Mechanisms of Common Ground in Human-Agent Interaction: A Systematic Review of Conversational Agent Research

Antonia Tolzin
University of Kassel, Germany
antonia.tolzin@uni-kassel.de

Andreas Janson
University St.Gallen, Switzerland
andreas.janson@unisg.ch

Abstract

Human-agent interaction is increasingly influencing our personal and work lives through the proliferation of conversational agents in various domains. As such, these agents combine intuitive natural language interactions by also delivering personalization through artificial intelligence capabilities. However, research on CAs as well as practical failures indicate that CA interaction oftentimes fails miserably. To reduce these failures, this paper introduces the concept of building common ground for more successful human-agent interactions. Based on a systematic review our analysis reveals five mechanisms for achieving common ground: (1) Embodiment, (2) Social Features, (3) Joint Action, (4) Knowledge Base, and (5) Mental Model of Conversational Agents. On this basis, we offer insights into grounding mechanisms and highlight the potentials when considering common ground in different human-agent interaction processes. Consequently, we secure further understanding and deeper insights of possible mechanisms of common ground in human-agent interaction in the future.

Keywords: common ground, conversational agent, human-agent interaction, systematic review

1. Introduction

Today, artificial intelligence (AI) is pervasive as it influences a lot of different areas in our private and working lives, for instance through AI-based conversational agents, which are applicated in various usage scenarios. AI mimics our human natural intelligence, in this context it can lead an intelligent conversation with a human counterpart (Elshan et al., 2022). Due to the COVID-19 pandemic, the use of intelligent assistants in business has significantly increased and Opus Research refers to 2021 as “The Year of the Ubiquitous Intelligent Assistants” (Miller, 2021). Software-based dialogue systems, also known as conversational agents (CAs) are frequently used assistants or moderators in our individual lives (Lieberman, 1997) but also proliferating into areas such as digital collaboration (Seeber et al., 2020). CAs are able to interact and communicate with other agents, like humans, to achieve goals by knowing their environment as well as memorizing the gained information and to improve their interaction skills through learning. Therefore, CAs can be seen as unique type of Information system (IS) entity, which are characterized by their intelligence and high level of interaction (Elshan et al., 2022). CAs are an increasing research field as traditional software-based systems can be enriched through CAs as an easy-to-use link between humans and computers. The perception of AI-based Systems and CAs is more and more shifting from tools to teammates (Seeber et al., 2020). Therefore, collaboration between humans and machines becomes important, indicating the need for investigating their conversational aspects.

Currently, CAs are employed in many different domains (Elshan et al., 2022) to provide direct interaction with users (Kim et al., 2019), foster engagement (Lundqvist et al., 2013), help users to reach their goals (Pérez et al., 2016) through natural conversational interaction (Cassell, 2000) and constant availability with immediate response times (Keyser et al., 2019). One central element of developing effective CAs is human-like design. Because of the perceived humanness and social presence of CAs, users react to these CAs as they would to a social actor, like another human (Nass & Moon, 2000). CAs typically show humanlike behavior and interact with users through natural language (Vassallo et al., 2010). CAs automatically and directly respond to users’ requests, acting independently without the support of a human counterpart (Spierling & Luderschmidt, 2018). They enable communication between humans and computers and aim to simulate human conversations (Bittner et al., 2019).

URI: https://hdl.handle.net/10125/102670
978-0-9981331-6-4

However, as communitive interactions are fragile, they can easily fail after misunderstandings during the dialogue (Benner et al., 2021) and impact the effectiveness between humans and computers (Luger & Sellen, 2016). CAs sometimes provide none, wrong, or incomprehensible responses, which leads to discomfort, annoyance, and questioning of the capabilities of the CA, which can end up in usage discontinuance and hinders the future success and spread of CAs (Chakrabarti & Luger, 2015; Weiler et al., 2021). Thus, improving the communication between humans and CAs is an important goal for researchers and practitioners (Meredith, 2017). We must keep in mind that dialogue understanding is an inherently interactive process. Understanding each other and the anticipated conversational grounding is a key element for language-based interactions between humans and computers (Rothwell et al., 2021). Despite the ambiguity of language, human dialogue partners understand each other remarkably well. This is because they establish common ground (H. H. Clark & Brennan, 1991). Nevertheless, identifying the needs and capabilities in the context of human-computer interaction (HCI) and developing presumptions about what the CA can do and understand is a great challenge for most people. On the other way around, it is also difficult for programmers and system designers to guess, how the human part of the dialogue will act (Koulouri et al., 2016).

There are several ambiguities concerning the definition and application of common ground, shared understanding, mutual understanding, or grounding in human-agent communication (Kontogiorgos, 2022). This might be due to the connection of two research fields (common ground from psychology and linguistics; and CAs from HCI) and leads to confusion about actual meaning and underlying mechanisms of common ground in human-agent interaction. To develop more effective CAs, it is crucial to understand, how two people can achieve common ground and how we can transfer these mechanisms in human-agent communication. Critically, to date, there is no systematic review examining how common ground is achieved in human-agent interaction. Therefore, we contribute to the literature of common ground by answering the following research question:

RQ: How is common ground in communication between a human and a conversational agent achieved?

We approach this question by conducting a structured literature review according to Webster and Watson (2002) and vom Brocke (2015) with the aim to discover common approaches, insights, and research foci. We will then sketch the most relevant findings, identify research gaps, and outline future research avenues.

2. Theoretical Background

2.1. Common Ground

Psycholinguistic research has examined how dialogue partners achieve mutual understanding and prevent misunderstandings and communicative breakdowns. Therefore, the social and collaborative aspects of human conversation is well established, for instance in the Interactive alignment theory (Pickering & Garrod, 2004) and the Communication accommodation theory (Giles & Powesland, 1997). To establish common ground, an interactive process to generate mutual understanding in conversations is needed (Koulouri et al., 2016). This process is fundamental for fruitful communication. Following, we define common ground as a shared understanding resulting from a coordination process between conversational partners (H. H. Clark, 1996; H. H. Clark & Brennan, 1991; Rothwell et al., 2021).

Successful grounding results in a shared context, guided comprehension, instant feedback of actions, and enhanced processes in conveying intent (Brennan, 1998). Dialogue partners form shared representations of what they are talking about, and then jointly modify these representations (Brennan, 1991). Through adding mutually understood contents, a continuously updated shared knowledge base is created and is available for subsequent use in the conversation (H. H. Clark & Wilkes-Gibbs, 1986). This joint process of adding and adapting contents of common ground from conversational turn to conversational turn has been formalized in the Contribution model (H. H. Clark & Schaefer, 1989). During the conversation, senders “check” that addressees understand what they are saying. In contrast, addressees provide different forms of evidence to let senders know about their understanding. For example, this could be implicit acknowledgements, like reacting to requests or active listening, or explicit feedback, like nodding, or saying “okay” or “Sorry…?” (H. H. Clark & Brennan, 1991; H. H. Clark & Schaefer, 1989). This verbal grounding does not bring new information or arguments to the conversation. It is more like a semantic mechanism to check that both dialogue partners received and understood the sender’s contribution. Responses can get interconnected and contingent on what has been said previously in the conversation by mutually grounding the conversational partner’s input (Sundar et al., 2010). This is a fundamental element of human communication and gives the sender a signal that the dialogue partner is actively listening (Ghose & Barua, 2013). It is important to differentiate between disagreement and miscommunication. Achieving common ground is not about dialogue partners agreeing with each other, but forming appropriate meta perspectives in relation to the conversational context (H. H. Clark & Brennan, 1991).

2.2. Conversational Agents

Common ground is not only a fundamental part of human-to-human interaction but also crucial in human-agent communication for successful interaction. Therefore, we need to establish common ground in human-agent interaction. This interaction can be realized through text, voice, or button applications. CAs include all software that allows people to have a conversation with a computer, for example chatbots, virtual agents, or artificial conversational entities. They can lead an intelligent conversation with a user via voice or textual methods (Knote et al., 2021). CAs have a long history, with memorable representatives like ELIZA, ALICE, Claude, and HeX. CAs are used for various reasons, including information retrieval or all areas of services, and in different contexts (Serban et al., 2017). They might include voice as an interaction channel (Schmitt et al., 2021), e.g., Amazon´s Alexa, and typically make use of natural language interfaces and machine learning techniques, which allow them to take on daily tasks more effectively, assisting the users (Budiu, 2018). In contrast, text-based CAs are rather based on a set of established rules or flow to react to queries posed by users. CAs can be technically distinguished into three types: messenger-like agents, voice-based agents, and embodied CAs (Hobert & Meyer von Wolff, 2019; Zierau et al., 2020). The first two types can be seen as interaction mode characteristics at a more mechanic dimension of CAs (Zierau et al., 2020). Messenger-like agents contain regular chat interfaces, known from Facebook Messenger or WhatsApp, whereas voice-based agents can recognize human speech and then respond with synthesized speech (Rothwell et al., 2021). Furthermore, embodied CAs include a "body", (e.g., avatars with full body or only face) representing a person in virtual environments (Nunamaker et al., 2011) and belong to the humanic personification dimension of CAs (Zierau et al., 2020). They are specifically conversational in all their behaviors and act human-like in dialogues.

3. Method

We performed an extensive search of eight major bibliographic databases, which were, AISeL, IEEE Xplore, ACM Digital Library, ProQuest, EBSCO, Science Direct, Wiley, and Springer, to include perspectives from the domains information systems, technical engineering, and communication. Queries were used, that combined the search terms related to the concepts of conversational agents and common ground (e.g., (chatbot OR "intelligent agent*" OR "intelligent assistant*" OR "intelligent personal assistant*" OR "virtual agent*" OR "virtual assistant*" OR "smart agent*" OR "smart assistant*" OR "conversational agent*" OR "conversational assistant*" OR "communicative agent*" OR "communicative AI") AND ("common ground" OR "shared understanding")). The database-specific search strings were semantically equivalent but formulated using the different syntaxes and technical support opportunities of the respective search engines. The search was conducted in May 2022 and includes all journal articles and conference paper in English language. We only searched within the title, abstracts, and keywords, except the databases Springer and ScienceDirect, where this was not possible. In these two databases, we conducted the search in the full-text and screened the title, abstracts, and keywords afterwards. There were no limits for the year of publication to ensure examining the maximum number of relevant articles. As the focus of this review is not on research outcomes, the search was not limited to empirical studies. Thus, we did not filter out non-empirical studies, theoretical or conceptual studies, or work in progress. Duplicates were removed and publications were initially screened based on the inclusion and exclusion criteria. The references were categorized as *matching*, *maybe matching*, and *not matching*. All papers that possibly meet inclusion criteria (*matching* or *maybe matching*) were retrieved as full text. Each step was conducted independently and afterwards matched to come to a joint result.

The database searches produced 20 records for title, abstract and keywords in six of the eight databases (18 after removal of duplicates) and additional 721 records (684 after removal of duplicates) for full text in the other two databases (Figure 1). After reviewing the abstracts, 35 full text versions of studies, 18 from the title, abstract and keyword screening in six of the databases and 17 from full text search in the other two databases, were

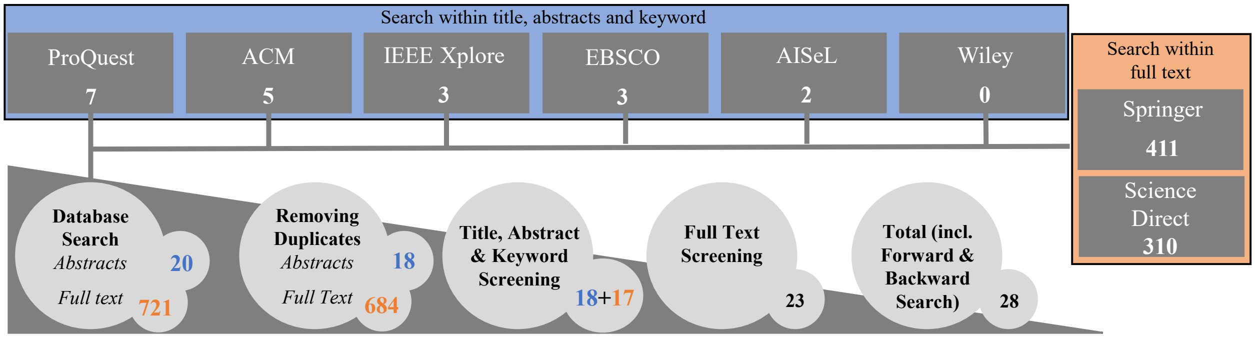

Figure 1. Literature search process

selected for further investigation. Finally, we excluded articles that do not explicitly fit within the scope of our literature review, applying three inclusion criteria: (1) the study must address the interaction, dialogue or communication between conversational agents and humans, (2) the study must focus on common ground theory or theories of shared understanding in communication, (3) the paper must be available in English. Articles examining communication and interaction more holistic (e.g., human-robot interaction, full conversational avatars in virtual reality environments) were also included, but only verbal and conversational aspects were investigated in this review. This led to a total inclusion of 23 studies. Forward and backward search with this set of relevant papers yielded five additional articles, resulting in a total of 28 papers for in-depth analysis. The final list of included papers can be found in the Appendix.

The papers included were critically reviewed by a formal narrative synthesis structured around the use of common ground in the interaction with CAs, and pre-defined criteria for data extraction were used (Popay et al., 2006). To derive possible mechanisms of common ground in human-agent interaction different aspects of common ground implementation were coded. We coded (1) in which context common ground was embedded (e.g., "Common ground and common interest are necessary components of engagement" (McKeown, 2015) was coded as *engagement*) and (2) the specific mechanism to reach common ground (e.g., "common ground was conceptualized as personalization, where information would be remembered by the agent to tailor their experience" (L. Clark et al., 2019) was coded as *personalization: remembering information*). In a second step, we put studies with similar contexts and specific mechanisms together and developed generic terms to describe these different groups of ways to achieve common ground. This results in five main mechanisms of common ground in human-agent interaction presented below.

4. Results

4.1. General Characteristics of Included Articles

The 28 included articles were all published over the preceding 20 years (2003-2022) with 22 (78.57%) papers published in the last ten years and seven (25%) papers published in 2021 or 2022. In general, an increasing interest in common ground theories in the application of CAs can be observed in the last few years. More than half of the papers examined had an empirical foundation (16, 57.14%), but there were also some theoretical (7, 25%) and some conceptual (5, 17,86%) articles. Looking more specific on the different characteristics of the investigated CAs, we see that most CAs are voice-based agents (11, 39.29%), some are text-based agents (7, 25%) and in few articles CAs have both, voice- and text-based elements (3, 10,71%). In four papers (14.29%) no specific information about the CA was given or there were only theoretical considerations about CAs in general. Because also articles examining the communication of humans and robots were included, three papers (10.71%) did not include a specific CA but investigated human-robot interaction and the communication therefore was voice-based. Next, the application context of CAs or robots was analyzed. The most common application context for CAs was a collaborative scenario or some kind of task performance (e.g., physical tasks or orientation in a virtual reality, cooking, schedule change) with eight articles (28.51%) included. Three CAs were used in a medical context (10.71%) and two CAs in education (7.14%). Two CAs were applicated in public spaces (7.14%), namely a guide in a museum providing background information and an orientation guide at an airport. Moreover, two CAs were used in the context of Aeronautics and spacecraft operations (7.14%). One CA each was applicated in the context of finance, product complaints, house inspection and marketing. Two papers (7.14%) contained more than one application context and tested or discussed the use of CAs in different domains. The last eight articles (28.57%) did not name an application context or examined the usage of CAs more general and only theoretical. Looking at the application of the CAs, we can identify various contexts, which on the one hand shows the broad application possibilities of CAs in all areas of working and private lives, but on the other hand demonstrates the importance to develop, test and evaluate suitable CAs for all specific application contexts with their individual requirements.

The theoretical foundation of common ground or shared understanding was broad, but most articles (18, 64.29%) were based on the common ground theory proposed by the psycholinguist Herbert H. Clark (1996) or on more than one theory including Clarks definition of common ground. Furthermore, articles were primary or secondary based on other theories, namely Wittgenstein (1967), Stalnaker (2002), Fusaroli et al. (2014), and Pickering and Garrod (2004). Only two (7.14%) papers considered shared understanding and not common ground, with one paper theoretical based on Dillenbourg (2008) and one paper without theoretical foundation. The remaining six (21.42%) articles had their own definition or no theoretical foundation of common ground. We see a dominance of the common ground theory of Clark, which is favorable as the ambiguous and heterogenous interpretations of common ground seem to have the same base and a widespread and general definition of common ground distinguished from everyday use is possible in future research. Lastly, it was differentiated whether common ground was directly or indirectly addressed and investigated in the included studies. Indirect investigations contain all articles where common ground was not the main focus auf research, only a small element of other primarily constructs or just a result of examinations not addressed beforehand. The exploration shows a well-balanced ratio of direct (15 paper, 53.57%) and indirect (13 paper, 46.43%) investigation of common ground.

4.2. Mechanisms of Common Ground

Based on the different implementation strategies for common ground and the contexts in which common ground elements were embedded, we extracted five different underlying mechanisms of common ground.

Embodiment. This mechanism describes the presentation of a personal assistant to the user, that is an identifiable conversational counterpart. People are more willing to put effort in establishing common ground if they interact with a CA under the same situational, social, and psychological conditions as a face-to-face interaction between two humans would take place (e.g., Corti & Gillespie, 2016, Pustejovsky & Krishnaswamy, 2021). This could be supported by a human-like body of the CA and presenting the CA as an autonomously communicating person. Through adding human-like characteristics, human-like treatment can be warranted, and authentic entity can be presented. This mechanism goes beyond the establishment of conversational common ground and contains non-verbal communication, aligning minds by being interesting, creative and humorous (see *Social Features* and *Mental Model of CA*).

A special case of common ground through embodiment is the virtual simulation environment, which refers to the context embeddedness of the interaction. Here, human and agent are situated within a virtual body in the same virtual environment and therefore share the same situation and perception. For instance, in a medical virtual 3D environment, where the user acts as a patient and the CA acts as a doctor, each represented with a full body avatar, they can refer to a visible object (e.g., stethoscope) in their environment because of shared situated references. This is how a shared perceptual and epistemic common ground could be created. In this context, voice plays an important role, because to reach a perfect human-like doctor-patient interaction, voice-based communication is crucial. Therefore, voice-based communication could foster embodiment and the human-like behavior.

Social Features. In human conversations different social features play a role and influence each other. Beside mutual understanding, facework, affective strategies, trust, humor and active listening, common ground is one of these important social features. To show authenticity, a CA needs to be transparent of its purpose, learn from experiences, show human-like behavior and coherence. Coherence, in this context, describes the social and context awareness and the capability to relate to common experiences and to establish common ground (e.g., Clark et al., 2019, Neururer et al., 2018). Moreover, these social features can create common ground and thereby elicit an appropriate mental model of the CA (see *Mental Model of CA*). Social features are especially important in task-orientated situations (see *Joint Action*) and in situations shared by human and CA. Through shared situations familiarity can be established. This could be realized through disclosure of personal information or talking about topics that already are in the common ground (e.g., weather, general political news).

Joint Action. Establishing and maintaining of common ground is a key element in human joint actions and therefore also in human-agent interaction. Joint action describes actions in which both conversational partners are involved, and they are aiming to achieve a shared goal and establish common ground. The sender must know that the recipient has the required context information in mind to interpret the utterance. A requirement for joint action is joint attention. For successful linguistic communication, joint attention to senders and recipient's intentions is crucial. It is a collective task, and therefore can only be achieved with other conversational partners, to represent the situation in a way that makes conversational common ground salient. But is still difficult for the CA to account for the intentions and motivations of a sender through joint attention (e.g., Bernard & Arnold, 2019, Pinhanez et al., 2018).

Common ground is what both conversational partners know in a transparent way. For example, in a problem-solving process, common ground can be achieved by sending relevant information to the conversational partner, verifying what each partner knows, establishing or negotiating shared meaning, requesting information or repair insufficiencies in shared knowledge. If common ground is established, coordination costs decrease as the source of information to foster the coordination of actions is unambiguous.

Knowledge Base. This mechanism is specifically relevant for task-oriented dialogue systems, as the CA has access beforehand to the whole knowledge about the task. The CA depends on different sources of knowledge, namely three: the context, the information evolving during the interaction and the beliefs of the dialogues stakeholders (Blache, 2017). In the second source common ground plays a crucial role and describes the initial user knowledge combined with the information instantiated during the dialogue. Important for the common ground in this context is, that all conversational partners suppose the others have access to the same information and knowledge. New information can be added in relation to information already existing in the common ground, which organize the knowledge base in a specific way. To achieve common ground, a CA needs to adapt to the user's level of knowledge and the level of common ground between user and CA (e.g., Blache, 2017).

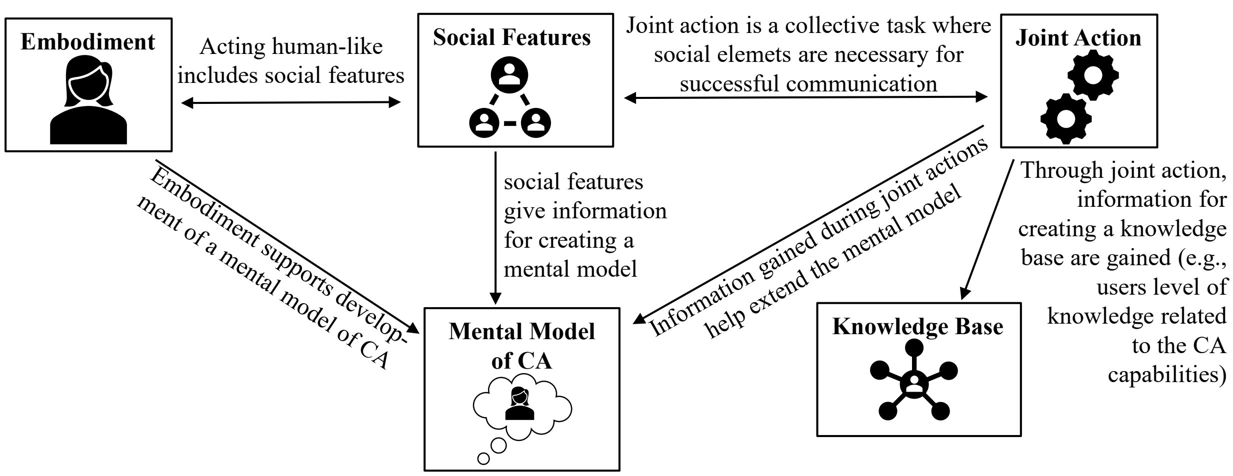

Figure 2. Relationships of the five Mechanisms to achieve Common Ground

Mental Model of Conversational Agent. People try to understand the intent and meaning behind what the conversational partner is saying beyond. This is important for exploring common interests, goals, given information beforehand and perception of the environment. During the conversation, people create a mental model of the CA. This mental model then influences their expectations of the CA and whether the CA is likely to interact effectively with them and therefore if it is worth it to put effort in establishing common ground (e.g., Frijns et al., 2021, Kiesler, 2005). The goal is that users develop an appropriate mental model of the CAs' abilities and intentions. Specifically, the CA should prompt users to make an appropriate estimate of the CAs role and what the CA knows. This mental model contains any anthropomorphism (see *Embodiment*) that has occurred, and it leads to expectations of the behavior of the system.

5. Discussion and Implications

Common ground is a key element of human dialogue and a necessary capability for CAs that use language. The purpose of this systematic review was to investigate how common ground in communication between humans and CAs is achieved. We analyzed extant research and synthesized five mechanisms of common ground in human-agent interaction. Their *relationships* and important implications for *future avenues (FA)* of CA research are discussed below.

The five mechanisms to achieve common ground are related to each other (Figure 2), especially *Mental Model of the CA* seems to be a key element as *Embodiment, Social Features* and *Joint Actions* foster the development of a mental model. Moreover, the mechanisms *Embodiment* and *Social Features* keep up human-like behavior of CAs, which makes it easier for users to find a common ground with the CA. For *Joint Actions* social elements are needed, which are important in task-orientated human-agent interactions. The mechanism *Knowledge Base* differs from the other mechanisms as it does not directly refer to the interaction, but describes how information gained during the interaction are organized within the CA.

In summary, an increasing interest in common ground in human-agent interaction can be observed in the last years. This seems to be a promising trend as we can see a shifting from the perception of conversational agents from tools to teammates. Therefore human-like conversations are gaining importance (Seeber et al., 2020). This can also be drawn from the fact, that the investigated studies have been carried out by researchers from a broad range of different disciplines (e.g., informatics, philosophy, psychology, linguistic, information systems, robotics). We observe that research streams become increasingly diverse. This is mainly because of the application contexts of CAs are extremely different which impedes the generalizability of results. The appropriate implementation and the underlying mechanisms of common ground could be very different in the specific domains. Therefore, future research could connect fundamental communication elements and CA research with interdisciplinary research approaches to expand the understanding of human-agent interaction *(FA #1)*. Hence, it is important to explicitly address the mechanisms of common ground and future research and focus on domain specific requirements for CA design.

Results showed, that about half of the reviewed papers directly addressed common ground and it is necessary to increase this part and focus on grounding effects, supporting, or restraining variables and underlying mechanisms of common ground in human-agent interaction. For instance, customer service as a domain is oftentimes characterized through short and one-time interactions. Thus, it is necessary to build common ground very quickly, e.g., through building consensus about the CA mental model or organizing knowledge. Moreover, as common ground is oftentimes studied in educational research, CAs in digital learning environments could provide through building common ground a more productive scaffolding for learning processes to ultimately improve learning outcomes (Winkler et al., 2020). Thus, we call on research that explicitly investigates common ground mechanisms in isolation and in combination with other boundary conditions *(FA #2).*

Furthermore, it is important to pay attention to the whole breadth of the human-machine interaction when it comes to terms of common ground *(FA #3)*. As the mechanisms *Embodiment, Social Features*, and *Mental Model of CA* showed, a holistic approach and investigation of the interaction of humans and agents is needed, including not only verbal and conversational elements, but also the presentation and social competencies of the CA, the environment, and the purpose of interaction. By widening our perspective of common ground in human-agent interaction, it is important to also consider human-robotic-interaction, as this is the most extreme implementation of embodiment. To sum up, the five mechanisms can be seen as appeal to investigate the mechanisms of common ground systematically and empirically in future research regarding interdisciplinary research approaches, boundary conditions, and the whole human-machine interaction.

However, the results should be interpreted with caution due to some limitations. First, due to broad search strategy, the systematic review tried to cover the full spectrum of application domains and foundation of the papers, rather than concentrating on a specific application context or specific empirically outcomes. But even if the search strategy was quite broad, search term specification may lack other relevant terms not considered. Furthermore, the literature search did only comprise searching scientific databases. Second, due to our broad search strategy and our chosen type of review, we did not assess the methodological and overall quality of included articles resulting in a less differentiated synthesis of our findings. Third, this paper builds on a cross-disciplinary, literature-based definition of common ground. Using a different understanding of common ground and shared understanding might lead to a different set of papers and potentially different results.

6. Conclusion

As understanding each other and the anticipated conversational grounding is essential for conversations between humans and CAs, it is fundamental to understand the underlying mechanisms of common ground in CA research. In this systematic review, we provide an initial understanding on common ground in human-agent interaction and identified five underlying mechanisms for achieving grounding for CA interaction processes. The mechanisms are related to each other

(Figure 2) but foster common ground in human-agent interaction in distinct ways. It is important to keep in mind that humans and agents can only have reduced common ground as compared to the common ground shared by humans. But to approach the goal of the best possible grounding in human-agent interaction it is crucial to put research effort in understanding the mechanisms of common ground, extracting boundary conditions and derive design principles for CAs based on common ground for successful human-agent communication and less conversational breakdowns.

Our work provides a theory of analysis and is a contribution to being able to better understand possible mechanisms of common ground in human-agent interaction in the future (Gregor, 2006). Overall, the results provide deeper insight into the different ways how common ground could be implemented in human-agent interaction and the possible underlying mechanisms. Thus, offering a broad spectrum for research in behavioral and design-oriented research.

7. Acknowledgements

The results presented in this article were developed in the research project Komp-HI funded by the German Federal Ministry of Education and Research (BMBF, grant 16DHBKI073). We thank the BMBF for supporting our research. The second author acknowledges funding from the Basic Research Fund of the University of St.Gallen.

8. References

Benner, D., Elshan, E., Schöbel, S., & Janson, A (2021). What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents. *International Conference on Information Systems*.

Bittner, E. A. C., Oeste-Reiß, S., & Leimeister, J. M (2019). Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. *HICSS*.

Blache, P. (2017). Dialogue management in task-oriented dialogue systems. In T. Chaminade & A. S. I. G. o. C.-H. Interaction (Eds.), *Proceedings of the 1st ACM SIGCHI*.

Brennan, S. E. (1991). Conversation with and through computers. *User Modeling and User-Adapted Interaction*(1), 67–86.

Brennan, S. E. (1998). The grounding problem in conversations with and through computers. *Social and Cognitive Approaches to Interpersonal Communication. Susan R. Fussell and Roger J. Kreuz (Eds.)*, 201–225.

Budiu, R. (2018). *The User Experience of Chatbots*. https://www.nngroup.com/articles/chatbots/

Cassell, J. (2000). Embodied conversational interface agents. *Communications ACM*, *43*(4), 70–78.

Chakrabarti, C., & Luger, G. F. (2015). Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics. *Expert Systems with Applications 42*(20), 6878–6897.

Clark, H. H. (1996). *Using language* (Sixth printing). Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). APA.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to Discourse. *Cognitive Science*, *13*(2), 259–294.

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). What Makes a Good Conversation? In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *CHI* (pp. 1–12).

Dillenbourg, P. (2008). *Collaborative learning: Cognitive and computational approaches*. *Advances in learning and instruction series*. Emerald Group Pub.

Elshan, E., Zierau, N., Engel, C., Janson, A., & Leimeister, J. M. (2022). Understanding the Design Elements Affecting User Acceptance of Intelligent Agents: Past, Present and Future. *Information Systems Frontiers : A Journal of Research and Innovation*, *24*(3), 699–730.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157.

Ghose, S., & Barua, J. J. (2013). Toward the Implementation of a Topic Specific Dialogue Based Natural Language Chatbot as an Undergraduate Advisor. *ICIEV 2013*, 1–5.

Giles, H., & Powesland, P. (1997). Accommodation Theory. In N. Coupland & A. Jaworski (Eds.), *Modern linguistics series. Sociolinguistics: A reader and coursebook* (pp. 232–239). Macmillan.

Gregor (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611.

Hobert, S., & Meyer von Wolff, R. (2019). Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents. *Wirtschaftsinformatik*.

Keyser, A. de, Köcher, S., Alkire, L., Verbeeck, C., & Kandampully, J. (2019). Frontline Service Technology infusion: conceptual archetypes and future research directions. *Journal of Service Management*, *30*(1), 156–183.

Kim, S., Lee, J., & Gweon, G. (2019). Comparing Data from Chatbot and Web Surveys, 1–12.

Knote, R., Janson, A, Söllner, M., & Leimeister, J. M (2021). Value Co-Creation in Smart Services: A Functional Affordances Perspective on Smart Personal Assistants. *Journal of the Association for Information Systems*, *22*(2), 418–458.

Kontogiorgos, D. (2022). *Mutual Understanding in Situated Interactions with Conversational User Interfaces: Theory, Studies, and Computation.*

Koulouri, T., Lauria, S., & Macredie, R. D. (2016). Do (and Say) as I Say: Linguistic Adaptation in Human–Computer Dialogs. *HCI*, *31*(1), 59–95.

Lieberman, H. (1997). Autonomous interface agents. In R. Bellamy & S. Pemberton (Eds.), *Chi 97: Conference on Human Factors in Computing* (pp. 67–74).

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA". In J. Kaye (Ed.), *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297).

Lundqvist, K. O., Pursey, G., & Williams, S. (2013). Design and Implementation of Conversational Agents for Harvesting Feedback in eLearning Systems. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, D. Hernández-Leo, T. Ley, R. Klamma, & A. Harrer (Eds.), *Lecture Notes in Computer Science. Scaling up Learning for Sustained Impact* (Vol. 8095, pp. 617–618). Springer Berlin.

McKeown, G. (2015). Turing's menagerie: Talking lions, virtual bats, electric sheep and analogical peacocks: Common ground and common interest are necessary components of engagement. In *ACII* (pp. 950–955).

Meredith, J. (2017). Analysing technological affordances of online interactions using conversation analysis. *Journal of Pragmatics*(115), 42–55.

Miller, D. (2021). *Welcome to 2021: The Year of Ubiquitous Intelligent Assistants.*

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied Conversational Agent-Based Kiosk for Automated Interviewing. *JMIS*, *28*(1), 17–48.

Pérez, J., Cerezo, E., Serón, F. J., & Rodríguez, L.-F. (2016). A cognitive-affective architecture for ECAs. *Biologically Inspired Cognitive Architectures*, *18*, 33–40.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, *27*(2), 169-90; discussion 190-226.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme*. Lancaster University.

Rothwell, C. D., Shalin, V. L., & Romigh, G. D. (2021). Comparison of Common Ground Models for Human--Computer Dialogue. *ACM*, *28*(2), 1–35.

Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2021). Voice as a Contemporary Frontier of Interaction Design. *ECIS 2021 Research Papers*.

Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2).

Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Pieper, M., Chandar, S., Ke, N. R., Rajeshwar, S., Brebisson, A. de, Sotelo, J. M. R., Suhubdy, D., Michalski, V., Nguyen, A., Pineau, J., & Bengio, Y. (2017). *A Deep Reinforcement Learning Chatbot*.

Spierling, U., & Luderschmidt, J. (2018). Chatbots und mediengestützte Konversation. In Kochhan (Ed.), *Media Management* (pp. 387–408). Springer Wiesbaden.

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, *25*(5/6), 701–721.

Sundar, S. S., Qian, X., & Bellur, S. (2010). Designing Interactivity in Media Interfaces: A Communications Perspective. *CHI,* 2247–2256.

Vassallo, G., Pilato, G., Augello, A., & Gaglio, S. (2010). Phase Coherence in Conceptual Spaces for Conversational Agents. In P. C.-Y. Sheu (Ed.), *Semantic computing* (pp. 357–371).

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. *CAIS*, *37*.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*(26).

Weiler, S., Matt, C., & Hess, T. (2021). Immunizing with information – Inoculation messages against conversational agents' response failures. *Electronic Markets.* Advance online publication.

Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. In R. Bernhaupt, F. '. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Proceedings of the 2020 CHI* (pp. 1–14). ACM.

Wittgenstein, L. (1967). *Philosophical investigations*. Blackwell.

Zierau, N., Wambsganss, T., Janson, A., & Schöbel, S. & Leimeister, J. M. (2020). The Anatomy of User Experience with Conversational Agents: A Taxonomy and Propositions of Service Clues. *International Conference on Information Systems*.

9. Appendix: List of Reviewed Paper

Bernard, D., & Arnold, A. (2019). Cognitive interaction with virtual assistants: From philosophical foundations to illustrative examples in aeronautics. *Computers in Industry*, *107*, 33–49.

Blache, P. (2017). Dialogue management in task-oriented dialogue systems. In T. Chaminade & A. S. I. G. o. C.-H. Interaction (Eds.), *Proceedings of the 1st ACM SIGCHI* (pp. 4–8).

Buschmeier, H., & Kopp, S. (2013). Co-constructing Grounded Symbols—Feedback and Incremental Adaptation in Human–Agent Dialogue. *KI - Künstliche Intelligenz*, *27*(2), 137–143.

Cassell, J. (2003). Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Modeling and User-Adapted Interaction*, *13*(1/2), 89–132.

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E, Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). What Makes a Good Conversation? In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI* (pp. 1–12).

Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, *58*, 431–442.

Frijns, H. A., Schürer, O., & Koeszegi, S. T. (2021). Communication Models in Human–Robot Interaction: An Asymmetric MODel of ALterity in Human–Robot Interaction (AMODAL-HRI). *International Journal of Social Robotics.*

Hanna, N.: Richards, D. (2014). Measuring the effect of personality on human-IVA shared understanding. In A. Bazzan (Ed.),. *Aamas '14 Proceedings* (pp. 1643–1644).

Hayashi, Y. (2020). Gaze awareness and metacognitive suggestions by a pedagogical conversational agent: an experimental investigation on interventions to support collaborative learning process and performance. *International Journal of Computer-Supported Collaborative Learning*, *15*(4), 469–498.

Jiang, H., Cheng, Y., Yang, J., & Gao, S. (2022). AI-powered chatbot communication with customers: Dialogic interactions, satisfaction, engagement, and customer behavior. *Computers in Human Behavior*, *134*, 107329.

Kiesler, S. (2005). Fostering common ground in human-robot interaction. In *Robot and Human Interactive Communication* (pp. 729–734).

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). Classifying Smart Personal Assistants: An Empirical Cluster Analysis. In T. Bui (Ed.), *HICCS Proceedings 2019.*

Kontogiorgos, D., Pereira, A., & Gustafson, J. (2021). Grounding behaviours with conversational interfaces: effects of embodiment and failures. *Journal on Multimodal User Interfaces*, *15*(2), 239–254.

Loisel, A., Chaignaud, N., & Kotowicz, J.-P. (2007). Designing a Human-Computer Dialog System for Medical Information Search. In *International Conferences on Web Intelligence and Intelligent Agent Technology* (pp. 350–353).

Lorini, E. (2020). Rethinking epistemic logic with belief bases. *Artificial Intelligence*, *282*, 103233.

McKeown, G. (2015). Turing's menagerie: Talking lions, virtual bats, electric sheep and analogical peacocks: Common ground and common interest are necessary components of engagement. In *2015 ACII* (pp. 950–955).

Montemayor, C. (2021). Language and Intelligence. *Minds and Machines*, *31*(4), 471–486.

Müller, R., Paul, D., & Li, Y. (2021). Reformulation of symptom descriptions in dialogue systems for fault diagnosis: How to ask for clarification? *Journal of Human-Computer Studies*, *145*, 102516.

Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on Authenticity in Chat Bots. *Multimodal Technologies and Interaction*, *2*(3), 60.

Pfau, J., Miller, T., & Sonenberg, L. (2014). Modelling and Using Common Ground in Human-agent Collaboration during Spacecraft Operations. In *SpaceOps Conference.*

Pinhanez, C. S., Candello, H., Pichiliani, M. C., Vasconcelos, M., Guerra, M., Bayser, M. G. de, & Cavalin, P. (2018). *Different but Equal: Comparing User Collaboration with Digital Personal Assistants vs. Teams of Expert Agents*.

Pustejovsky, J., & Krishnaswamy, N. (2021). Embodied Human Computer Interaction. *KI*, *35*(3-4), 307–327.

Rosen, Y. (2014). Comparability of Conflict Opportunities in Human-to-Human and Human-to-Agent Online Collaborative Problem Solving. *Technology, Knowledge and Learning*, *19*(1-2), 147–164.

Rothwell, C. D., Shalin, V. L., & Romigh, G. D. (2021). Comparison of Common Ground Models for Human--Computer Dialogue. *ACM Transactions on Computer-Human Interaction*, *28*(2), 1–35.

Tapie, J., Terrier, P., Perron, L., & Cellier, J.-M. (2006). Should remote collaborators be represented by avatars? A matter of common ground for collective medical decision-making. *AI & SOCIETY*, *20*(3), 331–350.

Vinciarelli, A., Esposito, A., André, E., Bonin, F., Chetouani, M., Cohn, J. F., Cristani, M., Fuhrmann, F., Gilmartin, E, Hammal, Z., Heylen, D., Kaiser, R., Koutsombogera, M., Potamianos, A., Renals, S., Riccardi, G., & Salah, A. A. (2015). Open Challenges in Modelling, Analysis and Synthesis of Human Behaviour in Human–Human and Human–Machine Interactions. *Cognitive Computation*, *7*(4), 397–413.

Wang, H. S., & Yim, C. K. (2019). Effects of dominance transitions on advice adherence in professional service conversations. *Journal of the Academy of Marketing Science*, *47*(5), 919–938.

Wrede, B., Kopp, S., Rohlfing, K., Lohse, M., & Muhl, C. (2010). Appropriate feedback in asymmetric interactions. *Journal of Pragmatics*, *42*(9), 2369–2384.