4.6 Breakout Group 5: Definition, Conceptualization and Measurement of Trust

Group 5

Contributors:
Group 5a (in person): Martin Porcheron, Minha Lee, Birthe Nesset, Frode Guribye
Group 5b (online): Margot van der Goot, Roger K. Moore, Ricardo Usbeck, Ana Paiva, Catherine Pelachaud, Elayne Ruane
Group 5c (in person): Björn Schuller, Guy Laban, Dimosthenis Kontogiorgos, Matthias Kraus, Asbjørn Følstad

4.6.1 Goal and key questions

Goal: Enable assessment and measurement of trust in CAs
Key question: How to define, conceptualize and measure trust in CA?
Relevant aspects: Modelling frameworks – antecedents / consequents; basis in knowledge on human-human communication; psychology of trust – over-trust / intuition

4.6.2 Key insights

Defining trust. Trust is addressed in different disciplines, both as a general concept within psychology, sociology, and management research (e.g., Rousseau, 1998; Mayer et al., 1995) and – more recently – as a term of relevance for users' perceptions of technology (e.g., Corritore et al., 2003; McKnight et al., 2011). A range of definitions exists for trust. There is variation in definitions concerning whether trust should be construed as a belief or attitude (Lewis et al., 2018; Lee & See 2004), and the degree to which there is a behavioural element in trust (Söllner et al., 2016; Malle & Ullman, 2021).

For conceptual clarity, it may be useful to consider trust an attitude which may be founded by trusting beliefs, and which may lead to trusting behaviour.

Trusting behaviour is determined by trust and may as such be an indicator of trust – provided users have a choice. Trusting behaviour is also moderated by environment, user group, and use case. An example of trusting behaviour is self-disclosure. Trust may also impact engagement level in behaviour and tendency to repeated use.

Developing trust through conversational interactions. The notion of trust in technology arguably is of particular relevance to CAs, due to their conversational interaction with users. Conversations are humanlike which has implications for users trusting beliefs and behaviours. Furthermore, conversations may be relational, leading to expectations of evolving capabilities in agent. Conversations may also be cooperative, leading to expectations of mutual adaptations in the user and conversational agent to achieve a common goal.

On this background, trust in CA may be considered as gradually built through conversations. In consequence, four trust concepts may be of particular relevance for CAs:

- *Initial trust*: trust required for users to initiate interaction. Initial trust corresponds to the notion of calculating trust in Rousseau et al. (1998)

- *Sharing trust*: trust required for sharing information with chatbot. The relevance of sharing trust may depend on varying levels of perceived sensitivity in the domain or topic of CA interaction.
- *Reliance trust*: trust required for relying on chatbot recommendations or decision support, that is, trust impacting user beliefs or behaviour beyond the context of the CA interaction.
- *Long-term trust*: trust required for repeated / routine use. Long term trust corresponds to the notion of relational trust in Rousseau et al. (1998)

Extending the trust model of Rousseau et al. (1998), the four trust concepts for CAs may be mapped out on a timeline of the evolving relation between user and CA as follows:

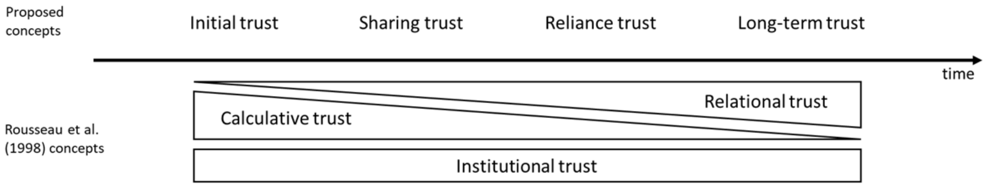

Figure 4 Extended trust model.

Balancing trust and trustworthiness. When considering trust, it is critical to distinguish between perceived trust and trustworthiness.

Perceived trust is held by the trustor, typically the user. Perceived trust and related trust beliefs may be measured through a range of self-report measurements, for example from information systems research (e.g. Lankton et al., 2015), social robotics (c.f. review in Hancock et al., 2020). Perceived trust may be impacted by the trustworthiness of the trustor. However, as information on this may not be available, other characteristics may impact trust. For CAs, anthropomorphism may be such a characteristic, as it may impact trust though not be correlated with trustworthiness.

Trustworthiness is a characteristic of the trustee, typically the service provider. Trustworthiness may depend on factors such as transparency, reliability, consistency, sincerity, honesty, integrity, benevolence, competence, and cooperation. These factors, though not necessarily static, may be considered observable characteristics in a trustee.

There is a need to study trustworthiness and perceived trust in parallel – to address potential overtrust (low trustworthiness and high perceived trust) and undertrust (high trustworthiness and low perceived trust). There is a lack of approaches or measurements for the integrated study of trustworthiness and trust.

Measuring trust by integrating self report measures and behavioural measures. In existing scales and measurements, trust is typically construed as personal, mainly available to researchers through self report measurements. Nevertheless, trust can be interpreted as reflected in and through people's behaviour, rather than merely a stance prior to the use of some device or system. Trust as reflected in trusting behaviour may enables trust to be measured also on the basis of user behaviour. There seem however to be a lack of distinct behaviour scales for trust assessment.

Possibly, trust may be measured by having a CA asking about sensitive information and monitor users' disclosing behaviour. Specifically, a tiered approach may be useful, based on asking questions of personal information of increasing level of sensitivity to infer a person's level or trust. However, the choice of behavioural measures of trust may depend on the context of the CA.

An integrated approach, combining self-report measures and tiered behavioural measures seems a promising approach for future research.

A proposed integrating framework for measuring trust and trusting behaviour. Following from the above, instruments and data sources for measuring trust may be divided into two broad groups: Subjective and objective measures:

- *Subjective measures* concern the measurement of trust determinants / trusting beliefs or behavioural intent (e.g., Lankton et al., 2015; Yagoda & Gillan, 2012). As a subjective measurement, perceptions of trust are expected to be explicit from the subject's report, corresponding to the subject's trust beliefs. Nevertheless, these might not be consistent with the subject's trusting behaviour due to personal perceptions and attitudes of the subject regarding the conversational AI system – e.g., due to scepticism of AI (Araujo et al., 2020).
- *Objective measures* include measures of physiological states, speech / voice, interaction with agent (e.g., sharing behaviour), changes in beliefs due to agent, behaviour in the world due to agent. Accordingly, the subject's behaviour would implicitly indicate higher or lower levels of trust. THe association between trusting behaviour and trust should be studied individually, depending on context, settings and task. Within the scope of conversational AI, behaviour such as self-disclosure (e.g., Laban et al., 2021a), reciprocity (e.g., Zonca et al., 2021), and changes in disclosure and expression over time (e.g., Laban et al., 2021b) could implicitly indicate changes in trust. These behaviours, however, might not be consistent with one's trust beliefs due to, for example, habits and needs (e.g., having the need to share, or being an impulsive individual) or affect-based factors of trust like the system's heuristics and demonstrated social cues (e.g., one might be more likely to share information with a more persuasive system despite not trusting it; e.g., Ghazali et al., 2019).

Subjective and objective measures may be included in a framework of trusting beliefs and trusting behaviour as follows:

4.6.3 Future Research

The following questions require further research effort to address:

- Developing a comprehensive framework to capture how trust evolves across long-term use.
- Refining the framework for trusting beliefs and trusting behaviour.
- Developing integrated approaches and measures for studying users perceived trust and the trustworthiness of service providers, to mitigate overtrust and undertrust.
- Developing integrated measures of trust and trusting behaviour, combining self report measures and tiered behavioural measures to support standardised measure for trust in conversational agents, and incorporating this in conversational systems.

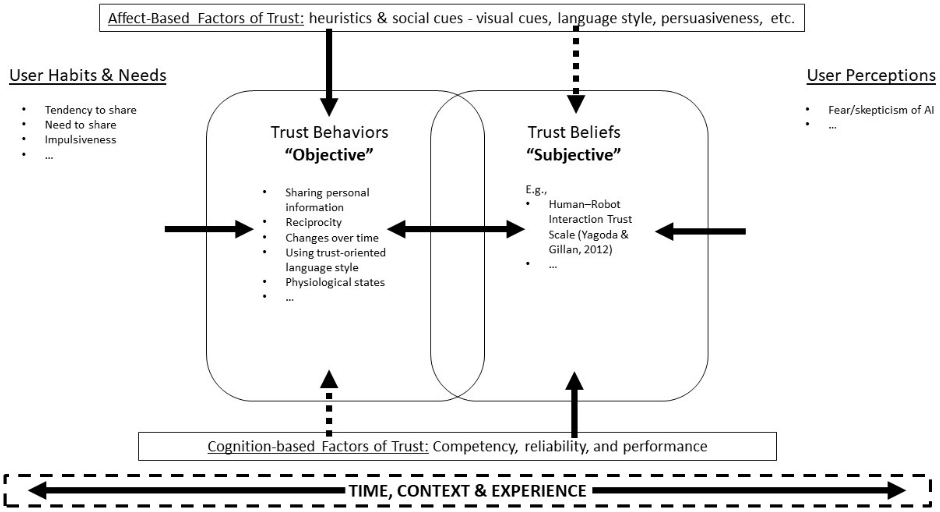
Affect-Based Factors of Trust: heuristics & social cues - visual cues, language style, persuasiveness, etc.
User Habits & Needs
Tendency to share
Need to share
Impulsiveness
...
Trust Behaviors
"Objective"
Sharing personal information
Reciprocity
Changes over time
Using trust-oriented language style
Physiological states
...
Trust Beliefs
"Subjective"
E.g.,
Human–Robot Interaction Trust Scale (Yagoda & Gillan, 2012)
...
User Perceptions
Fear/skepticism of AI
...
Cognition-based Factors of Trust: Competency, reliability, and performance
TIME, CONTEXT & EXPERIENCE

Figure 5 Framework of trusting beliefs and trusting behaviour.

References

1 Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI and Society, 35(3), 611 – 623.

2 Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. International Journal of Human-Computer Studies, 58(6), 737-758.

3 Ghazali, A. S., Ham, J., Barakova, E., & Markopoulos, P. (2019). Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance. Advanced Robotics, 33(7 – 8), 325 – 337.

4 Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2020). Evolving trust in robots: specification through sequential and comparative meta-analyses. Human factors, 0018720820922080.

5 Laban, G., George, J.-N., Morrison, V., & Cross, E. S. (2021a). Tell me more! Assessing interactions with social robots from speech. Paladyn, Journal of Behavioral Robotics, 12(1), 136 – 159.

6 Laban, G., Kappas, A., Morrison, V., & Cross, E. S. (2021b). Protocol for a Mediated Long-Term Experiment with a Social Robot. PsyArXiv.

7 Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. Journal of the Association for Information Systems, 16(10).

8 Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.

9 Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In Foundations of trusted autonomy (pp. 135-159). Springer, Cham.

10 Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In Trust in Human-Robot Interaction (pp. 3-25). Academic Press.

11 Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of management review, 20(3), 709-734.

12 McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. ACM Transactions on management information systems (TMIS), 2(2), 1-25.

13 Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23(3), 393-404.

14 Söllner, M., Hoffmann, A., & Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. European Journal of Information Systems, 25(3), 274-287.

15 Yagoda, R. E., & Gillan, D. J. (2012). You Want Me to Trust a ROBOT? The Development of a Human – Robot Interaction Trust Scale. International Journal of Social Robotics, 4(3), 235 – 248.

16 Zonca, J., Folsø, A., Sciutti, A. (2021). The role of reciprocity in human-robot social influence. iScience, 24(12), 103424.

4.7 Breakout Group 6: Interaction Design

Group 6

Contributors:
Group 6a (in person): Oliver Bendel, Sebastian Hobert, Ryan Schuetzler, Elisabeth André, Leigh Clark, Clayton Lewis, Stefan Schaffer, Eren Yildiz, Effie Law
Group 6b (online): Stefan Morana, Heloisa Candello, Christine Liebrecht, Zhou Yu, Dakuo Wang, Michelle Zhou, Ana Paula Chaves, Cosmin Munteanu, Soomin Kim

4.7.1 Goal and key questions

Goal: Identify interaction designs to strengthen trust in CA.
Key questions: How to design trusted conversational user interfaces?
Relevant aspects: UX – human-AI sociability; Human-in-the-loop; Evaluation – reliability/acceptance; Group interaction.

4.7.2 Key insights

4.7.2.1 Group 6a

The group started with reflecting on the following aspects of trust:

- Brand and UX: The producer of a chatbot affects our perception of trust; we trust certain products because we trust certain brands; the implication of UX-trust relation
- Group effect: If we trust people, and those people trust a chatbot, then we are more likely to trust it as well.
- Domain-dependency: Certain domains are more sensitive to trust fluctuation
- Modality: Intricate relations between modality, risk and trust

Next, the group focused on some specific aspects of conversational interactions. The multifaceted nature of trust and the numerous factors of people's interactions with CAs that can impact on perceptions and behaviours, complicating our understanding of how, why and when to design for trust and subsequently evaluate it. We present a discussion of some critical aspects of CA interactions and highlight the need for a holistic approach to creating trustworthy CAs.