Student Conference Proceedings
Vol 1, No 1, 2025, Article ID 2004, 4 Pages

Student Conference Lübeck

Proceedings Article

Evoking Promethean Shame – Exploring Emotional Reaction to AI Systems

Laura Duwe[a*] · Christiane Attig[b] · Tim Schrills[b] · Thomas Franke[b*]

[a]Student of Psychology – Cognitive Systems, Universität zu Lübeck, Lübeck, Germany
[b]Institute for Multimedia and Interactive Systems, Universität zu Lübeck, Lübeck, Germany
*Corresponding author, email: laura.duwe@student.uni-luebeck.de; thomas.franke@uni-luebeck.de

Received 05 February 2025; Accepted 03 June 2025; Published online 16 July 2025

Abstract

The present study aims to investigate the concept of Promethean Shame in the context of artificial intelligence (AI) usage and cooperative work with AI. Promethean Shame, termed by philosopher Günther Anders, refers to the experience of being confronted with human-made technology that outperforms human abilities. Understanding whether the concept influences perception towards AI is a crucial factor in creating effective human-AI collaboration. To assess participants' experience of Promethean Shame and analyze if the experience is influenced by AI use, a pre-post comparison study was conducted. Participants were asked to complete a text-correction task with the help of a Large Language Model-based chatbot. A questionnaire was used to assess multiple interaction factors before and after the task. Results indicate no significant change in Promethean Shame before and after the interaction as well as weak to moderate correlations to other concepts like trust and perceived intelligence.

I. Introduction

With the rapid development and exponential growing user base of Large Language Model (LLM)-based chatbots [1] such as ChatGPT, it becomes increasingly important for human-centered AI to better understand how users react to the AI systems when interacting with them. As AI technology is relatively new, there are large research gaps regarding the nature of human-AI interaction. In the field of engineering psychology, understanding the effects of technology interaction on a person's perception is essential; however, technology-related emotions such as shame have so far been understudied in human-AI interaction research. Understanding them could improve human-centered design and user satisfaction. LLM-based chatbots such as ChatGPT, a GPT-4o based chatbot, are increasingly integrated into both leisure and work environments due to their ability to perform a wide range of tasks efficiently. Their extensive knowledge and processing speed [1] can lead to situations where a chatbot might outperform the user in completing tasks or providing information. For example, a chatbot might generate a detailed project schedule in seconds, which would take the user hours. In the workspace, shame can have multiple possible triggers and various factors, including loss of power, loss of autonomy in goal setting or reaching a set goal and failure to realize opportunities for advancement [2].

Promethean Shame (PS), termed by Günther Anders [3], is characterized as the shame felt when human-made technology is superior to humans. The concept was developed before AI, as we know and use it today, existed. Therefore, the objective of this study was to determine whether the concept of PS has psychological relevance in the context of human-technology interaction, specifically AI.

To this end, we conducted a study to find out to what extent PS can be induced and assessed when interacting with technology that is portrayed as superior to the user. We developed a study where participants interacted with

an AI chatbot to solve a text-correction task. The interaction was designed with prompts to the chatbot and feedback mentioning the superiority of the AI to increase the potential of experiencing PS [2]. PS was measured before and after the interaction. Furthermore, we wanted to investigate exploratorily which subjective aspects of the interaction may influence the experience of PS. The pre-post comparison of PS, as well as attitudes toward and experience with technology, was chosen to show the extent to which the superiority of LLMs affected cooperation with humans. The findings could provide valuable guidance for programming chatbots to enhance the quality and comfort of human-technology interactions.

II. Method and Materials

II.I. Participants

To calculate the sample size, we conducted an a priori power analysis. For this, we used effect sizes ($d = 0.35$) of similar studies on shame. Results indicated the required sample size for a power of 0.80 being $N = 51$. In this study, $N = 49$ participants were included in the data analysis of the original 60 who completed the study. Seven participants had to be removed due to schematic answers or having an excessively short completion time, 4 had to be excluded due to the chatbot not working as intended. The study was conducted online via Prolific and participants received a compensation of 5,00€ for study completion. Age ranged between 21 and 63 years ($M = 34.7$, $Mdn = 32$). Participants had a relatively high level of prior LLM experience with 53% agreeing positively with the two items regarding use and cooperation experience (largely or completely agree). Twenty-five participants were male, 22 female, one non-binary and one participant preferred not to indicate their gender. Participants' educational level was relatively high, with 37 having a Bachelor's or Master's degree.

II.II. Measures

All measures assessed participants' responses on a 6-point Likert scale from *completely disagree* to *completely agree* unless otherwise stated. Cronbach's alpha was calculated for all scales to assess internal consistency as a reliability indicator. In the pre-task questionnaire, we assessed PS with the 10-item Promethean Shame Questionnaire [4]. Reliability was good ($\alpha = .88$). Technology optimism was assessed with the 9-items Technology Optimism Scale [5] measuring responses on a 5-point Likert scale from *strongly disagree* to *strongly agree*. Reliability was acceptable ($\alpha = .88$). Affinity with technology interaction (ATI) was measured with the 9-item ATI scale [6]. Reliability was acceptable ($\alpha = .78$).

In the post-task questionnaire, we assessed PS with the same questionnaire as in the pre-task survey. Internal consistency was excellent ($\alpha = .93$). Usefulness of the chatbot [7] was measured with 4 items. Reliability was acceptable ($\alpha = .78$). Perceived cooperativity was measured with the 15-item Perceived Cooperativity Scale (PCS [8]). Reliability was excellent ($\alpha = .92$). Perceived teaming was measured with the 9-item Teaming Perception Scale (TPS [8]). Internal consistency was good ($\alpha = .87$). Trust towards the chatbot was measured with 16 items from the human-computer trust questionnaire (HCT [9]) that contained the subscales reliability, understandability (i.e., cognitive trust), as well as faith and attachment (i.e., affective trust). Reliability of the full scale was excellent ($\alpha = .93$). Anthropomorphism was measured with 14 items of the Godspeed Questionnaire [10] that uses a 5-point semantic differential scale. We only used the subscales anthropomorphism, likeability, and perceived intelligence. Internal consistency for the full scale was excellent ($\alpha = .93$). Lastly, basic psychological need satisfaction was measured with the 12-item BPN-TU [11] on a 5-point Likert scale from *not at all* to *very much*. Reliability was excellent ($\alpha = .90$).

II.III. Design and Procedure

The study had a within-subjects design. After obtaining informed consent and presenting the study description, participants were asked to respond to a pre-task questionnaire before they received a vignette. They were asked to imagine they worked in a magazine editorial and were tasked with fact-checking encyclopedic articles with the help of an LLM-based chatbot.

For the experimental environment, the chatbot was created with Gradio, a Python library. The chatbot was created with the LLM GPT-3.5-turbo and system-prompted to help the user find the specified errors in the text, correct them, and give factual information on the corrected words. The chatbot's instruction defined the errors beforehand to ensure a consistent and correct output and avoid errors in the output. It was implemented in LimeSurvey with a user interface to be directly interacted with.

As material for the text correction task, articles of three different online encyclopedias (newworldencyclopedia.org, britannica.com, encyclopedia.com) were summarized by ChatGPT-4o and manually modified to contain three factual errors. They were solvable through the context of the article or general knowledge. The short encyclopedic articles (word count: M= 379; Min= 351; Max= 417) ranged over six different topics which were inspired by the trivia game trivial pursuit [12], to cover a wide range of general knowledge: geography (glacier and fjord), entertainment (cartoon and rock music), history (Berlin wall and French revolution), art and literature (science fiction and Salvador Dalí), science and nature (deer and Space Shuttle), and sport and leisure (surfing and hiking).

10.18416/SCP.2025.2004

Participants were asked to choose one out of the six topics. They were then tasked with correcting two short encyclopedic articles within their chosen topic with the help from the chatbot. They were instructed to start the interaction by typing the prompt “Please find all factual errors in the text” as well as to provide the article text to the chatbot. After correcting the two articles, they received standardized feedback within LimeSurvey informing them that the chatbot would have worked faster and more reliable if it had worked on its own. Afterwards, participants filled out the post-task questionnaire and were able to provide feedback on both the chatbot functionality and the study. Finally, participants received a debriefing on the means of the study and the reason behind the feedback, explaining that it in no way reflected their individual performance and instead was scripted to analyze the influence of the interaction and feedback on PS.

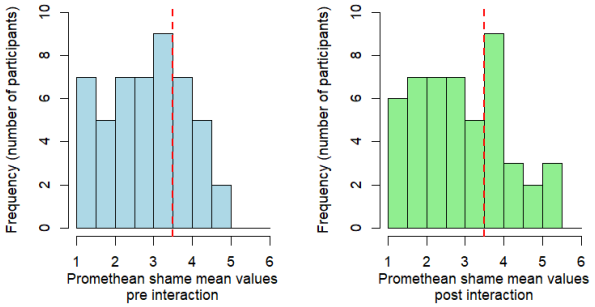

Figure 1: Sample distribution of Promethean Shame mean values pre- and post-interaction. The red dashed line represents the scale midpoint (3.5).

III. Results

The statistical analysis was performed with R and RStudio (posit.co). First, descriptive statistics showed that the sample distribution mean values differed slightly between the pre- ($M = 2.90$; $SD = 1.03$) and post- ($M = 2.96$; $SD = 1.16$) measurements for PS (see Fig. 1). However, using a paired t-test ($t_{48} = 0.40$; $p < .694$), no significant effect ($d = .057$) could be found.

Second, to shed further light on the stability of PS scores, a Pearson correlation analysis was conducted, revealing a high correlation between the pre- and post-interaction scores ($r = .67$; $p < .001$). We also tested differences on item level, which revealed a significant increase for Item 1 ($M_{pre} = 2.94$; $M_{post} = 3.43$, $t_{48} = 2.50$, $p = .016$) and a significant decrease for Item 9 ($M_{pre} = 3.63$; $M_{post} = 3.16$, $t_{48} = 2.45$, $p = .018$).

Third, to explore correlations between PS and user experience factors, Spearman correlation analyses were conducted and visualized (see Table 1). Spearman's correlation was chosen due to non-normal variable distributions. Significant, moderate positive correlations were found between pre-interaction PS and affective trust (i.e., subscales faith and attachment) as well as perceived intelligence. Weak but insignificant positive correlations were found for PCS, TPS, anthropomorphism, likeability, and basic psychological needs satisfaction. All correlations between post-interaction PS and other study variables were insignificant. Weak but insignificant positive correlations were found for affective trust, all anthropomorphism subscales, and basic psychological needs satisfaction. A weak, insignificant negative correlation was found for usefulness.

Table 1: Correlation coefficients of Promethean Shame pre- and post-interaction with the examined variables and subscales.

Variable	Pre-Interaction		Post-Interaction	
	r	p	r	p
Trust	.31	.029	.16	.271
Reliability	.08	.579	-.02	.907
Understandab.	.22	.134	.11	.444
Faith	.31	.029	.19	.203
Attachment	.34	.018	.17	.229
PCS	.23	.117	.04	.790
TPS	.16	.278	.03	.863
Godspeed	.27	.061	.15	.317
Anthropom.	.24	.095	.24	.093
Likability	.12	.410	.10	.632
Perc. Intell.	.35	.015	.14	.348
Usefulness	-.02	.909	-.15	.301
BPN-TU	.23	.106	.19	.198
Technology optimism	.01	.920	-.02	.879
ATI	-.09	.537	-.04	.788

IV. Discussion

In the present study we examined the concept of Promethean Shame (PS) within a text-correction task assisted by an LLM-based chatbot. The analysis revealed no significant increase in experienced PS after the interaction. On item level, significant differences were found for two items: While the mean score for Item 1 ("I feel less capable than the technical systems around me") increased after the interaction, the mean score for Item 9 ("I constantly try to prove myself in areas where I feel less competent than technical systems") decreased. Participants may feel more comfortable with their own ability to use the chatbot correctly. Yet their need to prove themselves in areas they feel less competent in decreased after

interacting with the chatbot. This suggests that the effects, assessed using the PS questionnaire, may be more multifaceted or possibly obscured by the variation in change across different items. The interaction may target specific dimensions of self-perception and behavior rather than influencing the overall assessment. At the same time, it suggests that the observed changes may not have been strong enough to be consistently noticeable across all dimensions. The low correlations between PS and variables like trust, basic psychological needs satisfaction, and anthropomorphism indicate possible influences that warrant further exploration. The concept of PS in the context of AI usage and psychological research remains largely unexplored although it should be considered systematically to better understand human-AI interaction. While the concept of PS is based in philosophy, it handles related questions as psychological research and should therefore be implied more in scientific studies. The questionnaires used in this study showed very high reliability.

The system prompts determining the chatbot's responses had been specifically created for this study and not been tested in prior studies. Therefore, the chatbot's responses and predetermined answer format were rigid, with little room for spontaneous interaction. This could have affected participants' reaction to the chatbot that might not reflect actual use of LLM-based chatbots. Future research should use a between-subject design to comparatively assess the effects. Comparing negative feedback with no feedback can isolate the impact on PS, as higher shame levels in the negative feedback group would clearly separate the response to the feedback and the other factors. Furthermore, the existence of possible effects of interpersonal characteristics on the perception of PS should be further investigated. It would be important to know if there are effects that cause people to be more susceptible to feel PS than others (in our study, ATI had no effect). For example, personality traits like neuroticism might be linked to higher PS due to the heightened shame experience. In a follow-up study, a better instructed and less rigid chatbot should be used, capable of answering more questions and fact-checking the users own inputs.

V. Conclusion

This pilot study explored the measurability and relevance of the construct Promethean Shame to enhance the understanding of human-AI interaction. Results showed no significant changes after the interaction with an LLM-based chatbot, but we found differences among single items of the questionnaire which could refer to specific dimensions of self-perception and behavior.

Acknowledgments

The work has been carried out and supervised by the Institut für Multimediale und Interaktive Systeme, Universität zu Lübeck. Furthermore we would like to thank Luisa Winzer and Maged Mortaga for their support in developing the study environment.

Author's statement

Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the Ethics committee of the University zu Lübeck (2025-019).

References

[1] A. Nazir and Z. Wang. A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges. *Meta-Radiology*, 1(2):100022, 2023, doi:10.1016/j.metrad.2023.100022.

[2] C. Poulson. Shame and work. *Emotions in the workplace: Research, theory, and practice*, pp. 250–271, 2000.

[3] G. Anders, Die Antiquiertheit des Menschen Band 1, 7th ed. C.H. Beck München, 1985, 1.

[4] T. Schrills and T. Franke, The Antiquatedness of Men: Assessing Propensity to Promethean shame in Human-AI Interaction, presented at the DGPs-Kongress, 2024.

[5] G. L. Coelho, R. Vilar, P. H. Hanel, R. P. Monteiro, M. G. Ribeiro, and V. V. Gouveia. Optimism scale: Evidence of psychometric validity in two countries and correlations with personality. *Personality and Individual Differences*, 134:245–251, 2018, doi:10.1016/j.paid.2018.06.030.

[6] T. Franke, C. Attig, and D. Wessel. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction*, 35(6):456–467, 2018, doi:10.1080/10447318.2018.1456150.

[7] C. Attig, L. Winzer, T. Schrills, M. Zoubir, M. Mortaga, P. Wollstadt, C. Wiebel-Herboth, and T. Franke, Understanding successful human-ai teaming: The role of goal alignment and ai autonomy for social perception of llm-based chatbots, unpublished.

[8] C. Attig, P. Wollstadt, T. Schrills, T. Franke, and C. B. Wiebel-Herboth. More than Task Performance: Developing New Criteria for Successful Human-AI Teaming Using the Cooperative Card Game Hanabi. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24). Association for Computing Machinery*, (245)pp. 1–11, 2024, doi:10.1145/3613905.3650853.

[9] M. Madsen and S. Gregor. Measuring human-computer trust, 2000.

[10] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, Godspeed Questionnaire Series, 2009. doi:10.1037/t70855-000.

[11] L. Moradbakhti, B. Leichtmann, and M. Mara. Development and Validation of a Basic Psychological Needs Scale for Technology Use. *Psychological Test Adaptation and Development*, 5(1):26–45, 2024, doi:10.1027/2698-1866/a000062.

[12] A. Haney Chris Scott, Trivial pursuit, [Board game], 1981.