Detecting bluffing in a two-player game with passive brain-computer interfaces: implications for human-machine interaction

Diana E. Gherman*[1], Laurens R. Krol[1], Thorsten O. Zander[1]

[1]*Brandenburg University of Technology Cottbus–Senftenberg, Germany*

*P.O. Box 101344, Postfach 101344, 03013 Cottbus. E-mail: diana.gherman@b-tu.de

Introduction: Machines are becoming a ubiquitous presence in human day-to-day life and autonomous systems are increasingly making decisions for us. Nevertheless, our interaction with these systems feels incomplete as the implicit and non-verbal cues that are crucial in human communication are overlooked [1]. A solution to this problem could be brought by passive brain-computer interfaces (pBCIs), which have proven valuable in decoding cognitive and affective states from brain activity [2]. With this study, we show that pBCIs are able to distinguish truths from bluffs more accurately than human participants. Our findings provide deeper understanding of the significance and potential contributions of pBCIs in contexts that involve social interaction.

Material, Methods and Results: The study included 6 pairs of participants that played 8 rounds of a dice game that involves both bluffing and truth-telling, illustrated in *Figure 1*.

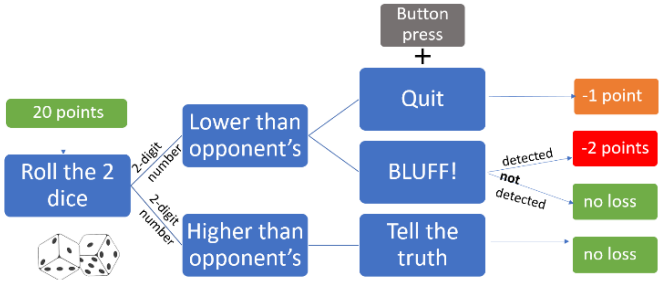

Figure 1.Scheme of the bluffing game's rules

All players' brain activity was recorded with 128 EEG electrodes. Final game points were converted into monetary rewards. Hence, players were motivated to bluff and detect the opponents' bluffs. To contrast bluffs versus truths, we extracted samples of data time-locked to the button presses that preceded each player's announcement of either the true number on their dice, or their bluffs. We excluded quitting trials. The pre-processing and feature extraction method followed a windowed means approach [3] and a bandpass filter (0.1 – 8 Hz). For the training and testing classification, we used a regularized linear discriminant analysis (rLDA) and a 5x5 cross-validation method. The pBCI system managed to distinguish truths from bluffs with an accuracy of up to 76%, significantly higher than the overall human opponents' accuracy of 66%.

Discussion: Our classifier accurately distinguished bluffs from truths, showing ability to detect complex mental states from EEG that surpasses a human's ability to do the same based on facial, social, and contextual cues. An analysis into the neural sources that contributed to classification indicates that cortical sources including the anterior cingulate cortex (ACC) play a key role in the deception mechanism, in line with neuroscientific studies on the subject [5].

Significance: Despite the lack of access to hidden states, recent advancements show AIs can beat humans at poker [5], demonstrating that some AIs can handle real-world, social scenarios. But what if AI systems would have access to such hidden information? Would a caregiver robot understand its elder patient's needs more and empathize better? Or would an AI algorithm be able to contribute to difficult negotiation sessions? Although we cannot answer these questions, this study lays the foundation for understanding how to improve human-machine relationships with pBCI.

References:

[1] Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005, August). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In 2005 IEEE/RSJ international conference on intelligent robots and systems (pp. 708-713). IEEE.
[2] Zander, T. O., & Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. Journal of neural engineering, 8(2), 025005.
[3] Blankertz, B., Schäfer, C., Dornhege, G., & Curio, G. (2002, August). Single trial detection of EEG error potentials: A tool for increasing BCI transmission rates. In International Conference on Artificial Neural Networks (pp. 1137-1143). Springer, Berlin, Heidelberg.
[4] Brown, N., & Sandholm, T. (2019). Superhuman AI for multiplayer poker. Science, 365(6456), 885-890.
[5] Lisofsky, N., Kazzer, P., Heekeren, H. R., & Prehn, K. (2014). Investigating socio-cognitive processes in deception: a quantitative meta-analysis of neuroimaging studies. Neuropsychologia, 61, 113-122.