. *Frontiers in Psychiatry*, *13*, 963457. https://doi.org/10.3389/fpsyt.2022.963457

Fittipaldi, S., Legaz, A., Maito, M., Hernandez, H., Altschuler, F., Canziani, V., Moguilner, S., Gillan, C. M., Castillo, J., Lillo, P., Custodio, N., Avila-Funes, J. A., Cardona, J. F., Slachevsky, A., Henriquez, F., Fraile-Vazquez, M., Cruz de Souza, L., Borroni, B., Hornberger, M., … Ibanez, A. (2024). Heterogeneous factors influence social cognition across diverse settings in brain health and age-related diseases. *Nature Mental Health*, *2*(1), 63–75. https://doi.org/10.1038/s44220-023-00164-3

Grossman, M., Seeley, W. W., Boxer, A. L., Hillis, A. E., Knopman, D. S., Ljubenov, P. A., Miller, B., Piguet, O., Rademakers, R., Whitwell, J. L., Zetterberg, H., & Van Swieten, J. C. (2023). Frontotemporal lobar degeneration. *Nature Reviews Disease Primers*, *9*(1), 40. https://doi.org/10.1038/s41572-023-00447-0

Ibañez, A., & Manes, F. (2012). Contextual social cognition and the behavioral variant of frontotemporal dementia. *Neurology*, *78*(17), 1354–1362. https://doi.org/10.1212/WNL.0b013e3182518375

Kragel, P. A., Reddan, M. C., LaBar, K. S., & Wager, T. D. (2019). Emotion schemas are embedded in the human visual system. *SCIENCE ADVANCES*.

Kumfor, F., Ibañez, A., Hutchings, R., Hazelton, J. L., Hodges, J. R., & Piguet, O. (2018). Beyond the face: How context modulates emotion processing in frontotemporal dementia subtypes. *Brain*, *141*(4), 1172–1185. https://doi.org/10.1093/brain/awy002

Marafioti, G., Cardile, D., Culicetto, L., Quartarone, A., & Lo Buono, V. (2024). The Impact of Social Cognition Deficits on Quality of Life in Multiple Sclerosis: A Scoping Review. *Brain Sciences*, *14*(7), 691. https://doi.org/10.3390/brainsci14070691

Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., van Swieten, J. C., Seelaar, H., Dopper, E. G. P., Onyike, C. U., Hillis, A. E., Josephs, K. A., Boeve, B. F., Kertesz, A., Seeley, W. W., Rankin, K. P., Johnson, J. K., Gorno-Tempini, M.-L., Rosen, H., … Miller, B. L. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, *134*(9), 2456–2477. https://doi.org/10.1093/brain/awr179

Yargholi, E., Mertens, L., Van Hove, L. I., Gunay, E. M., Vennekens, J., Van den Stock, J., & Op de Beeck, H. (In prep.). *Distributed high-level processing supports the emotional assessment of complex social scenes, resulting in alignment between the human brain, behavior, and AI models*.

Yogarajah, M., & Mula, M. (2019). Social cognition, psychiatric comorbidities, and quality of life in adults with epilepsy. *Epilepsy & Behavior*, *100*, 106321. https://doi.org/10.1016/j.yebeh.2019.05.017