

Smiling in a Wizard of Oz Experiment: Emotional vs. Social Smiles, General Effects and Sex Differences

Martin Krippel¹(✉), Matthias Haase², Julia Krüger²,
and Jörg Frommer²

¹ Institute for Psychology, Department Methodology,
Psychodiagnostics and Evaluation Research,
Otto-von-Guericke University of Magdeburg, Magdeburg, Germany
martin.krippel@ovgu.de

² Department of Psychosomatic Medicine and Psychotherapy, Medical Faculty,
Otto-von-Guericke University Magdeburg, Magdeburg, Germany

Abstract. Derks et al. (2008) showed the similarity of user's emotional involvement in human-human-interaction compared to human-computer-interaction. This implies display rules and gender norms on display rules. In a Wizard of Oz Experiment users had to overcome two challenging situations, where emotions should be induced. Emotional smiles and social smiles were categorized after facial movements were coded with the Facial Action Coding System (Ekman and Friesen 1978). Emotional smiles were more frequent during one of the challenges compared to the baseline, showing that emotion induction was successful. Social smiles were more frequent at the start of the experiment compared to the challenges, but only for woman. This supports the assumption of Derks et al. (2008) that the display rule "smile at the start of an interaction" is only valid for woman. Together the results back the idea that computers are seen as human-like counterparts.

Keywords: Smiling · Emotional smile · Social smile · FACS · Sex differences · Display rules · Gender norms

1 Introduction

1.1 Social Versus Emotional Smiles

Derks et al. (2008) show that the users' emotional involvement in computer-mediated interaction is comparable to that in face-to-face interaction. Thus, it might be suggested that the computer/technical system is perceived as a virtual person by the user and therefore he/she behaves as if the computer/technical system would be a human-like counterpart.

This implies the usage of display rules (Ekman and Friesen 1982; Ekman et al. 1969). One of these display rules is to smile at the start of a communication with a new interaction partner. Such an intentional smile is a social smile (Ekman and Friesen 1982), telling the other person that one will not harm him/her.

To be able to distinguish between different kinds of smiles and facial actions in general, one needs a coding system, which objectively describes facial movements. The gold standard here is the Facial Action Coding System (FACS, Ekman and Friesen 1978; Ekman et al. 2002), where movements are anatomically defined as so called Action Units (AUs). According FACS the social smile involves only the zygomaticus major muscle pulling the lip corners up (AU 12), whereas the emotional smile also involves the orbicularis oris muscle (AU 6), which induces crow feet wrinkles (Fig. 1). Besides such initial interaction situation, it could furthermore be expected that emotional smiles also occur during events which induce negative emotions. This could be because events that induce negative emotions could the same time also induce positive emotions. Additionally as mentioned above smiling is not only a sign of joy, but also one of a mere social signal. This could be shown for human-human-interaction (Ekman et al. 1969) and for human-computer interaction (HCI) contexts (Hoque and Picard 2011).

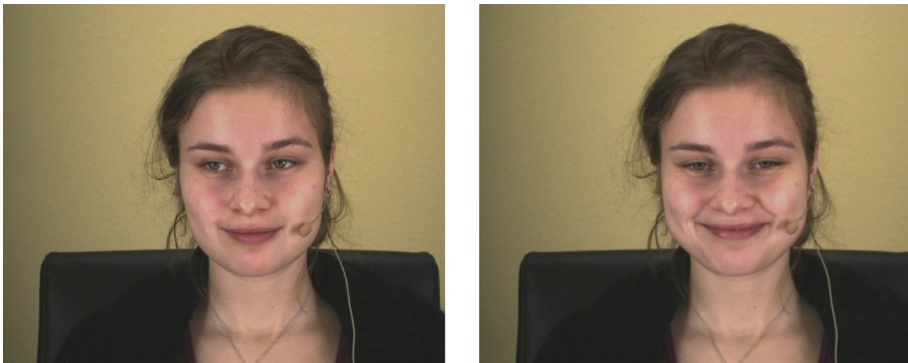


Fig. 1. Social smile (left) and emotional smile (right)

Another interesting research question in HCI and in emotion research is whether and which differences in social and emotional smiling occur between genders. In a meta-analysis, LaFrance et al. (2003) showed that in general woman smile more often than men. We want to check whether this difference can also be found in HCI, suggesting the assumption that the technical system is treated like a person, and whether the sex difference in smiling frequency is based on social or emotional smiles.

1.2 Current Experiment

In the current Wizard-of-Oz (WOz) experiment (Rösner et al. 2012), a speech based system was simulated. The user had to create a solution for a time sensitive task. During the task the users were had to overcome challenges. The subjects got to know the simulated system by a system's self-introduction at the beginning (*baseline*). Afterwards, they had to pack their baggage for a holiday trip. At one moment the weight of their baggage transgressed a not defined amount (*Weight limit barrier*) and

subjects had to change their baggage. Another challenge appeared when participants were told that the holidays were not summer, but winter holidays (*Waiuku barrier*). The German participants thought they would pack for summer holidays as suggested in the cover story of the experiment. During the experiment, they got the information that their holiday's destination will be Waiuku, a city in New Zealand, where the temperature is obviously lower than in Germany at the travel time. Accordingly, participants had to change the clothes in their baggage to adapt to the new situation. Therefore, they were given only a few minutes.

Regarding facial activity, both challenge situations (*weight limit barrier* as well as *Waiuku barrier*) were interesting regarding a comparison to the facial activity during the *baseline* (system's self-introduction).

1.3 Questions

1. Does social smiling frequency differ between baseline and challenge situations?
2. Does emotional smiling frequency differ between baseline and challenge situations?
3. Is there a sex difference in social smiling frequency in general or in only a part of the conditions?
4. Is there a sex difference in emotional smiling frequency in general or in only a part of the conditions?

2 Methods

2.1 Sample

We gathered participants from two different age groups (aged between 19 and 29 as well as above 60). Efforts were made to achieve an equal distribution of participants with regard to age group, gender and level of education. We acquired a total of 135 participants. Data from five participants was excluded due to technical problems regarding the records and/or absence from the second appointment despite several attempts to establish contact. Therefore, the total sample consisted of 130 participants.

The facial activity data of 80 participants in the total sample were used for FACS coding and statistical analysis. Two participants were used to check interrater agreement. The rest of the total sample could not be used because of technical problems with regard to the synchronization of videos with sound stimuli and time markers. The final sample consisted of 37 men and 43 women (see Table 1). The age range was from 18 to 81 with a mean of 49.23 (SD = 22.78). Participants were part of either the younger (18 to 28) or the older group (60 to 81).

2.2 Design and Procedure

Implementing a fully automated technical system is difficult, however simulating one by the use of the WOz experiments allows examining research questions and provide a

Table 1. Sample parameters

	N	Women	Age mean	Age SD
Young	34	20	23.06	2.806
Old	46	23	68.28	5.06
All	80	43	49.06	22.89

deeper understanding regarding the interaction between users and computer systems. In such experiments an interface is controlled by a human operator which simulates the computer system.

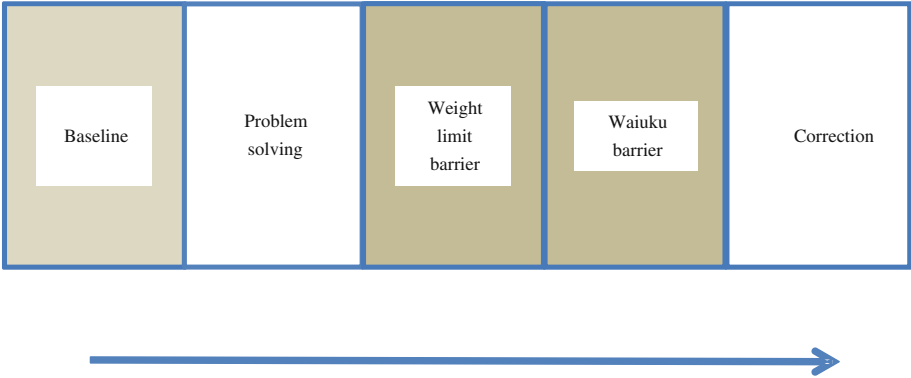


Fig. 2. Design

Participants were instructed to communicate with the computer system via speech input and output. The simulated system asked for personal information from users under the pretext of needing to conduct an individual calibration. This served the underlying purpose of having participants adjust to the speech control and functionality of the simulated system. This section was defined as baseline (Fig. 2). Subsequently, participants were told to prepare baggage for a trip with the aid of the system and were informed about the time limit for this task. In the following, participants worked through twelve categories (e.g., jackets, tops, trousers, shoes, etc.) to gather items for their baggage.

It was suggested that participants pack for a summer vacation. They began by choosing items out of each category. This section was defined as the baseline. The first limitation was introduced after solving the eighth category. Participants were informed that they had exceeded the weight limit for their baggage, independently of the number of items they had gathered so far. This required the removal of some items in order to add more. The system advised participants to change their strategy. This section was called weight limit barrier. During the third section (Waiuku barrier), which also represented a challenge, it was disclosed that participant’s holiday destination is a summer location but a winter one. At this point, a randomly pre-selected subsample received a system-initiated intervention based on psychotherapeutic objectives. The intervention focused on problem actualization and the activation of resources.

System-initiated communication shifted to a more interpersonal level. All participants had to re-adapt their strategy without clearly defined time limitations. Finally, they all had the opportunity to modulate their choices under time restrictions. This section simultaneously represented the last challenge as well as the experimental part of the WOz trial.

Self-report measures of emotions were not used in the experiment because they bring users out of the flow of the experience, changing the physiological features of and the experience of emotion (Kassam and Mendes 2013). More complete information with regard to the resulting data corpus are presented in Frommer et al. (2012) and Rösner et al. (2012).

2.3 FACS Coding

FACS (Ekman and Friesen 1978; Ekman et al. 2002) has been the gold standard of facial movement categorization ever since its publication. FACS describes every facial movement independently of emotion categorization.

Four certified FACS coders coded the videos in groups of two, meaning that each video was coded by two coders. In the case of disagreement, the coders discussed their codes until they arrived at a joint decision. Data from two participants who were not part of the statistical analysis formed the basis of the interrater agreement computation. Agreement was computed via a formula by Wexler (Ekman and Friesen 1978). Agreement scores between the group of two standard coders and an additional fifth coder (first author) were between .71 and .89 for all AUs.

For each experimental condition, the five seconds immediately following the conveyance of the relevant information (in the baseline at the start of the experiment) were coded. The frequency of each emotion served as the dependent variable.

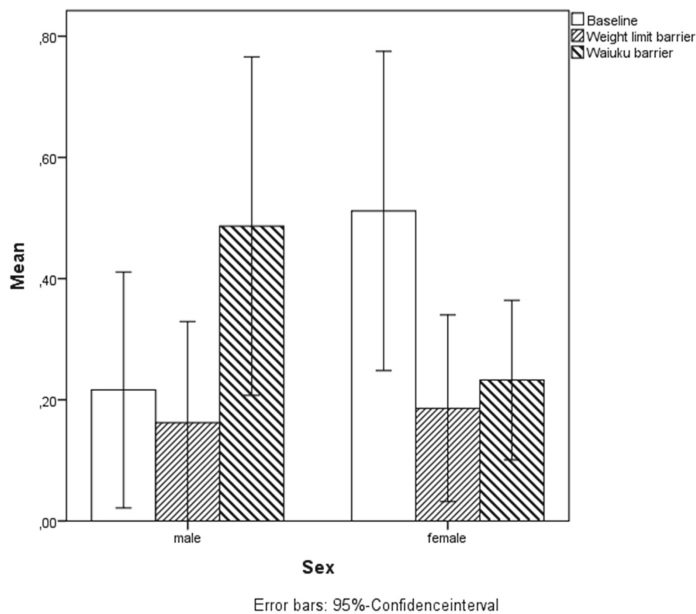
Although AUs were classified into several emotion categories, we concentrated on emotional and social smiles in this paper.

3 Results

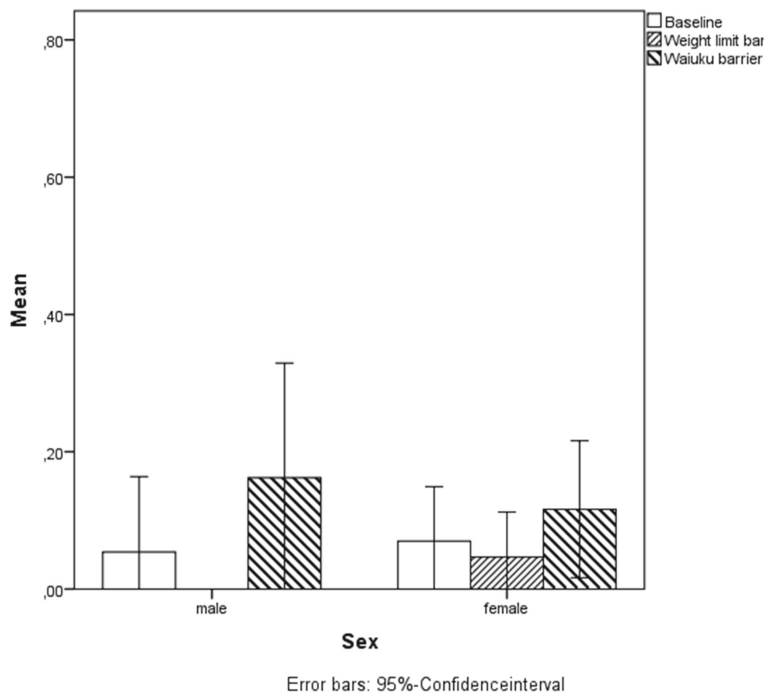
In order to answer the research questions, we conducted a mixed ANOVA with the experimental conditions as repeated measurement factor and sex as between-subjects factor for both, the ‘social smile’ (Fig. 3) and the ‘emotional’ smile (Fig. 3b).

Whereas for the dependant variable ‘social smile’ no main effect on the experimental conditions ($p = .10$) and the main effect sex ($p = .797$) showed up, the interaction of experimental conditions with sex was significant ($p = .026$).

For the dependant variable ‘emotional smile’ the main effect on experimental conditions was significant ($p = .030$), whereas no main effect sex ($p = .907$) and no interaction effect ($p = .567$) showed up.



a) a) Social Smile



b) Emotional Smile

Fig. 3. Mean frequencies of social (a) and emotional (b) smile during the three conditions, separated by sex

4 Discussion

As assumed, a considerable amount of participants showed smiling, especially social smiling during the *baseline* condition. Furthermore, we also found out that woman showed more social smiling during *baseline* but less during the *Waiuku barrier*, whereas in men it was the opposite.

This supports the idea that social smiling during baseline is based on display rules. The social smile during baseline probably represents a greeting sign, signaling to the computer that one will not harm him. The higher frequency of the social smiles in woman maybe based on gender norms LaFrance et al. (2003), a specific form of display rules, indicating that woman should smile and men need not. The application of display rules and gender norms suggests that users attribute human-like characteristics to the technical system.

Results on emotional smiles support the findings by Hoque and Picard (2011) in a way that emotional smiles can even be induced by negative events. As described by Papa and Bonanno (2008), this could be understood as a kind of self-regulation, which is part of a dissociating reaction for being fooled. By dissociating from the situation (which implies a reduction of ego-involvement), users feel amused about being fooled by the technical system.

All in all, results on social and emotional smiles indicate that users interact with a technical system as if it would be a human-like counterpart. This is in line with findings based on semi-structured interviews with the users after the experiment (Krüger et al. 2015). The results show a phenomenon called anthropomorphism. It means the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions” (p. 864, Epley et al. 2007). As Epley et al. (2007) point out, three components are important for anthropomorphism: “the accessibility and applicability of anthropocentric knowledge (elicited agent knowledge), the motivation to explain and understand the behavior of other agents (effectance motivation), and the desire for social contact and affiliation (sociality motivation)” (p. 864, Epley et al. 2007). Although we did not measure these aspects, they should have been relevant in our experimental situation. This should be the case especially for the desire for social contact and affiliation, because participants were alone with the computer.

References

- Derks, D., Fischer, A.H., Bos, A.E.R.: The role of communication in computer-mediated communication: A review. *Comput. Hum. Behav.* **24**, 766–785 (2008)
- Ekman, P., Friesen, W.V.: *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto (1978)
- Ekman, P., Friesen, W.V.: Felt, false, and miserable smiles. *J. Nonverbal Behav.* **6**, 238–258 (1982)
- Ekman, P., Friesen, W.V., Hager, J.C.: *Facial Action Coding System - The Manual on CD ROM*. Prentice-Hall, Englewood Cliffs (2002)

- Ekman, P., Sorensen, E.R., Friesen, W.V.: Pan-cultural elements in facial displays of emotion. *Science* **164**(3875), 86–88 (1969)
- Epley, N., Waytz, A., Cacioppo, J.T.: On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev.* **114**(4), 864–886 (2007)
- Frommer, J., Michaelis, B., Rösner, D., Wendemuth, A., Friesen, R., Haase, M., Siegert, I.: Towards emotion and affect detection in the multimodal LAST MINUTE corpus. In: Paper Presented at the LREC, Eighth International Conference on Language Resources and Evaluation, Istanbul (2012)
- Hoque, M., Picard, R.W.: Acted vs. natural frustration and delight: many people smile in natural frustration. In: Paper Presented at the 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011) (2011)
- Kassam, K.S., Mendes, W.B.: The effects of measuring emotion: physiological reactions to emotional situations depend on whether someone is asking. *PLoS ONE* **8**(6), e64959 (2013). doi:[10.1371/journal.pone.0064959](https://doi.org/10.1371/journal.pone.0064959)
- Krüger, J., Wahl, M., Frommer, J.: Making the system a relational partner: users' ascriptions in individualization focused interactions with companion-systems. In: Paper Presented at the CENTRIC 2015: The Eighth International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services, Barcelona (2015)
- LaFrance, M., Hecht, M.A., Paluck, E.L.: The contingent smile: a meta-analysis of sex differences in smiling. *Psychol. Bull.* **129**, 305–334 (2003)
- Papa, A., Bonanno, G.A.: Smiling in the face of adversity: The interpersonal and intrapersonal functions of smiling. *Emotion* **8**(1), 1–12 (2008)
- Rösner, D., Frommer, J., Friesen, R., Haase, M., Lange, J., Otto, M.: LAST MINUTE: a multimodal corpus of speech-based user-companion interactions. In: Paper Presented at the Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), ELRA. http://lrec.elra.info/proceedings/lrec2012/pdf/550_Paper.pdf (2012). Accessed Sept 2015