



Beyond Likeability: Investigating Social Interactions with Artificial Agents and Objective Metrics

Sebastian Loth*

Social Cognitive Systems and Psycholinguistics, Centre of Excellence on Cognitive Interaction Technology (CITEC), Faculty of Technology, Bielefeld University, Bielefeld, Germany

Keywords: virtual agents, social signals, interaction design, gaze aversion, social eye gaze, dialogue capability, facial animation

A commentary on

Comprehension and engagement in survey interviews with virtual agents

by Conrad, F. G., Schober, M. F., Jans, M., Orłowski, R. A., Nielsen, D., and Levenstein, R. (2015). *Front. Psychol.* 6:1578. doi: 10.3389/fpsyg.2015.01578

Numerous studies have investigated whether additional abilities of artificial agents improve the interaction with their users. Typically, a specific capability was added to a baseline system and participants interacted with both versions. Questionnaires (e.g., GOODSPEED, Bartneck et al., 2009) or time-related metrics combined with a questionnaire (e.g., PARADISE, Walker et al., 2000) were then evaluated. In most published work, the new capability demonstrably improved the user ratings. For example, agents with the capability to praise (Fasola and Mataric, 2012), err (Salem et al., 2013), and blink pleasantly (Takashima et al., 2008) were more likeable than their counterparts without this capability. However, subjective ratings can be unreliable (Cahill, 2009; Belz and Kow, 2010) and hardly address questions beyond likeability, e.g., whether an action is perceived as social or task-oriented, and when and why an agent is more or less comprehensible. But investigating human social behavior with controlled and ecologically valid experiments is notoriously difficult given the variance in natural interactions (e.g., Bergmann et al., 2010; Sciutti et al., 2015). Artificial agents can precisely reproduce interactive behavior in real-time but are still rarely used for investigations beyond likeability.

Conrad et al. (2015) used subjective ratings and objective metrics (gaze behavior, response accuracy, number of clarification questions and backchannels) for investigating how the users' comprehension of straightforward and complicated survey questions¹ was affected by the virtual agent's facial expressiveness² and dialogue capability³. As expected, the subjective ratings improved with more advanced agent capabilities. Objective metrics such as response accuracy and eye tracking data provided further insights, e.g., whether interviewees avoid eye contact with their interviewer (gaze aversion). Research in human-human interaction indicated that gaze aversion was not social behavior but that it reduced distraction and directed all the interviewee's cognitive

OPEN ACCESS

Edited and reviewed by:

Claire Marie Fletcher-Flinn,
University of Auckland, New Zealand

Reviewed by:

Vero Vanden Abeele,
KU Leuven, Belgium

***Correspondence:**

Sebastian Loth
sebastian.loth@uni-bielefeld.de

Received: 21 July 2016

Accepted: 11 September 2017

Published: 22 September 2017

Citation:

Loth S (2017) Beyond Likeability: Investigating Social Interactions with Artificial Agents and Objective Metrics. *Front. Psychol.* 8:1662. doi: 10.3389/fpsyg.2017.01662

¹Questions from the US current population survey (monthly survey of 60,000 households estimating, e.g., unemployment rate, see United States Census Bureau, 2016) were "complicated" if they were ambiguous (e.g., is a floor lamp a piece of furniture or an electrical appliance) and likely triggered clarification questions.

²Facial animation was either low (head and face did not move, the eyes did not blink and the mouth was only opened and closed indicating whether the agent was speaking) or high (21 channel-motion tracking of a human face, appropriate lip movements, blinking behavior at all times during the interview).

³Dialogue capability was either low (the agent responded to user questions with a neutral probe or repeated the interview question) or high (the agent tried to provide helpful answers such as clarifying terms). In both conditions, the agent read out the interview questions and repeated them if needed.

resources toward answering the question (Glenberg et al., 1998; Doherty-Sneddon et al., 2002; Doherty-Sneddon and Phelps, 2005). But the interviewees' gaze patterns in Conrad et al.'s (2015) study were independent of the question's difficulty if the agent had high dialogue capabilities. This implies a social component in gaze aversion that cannot be ruled out by the original studies that used (a) human questioners such that gaze aversion could be social (Glenberg et al., 1998; Doherty-Sneddon et al., 2002), (b) no questioner such that truly social gaze behavior was impossible to observe (Glenberg et al., 1998), or (c) memory tasks (Glenberg et al., 1998; Doherty-Sneddon and Phelps, 2005) that are sensitive to any intervening information such as faces (Posner and Konick, 1966; Dale, 1973; West, 1999; de Fockert et al., 2001). A social component in gaze aversion is not surprising given the social nature of other behaviors. For example, facial expressions of surprise were more pronounced if other humans were present in the room (Schützwohl and Reisenzein, 2012). Even the electrodermal activity of truth tellers and deceivers differed only if they believed their virtual interviewer was controlled by human but not by artificial intelligence (Ströfer et al., 2016). Thus, gaze aversion appears to be at least partly social. Conrad et al.'s (2015) findings illustrate that objective metrics in HMI reveal new and additional insights (cf. Zarrieß et al., 2015) providing a valuable tool for understanding social interaction.

Designing artificial agents driven by likeability tends to accumulate capabilities regardless of their contribution to the interaction. First, (likeable) capabilities contribute differently to an interaction and do not necessarily increase an artificial agent's comprehensibility. Conrad et al. (2015) tested the effect of each capability on their participants' behavior. For example, high dialogue capabilities produced the same response accuracy irrespectively of the agent's facial expressiveness. On the other hand, facial expressiveness influenced the participants' back-channeling behavior which could smoothen the interaction (cf. Clark and Brennan, 1991; Clark and Krych, 2004). Secondly, simply combining likeable behaviors does not necessarily result in a human-like, comprehensible artificial agent. For example, humans have difficulty integrating human beat gestures and speech if enacted by a robot (Bremner and Leonards, 2015). Also, averaging across behaviors of individuals may result in

an unintelligible chimera whereas humans prefer the behavior of one individual (Bergmann et al., 2010; Kelly et al., 2010). Arguably, a fictional character (or artificial agent) gains appeal by focussing on essential communicative behaviors that form a coherent, expressive and comprehensible individual (Thomas and Johnston, 1995). Thirdly, not being perceived as human may not be likeable but advantageous because users would assume an automatic agent operates consistently rather than having bad intentions or preferring someone over someone else. Also, this intuitively differentiates between humans and artificial agents (cf. Stein and Ohler, 2017). In contrast to being all human-like, using human cognitive principles for maximizing an agent's capability to interact with its users is highly desirable. Thus, objective metrics and insightfully designed experiments are more important than likeable capabilities.

In sum, Conrad et al. (2015) estimated the communicative relevance of an artificial agent's social behaviors given a specific task with objective metrics such as response accuracy and gaze patterns. Their study demonstrated that the difficult research in social interaction can benefit from artificial agents if combined with insightful, reliable, replicable and objective dependent metrics.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

This research/work was supported by the Cluster of Excellence Cognitive Interaction Technology "CITEC" (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

ACKNOWLEDGMENTS

I would like to thank my colleagues Christina Unger, Hendrik Buschmeier, and Julian Hough for providing their comments on this manuscript.

REFERENCES

- Bartneck, C., Kulic, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 71–81. doi: 10.1007/s12369-008-0001-3
- Belz, A., and Kow, E. (2010). "Comparing rating scales and preference judgements in language evaluation," in *Proceedings of the 6th International Natural Language Generation Conference* (Dublin: Association for Computational Linguistics), 7–15.
- Bergmann, K., Kopp, S., and Eyssel, F. (2010). "Individualized gesturing outperforms average gesturing - evaluating gesture production in virtual humans," in *Proceedings of the 10th International Conference on Intelligent Virtual Agents (LNCS 6356)*, eds J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova (Berlin; Heidelberg: Springer International Publishing), 104–117.
- Bremner, P., and Leonards, U. (2015). "Speech and gesture emphasis effects for robotic and human communicators: a direct comparison," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: ACM Press), 255–262. doi: 10.1145/2696454.2696496
- Cahill, A. (2009). "Correlating human and automatic evaluation of a German surface realiser," in *ACLShort '09 Proceedings of the ACL-IJCNLP 2009* (Singapore: Association for Computational Linguistics), 97–100.
- Clark, H. H., and Brennan, S. E. (1991). "Grounding in communication," in *Perspectives on Socially Shared Cognition, 1st Edn.*, eds L. B. Resnick, J. M. Levine, and S. D. Teasley (Washington, DC: American Psychological Association), 127–149.
- Clark, H. H., and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *J. Mem. Lang.* 50, 62–81. doi: 10.1016/j.jml.2003.08.004
- Conrad, F. G., Schober, M. F., Jans, M., Orłowski, R. A., Nielsen, D., and Levenstein, R. (2015). Comprehension and engagement in survey

- interviews with virtual agents. *Front. Psychol.* 6:1578. doi: 10.3389/fpsyg.2015.01578
- Dale, H. C. A. (1973). Short-term memory for visual information. *Brit. J. Psychol.* 64, 1–8. doi: 10.1111/j.2044-8295.1973.tb01320.x
- de Fockert, J. W., Rees, G., Frith, C. D., and Lavie, N. (2001). The role of working memory in visual selective attention. *Science* 291, 1803–1806. doi: 10.1126/science.1056496
- Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., and Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Dev. Psychol.* 38, 438–445. doi: 10.1037/0012-1649.38.3.438
- Doherty-Sneddon, G., and Phelps, F. G. (2005). Gaze aversion: a response to cognitive or social difficulty? *Mem. Cogn.* 33, 727–733. doi: 10.3758/BF03195338
- Fasola, J., and Mataric, M. J. (2012). Using socially assistive human-robot interaction to motivate physical exercise for older adults. *Proc. IEEE* 100, 2512–2526. doi: 10.1109/JPROC.2012.2200539
- Glenberg, A. M., Schroeder, J. L., and Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Mem. Cogn.* 26, 651–658. doi: 10.3758/BF03211385
- Kelly, S. D., Creigh, P., and Bartolotti, J. (2010). Integrating speech and iconic gestures in a stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254
- Posner, M. I., and Konick, A. F. (1966). Short-term retention of visual and kinesthetic information. *Organ. Behav. Hum. Perform.* 1, 71–86. doi: 10.1016/0030-5073(66)90006-7
- Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., and Joublin, F. (2013). To ERR is human(-like): effects of robot gesture on perceived anthropomorphism and likability. *Int. J. Soc. Robot.* 5, 313–323. doi: 10.1007/s12369-013-0196-9
- Schützwohl, A., and Reisenzein, R. (2012). Facial expressions in response to a highly surprising event exceeding the field of vision: a test of Darwin's theory of surprise. *Evol. Hum. Behav.* 33, 657–664. doi: 10.1016/j.evolhumbehav.2012.04.003
- Sciutti, A., Ansuini, C., Becchio, C., and Sandini, G. (2015). Investigating the ability to read others' intentions using humanoid robots. *Front. Psychol.* 6:1362. doi: 10.3389/fpsyg.2015.01362
- Stein, J.-P., and Ohler, P. (2017). Venturing into the uncanny valley of mind-The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160, 43–50. doi: 10.1016/j.cognition.2016.12.010
- Ströfer, S., Ufkes, E. G., Bruijnes, M., Giebels, E., and Noordzij, M. L. (2016). Interviewing suspects with avatars: avatars are more effective when perceived as human. *Front. Psychol.* 7:545. doi: 10.3389/fpsyg.2016.00545
- Takashima, K., Omori, Y., Yoshimoto, Y., Itoh, Y., Kitamura, Y., and Kishino, F. (2008). "Effects of avatar's blinking animation on person impressions," in *GI '08 Proceedings of Graphics Interface 2008*, vol. 169–176 (Windsor, ON: Canadian Information Processing Society).
- Thomas, F., and Johnston, O. (1995). *The Illusion of Life: Disney Animation*, 1st Hyperion Edn. New York, NY: Hyperion.
- United States Census Bureau (2016). Current Population Survey (CPS).
- Walker, M., Kamm, C., and Litman, D. (2000). Towards developing general models of usability with PARADISE. *Nat. Lang. Eng.* 6, 363–377. doi: 10.1017/S1351324900002503
- West, R. (1999). Visual distraction, working memory, and aging. *Mem. Cogn.* 27, 1064–1072. doi: 10.3758/BF03201235
- Zarrieß, S., Loth, S., and Schlangen, D. (2015). "Reading times predict the quality of generated text above and beyond human ratings," in *Proceedings of the 15th European Workshop on Natural Language Generation*, eds A. Belz, A. Gatt, F. Portet, and M. Purver (Brighton: The Association for Computational Linguistics), 38–47.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Loth. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.