# Evaluated by a Machine. Effects of Negative Feedback by a Computer or Human Boss

Nicole C. Krämer[1]✉, Lilly-Marie Leiße[1], Andrea Hollingshead[2] & Jonathan Gratch[2]

[1]University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany; [2]University of Southern California, USA
Nicole.kraemer Lilly-Marie.leisse @uni-due.de; aholling@usc.edu;
gratch@ict.usc.edu

**Abstract.** In today´s remote working environments that include tasks given and performed via the Internet, people will encounter computer bosses that supervise their work. There is no knowledge on whether people will accept (negative) feedback that is given by an autonomous agent instead of a human. In a 2x2 between subject online experiment 183 participants performed a proofreading task and received either emotional or factual feedback by a human or computer boss. Results indicate that while the bosses´ behavior affects perceived warmness, human likeness and perceived psychological safety in the sense that factual feedback is perceived as more positive, there was only one significant result for the manipulation of the boss with regard to the perception of human-likeness.

**Keywords:** Feedback, media equation, CASA, work, computer boss

## 1 Differential effects by Human and Computer Bosses?

Intelligent technical applications are entering our lives: Smart home applications, robots for medical healthcare or as co-workers in our jobs are only a few examples. Especially the emerging application of artificial entities in organizational and operational contexts and their effects when they take the role of a supervisor needs to be scrutinized. It has to be examined whether from the perspective of the employee, a computer boss is suited better or worse to master fragile interactions such as giving negative feedback. As early as 1999, it was argued that interactions with computers could result in beneficial consequences such as increased psychological safety (i.e., the certainty that the other will respond in a favorable way when one reports a mistake) [1].

Several findings demonstrate differences between the interaction with a computer and a real human. For example, some studies [2] reveal that people self-disclose more towards a computer than towards a real person. The fear of being judged is lower and people deliver more honest answers. Further studies show that participants experience less negative emotions and report less desire for revenge and retaliation [3]. On the

other hand, numerous studies following the media equation approach find that people´s social reactions to computers and humans are very much alike [4]. Additionally, other findings indicate that a social reaction towards a computer is rather dependent on the amount of social cues displayed by the computer than on the knowledge whether one interacts with a human or a computer [5]. In order to contribute to the controversy on computer agents´ abilities to evoke social effects and to provide insights for the emerging applied area of computer bosses, the present research analyzes whether computer and human bosses evoke similar or different effects and/or whether it is their behavior that matters.

## 2      Method

**Participants.** 183 participants (74 females, Mage = 34.20, age range: 20-69 years) were considered for analysis (after excluding 19 due to lack of understanding).

**Design.** A 2x2 between-subjects design was used. The first manipulation concerned the nature of the supervisor: in one condition participants were instructed that they were interacting with a mediated real human boss, people in the other condition were led to believe their boss was an autonomously acting computer agent. The second manipulation was a distinction between factual and emotional feedback in reaction to the performance in a proofreading task. In all conditions, participants received the same negative, bogus feedback. The communication between the employee and the supervisor proceeded only via text messages. A picture of the particular boss was included in order to remind the participants of their interlocutor during the procedure and increase social presence (Figure 1). Two different pictures were used in each condition to guarantee generalizability.



**Fig. 1.** Fictitious pictures of the human bosses (left) and computer bosses (right)

**Procedure.** The whole study was processed online. Depending on condition, participants received different instructions: "Today, you are being supervised by another person over the web [computer boss]. This is another Turker [an artificially intelligent program] that has been trained to execute all the tasks and functions that a normal boss would do. […] Please attend to your [computer] bosses instructions."

After this description of the situation, the remainder of the survey was supervised by the [computer] boss. The participants received a text with numerous spelling errors and were asked to detect all errors. After this, the participants received a graphical presentation of their performance on a scale (with a low score of 3.5 of 10 possible points). Additionally, the boss commented on the performance either in an emotional or factual way. "[…] As you can already see at the chart, your performance was not

that good at all. *It is vividly shown there.* [*I am really upset.*] Actually you are just above the lower third. […] *Many things went wrong.* [*You disappoint me a lot.*] Seems you did not work hard enough and *that was not the aim of this task* [*that is embarrassing*]. […]. *Take my advice* [*Don't blame yourself.*]." Afterwards, participants filled in the questionnaires.

**Measures.** All questions were directly administered by the boss.

*Evaluation of the boss.* Two questionnaires were included to gain information about the perception concerning the boss. Twelve items measured the leading ability of the boss (e.g., "I create a good working atmosphere for you".) The items were rated on a seven-point scale from strongly disagree (1) to strongly agree (7) ($\alpha$ = .961; M = 2.61, SD = 1.257). In the second questionnaire, participants were asked to rate the personality of their boss with 13 item pairs (e.g., warm – cold, helpful – hindering; $\alpha$ = .891; M = 4.85, SD = 1.035). A factor analysis yielded three factors (warmness, efficiency, human-likeness, 70.80% of variance).

*Social presence.* The interaction were evaluated by four statements of the Social Presence Scale [6] such as: "During the interaction, how much did you feel as if someone was talking to you?" The statements were rated on a nine-point Likert scale from not at all (1) to very much (9) ($\alpha$ = .768; M = 6.24, SD = 1.704).

*Psychological safety.* A slightly reworded version of the Psychological Safety Scale [7] was included (e.g., "I value others' unique skills and talents"). Each statement was evaluated on a seven-point scale from strongly disagree (1) to strongly agree (7) ($\alpha$ = .865; M = 3.05, SD = 1.144).

## 3 Results

Concerning the evaluation of the boss, there was no significant main of the nature of the boss nor of the nature of the feedback (nor an interaction effect). There was, however, an effect on the perceived personality of the boss: Concerning the first factor "warmness", one significant main effect was found for the feedback condition ($F_{(1, 163)}$ = 7.497, p = .007, $\eta2$ = .044): Participants in the emotional feedback condition (M = 4.93, SD = 1.27) rated the warmness of the boss significantly lower than in the factual feedback condition (M = 5.46, SD = 1.06). However, no significant main effect was found for the boss condition. Also, there were two main effects for the factor "human likeness": The first one ($F_{(1, 163)}$ = 27.557, p < .001 , $\eta2$ = .145) showed that the computer boss (M = 5.35, SD = 1.387) was rated significantly less human-like than the human boss (M = 4.21, SD = 1.520). The second main effect referred to a significant difference between the two kinds of feedback ($F_{(1, 163)}$ = 6.224, p = .014, $\eta2$ = .037). Participants in the factual feedback condition (M = 4.5, SD = 1.61) rated the human likeness of the boss significant lower than participants in the emotional feedback condition (M = 4.99, SD = 1.49).

For psychological safety, a two-way independent analysis of variance revealed a significant main effect for the feedback condition ($F_{(1, 166)}$ = 6.649, p = .011, $\eta2$ = .039). Participants in the factual feedback condition (M = 3.27, SD = 1.212) reported a significant higher amount of psychological safety compared to participants in the emotional feedback condition (M = 2.82, SD = 1.024). There was no main effect for the boss condition nor an interaction effect.

Both, nature of the boss and nature of the feedback had no influence on perceived social presence. Finally, the nature of the boss did not influence participants´ performance.

## 4 Discussion

The present study focused on the question whether a computer boss supervising a task and providing negative feedback will be perceived differently from a human boss. To this aim we had people in an M-Turk setting perform a task on which they were subsequently given negative feedback in either an emotional or factual way by either a human or computer boss. As was already observed in prior studies [5] the information on whether the boss was a human or a computer did not make a difference. After the reception of negative feedback the human and the computer boss were evaluated similarly – in general and also with regard to the sub-factors warmth and efficiency. Also, the perceived psychological safety and performance did not differ, indicating that a human boss is not per se perceived as providing more psychologically safe behavior in times of need. However, this also implies that a computer boss is not automatically perceived as a "better boss" as other prior studies seemed to suggest. The study again demonstrated that behavior matters and yields more effects than the nature of the interlocutor. Results indicate that independent of whether the boss is human or artificial factual feedback is perceived as more beneficial as emotional feedback – in the sense that people perceived the boss warmer, more human-like and felt more psychological safety.

## References

1. Edmondson, A.: Psychological Safety and Learning Behavior in Work Teams. In: Administrative Science Quarterly, 44(2), 350 (1999)
2. Lucas, G. M., Gratch, J., King, A., & Morency, L.-P.: It's only a computer: Virtual humans increase willingness to disclose. In: Computers in Human Behavior, 37, 94-100 (2014)
3. Kim, D., Frank, M. G., & Kim, S. T.: Emotional display behavior in different forms of Computer Mediated Communication. In: Computers in Human Behavior, 30, 222-229 (2014)
4. Reeves, B., & Nass, C. I.: The media equation: How people treat computers, television, and new media like real people and places. Stanford, Calif., New York: CSLI Publications; Cambridge University Press (1996)
5. von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H.: "It doesn't matter what you are!" Explaining social effects of agents and avatars. In: Computers in Human Behavior, 26(6), 1641-1650 (2010)
6. Lee, K. M., & Nass, C.: Designing social presence of social actors in human computer interaction. Conference on Human Factors in Computing Systems - Proceedings, 289-296 (2003)
7. Baer, M., & Frese, M.: Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. In: Journal of Organizational Behavior, 24(1), 45-68 (2003)