Measuring physiological and behavioural responses to potential negative consequences inflicted on anthropomorphized agents.

Pietro Cau, Marthe Post, Anna Maschek

May 8, 2024

Abstract

Does anthropomorphization of a non-human agent affect people's stress levels and willingness to take actions that risk damaging that non-human agent?

We present a method of studying the effect of potential negative consequences inflicted on anthropomorphized agents on people's behaviour and psychological state.

We propose the use of a modified game of Jenga and an anthropomorphized egg in danger of falling to induce the desired state, and measure behavioural responses (task performance) as well as physiological signals (ECG and EDA) as means of making inferences on psychological effects and ANS activation.

We then share the results of a pilot study performed on a small number of subjects in order to validate our method, and the data collection and analysis pipeline.

Background

It has long been studied and theoretically proven that humans have an innate tendency to attribute human characteristics to nonhuman objects. This is referred to as anthropomorphism Epley et al. (2007), which has been prevalent across cultures and ages and it extends beyond just entertainment; it significantly influences the way people interact with and perceive various entities. In a study by Vaes et al. (2016), they researched whether attributing minimal humanity cues (such as giving a name) could induce a sense of empathy towards non-human entities which were stimulated painfully. The researchers found that by only giving a name to such an entity, this already induced an empathic response when the anthropomorphized agent was painfully stimulated. Multiple other studies have also shown that by attributing humanlike features to nonhuman objects, empathy for those objects increased (Riek et al., 2009; Waytz et al., 2010; Epley et al., 2007; Vaes et al., 2016; Mattiassi et al., 2021). More specifically, Waytz et al. (2010) argue that when people perceive an agent to be human, the agent will be acted towards with moral care and consideration.

Feeling empathy towards other human beings is a natural phenomenon. Mattiassi et al. (2021) initially referred to it as the capacity of humans to project themselves into objects and feel their subjective quality, in other words, the ability to understand others. One aspect related to the ability to feel empathy is related to others' being in pain. People generally feel uneasy when they have to watch other people getting hurt or being in pain. Furthermore, the study by Vaes et al. (2016) tested how the ability to empathise was influenced by other factors related to anthropomorphism, including the ability to perceive physical pain in nonhuman agents. They showed that painfully stimulating nonhuman agents which were attributed humanity cues were perceived differently compared to when they were not attributed humanity cues.

Humans are generally averse to situations in which they could potentially cause other human beings to get hurt Crockett et al. (2014). In the case of having this as the consequence of failing a specific task, the participant is therefore expected to take the approach that minimises the outcome of the human getting hurt as a consequence of them failing the task. This would in turn increase the pressure on the performance on that task. McCoy et al. (2013) showed that increased pressure to perform well on a task induces more stress on the participant while engaged in the task, noticeable by, for example, increased heart rate. However, to the best of our knowledge, it has not yet been studied how feeling empathy towards anthropomorphized agents influences the performance and behaviour of people in a task where the consequences involve the anthropomorphized agent getting painfully stimulated, as opposed to a real human being.

This research could give a first insight into how minimal humanity cues affect people's attitude and behaviour towards non-human agents, these insights can lead to considerations useful when designing anthropomorphized agents. Potential applications range from marketing to social robots and beyond.

Research Question

We selected the following research question as the focus of this study: Does anthropomorphization of a non-human agent affect people's stress levels and willingness to take actions that risk damaging that non-human agent?

Experiment Design

The experiment is designed to be carried out as a between subjects study. Participants will be divided into two groups; a control group and a manipulation group. In each condition, two rounds of a modified game of Jenga will be played by the participant: a baseline session and an experimental session.

In the experimental session, for the manipulation group, an egg will be placed on top of a Jenga tower, featuring certain humanity cues (i.e. giving it a name, having a face drawn on it) to enhance the anthropomorphization effect. In the control group, an egg will also be placed on top of the tower but will have none of these humanity cues. Also, the groups will be provided with different prompts before the start of the experimental session. The control group will be told that: "An egg has been placed on top of the tower. Do not make it fall" While the manipulation group will be told that: "Bob has climbed to the top of the tower, he is afraid of heights. Do not make him fall"

At the beginning of the experiment, for both groups, the experimental session will be preceded by a baseline session, where we will get baseline measurements for ECG and EDA. This will be done by connecting the necessary sensors from a BitAlino to the participant (chest for ECG, non-dominant hand for EDA).

In the baseline session participants in both groups will play a game of Jenga with modified rules. They will be instructed to remove blocks, without stacking them on top of the tower, until they think no block can be safely removed without the tower collapsing, at which point they can indicate their desire to stop. They will be instructed not to remove blocks from the top layer. Also, in order to incentivize better performance and "risk-taking" participants will be told that amongst all the participants, the one that will remove the most blocks will win a small monetary prize. We will also record the time it takes the participant to complete the baseline task, and the number of blocks removed.

Then, in the experimental session, both groups will be asked to play another round of Jenga with the same modified rules, and will again be told that there is a small monetary prize to be won for the participant with the best performance, this time however, an egg will be placed upright on top of the tower.

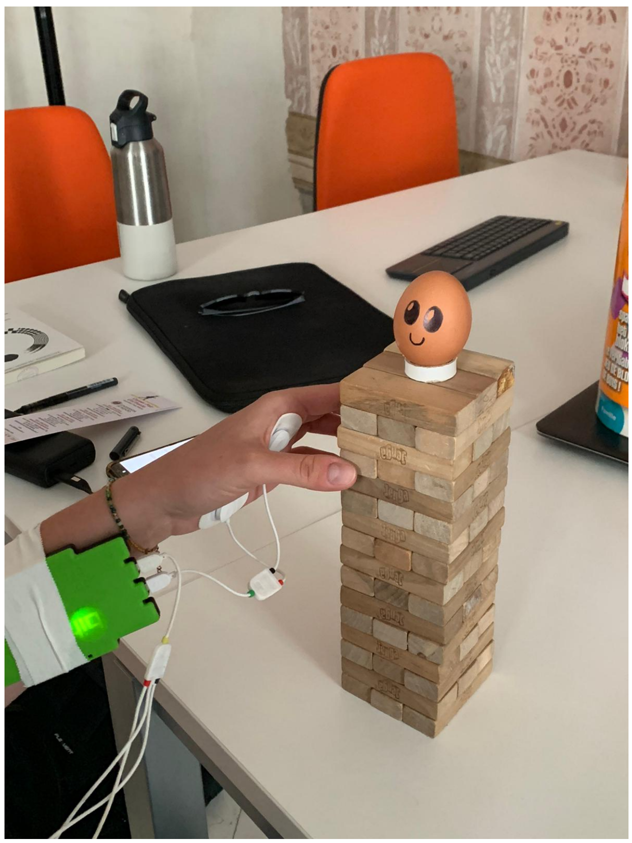

Figure 1: Experiment setup

During the experimental session, we will again record the heart rate of the participants using ECG, as well as electrodermal activity using EDA. Furthermore, the time the round took, the amount of blocks removed from the Jenga tower, as well as whether the tower fell or not in the end will be noted.
At the end of the experiment, for each participant we will have the following data:

- Physiological Signals:

  - Baseline (simple Jenga) HRV
  - Baseline phasic EDA
  - EggJenga HRV
  - EggJenga phasic EDA

- Performance Measures:
  - Baseline time on task
  - Baseline number of blocks removed
  - EggJenga time on task
  - EggJenga number of blocks removed

For the data analysis, we will compare the delta between baseline session measurements and experimental session measurements for both groups, in order to assess if there is a significant difference between them.

To analyse the collected physiological signals, we will use mean IBI, Inter-Beat Interval, but other HRV variables such as the RMSSD, SDNN, and LF/HF, which are known to be good proxies for measuring stress (Peabody et al., 2023; Karthikeyan et al., 2013) could also be used. Analysis of the phasic component of EDA signal is also known to be a reliable indicator of psychological stress Giannakakis et al. (2019).

Hypothesis

We expect potential negative consequences for an anthropomorphized agent to affect people's stress levels and willingness to take risks.

It is expected that there will be a greater increase in stress level (detected from HRV and the phasic component of the EDA signal) going from the baseline to the experimental session for the participants in the manipulation group (anthropomorphized egg) compared to subjects in the control group.
Furthermore, it is hypothesised that people in the condition with the anthropomorphized egg, will take fewer risks compared to the controls, resulting in a lower number of blocks removed before stopping, and that they will put more care into removing blocks, resulting in a longer time (on average) for removing each block.

Some of the existing literature supporting this hypothesis includes the research from Vaes et al. (2016), the research from Mattiassi et al. (2021), and the research from Waytz et al. (2010), all these studies indicate the human predisposition to empathise with anthropomorphized agents, suggesting that in a situation in which the subject's action could cause harm to an anthropomorphized agent, the subject will be more careful and risk-averse.

Pilot Study: Data Analysis and Results

We performed a pilot study with 5 subjects to test the concept of the experiment, and the data collection and analysis pipeline.

Data Analysis

During the pilot study we were unfortunately unable to collect a usable EDA, so we had to limit the data analysis to ECG data and performance measures.

To analyse the ECG measurements, we looked at the difference in mean IBI (Inter-Beat Interval) between the baseline condition and the experimental condition for each participant, and then compared the values of the participants in the different groups (control/manipulation). We did the same thing for the difference in average "time per block" between the baseline and experimental condition of both groups.

Results

It has to be noted that the following results are not to be considered as valid, as they are only an observation of the data collected from a pilot study with a very limited number of subjects.

Difference in mean IBI between baseline and experimental session for control versus manipulation group

The graph below shows the results of the data analysis on the difference in heart rate between the baseline session and the experimental session for the control group and the manipulation group. It can be observed that the manipulation group (the group which had the anthropomorphized egg on the tower), on average, had a bigger range in their heart rate for the experimental session compared to the baseline session. This range between the two sessions was much smaller for the control group (the group who has a normal egg on top of the tower). In other words, in the manipulation group, there were some participants that had an increased heart rate on the experimental task compared to the baseline, but there were also participants who had a slower heart rate on the experimental session compared to the baseline. Taken altogether the results are, as expected, inconclusive, since the p-value does not reach statistical significance (p=0.64). No conclusions can be drawn from these results.

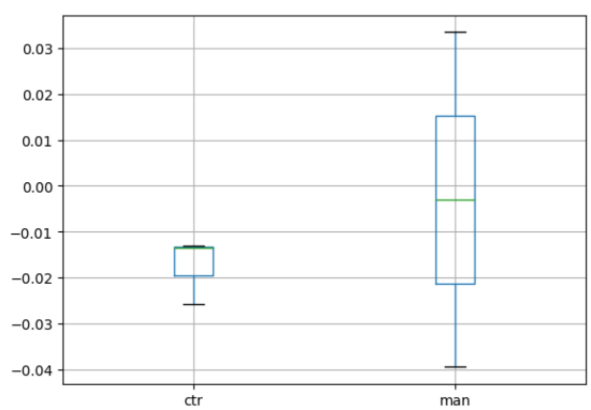

Figure 2: Difference in mean IBI between sessions

t-test result: (statistic=-0.521, pvalue=0.638, df=3.0)

Difference in time spent per block between baseline and experimental session for control versus manipulation group

Secondly, we analysed the average time taken to remove each block, calculated as the time on task divided by the number of blocks removed. The time spent on each block for the experimental session was compared with that of the baseline session, between participants in the control group and participants in the manipulation group. The graph below shows the result of this analysis, where the values were calculated according to

$$\frac{t_b}{n_b} - \frac{t_e}{n_e}$$

Where t_b, t_e are the time on task for the baseline and experimental sessions, and n_b, n_e are the number of blocks removed for the baseline and experimental sessions.

It can be seen that most participants reflect a positive value, meaning they were faster in the experimental condition compared to the baseline session. Furthermore, the controls reflect some higher values than the manipulation group, meaning that the controls had a bigger difference in average time spent per block between the two sessions, and that this difference was smaller for the manipulation group. Again, as expected, the p-value does not reflect statistical significance with a value of p=0.72.

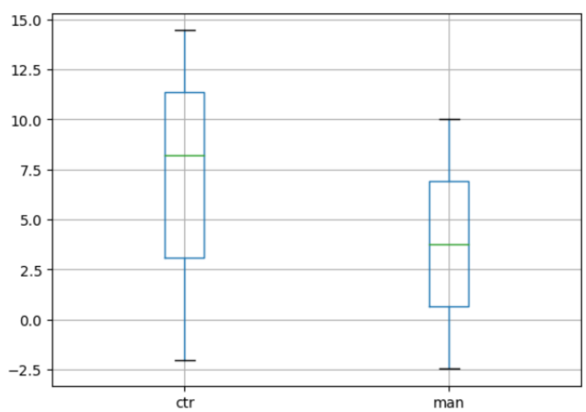

Figure 3: Difference in time per block between sessions

t-test result(statistic=0.399, pvalue=0.716, df=3.0)

Difference in number of blocks removes in experimental session compared to baseline session for control versus manipulation group

Lastly, we looked into the total number of blocks removed from the tower in the experimental session compared to the baseline session for both the control group and the manipulation group. The graph below shows the result. It can be seen that the participants in the manipulation group, on average, took out more blocks in the experimental session compared to the baseline session. The controls, on the other hand, reflect a value of 0 or a negative value, implying that they either took out the exact same number of blocks or less in the experimental session compared to the baseline session, thus the opposite of what we expected to see. The test did not reach statistical significance with a p-value of p=0.10.

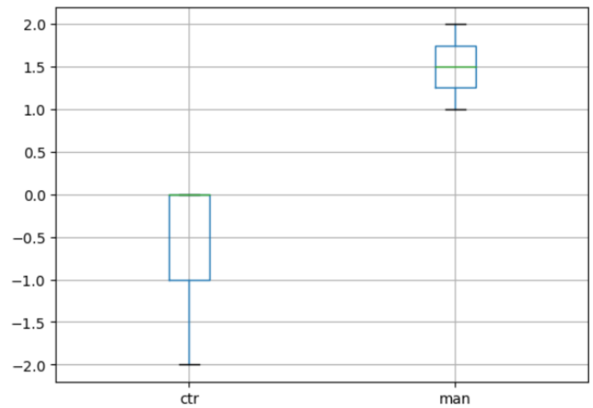

Figure 4: Difference in number or blocks removed between sessions

t-test result(statistic=-2.310, pvalue=0.104, df=3.0)

Conclusion

We believe the proposed method to be effective for studying the effects of anthropomorphization on people's behavior and psychological stress. More work is needed to fully establish its validity and limitations, and a wider study needs to be conducted in order to be able to reach statistically significant conclusions that could accept or refuse our hypothesis.

References

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 111(48), 17320–17325. https://doi.org/10.1073/pnas.1408988111

Epley, N., Waytz, A., Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. https://doi.org/10.1037/0033-295x.114.4.864

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., Tsiknakis, M. (2019). *Review on psychological stress detection using biosignals. IEEE transactions on affective computing*, 13(1), 440-460.

Karthikeyan, P., Murugappan, M., Yaacob, S. (2013). Detection of Human Stress Using Short-term ECG and HRV Signals. *Journal of Mechanics in Medicine and Biology*, 13(02), 1350038. https://doi.org/:10.1142/S0219519413500383

Mattiassi, A. D. A., Sarrica, M., Cavallo, F., Fortunati, L. (2021). What do humans feel with mistreated humans, animals, robots, and objects? Exploring the role of cognitive empathy. *Motivation And Emotion*, 45(4), 543–555. https://doi.org/10.1007/s11031-021-09886-2

McCoy, S. K., Hutchinson, S. B., Hawthorne, L., Cosley, B. J., Ell, S. W. (2013). Is pressure stressful? The impact of pressure on the stress response and category learning. *Cognitive, Affective*

Behavioral Neuroscience, 14(2), 769–781. https://doi.org/10.3758/s13415-013-0215-

Peabody JE, Ryznar R, Ziesmann MT, Gillman L. A Systematic Review of Heart Rate Variability as a Measure of Stress in Medical Professionals. *Cureus.* 2023

Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., Robinson, P. (2009, March). How anthropomorphism affects empathy toward robots. *In Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 245-246).

Vaes, J., Meconi, F., Sessa, P., Olechowski, M. (2016). Minimal humanity cues induce neural empathic reactions towards non-human entities. *Neuropsychologia*, 89, 132–140. https://doi.org/10.1016/j.neuropsych

Waytz, A., Epley, N., Cacioppo, J. T. (2010). Social cognition unbound. *Current Directions in Psychological Science*, 19(1), 58–62. https://doi.org/10.1177/0963721409359302

Waytz, A., Cacioppo, J., Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219-232.