International Journal of Human-Computer Studies 132 (2019) 138–161

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

A Taxonomy of Social Cues for Conversational Agents

Jasper Feine*, Ulrich Gnewuch, Stefan Morana, Alexander Maedche

Institute of Information Systems and Marketing (IISM), Karlsruhe Institute of Technology (KIT), Kaiserstraße 89, 76133 Karlsruhe, Germany

ARTICLE INFO

Keywords:
Conversational agent
Chatbot
Social cue
Computers are social actors
Taxonomy
Classification
Literature review

ABSTRACT

Conversational agents (CAs) are software-based systems designed to interact with humans using natural language and have attracted considerable research interest in recent years. Following the Computers Are Social Actors paradigm, many studies have shown that humans react socially to CAs when they display social cues such as small talk, gender, age, gestures, or facial expressions. However, research on social cues for CAs is scattered across different fields, often using their specific terminology, which makes it challenging to identify, classify, and accumulate existing knowledge. To address this problem, we conducted a systematic literature review to identify an initial set of social cues of CAs from existing research. Building on classifications from interpersonal communication theory, we developed a taxonomy that classifies the identified social cues into four major categories (i.e., verbal, visual, auditory, invisible) and ten subcategories. Subsequently, we evaluated the mapping between the identified social cues and the categories using a card sorting approach in order to verify that the taxonomy is natural, simple, and parsimonious. Finally, we demonstrate the usefulness of the taxonomy by classifying a broader and more generic set of social cues of CAs from existing research and practice. Our main contribution is a comprehensive taxonomy of social cues for CAs. For researchers, the taxonomy helps to systematically classify research about social cues into one of the taxonomy's categories and corresponding subcategories. Therefore, it builds a bridge between different research fields and provides a starting point for interdisciplinary research and knowledge accumulation. For practitioners, the taxonomy provides a systematic overview of relevant categories of social cues in order to identify, implement, and test their effects in the design of a CA.

1. Introduction

Conversational agents (CAs) are software-based systems designed to interact with humans using natural language (Dale, 2016; McTear et al., 2016). They are currently attracting much attention and are considered to have great potential in many application domains such as retail, healthcare, and education (Følstad and Brandtzæg, 2017; Gartner, 2017). Recent technological advances in artificial intelligence have led to great interest by organizations in using CAs to support users in finding relevant information about products and services as well as performing routine tasks (Gartner, 2018; Larivière et al., 2017; Maedche et al., 2019). Today, text-based CAs or chatbots are increasingly being implemented on messaging platforms and websites (Araujo, 2018). For example, over 100,000 chatbots have been created in less than one year on Facebook Messenger (Johnson, 2017). Also, voice-based CAs can be found on PCs and many mobile phones (e.g., Apple's Siri, Microsoft's Cortana), and other types of physical devices (e.g., Google's HomePod, Amazon's Echo Dot) (Maedche et al., 2016). Furthermore, considerable research has been devoted to developing lifelike 3D animated and embodied CAs (ECA) that can interact with humans through realistic socio-emotional multimodal behaviors (Cassell, 2000a; Pelachaud, 2017). They are successfully used in several domains such as health care and education (Bickmore and Gruber, 2010; Zhang et al., 2017).

The overall idea of interacting with computers using natural language dates back to the 1960s (McTear et al., 2016). Since the first text-based CAs, such as ELIZA (Weizenbaum, 1966), were developed, much research has been conducted on CAs in the fields of computer science (CS) information systems (IS), and human-computer interaction (HCI). Over the years, researchers from these fields have used various names for this class of systems (e.g., CA, ECA, chatbot, virtual assistant, digital assistant), making it difficult to compare and interpret the results of their studies (Dale, 2016; McTear, 2017). Nevertheless, there is a consensus among researchers that the design and evaluation of CAs need to consider both their technical and social aspects (Araujo, 2018; Bickmore and Cassell, 2005; Go and Sundar, 2019; Louwerse et al., 2005; Pelachaud, 2017). Since CAs enable users to interact with computers using natural language (i.e., a central human quality) and are capable of sensing and expressing several multimodal verbal and nonverbal characteristics usually

* Corresponding author.
E-mail addresses: jasper.feine@kit.edu (J. Feine), ulrich.gnewuch@kit.edu (U. Gnewuch), stefan.morana@kit.edu (S. Morana), alexander.maedche@kit.edu (A. Maedche).

https://doi.org/10.1016/j.ijhcs.2019.07.009
Received 6 September 2018; Received in revised form 27 June 2019; Accepted 27 July 2019
Available online 29 July 2019

associated with humans (e.g., joke, gender, gestures, facial expressions, response delay), users often react socially to them (e.g., Go and Sundar, 2019; Krämer, 2008b; Louwerse et al., 2005; Niewiadomski and Pelachaud, 2010). While these characteristics have also been given various names in different research streams and fields (e.g., social cues, anthropomorphic features, human-like characteristics, human-like behavior, multimodal behavior), many studies have shown that they have a significant impact on how users perceive and interact with CAs (e.g., Araujo, 2018; Bickmore and Picard, 2005; Gnewuch et al., 2018a; Li et al., 2017). To explain this phenomenon, most studies build on the Computers are Social Actors (CASA) paradigm (Nass et al., 1994; Nass and Moon, 2000), which states that humans interacting with computers exhibit social reactions that are similar to those observed in interpersonal communication. More specifically, humans tend to react subconsciously to social cues of computers, no matter how rudimentary these cues are (Nass et al., 1994; Nass and Moon, 2000). Therefore, social cues can positively influence various CA-related outcomes such as perceptions of a CA's social presence (Araujo, 2018; Puetten et al., 2010), trust in a CA (Visser et al., 2016), or user satisfaction with a CA (Verhagen et al., 2014). Moreover, social cues determine the credibility of a CA (Demeure et al., 2011), the believability of a CA (Carolis et al., 2004; Demeure et al., 2011; Pelachaud and Bilvi, 2003), and the success of a long-term relationship between a CA and a human (Bickmore and Picard, 2005). However, social cues have also been associated with adverse effects (Brandtzaeg and Følstad, 2018; Fogg, 2002; Ghazali et al., 2018; Wallis and Norling, 2005), which may impede the adoption and use of CAs (Mimoun et al., 2012). Consequently, it is essential for researchers and practitioners to have a comprehensive understanding of the different types of social cues of CAs (Fogg, 2002; Nass and Moon, 2000).

Existing research on social cues of CAs is scattered across different fields of research, each using its specific terminology in order to reflect the increasing specialization of its scientific discipline (Pantic et al., 2011). This hinders the interdisciplinary exchange of researchers and makes it difficult to classify and accumulate knowledge from their own as well as related fields. Therefore, a shared understanding of social cues of CAs can support researchers as well as practitioners to understand and extend the existing body of knowledge. However, to the best of our knowledge, no study aims to derive a comprehensive classification that seeks to integrate existing research on social cues of CAs. Efforts have been made to classify social cues of computers (Fogg, 2002), nonverbal cues of ECAs (Cassell et al., 1994; Cowell and Stanney, 2005), social cues of service agents (Wuenderlich and Paluch, 2017), and social cues of physical robots (e.g., Fiore et al., 2013; Hegel et al., 2011; Wiltshire et al., 2014). Moreover, computational models have been proposed that describe how ECAs can express several social cues using a multimodal realization (e.g., verbal cues combined with gestures and facial expressions in a specific temporal arrangement) that interact with each other in order to convey a specific meaning (Bevacqua et al., 2010; Carolis et al., 2004; Pelachaud, 2009a, 2005).

While these classifications and models each provide valuable knowledge on social cues and their multimodal realization for a specific domain, a comprehensive classification of social cues of CAs combining research findings from several domains is lacking. Currently, researchers and practitioners, such as CA designers, are "*in danger of re-inventing the wheel*" (p. 46) by neglecting or being unaware of the rich body of scientific work on social cues of CAs of the past decades (McTear, 2017). To structure and organize a large body of knowledge, researchers often use taxonomies to classify objects based on their similarity (Nickerson et al., 2013). Taxonomies can not only bring order to complex research domains, but also offer guidance for researchers and practitioners (Nickerson et al., 2013). Hence, we address the following research question:

How to build a taxonomy of social cues for conversational agents?

To address the research question, we conduct a systematic literature review (SLR) to identify an initial set of social cues of CAs, classify them in a taxonomy, evaluate the taxonomy using a card sorting approach, and finally apply the taxonomy to investigate additional social cues of CAs from research and practice. This article contributes by providing a comprehensive taxonomy of social cues of CAs that comprises four categories (i.e., verbal, visual, auditory, and invisible cues) based on Leathers' (1976) classification of the human communication systems with ten additional subcategories based on other well established classifications (Burgoon et al., 2010; Leathers, 1976; Trenholm and Jensen, 2011). The taxonomy extends existing classifications of social cues of CAs by integrating the four communication systems responsible for creating and transmitting messages in interpersonal communication (Leathers, 1976) into a representation that applies to CAs. Our application of the taxonomy to existing research beyond our initial set of identified social cues as well as to three real-world examples (i.e., text-based CA, voice-based CA, ECA) demonstrates its usefulness in classifying social cues of different types of CAs. Consequently, researchers can apply the taxonomy to systematically classify existing and future research about social cue phenomena into one of the taxonomy's categories and corresponding subcategories. Practitioners can use the systematic overview of relevant categories of social cues in order to identify, implement, and test the effects of social cues in the design of their CAs.

The remainder of this article is organized as follows. First, we introduce related work on CAs, define social cues of CAs, and review existing classifications. Next, we describe our three-step research methodology. Subsequently, we outline the results of the literature review and present the development of the taxonomy of social cues for CAs. Finally, we discuss the results by showcasing the usefulness of the taxonomy for research and practice and outline the limitations and potential avenues for future research.

2. Related work and theoretical foundations

2.1. Conversational agents

The first CA, ELIZA, was developed in 1966 by Joseph Weizenbaum as a computer program that "*makes natural language conversation with a computer possible*" (Weizenbaum, 1966, p. 36). In the 1980s, this was followed by the appearance of voice-based dialog systems, voice user interfaces, ECAs, and social robots (McTear et al., 2016). Despite the large number of different terms used to describe this technology (e.g., CA, ECA, chatbot, dialog systems, companions, virtual assistant, digital assistant), all CAs build on the idea of communicating via natural language (Dale, 2016). In order to cover several different types of systems, we consider a CA as a software-based system designed to interact with humans using natural language (Dale, 2016; McTear et al., 2016). This means that the user and CA interact in a voice or text-based conversation without using restricted command phrases or a predefined set of keywords (McTear, 2017). Although most CAs build on similar technology (i.e., natural language processing), they differ considerably in their design and application purposes (Dale, 2016).

Text-based CAs (i.e., chatbots) are often implemented on websites and messenger platforms (e.g., Facebook Messenger, WeChat) in order to provide customer service (Brandtzaeg and Følstad, 2018; Feine et al., 2019a; Gartner, 2018). In addition, chatbots can be implemented in various other domains such as for tutoring (Kerly et al., 2007), to provide energy feedback (Gnewuch et al., 2018b), to provide library information (Allison, 2012), or to support collaboration at the workplace (Frommert et al., 2018).

Research on voice-based CAs, which are often referred to as spoken-dialog systems, voice-user interfaces, or interactive voice response systems, began in the late 1980s (McTear et al., 2016). One of the most prominent spoken-dialog system projects were ATIS (Air Travel Information Service) in the USA, SUNDIAL in Europe, and HMIHY (How May I Help You) from AT&T (Gorin et al., 1997; McTear et al., 2016). These voice-user interfaces were introduced in the 1990s in order to automate self-service tasks and call routing (McTear et al., 2016). Nowadays, voice-based CAs are dominated by major technology companies and often take on the role of personal assistants on devices such as smartphones (e.g., Apple's Siri, Google's Assistant, Samsung's Bixby), smart speakers (e.g., Amazon's Alexa), and PCs (e.g., Microsoft Cortana) (McTear, 2017).

In addition, research has deeply investigated ECAs that use verbal and nonverbal communication to realize realistic human-like conversational behavior, express social competence, and impact the user's decision making and situation awareness (Cassell, 2000a, 2000b). ECAs typically have a visual representation (e.g., a 3D avatar) and can display various multimodal verbal and nonverbal behavior such as believable human-like movements, mimicry, gaze behavior, spoken intonation, and facial expressions (Carolis et al., 2004; Cassell et al., 2000; Pelachaud, 2017, 2009b). Moreover, many ECAs are capable of detecting and interpreting communicative signals from their human interlocutors (Pelachaud, 2009a). Thus, they can communicate in a realistic, human-like, and socially aware manner. Research has shown that ECAs can serve as a tourist information point (Garrido et al., 2017), relational clinical agent (Bickmore and Gruber, 2010), as an automatic interviewing kiosk (Nunamaker et al., 2011), or even as a personal assistant for conferences attendees (Cassell, 2019).

In recent years, the technical capabilities of CAs have increased considerably (McTear et al., 2016) and many CAs have been introduced into the market (Dale, 2016; Klopfenstein et al., 2017). However, many researchers argue that CAs need more than just sophisticated technical capabilities to succeed (Wallis and Norling, 2005). CAs must act socially (Fogg, 2002; Go and Sundar, 2019; Shechtman and Horowitz, 2003; Wallis and Norling, 2005) and should display authentic and expressive behaviors (Carolis et al., 2004; Pelachaud, 2009b). However, researchers indicate a lack of design knowledge across different fields in order to design a successful CA from a social point of view (McTear et al., 2016; Reeves, 2017). Besides high-level suggestions and domain-specific design advice, there are no general design guidelines for social CAs (McTear, 2017). As a result, many CAs fail to meet user expectations (Mimoun et al., 2012), causing many CAs to confuse, frustrate, and sometimes even annoy users (Chakrabarti and Luger, 2015; Moore, 2013; Wallis and Norling, 2005). Consequently, it is crucial to pay attention to the various social design features of a CA as they, for example, affect user satisfaction (Verhagen et al., 2014), working alliance (Bickmore and Picard, 2005), perceived interpersonal stances (Ochs et al., 2017), or trustworthiness of the CA (Cassell and Bickmore, 2000).

2.2. Conversational agents are social actors

Since CAs use natural language and can express a variety of human-like verbal and nonverbal behaviors, interaction with them often feels similar to the interaction with real human beings (Gnewuch et al., 2017). This can be traced back to the phenomenon that computers are treated as social entities and that humans attribute human characteristics towards computers, which do not warrant any human attributions (i.e., a computer program is not a human) (Nass and Moon, 2000).

For example, Nass and colleagues showed that users perceive a computer with two different voices as two distinct social actors and that users apply gender stereotypes towards a computer dependent on its voice (Nass et al., 1997, 1994). Furthermore, they found that participants ascribe a personality to a computer depending on its strength of language, the interaction order and the expressed confidence level (Moon and Nass, 1996; Nass et al., 1995). They further discovered that a computer could be affiliated as a team member (Nass et al., 1996) and that help offered from a computer results in increased motivation to reciprocally help the computer (Fogg and Nass, 1997). Hence, computers trigger the user to exhibit emotional, cognitive, or behavioral reactions similar to reactions shown during interpersonal communication (Krämer, 2005). However, "*no studies have shown exactly how computing products trigger social responses in humans*" (Fogg, 2002, p. 89).

Particularly in the field of HCI, many studies have used the Computer Are Social Actors (CASA) paradigm as their theoretical foundation to explain the social reactions of humans towards computers (Nass et al., 1994; Nass and Moon, 2000). According to the CASA paradigm, humans turn their conscious attention to a subset of cues from a computer (e.g., female avatar) that cause them to categorize a computer as a relevant social entity (e.g., computer is female) while ignoring that the computer does not warrant human attributions (e.g., a computer cannot be biologically female) (Nass and Moon, 2000). Therefore, humans automatically apply social rules, expectations, and scripts known from interpersonal communication and apply it to the computer (e.g., apply gender stereotypes to computer) (Nass et al., 1994; Nass and Moon, 2000). Nass and colleagues argue that, from an evolutionary perspective, the human brain was developed at a time when only humans showed social behavior (Nass and Moon, 2000). In order to deal with the daily life, the brain developed automatic social responses to react to other social entities. Therefore, humans are hardwired to respond to anything that seems alive in some way (Fogg, 2002). This happens subconsciously and instinctively rather than rationally so that people often do not even notice that they have reacted in a social manner towards a computer (e.g., humans may not realize that they applied gender stereotypes to computers) (Nass et al., 1994; Nass and Moon, 2000). As a consequence, cues of a computer that lead to a social attribution are often called social cues (Araujo, 2018; Baur et al., 2015; Puetten et al., 2010; Reidsma et al., 2013) which are defined in more detail in the following section.

2.3. Social cues of conversational agents

To understand humans (e.g., emotional states, innate abilities), humans rely on many perceivable cues (e.g., gender, smile, gesture, voice variations) during an interpersonal interaction (Donath, 2007). Due to the similarity of interpersonal communication and the interaction with CAs, cues are also important design features of CAs (Nass and Moon, 2000). However, cues of CAs are often referred to in many different ways: cues, signals, social cues, social signals, but also anthropomorphic features or human-like characteristics (Donath, 2007; Pantic et al., 2011). To clarify the terminology, we outline existing definitions of cues, signals, social cues, and social signals as well as provide our conceptualization of social cues of CAs below.

In order to distinguish between a cue and a signal, Smith and Harper (2003) argue from an ethological perspective that any communicative sign can be divided into a cue and a signal. Whereas a cue can be defined as "*any feature of the world, animate or inanimate, that can be used by an animal as a guide to future action*" (p. 3), a signal can be seen "*as any act or structure which alters the behavior of other organism, which evolved of that effect*" (Smith and Harper, 2003, p. 3). In another ethological definition, Hauser (1996) states that cues and signals both represent information but "*cues tend to be permanently ON, whereas signals are more plastic and can be in an ON and OFF state*" (p. 9). From a psychological perspective, cues can be defined as stimuli which serve "*as a sign or signal of something else and this connection must have been previously learned*" (Pantic et al., 2011, p. 517). Thus, cues function as indicators that "*once received as a percept, are attributed information through a decoding process*" (Vinciarelli et al., 2012, p. 71). Besides, Donath (2007) proposes that "*everything that we use to infer a hidden quality is a cue. A cue is a signal only if it is intended to provide that information*" (p. 2). Summarizing these thoughts, Pantic et al. (2011) argue that a signal is any perceivable stimulus from which the receiver may draw some information.

In the next step, we introduce the two terms *social* cues and *social* signals. These terms are often used for cues or signals that do not only convey information but are essential to interpret, understand, and engage in a meaningful social interaction (Vinciarelli et al., 2009). Therefore, Vinciarelli et al. (2009) argue that behavioral social cues are relevant for producing social awareness and can be operationalized as "*temporal changes in neuromuscular and physiological activity*" (p. 1744). In the context of persuasive computers, Fogg (2002) considers social cues as cues of computers "*that elicit social responses from their human users*" (p. 89). In addition, Nass and colleagues state that social cues are those cues that trigger subconscious social reactions (Nass and Moon, 2000). In the context of human-robot interaction (HRI), Lobato et al. (2015) define social cues as features that "*act as channels of social information*" (p. 62). Similarly, Fiore et al. (2013) define social cues as "*biologically and physically determined features salient to observers because of their potential as channels of useful information*" (p. 2). On the other hand, social signals are considered as the ``*expression of ones attitude towards social situation and interplay, and they are manifested through a multiplicity of non-verbal behavioural cues*" (Vinciarelli et al., 2009, p. 1743). Pantic et al. (2011) define social signals as signals that provide ``*information about `social facts', i.e., about social interactions, social emotions, social attitudes, or social relations*" (p. 519). In the context of HRI, social signals can be defined as combinations of social cues that are "*conveying the perceived underlying meaning*" (Lobato et al., 2015, p. 62). Thus, social signals can be seen as the "*meaningful interpretations of cues in the form of attributions of an agent's mental state or attitudes*" (Wiltshire et al., 2014). Moreover, Fiore et al. (2013) argue that social signals are "*semantically higher than social cues*" (p. 2) and "*can be operationalized as meaningful interpretations based on mental states and attitudes attributed to another agent*" (p. 2).

In order to clearly distinguish between the terms cues and signals in this article, we follow Donath (2007) in arguing that a signal evolves from cues when they are created to have a communicative meaning or the receiver attributes an informative meaning to them. Therefore, we define a cue of a CA as any design feature of a CA salient to the user that presents a source of information (e.g., nodding) (Smith and Harper, 2003). Thus, cues are antecedents of signals and comprise all perceptible design features of a CA. Subsequently, cues can evolve into a *social signal* (Smith and Harper, 2003) through the attribution of socialness towards the CA (i.e., nodding of a CA is perceived as a signal of agreement) (Nass and Moon, 2000; Wiltshire et al., 2014). This attribution is the result of a conscious or subconscious interpretation of the cues, which ultimately triggers a social reaction of the user (e.g., user reacts to the CA's nodding) (Knapp et al., 2013; Nass and Moon, 2000; Vinciarelli et al., 2012). These social reactions of a user are considered social "*if a participant's emotional, cognitive, or behavioral reactions are similar to reactions shown during interactions with other human beings*" (Krämer, 2005, p. 443). Thus, we define the term *social cue* as a *cue* that triggers a *social reaction* towards the emitter of the cue (Fogg, 2002; Nass and Moon, 2000). Table 1 summarizes the definitions of the key concepts of this article.

Fig. 1 outlines the process of how a (social) cue evolves into a social signal and subsequently triggers a social reaction based on one example. Nass et al. (1997) showed that a CA's gender of voice (i.e., a cue) leads humans to attribute a biological gender towards a CA (i.e., a social signal). This triggers the user to express gender-based stereotypic responses towards the CA (i.e., a social response). As this reaction towards a CA is similar to human behavior in interpersonal communication, the cue called *gender of voice* can be considered as a social cue.

Finally, Table 2 illustrates several examples of social cues, social signals, as well as their corresponding social reactions. To provide an exemplary overview, we selected two examples for each type of CA (i.e., text-based, voice-based, and ECAs) from literature.

It must be noted that relationships between social cues, their corresponding social signals, and the resulting social reactions are not deterministic cause and effect relationships (i.e., a single social cue does not always lead to a single social signal). Instead, a single social cue can lead to many different social signals (Carolis et al., 2004). For example, a smile of a CA can be perceived as the social signal of friendliness, the emotion of joy, or as a dominant or a submissive personality (Carolis et al., 2004; Youssef et al., 2015). Moreover, the relationship between social cues and social signals is highly context-dependent (Lamolle et al., 2005). In most Western cultures, vertical nodding is generally perceived as agreement, whereas in Bulgaria, this social cue is interpreted differently and means disagreement (Andonova and Taylor, 2012). Moreover, one social signal (e.g., agreement) is usually the result of a complex interplay of several and sometimes multimodal single social cues (e.g., greeting, nodding, smile, and gesture) (Bevacqua et al., 2010; Pelachaud, 2009a). Therefore, social cues usually do not occur in isolation and need to be considered together in order to create an expressive, natural, and believable social behavior (Bevacqua et al., 2010; Caridakis et al., 2007; Pelachaud, 2005). As a consequence, researchers describe communicative functions conveyed by a CA usually as pairs of the desired meaning and their corresponding operationalization through social cues (Carolis et al., 2004). The combination of several single social cues at the same time, however, can also lead to conflicts and to abnormal behaviors (e.g., frown and a simultaneous raising of the eyebrows) (Pelachaud, 2009a, 2005). Moreover, a smile can signal friendliness, whereas a smile followed by gaze and head aversion can create the social signal of embarrassment (Chollet et al., 2014; Pelachaud, 2009a). Therefore, it is important to consider the sequence, length, and temporal arrangement of single social cues since social signals evolve dynamically over time (Vinciarelli et al., 2012). Finally, a smile usually responds to another smile, and a posture is usually followed by another posture (Pelachaud, 2017). Therefore, the imitation and reciprocal adaptation of social cues (e.g., smile of a CA as reaction to a smile of a user, repetition of user utterances by the CA) also impacts the conveyed social signals of a CA (Campano et al., 2015; Lamolle et al., 2005; Prepin et al., 2013; Youssef et al., 2015).

Since different social signals are created through the co-occurring, temporal arrangement, multimodal realization, and reciprocal adaptation of several single social cues, we argue that a classification of single social cues on the lowest level of complexity would provide a good starting point for researchers from different domains as well as different contexts and cultures. Although we are aware that single social cues usually do not occur in isolation, we focus on classifying single social cues since researchers and practitioners should have a clear understanding of the different types of social cues of CAs. This understanding then serves as a foundation to investigate their context-dependent outcomes and decide how specific social signals should be operationalized. Thus, in this article, we use the term *social cues* to refer to single social cues of CAs.

2.4. Existing classifications of social cues

In order to distinguish between different social cues, interpersonal communication theory already provides several useful starting points. Burgoon et al. (2011) classify nonverbal communication cues into eight major codes that constitute the way they are created, transmitted, perceived, and interpreted. These are called kinesics, vocalics, physical appearance, proxemics, haptics, chronemics, environment and artifacts,

Table 1
Definitions of key concepts.

Concept	Definition
Cue	A *cue* is any design feature of a CA salient to the user that presents a source of information (Smith and Harper, 2003).
Social Signal	A *social signal* is the conscious or subconscious interpretation of cues in the form of attributions of mental state or attitudes towards the CA (Nass and Moon, 2000; Wiltshire et al., 2014).
Social Reaction	A *social reaction* is an emotional, cognitive, or behavioral reaction of the user towards a CA that is considered appropriate when directed at other humans beings (Krämer, 2005).
Social Cue	A *social cue* is a *cue* of a CA that triggers a social reaction of the user towards the CA (Fogg, 2002; Nass and Moon, 2000).

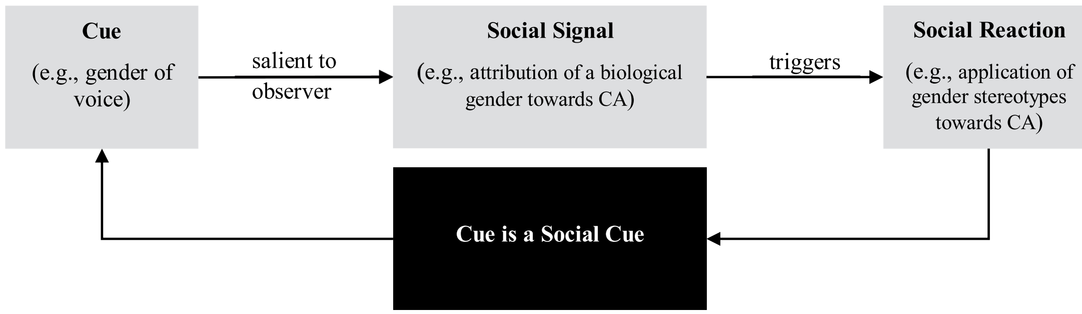


Fig. 1. The emergence of a social reaction towards a cue of a CA defines a social cue (example based on Nass et al., 1997).

Table 2
Examples of social cues of CAs.

Social cue(s)	Social signal	Social reaction	References
Choice of words	Perceived politeness of CA.	Impact on user's learning endeavors.	Mayer et al. (2006)
Excuse	Perceived empathy of CA.	Users spend more time interacting with a CA.	Klein et al. (2002)
Interaction order, strength of language, confidence	A CA that uses a strong (weak) language, always replies first (last), and has a high (low) confidence is perceived as being dominant (submissive).	Users perceive a CA as more satisfactory and beneficial when its personality matches their personality.	Nass et al. (1995)
Gender of voice	Attribution of a biological gender towards CA.	Application of gender stereotypes towards CA.	Nass et al. (1997)
Head movement, facial expression, eye movement, gestures	Communicative functions about the CA's beliefs, intentions, affective state, and mental state.	Perception and identification of CA's expressive behavior.	Pelachaud (2005)
Head movement, smile, facial expression, eye movement, vocal segregates, vocalization, voice tempo, pitch range	Attribution of meaning to multimodal backchannels (e.g., agreement, refusal).	Understanding the conveyed meaning of backchannels.	Bevacqua et al. (2010)

and olfactics (Burgoon et al., 2011). Leathers (1976) classifies the interpersonal communication system into four subsystems that transfer meaning either each on its own or by interacting, reinforcing, and conflicting with the other systems. The four systems are called verbal, visual, auditory, and invisible (Leathers, 1976). Furthermore, Trager (1958), Crystal (1969), and Laver (1980) provide influential classifications of nonverbal vocal cues.

Reviewing related work in HCI and HRI, we identified existing classifications that provide valuable insights about different types of social cues for specific technology domains (e.g., email, robots, computers in general) or for specific application contexts (e.g., digital services). For example, Walther (2006) classifies nonverbal cues transmitted in computer-mediated communication and structures them into cues that either remain from interpersonal communication (e.g., chronemics) or are reintroduced by technology (e.g., 2D avatars, anthropomorphic icons). In another classification, Fogg (2002) provides an overview of different types of social cues of computers that can be used to create persuasive technology products. He proposes five primary categories of social cues: physical, psychological, language, social dynamics, and social roles. Cassell et al. (2000) classify the design dimensions along which the embodiment of a CA can vary. They distinguish whether the appearance of the ECA is animated, photorealistic, stable, 2D or 3D, or humanoid. Cowell and Stanney (2005) review several empirical studies that investigate non-verbal cues. They categorize nonverbal cues from ECAs that influence the perceived credibility of the character dependent on the origin (i.e., non-behavioral, behavioral) and individual control of the social cue (i.e., low, high). In an extensive review written in German, Krämer (2008a) analyses various theories and empirical studies covering social responses to CAs. She concludes that two types of human-like cues are responsible for a subliminal attribution of socialness to a CA: behavior cues (e.g., interactivity, movements, actions) and outer cues (e.g., eyes). These cues trigger social responses irrespective of whether the user judges the agent as being human or not (Krämer, 2008a). Wuenderlich and Paluch (2017) analyze how social cues affect the authenticity perceptions of service agents. They categorize social cues into agent-related cues and communication-related cues. Agent-related cues refer "*to the user's evaluation of the service agent*" (Wuenderlich and Paluch, 2017, p. 7). They consist of visual (e.g., picture of the agent) and audio cues (e.g., voice of the agent), as well as identity cues (e.g., display name of the agent). Communication-related cues include the communication styles of the agent, which influences "*how users evaluate the quality of the communication*" (Wuenderlich and Paluch, p. 7). They include variations of the use of language such as empty phrases, colloquial language, emotions, attentiveness, and personalization.

In the domain of Social Signal Processing (SSP), Vinciarelli et al. (2009) distinguish the most critical behavioral cues necessary to understand social interactions. Therefore, they separate behavioral social cues in physical appearance (i.e., height, body shape, attractiveness, body shape), gesture and posture (i.e., hand gestures, posture, walking), face and eye behavior (i.e., facial expression, gaze behavior, focus of intention), and space and environment (i.e., distance, seating arrangements) (Vinciarelli et al., 2009). Akhtar and Falk (2017) derive a taxonomy of social cues from the observation that SSP methods generally use two kinds of cues: cues including words (e.g., the semantic linguistic content of speech), and wordless and visual cues (e.g., gestures). Verbal cues account for "*what is being said and include descriptive verbal messages of spoken communication*" (Akhtar and Falk, 2017, p. 1). Non-verbal cues are expressed through "*temporal changes in neuromuscular and physiological activities*", which can be further separated in several subgroups (e.g., vocal, visual, sensor/device, neurological) (Akhtar and Falk, 2017, p. 1).

In addition, much research on HRI has been dedicated to understanding and modeling social cues of robots. For example, Hegel et al. (2011) propose a multidimensional taxonomy of social cues for robots which distinguishes social cues according to their sign typology (i.e., signal, cue), the designer's intention (i.e., explicit, implicit), source of sign (i.e., human, artificial), perceptual type (i.e., appearance, auditive, olfactory, tactile, motion). In another HRI classification, Fiore et al. (2013) build on SSP and distinguish between physical and behavioral social cues of robots. Physical cues consist of "*aspects of physical appearance and environmental factors, such as the distance between a social agent and an observer*" (p. 2) and behavioral cues consist of "*non-verbal movements, actions, and gestures as well as verbal vocalizations and expressions using the body and face*" (p. 2) such as gestures, laughers, and smiles (Fiore et al., 2013). Moreover, Wiltshire et al. (2014) categorize social cues of robots into paralinguistic cues, gaze cues, and proxemic cues.

While the classifications mentioned above provide valuable insights on how to differentiate types of social cues dependent on the specific technology or application context, a comprehensive overview and classification of social cues of CAs from various research domains is lacking.

3. Methodology

In this section, we outline our methodology to review existing research on social cues of CAs and to develop a taxonomy. As shown in Fig. 2, our methodology comprises three steps. First, we conducted a SLR on social cues of CAs following established guidelines (Kitchenham, 2004; Webster and Watson, 2002; Wolfswinkel et al., 2013). Then, we used the identified social cues in the selected publications as input to develop a taxonomy of social cues for CAs based on the approach by Nickerson et al. (2013). Subsequently, we evaluated the taxonomy using a card sorting procedure based on Moore and Benbasat (1991).

3.1. Step 1: literature review

As a first step, we conducted a SLR to identify and analyze existing research on social cues of CAs based on the guidelines of Kitchenham (2004) and Webster and Watson (2002). Since research on this topic is scattered across different areas, we selected three databases covering relevant literature in CS and IS, namely IEEE Xplore Digital Library, ACM Digital Library, and EBSCOhost. To account for different names used to describe CAs (e.g., CA, ECA, chatbot, virtual assistant, digital assistant) and social cues (e.g., social cues, anthropomorphic features, human-like characteristics), we conducted an exploratory search in all three databases to identify relevant keywords and synonyms to build our search term. We decided to perform a full-text search to include relevant publications that do not explicitly mention CAs and social cues in their abstracts, titles, or keywords. Subsequently, all publications were assessed with respect to the following inclusion criteria: first, publications had to be original, peer-reviewed, and written in English. Second, publications had to refer to any type of CA (e.g., voice-based, text-based, embodied) and analyze social cues of a CA that led to social reactions by the users. The complete search strategy is shown in Fig. 3. In addition, we conducted a backward/forward search to identify further publications (Webster and Watson, 2002).

Subsequently, all selected publications were coded according to the guidelines of Wolfswinkel et al. (2013). We reviewed all selected publications and identified and labeled all excerpts dealing with social cues. Next, we systematically differentiated, partitioned, and integrated these excerpts in several iterative adjustment cycles to identify relevant social cues. The results of step 1 served as the initial input for our subsequent taxonomy development process.

3.2. Step 2: taxonomy development

In literature, the terms "taxonomy", "classification", and "typology" have been used interchangeably (Gregor, 2006; Nickerson et al., 2013). While a discussion of their individual differences is beyond the scope of this article (for a detailed discussion, see Lakoff, 1987), the general process of classification is the assignment of objects to categories based on their similarity (Bailey, 1994). In developing our taxonomy, we followed the method by Nickerson et al. (2013). Their method integrates two development approaches (i.e., a conceptual-to-empirical and an empirical-to-conceptual approach) into a single iterative approach. The conceptual-to-empirical approach is a top-down approach that subdivides a general category based on theory foundation and not on empirical findings (Gerber et al., 2017; Nickerson et al., 2013). The empirical-to-conceptual approach is a bottom-up approach that groups objects into categories based on their perceived similarities (Gerber et al., 2017; Nickerson et al., 2013). The method proposed by Nickerson et al. (2013) combines the advantages of both approaches and allows researchers to modify the taxonomy in a more flexible manner. In addition, we extended the method of Nickerson et al. (2013) with hierarchical categories and subcategories as described in Prat et al. (2015).

Nickerson et al. (2013) suggest to define objective and subjective ending conditions that determine the ending of the iterative development cycles. The objective ending conditions are met when the taxonomy is mutually exclusive and collectively exhaustive (Nickerson et al., 2013). This means that the classification consists of enough categories to assign each object to a category (collectively exhaustive), and each object is assigned to one and only one category (mutually exclusive). Thus, there is exactly one category for each object (Bailey, 1994; Nickerson et al., 2013). The subjective ending conditions are met when the taxonomy is concise (i.e., meaningful number of categories), robust (i.e., categories provide a sufficient differentiation among the social cues), comprehensive (i.e., includes all social cue categories of interest), extensible (i.e., other not yet mentioned social cue categories could be easily added), and

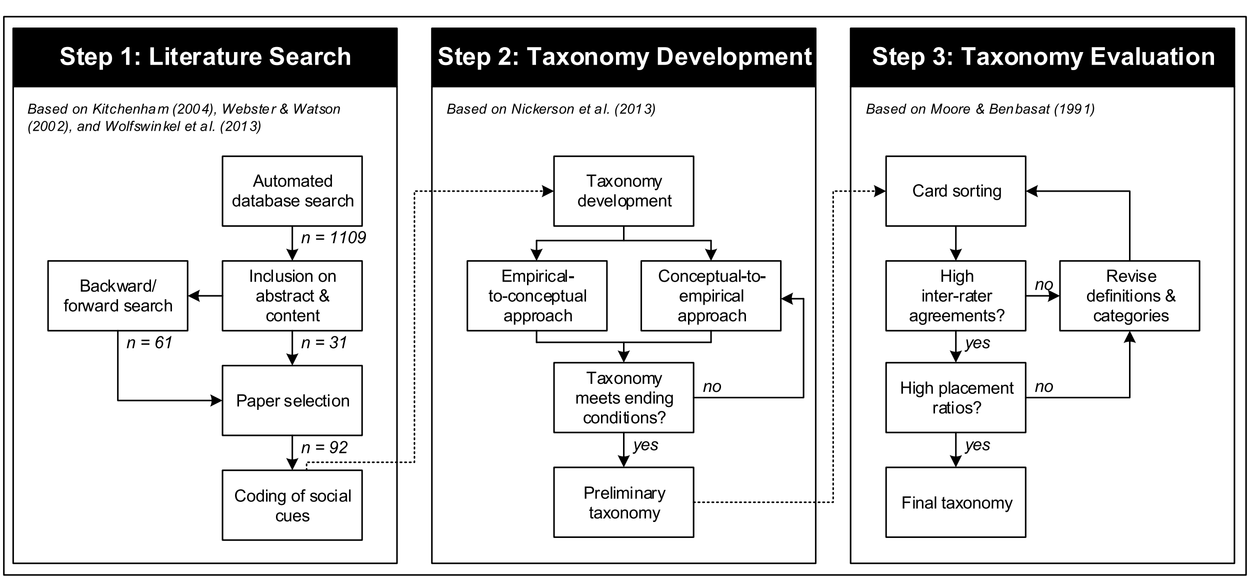


Fig. 2. Research methodology.

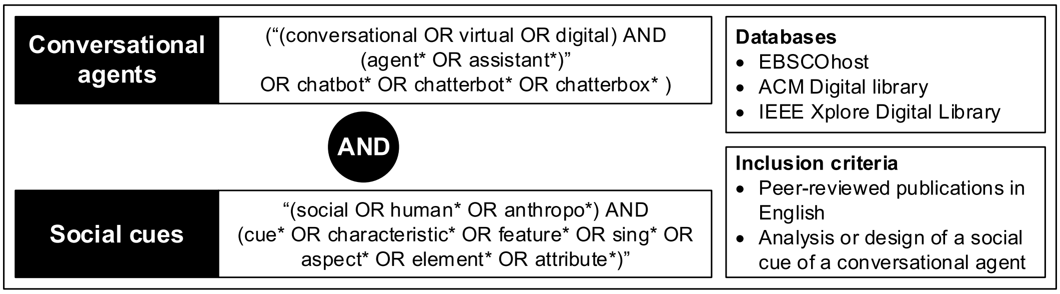


Fig. 3. Search strategy.

explanatory (i.e., provides useful explanations of the nature of social cues) (Nickerson et al., 2013). The iterative development process ends when all these conditions are met.

3.3. Step 3: taxonomy evaluation

A taxonomy needs to benefit its users (Nickerson et al., 2013) and is only as good as the categories on which it is based (Bailey, 1994). Therefore, all categories must be easy to understand, all category names and definitions must be meaningful and natural, and the logic used to assign the objects to categories must be clear, simple, and parsimonious (Gregor, 2006). We selected a card sorting procedure to evaluate our preliminary taxonomy in order to assess how potential users of the taxonomy (i.e., CA researchers and practitioners) understand the categories, subcategories, and definitions (Moore and Benbasat, 1991).

Our card sorting procedure was divided into three consecutive iterations, each with new participants and a refined version of the taxonomy. The participants recruited for these sessions were potential users of the taxonomy, namely CA researchers and practitioners. In each session, each individual participant was introduced to the topic and received the definitions for each social cue category. Then, all cards were handed out, containing all relevant information about each social cue (i.e., name, detailed description, examples, see example cards in Appendix Fig. A1) (Rugg and McGeorge, 2005). Finally, the participants were asked to sort each social cue card to one of the categories and subsequently to one of the subcategories. All card sorting sessions were audio recorded with participants' consent. The card sorting process iterates until two measures confirm that the taxonomy is perceived as meaningful and natural (Moore and Benbasat, 1991). First, a high inter-rater agreement indicates a high reliability of the sorting sessions, which suggests that different users of the taxonomy understand the categories in a similar way. This is measured using Cohen's Kappa, which is the chance corrected coefficient of agreement (Cohen, 1960). Cohen's Kappa can only be applied to two sorters, so it was calculated for each pair of participants. Additionally, Fleiss' Kappa was calculated, an extended version of Cohen's Kappa for more than two sorters (Fleiss, 1971). Moore and Benbasat (1991) consider an agreement above a Kappa value of 0.65 to be acceptable. Others consider a value above 0.81 (Landis and Koch, 1977) or 0.91 (LeBreton and Senter, 2008) as a very strong agreement. Second, we calculated social cue placement ratios that indicate how many social cues are placed in our intended target category (Moore and Benbasat, 1991). A category with a high degree of correct social cue placements indicates that the categories are well understood. However, there is no measure for "good" placement ratios, as this method is rather a qualitative analysis used to identify problem areas (Moore and Benbasat, 1991). In addition to these two measures, we audio recorded all sessions to better understand the thoughts and concerns of the participants (Rugg and McGeorge, 2005).

4. Results

4.1. Literature review results

The literature search was conducted in January 2018 and yielded a total of 1.109 results in three databases. First, we removed publications based on formal criteria (i.e., duplicates, non-English and not peer-reviewed publications). Second, we assessed titles and abstracts of all retrieved publications for relevance. We excluded publications that were not concerned with our research focus. For example, many publications focused on the architecture and technical implementation of CAs or analyzed social cues of different technologies such as physical robots. Third, we retrieved the full text of all remaining publications and analyzed them based on the following criteria. We excluded all publications that did not investigate or design social cues of a CA. This led to the selection of 31 relevant publications. Finally, we performed an additional backward/forward search (Webster and Watson, 2002), which identified further 61 relevant publications. The high number of additional retrieved publications indicates that social cues of CAs are often investigated, but not always explicitly mentioned. Finally, the SLR ended and identified a total of 92 relevant publications.

In the next step, we labeled all study excerpts related to social cues following the method of Wolfswinkel et al. (2013). Therefore, we conducted iterative abstraction and integration cycles and derived 48 distinct social cues. To achieve a consistent level of abstraction, we oriented ourselves at the level of abstraction of well-established classifications and communicative codes in intermediate communication theory (e.g., Burgoon et al., 2010; Knapp and Daly, 2011; Trenholm and Jensen, 2011) and in prosody and paralanguage in speech (Crystal, 1969; e.g., Trager, 1958). This final list of social cues is summarized in Table 3. As with any other literature review, we do not argue that this list is exhaustive. However, 48 social cues deem suitable to serve as an initial starting point for developing a taxonomy.

4.2. Taxonomy development results

We started the taxonomy development process by defining a high-level meta-characteristic as a basis for the classification of social cues (Nickerson et al., 2013). Any subsequently identified category should be a logical consequence of this meta-characteristic to avoid naive empiricism and thus, should be based on the expected needs of potential users (Nickerson et al., 2013). Due to the different interpretations of social cues resulting from their interplay and the influence of context, we argue that researchers and practitioners must first understand the different types of social cues before conclusions about their outcomes and their operationalization can be drawn. Therefore, we aim to investigate and classify the different types of social cues that can be implemented as design features in a CA. Thus, the meta-characteristic for the taxonomy development process is the type of social cues of CAs.

We decided to start with an empirical-to-conceptual taxonomy development approach, as we identified 48 social cues in the SLR that served as our initial empirical basis (Nickerson et al., 2013). The first

Table 3
List of identified social cues (number of publications)[a].

[SC 1]	2D-/3D-agent visualization (n = 1)	[SC 17]	First turn (n = 5)	[SC 33]	Refer to past (n = 3)
[SC 2]	Abbreviation (n = 1)	[SC 18]	Formality (n = 4)	[SC 34]	Response time (n = 4)
[SC 3]	Age (n = 3)	[SC 19]	Gender (n = 11)	[SC 35]	Self-disclosure (n = 5)
[SC 4]	Arm and hand gesture (n = 9)	[SC 20]	Gender of voice (n = 5)	[SC 36]	Self-focused question (n = 4)
[SC 5]	Ask to start/ pursue dialog (n = 1)	[SC 21]	Greetings and farewells (n = 4)	[SC 37]	Sentence complexity (n = 2)
[SC 6]	Attractiveness (n = 4)	[SC 22]	Grunts and moans (n = 1)	[SC 38]	Small talk (n = 6)
[SC 7]	Background (n = 1)	[SC 23]	Head movement (n = 12)	[SC 39]	Strength of language (n = 8)
[SC 8]	Clothing (n = 4)	[SC 24]	Joke (n = 4)	[SC 40]	Tactile touch (n = 2)
[SC 9]	Color of agent (n = 6)	[SC 25]	Laughing (n = 3)	[SC 41]	Temperature (n = 1)
[SC 10]	Conversational distance (n = 3)	[SC 26]	Lexical diversity (n = 1)	[SC 42]	Thanking (n = 2)
[SC 11]	Degree of human-likeness (n = 17)	[SC 27]	Name tag (n = 3)	[SC 43]	Tips and advice (n = 4)
[SC 12]	Emoticons (n = 2)	[SC 28]	Opinion conformity (n = 3)	[SC 44]	Typeface (n = 1)
[SC 13]	Excuse (n = 5)	[SC 29]	Photorealism (n = 4)	[SC 45]	Vocal segregate (n = 4)
[SC 14]	Eye movement (n = 16)	[SC 30]	Pitch range (n = 5)	[SC 46]	Voice tempo (n = 6)
[SC 15]	Facial expression (n = 25)	[SC 31]	Posture shift (n = 10)	[SC 47]	Volume (n = 2)
[SC 16]	Facial feature (n = 2)	[SC 32]	Praise (n = 6)	[SC 48]	Yawn (n = 1)

[a] A description and examples for each social cue are provided in Table A1 in the appendix.

iteration cycle aimed to sort all social cues into categories at the highest possible level (Gregor, 2006) and to define the general types that determine how social cues are created. By scanning the data, we classified the social cues of CAs into the two fundamentally different ways they are created in interpersonal communication. Overall, 17 social cues are created by written or spoken words, and 31 social cues are not associated with the use of words. The distinction in verbal and nonverbal cues is often applied in interpersonal communication theory since a dialogue is an ensemble of verbal and nonverbal communication (DeVito, 2013; Fernández-Dols, 2013; Guerrero et al., 1999). Therefore, verbal means "*expressed with words*" (Fernández-Dols, 2013, p. 79) and nonverbal "*expressed by non-linguistic means*" (Gamble and Gamble, 2014, p. 152), which is often referred to as paralanguage in voice-based communication (Poyatos, 1991; Schötz, 2002). The assignment of all 48 social cues to these two categories fulfills the objective ending criteria for building a flat and one-dimensional, mutually exclusive, and collectively exhaustive taxonomy. However, a classification based on trivial categories creates a trivial taxonomy (Bailey, 1994). Consequently, we did not perceive this initial taxonomy as concise (i.e., as it has only two categories) and decided to conduct a further iteration.

In the next iteration, we switched to a conceptual-to-empirical classification approach to reveal more concise categories. After examining communication literature, we decided to classify the social cues based on the human communication systems described by Leathers (1976). This provides a "*holistic, comprehensive, and realistic picture of the complex set of behaviors that interact to make up human communication*" (p. 11). Therefore, we argue that it also provides a valuable starting point for classifying social cues of CAs. Leathers states that the human communication system consists of the verbal and nonverbal communication systems (in accordance with the taxonomy of the first iteration). He further categorizes the nonverbal communication system into three subsystems, namely visual, auditory, and invisible (Leathers, 1976; Leathers and Eaves, 2015). Each of the four communication systems is responsible for creating and transmitting different messages in interpersonal communication and thus, seems appropriate for the classification of social cues of CAs. Therefore, we assigned all 31 nonverbal social cues to one of the three corresponding nonverbal communication systems. 19 social cues were assigned to visual cues, which relate to all nonverbal cues that are created through visual channels and are decoded by sight (Leathers, 1976; Leathers and Eaves, 2015). Eight social cues were assigned to the auditory cues, which are created through nonverbal sounds and are decoded by hearing (Leathers, 1976; Leathers and Eaves, 2015). Finally, four cues were assigned to the invisible cues, which are transmitted in the absence of any visualizations or sounds (e.g., through the use of time, through odors, through touch) (Leathers, 1976; Leathers and Eaves, 2015). This results in four mutually exclusive and collectively exhaustive social cue categories that provide a complete and holistic differentiation of the channels through which social cues are created. In the next step, we divided these categories into subcategories to identify more specific ways to create social cues. Therefore, we performed additional conceptual-to-empirical development iterations to subdivide each of the four categories, which are described below.

Verbal cues refer to all social cues created by words. What people say or write with words belongs to the discourse of an interaction which can be defined as the "*social action made visible in language*" (Antaki, 2008, p. 2). In order to analyze the discourse of an interaction, various approaches have been developed (Antaki, 2008). However, each method faces a variety of challenges (Antaki et al., 2003). One reason is the complex structure of the human language that can result in under-analysis of the diverse facets of the human language. Following Trenholm and Jensen (2011), we can analyze language according to the codes that constitute it (e.g., discrete and separable units), the function it conveys (e.g., express and control emotion), or the structure of language (e.g., semantic, syntactic, pragmatic). Moreover, verbal cues can be produced on different layers such as conversational behavior, topic selection, style, syntax, lexicon, and speech (Mairesse et al., 2007). In addition, verbal communication can be analyzed according to the various facets of content analysis procedures, which distinguish the syntactic, syntactic-semantic, semantic, semantic-pragmatic, syntactic-pragmatic, semantic-pragmatic, and pragmatic level of analysis (Titscher et al., 2000). Taking these dimensions into consideration, it becomes prevalent that "*in the study of language, as in any other systematic approach, there is no neutral terminology*" (Searle et al., 1980, p. vii). In order to ensure that the dimensions of the taxonomy remain natural, simple, and parsimonious (Gregor, 2006), we follow Walther (2008) that language cues can engender social functions depending on the "*style and the verbal content of the articulated message*" (Walther, 2008, p. 394). Other researchers divide verbal cues into similar categories (Collier, 2014; Tannen, 1984; Thomas et al., 2018). Thus, it can be assumed that the same verbal content (i.e., what is said) can be expressed in many different styles (i.e., how something is said) (Collier, 2014). Thus, content cues refer to all aspects of the language that remain after a message has been transcribed and paraphrased and contains the strict and literal meaning itself (Collier, 2014). Moreover, everything said must be said somehow (Tannen, 1984). Language can create different social meanings which are transmitted on different linguistic levels such as phonology, syntax, semantics, or lexicon (Bell, 1997). Therefore, style cues refer to the meaningful deployment of language variation in a message (Selting, 2009). Since both content and style elements of the articulated message generate social reactions (Walther, 2006, 1992), we assigned eleven social cues to content cues and six to style cues.

Visual cues refer to all nonverbal social cues that are visually perceptible and can be created in three different ways: kinesics (i.e., body movement and gestures representing body language), proxemics (i.e., use of space, distance, and territory), or artifacts (i.e., appearance, clothing, and accessories) (Leathers, 1976; Leathers and Eaves, 2015; Trenholm and Jensen, 2011). Since Leathers' (1976) artifactual communication system is directly derived from the human appearance, it consists of the fixed biological appearance and its manipulation. Since the visual appearance of a CA can be designed in almost all possible ways, it does not seem reasonable to differentiate between fixed and variable appearance forms. Therefore, the term "agent appearance" was used for this category, which contains all social cues related to the visual representation of a CA. Finally, we assigned all visual cues to one of the three subcategories. This resulted in the assignment of ten social cues to agent appearance cues, five to kinesic cues, and two to proxemic cues. Nevertheless, two social cues could not be assigned to one of these three subcategories, namely typefaces (Candello et al., 2017) and emoticons (Brandão et al., 2013; Li et al., 2017). Therefore, we switched to an empirical-to-conceptual approach to analyze these two social cues. We identified that these two social cues do not fit into interpersonal communication theory since emoticons and typefaces are not present in human face-to-face communication. Instead, they are specific features of computer-mediated communication (CMC) (Liebman and Gergle, 2016), in which they are used as visual cues to expand the meaning of text messages (Rezabek and Cochenour, 1998; Walther, 2006). In the literature, they are often referred to as CMC cues (e.g., Kalman and Gergle, 2014, 2010) or CMC features (e.g., Hill et al., 2015). Thus, we followed these propositions and assigned all social cues created by visual and text-based elements, such as typefaces and emoticons, to the newly developed fourth visual cue category called CMC cues.

Auditory cues refer to all social cues created through nonverbal sounds, which are also often referred to as vocalics, paralanguage, or prosody (Burgoon et al., 2011). Various authors have provided classifications in order to distinguish nonverbal vocal cues which surround speech behavior (Knapp et al., 2013). For example, some distinguish them by primary qualities (e.g., pitch, tempo) and voice qualifiers (Poyatos, 1991). Others refer to paralinguistic (affective information) and extralinguistic (i.e., voice qualities) information in speech (Laver, 1980). Crystal (1969) analyzed spontaneous speech and provided an influential distinction of the English tone of voice. He distinguishes the non-linguistic vocal effects, semiotic frame, and the vocal-auditory components which are further separated into segmental verbal (e.g., vocalizations), pause phenomena, and non-segmental features which consists of prosodic features (e.g., tone, pitch-range, loudness) and paralinguistic features (e.g. falsetto, chest) (Crystal, 1969). One of the first systematic and most influential studies in this field was the taxonomy proposed by Trager (1958) (Nöth, 1995). Following his taxonomy, auditory cues include voice set, voice qualities, and vocalizations (Trager, 1958). Voice set refers to the idiosyncratic background of speech (Trager, 1958). These include permanent or quasi-permanent physical and physiological characteristics of the voice such as gender, age, and health (Nöth, 1995; Trager, 1958). Voice qualities include all recognizable and adjustable characteristics of the voice along a continuum such as the acceleration or deceleration of speech speed or the narrowing or spreading of the pitch range (Burgoon et al., 2010; Trager, 1958). Vocalizations refer to the nonlinguistic vocal sounds or noises which do not belong to the background characteristic of speech (Trager, 1958). They are remote from any linguistic relevance (James, 2017) and include vocal features like laughing and crying, vocal qualifiers in terms of overloud or oversoft, as well as vocal segregates such as segmental sounds like "uh-huh" and "mhm" (Nöth, 1995; Trager, 1958). Using the taxonomy by Trager (1958), we assigned one social cue to voice set, three to voice qualities, and four to vocalizations.

Invisible cues refer to all social cues which we cannot see or hear (Knapp et al., 2013; Leathers, 1976; Leathers and Eaves, 2015). Due to the invisible character of these cues, invisible cues constitute "*the silent language*" (Hall, 1990) in communication and comprise chronemic, haptic, and olfactory cues (Leathers, 1976; Leathers and Eaves, 2015; Trenholm and Jensen, 2011). Chronemics describes the function of time and timing in communication such as waiting times, lead times, or tempo (Burgoon et al., 2011, 2010; Hall, 1990). Haptics - also referred to as tactile communication (Leathers, 1976; Leathers and Eaves, 2015) - encompasses the perception and use of touch (Burgoon et al., 2010). This includes various forms of touch (e.g., slaps, kisses, kicks), their intensity, position, and the body parts that perform the touch (Burgoon et al., 2011). Haptic cues may be visible, but they "*communicate powerful meanings in the absence of any illumination and* […] *the decoder relies on cutaneous receptors rather than eyesight to decode them*" (Leathers and Eaves, 2015, p. 13). Finally, olfactory communication refers to all communication elements that are created through the use of odors and smells (Burgoon et al., 2011). Subsequently, all invisible social cues were assigned to one of the three subcategories. Hence, we assigned two social cues each to chronemics and haptics, but no olfactory cue was identified. This violates one of the objective ending conditions of Nickerson et al. (2013), which states that at least one object must be assigned to each category. Thus, we excluded olfactory cues.

Finally, all 48 social cues could be assigned to one of the four identified social cue categories and subsequently, to one of their ten subcategories. The taxonomy is exclusive and exhaustive because every social cue was assigned to exactly one category and later to exactly one subcategory. As all objective and subjective ending conditions were met (i.e., concise, robust, comprehensive, expendable, explanatory), the taxonomy development process ended at this point.

4.3. Taxonomy evaluation results

To evaluate whether the taxonomy appears clear, simple, and parsimonious (Gregor, 2006), we conducted a series of card sorting evaluation rounds (Moore and Benbasat, 1991). In three consecutive weeks, three sessions were conducted with five participants each (novice CA designers such as graduate students (n = 7) and Ph.D. students (n = 5), practitioners: n = 3), 11 men and 4 women, with an average age of 26 years, SD = 1.77). The participants had varying usage experience with CAs (daily interaction (n = 4), several times a week (n = 7), a couple of times a month (n = 4)). None of the participants were involved in the taxonomy development. All participants sorted each of the 48 social cue cards individually to one of the four categories and then to one of the ten social cue subcategories. This resulted in a total of 240 social cue placements per evaluation round. Each session lasted on average 58 min ($SD_{duration}$ = 6 min). Different agreement measures were calculated for each card sorting round. These include Cohen's Kappa (Cohen, 1960), Fleiss' Kappa (Fleiss, 1971), and the placement ratios that indicate how often a social cue is placed in the target category (Moore and Benbasat, 1991). Table 4 shows all agreement measures, all placement ratios, and the taxonomy refinements between the rounds. Appendix Table A2, Table A3, and Table A4 provide detailed placement information for each sorting round.

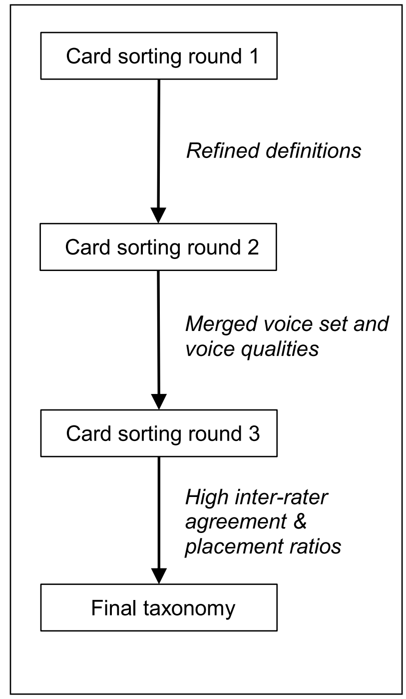
Round 1: The first card sorting round provided insights on how users of the taxonomy perceive the categories. Average raw agreement scores (0.90), averaged Cohen's Kappa (0.88), and Fleiss' Kappa (0.88) revealed a strong inter-rater agreement (according to LeBreton and Senter, 2008; Landis and Koch, 1977). Comparing the actual sorting results of all five participants with the intended assignment of the research team showed that the five participants achieved an average correct assignment in 94% of the placements. More specifically, the participants assigned the social cues correctly for six categories. Furthermore, social cues of the remaining categories were correctly assigned in more than 83% of the cases. Only the voice set category performed worse with an average placement rate of 56%. The analysis of the sessions' audio recordings revealed that several participants struggled with the specific definitions of the categories and descriptions of some social cues (e.g., auditory cues and CMC cues). Hence, we analyzed their feedback and refined several definitions.

Round 2: The second card sorting round was performed with five different participants. The averaged raw agreement score (0.92),

Table 4
Card sorting process and results.

Agreement measure	Round 1	Round 2	Round 3
Averaged raw agreement	0.90	0.92	0.96
Averaged Cohen's Kappa	0.88	0.91	0.95
Fleiss' Kappa	0.88	0.91	0.95

Placement ratio summary		Round 1	Round 2	Round 3
Verbal	Content	0.95	0.98	1.00
	Style	1.00	0.89	0.93
Visual	Kinesics	1.00	1.00	1.00
	Proxemics	1.00	1.00	1.00
	Agent appearance	0.92	0.96	0.98
	CMC	0.83	0.82	1.00
Auditory	Voice set	0.56	0.71	-
	Voice qualities	1.00	1.00	1.00
	Vocalizations	1.00	1.00	0.95
Invisible	Chronemics	0.90	1.00	1.00
	Haptics	1.00	1.00	1.00
Average		**0.94**	**0.96**	**0.98**



averaged Cohen's Kappa (0.91), and Fleiss' Kappa (0.91) further increased, indicating a stronger agreement compared to the first round. The analysis of the placement ratios showed that social cues from six categories were correctly assigned to the target categories. Furthermore, three categories had a placement ratio above 89% and voice set increased to 71% but remained lowest. The CMC category did not improve and resulted in slightly lower Kappa values. The analysis of the placement ratios and audio recordings revealed that the participants struggled to assign specific social cues to the group of voice set (i.e., gender of voice) and voice qualities (i.e., pitch range). Although they understood the definitions and differences correctly, one participant stated, "*a conversational agent has no permanent vocal characteristics because the developers are able to change everything like gender and pitch range*". Other participants argued that "*gender and pitch belong together*" and another participant mentioned, "*it is technically possible to change the gender, so I put it to voice qualities*". These comments indicated that users might not be able to distinguish between voice set cues and voice quality cues. This distinction seems to be unsuitable for the design of CAs since all voice characteristics can be individually modified. Thus, we decided to merge these categories. This is supported by literature since not all researchers followed the three group distinction of Trager (1958) from which these categories were originally derived. Nöth (1995) notes that "*the domain of voice set is not always distinguished from that of voice quality*" (p. 250). Trenholm and Jensen (2011) also refer only to voice qualities and Trager (1958) himself states that both, voice set and voice qualities, are the "*background characteristic of the voice*" (p. 5).

Round 3: The third card sorting round was performed with another five participants. Averaged raw agreement (0.96), as well as Kappa values (0.95), rose to a stronger level of agreement (Landis and Koch, 1977; LeBreton and Senter, 2008), as only a Kappa value remained at 0.9. The average placement ratios further improved to 98% and the single placement ratios revealed complete conformance in seven out of ten categories (Moore and Benbasat, 1991). It was evident that the merging of the two categories voice set and voice qualities resulted in a substantial improvement of correct placements. The analysis of the audio recordings revealed that no participant was confused by the auditory categories anymore. However, we identified a minor issue during the audio recording analysis. Two participants had problems in understanding CMC cues and assigned some CMC cues to other categories. One participant stated, "*emoticons are closely linked to verbal cues*". However, CMC cues appear visually as "*they look fundamentally different than printed linguistic text*" (Garrison et al., 2011, p. 123). Another participant mentioned that he was "*not sure if typefaces can augment or modify a meaning of a message*". Therefore, he was not able to assign this social cue correctly. After interviewing all participants and discussing the meaning of CMC cues, they agreed that it is a valuable category, but "*at first glance, it seemed somewhat abstract*".

The final taxonomy classifies all social cues identified in the SLR in mutually exclusive and collectively exhaustive categories. All categories of the taxonomy were drawn from existing communication theories and consists of four categories on the first hierarchical level and ten subcategories on the second hierarchical level. Table 5 summarizes the definitions of all categories and subcategories and displays the corresponding theoretical references.

Additionally, Fig. 4 depicts the taxonomy of social cues for CAs and the mapping of all 48 identified social cues (and their assigned IDs in square brackets) to their categories and subcategories.

5. Discussion

To answer our research question, we followed a three-step research approach. First, we identified and analyzed existing research on social cues of CAs by conducting a SLR. Second, we used the social cues identified in the SLR as the input for an iterative taxonomy development process in order to develop a taxonomy that classifies social cues into theoretically sound categories and subcategories. Third, we evaluated the mapping of social cues to one of the categories of the taxonomy and verified that categories are natural, simple, and parsimonious.

The taxonomy contributes to the literature by extending existing classifications of social cues of CAs by integrating the four communication systems responsible for creating and transmitting messages in interpersonal communication into a representation that applies to CAs. The taxonomy supports researchers in classifying existing and future research on social cues of CAs and supports practitioners in identifying, implementing, and testing their effects in the design of a CA. To

Table 5
Definitions of taxonomy categories and subcategories.

Category	Definition
Verbal	**Verbal cues refer to cues expressed with written or spoken words** (Knapp et al., 2013; Leathers, 1976; Leathers and Eaves, 2015).
Content	Content cues refer to the strict and literal meaning of a message (i.e., *what is said*) (Collier, 2014; Recanati, 2001).
Style	Style cues refer to the meaningful deployment of language variation in a message (i.e., *how something is said*) (Collier, 2014; Selting, 2009; Tannen, 1984).
Visual	**Visual cues refer to cues that can be seen (except words themselves)** (Leathers, 1976; Leathers and Eaves, 2015; Trenholm and Jensen, 2011).
Kinesics	Kinesic cues refer to all body movements of the agent (Burgoon et al., 2010; Leathers, 1976; Leathers and Eaves, 2015).
Proxemics	Proxemic cues refer to the role of space, distance, and territory in communication (Burgoon et al., 2010; Leathers, 1976; Leathers and Eaves, 2015).
Agent appearance	Agent appearance cues refer to an agent's graphical representation (Burgoon et al., 2010; Leathers, 1976; Leathers and Eaves, 2015).
CMC	Computer-mediated communication (CMC) cues refer to visual elements that can augment or modify the meaning of a text-based message (Kalman and Gergle, 2014; Rezabek and Cochenour, 1998; Walther and Tidwell, 1995).
Auditory	**Auditory cues refer to cues that can be heard (except words themselves)** (Leathers, 1976; Leathers and Eaves, 2015).
Voice qualities	Voice qualities refer to permanent and adjustable characteristics of speech (Burgoon et al., 2010; Nöth, 1995; Trager, 1958).
Vocalizations	Vocalizations refers to nonlinguistic vocal sounds or noises (Burgoon et al., 2010; Nöth, 1995; Trager, 1958).
Invisible	**Invisible cues refer to cues that cannot be seen or heard** (Leathers, 1976; Leathers and Eaves, 2015; Trenholm and Jensen, 2011).
Chronemics	Chronemic cues refer to the role of time and timing in communication (Burgoon et al., 2010; Trenholm and Jensen, 2011; Walther and Tidwell, 1995).
Haptics	Haptic cues refer to tactile sensations on the user's body (Burgoon et al., 2010; Trenholm and Jensen, 2011).

demonstrate that the taxonomy of social cues for CAs is useful, generalizable, and can be applied to classify and identify social cues beyond the initial set of social cues, we present the application of the taxonomy to (1) existing research and (2) three real-world examples of CAs in the next sections. Finally, we discuss limitations of our work and provide avenues for future research.

5.1. Applying the taxonomy to analyze existing research

In order to demonstrate the usefulness and generalizability of the proposed taxonomy, we analyzed existing research on social cues of CAs. First, we applied the taxonomy as an analytical framework to investigate the different types of social cues identified in our initial literature review. Second, we demonstrated that the taxonomy can be applied to classify additional social cues in publications beyond our initial literature review (i.e., additional narrative literature review about ECAs).

To apply the taxonomy as an analytical framework, we used the results of the literature review and analyzed the mapping of each of the 48 social cues to the corresponding 92 publications. This assignment was described in Section 4.1. Moreover, we relied on the social cue-to-category/subcategory mapping of the taxonomy (see Fig. 4), which was also carried out by the authors of this article and further evaluated by 15 participants in the card sorting procedure described in Section 4.3. The assignment of all 92 publications to the corresponding social cue categories and subcategories, depending on whether they analyzed such social cues or not, is depicted in Table A5 in the Appendix. The analysis of this assignment showed that the identified social cues of CAs are dominated by a few social cue categories and subcategories (see Fig. 5). While most identified publications analyze visual ($n = 61$ publications) and verbal cues ($n = 42$), only 19 publications analyze auditory and 12 invisible cues. Moreover, the results show that the following three social cue subcategories are extensively researched: appearance cues ($n = 25$), content cues ($n = 31$), and kinesic cues ($n = 31$). In contrast, certain social cue subcategories are largely underrepresented in our sample such as proxemic cues ($n = 3$), haptic cues ($n = 3$), and CMC cues ($n = 3$). Thus, by following this approach, researchers can use the taxonomy as a framework to systematically classify their findings of social cue phenomena into one of the social cue categories (i.e., verbal, visual, auditory, invisible) and subcategories. This supports researchers in overcoming different terminology and domain restrictions and facilitating discussions.

To demonstrate that the taxonomy is valuable for classifying social cues beyond the initial set of identified publications, we reviewed and classified additional publications investigating social cues of CAs. Therefore, we focused on research about the most comprehensive form of a CA (i.e., ECAs). As ECAs support a broad bandwidth and multimodal realization of different types of social cues, publications about ECAs provide a great source of additional social cues of CAs to test our taxonomy. Therefore, we conducted an additional narrative literature review (Paré et al., 2015) in order to synthesize prior study findings that investigate social cues of ECAs. Our search strategy was to retrieve the ten most cited publications in Google Scholar by using the search term "*embodied conversational agent*". Therefore, we searched Google Scholar and ordered publications by citations.[1] Then, we excluded five books, three publications that were already included in our literature review (i.e., Bickmore and Cassell, 2001; Bickmore and Picard, 2005; Cassell et al., 1999), two publications that investigate physical robots, and one editorial comment. Finally, we selected the remaining ten publications with the highest number of citations. Each author read the publications separately to identify their investigated social cues of ECAs. After agreeing on a list of social cues, each author assigned them separately to the corresponding categories and subcategories of the taxonomy. Social cues on which there was disagreement were discussed and placed in mutually agreeable categories with the moderation of another researcher not involved in this study.

As depicted in Table 6, we identified a large number of social cues of ECAs and were able to use the taxonomy to classify each of the identified social cues into one of the corresponding social cue categories and subcategories of the taxonomy. For example, Rosis et al. (2003) investigate how an ECA can communicate complex information through the facial features, facial expressions, head movements, and eye movements and investigates the impact on believability and persuasion of the CA. Also, we were able to identify additional social cues that were not covered in the initial literature review and can now be added to the knowledge base. For example, we found additional verbal content cues: Cassell and Thorisson (1999) investigate verbal acknowledgment (i.e., state "*okey-dokey*", "*let's go to Jupiter*" as a part of an action) and confused expressions (i.e., expressions when the CA does not understand the message of the user). Ryokai et al. (2003) investigate the impact of decontextualized language (i.e., quoted speech

[1] We used the tool "*publish or perish 6*" to query Google scholar and to sort by citations.

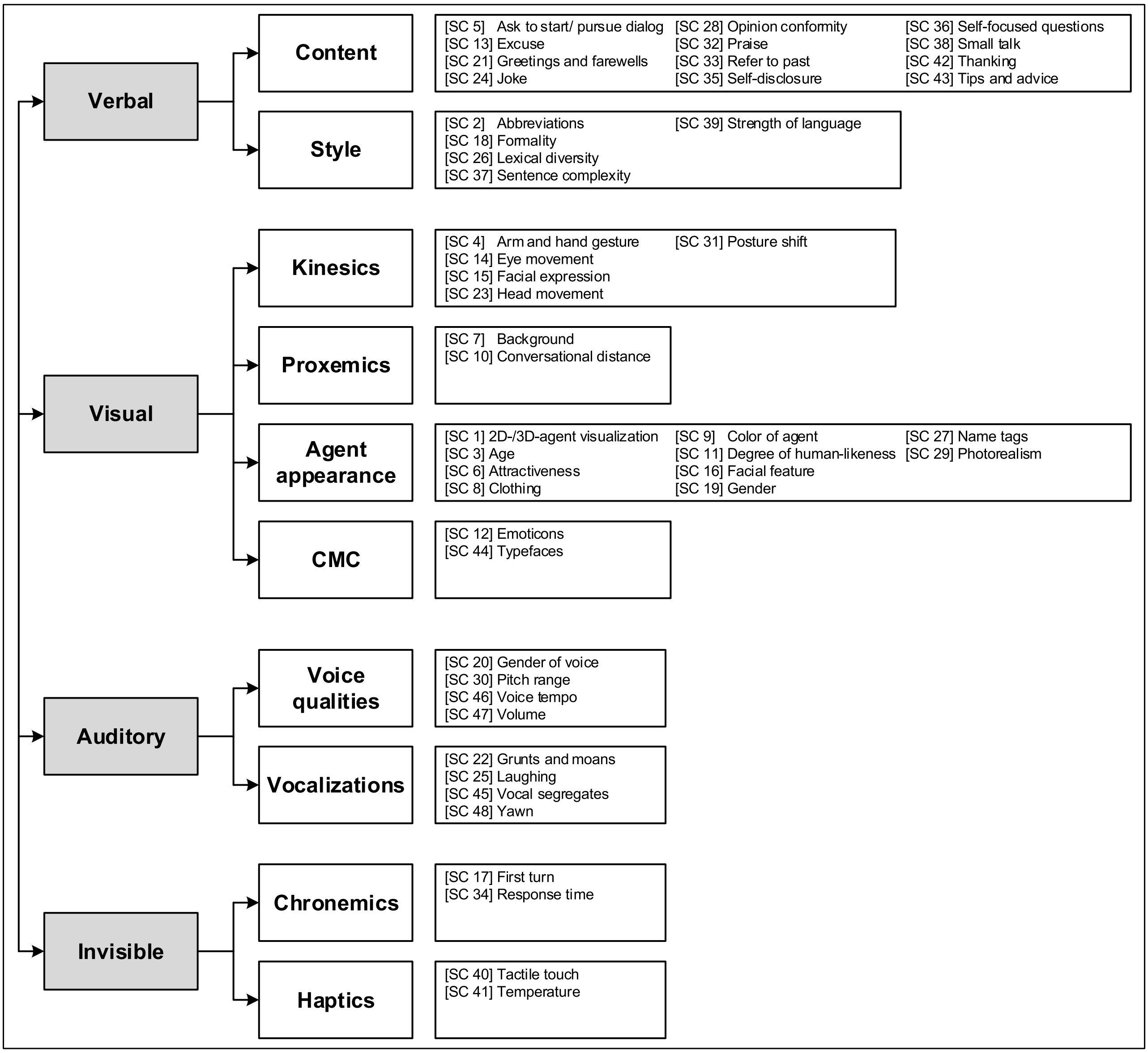


Fig. 4. Taxonomy of social cues for conversational agents.[4]

such as "*Oh, sheriff*"), temporal expressions (e.g. "*today I'm going to…*"), and spatial expressions (e.g. "*from the other side of the forest*"). Since we were able to classify all identified social cues to one of the taxonomy's categories, we argue that the taxonomy can be used to classify and accumulate research about social cue of CAs beyond the initial list of identified social cues. Therefore, the categories of the taxonomy seem to be a suitable starting point also to classify the large number of social cues implemented in ECAs.

5.2. Applying the taxonomy to analyze three real-world examples

To further illustrate the taxonomy's usefulness in identifying social cues, we applied it to exemplarily analyze the different social cues embedded in the design of three real-world CAs. Therefore, we investigated the social cues of (1) a text-based CA (Poncho on Facebook Messenger) (D'Arcy, 2016; Heath, 2018), (2) a voice-based CA (Amazon's Alexa[2]), and (3) a comprehensive ECA (SARA[3]). We selected these three examples as they represent typical instantiations of different types of CAs. We chose Poncho because it has been one of the earliest CAs on Facebook Messenger (D'Arcy, 2016). We chose Alexa because it currently has the largest market share in the smart speaker market (Forbes, 2018). Finally, we chose SARA as it is one of the most advanced ECAs developed at Carnegie Mellon University's ArticuLab (Cassell, 2019). Again, the analysis was carried out by all authors of this article separately and all disagreements in identified social cues were resolved by discussion. However, it must be noted that our analysis is non-exhaustive and primarily serves to demonstrate how the taxonomy can be used to identify implemented social cues of existing CAs.

First, we investigated Poncho, a text-based CA (i.e., chatbot) on Facebook Messenger that provides weather information and sends daily weather forecasts (Heath, 2018). Since Poncho does not communicate via voice and only has a static profile and background picture, we excluded irrelevant social cue categories for Poncho's current design, namely all auditory and kinesic cues. Consequently, the taxonomy enabled us to systematically identify three categories and seven

[2] https://developer.amazon.com/de/documentation/, last accessed on 25.06.2019
[3] http://articulab.hcii.cs.cmu.edu/projects/sara/, last accessed on 25.06.2019

[4] The description of each social cue is provided in Table A1 in the appendix.

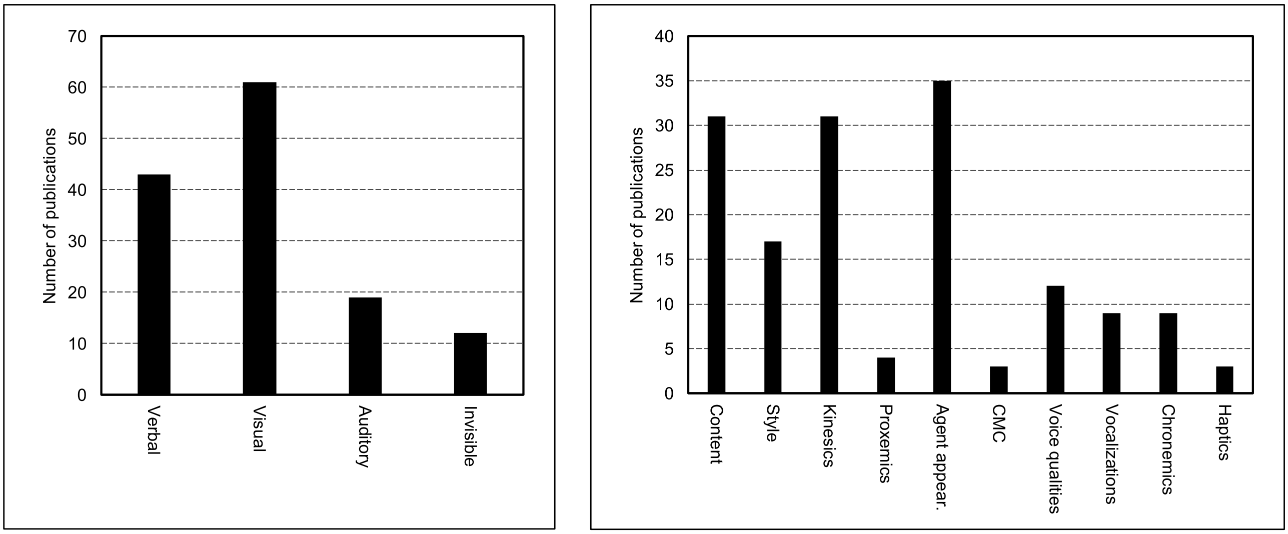


Fig. 5. Overview of social cues investigated in publications identified in the literature review (multiple assignments of one publication to several groups is possible).

subcategories of social cues that Poncho may exhibit. Next, we identified many visual and verbal social cues, even before we started a conversation with Poncho. For example, Poncho exhibits several visual cues like a name tag and a comic-like profile picture (i.e., a low degree of photorealism) showing a cat with a smiling face, a yellow raincoat, and some blue and yellow background. Furthermore, Poncho uses a neutral typeface and introduces itself to the user with a short statement that includes verbal cues (i.e., greetings and an informal conversation style). After a short conversation, Poncho uses four additional verbal cues as Poncho refers to the past, tells a joke, engages in small talk, and uses chronemic cues by delaying its responses (i.e., using different response times). In addition, Poncho's messages have a rather low sentence complexity. In summary, we could exemplary identify a total of 11 social cues that were (either intentionally or unintentionally) implemented in Poncho's design.

Second, we investigated Alexa, a disembodied voice-based CA that serves as a personal assistant on Amazon's Echo devices. In our analysis, we focused on the original Echo devices without a screen in the English language. Since Alexa does not have a visual representation and does not use text-based communication, we first excluded all visual cues and CMC cues as users can only see the physical device itself. Consequently, the taxonomy enabled us to systematically identify three categories and six subcategories of social cues that Alexa may exhibit. Next, we identified a wide range of verbal cues. For example, Alexa can tell jokes, greetings and farewells, as well as engage in small talk. Developers can also build skills that convey additional content cues such as self-disclosure or self-focused questions. Regarding verbal style, Alexa adopts a rather informal style and aims to avoid complex sentences and abbreviations (Amazon, 2019c). Nevertheless, developers have many options to implement additional content and style cues when developing a skill. Moreover, when interacting with Alexa, many auditory and verbal cues can be identified. For example, Alexa has a female voice and although its name can be changed, most users prefer to call "*her*" by the female name Alexa (Gao et al., 2018). From an auditory perspective, Alexa comes in its standard configuration with a specific pitch range and voice tempo, which can be further customized by the skill developers. More specifically, they can customize Alexa's volume, pitch range, and voice tempo using the Speech Synthesis Markup Language (SSML) (Amazon, 2019a). However, in order to avoid that "*Alexa sound (s) like ET*", the amount of change applied to these voice qualities is limited (Hermann, 2019; Myers, 2017; Perez, 2017). For example, using the "*Whisper Mode*", users can whisper to Alexa and it whispers back. Moreover, Alexa uses vocalizations as, for example, it can laugh on command (``*Alexa, laugh*") (Chokshi, 2018) or responds with "*Hmm, I don't know that*" (Amazon, 2019b). Finally, we reviewed the invisible cues and identified that Alexa can use chronemic cues. Although it automatically pauses after a period, developers can implement delays of up to 10 s to customize Alexa's response time (Amazon, 2019a). In summary, we could exemplary identify a set of 16 social cues that were (either intentionally or unintentionally) implemented in Alexa's design. However, many more can be added by developers of Alexa skills (e.g., verbal content and style cues).

Third, we investigated SARA (Socially-Aware Robot Assistant), an ECA that serves as a personal assistant for conference attendees (e.g., at the World Economic Forum annual meeting). SARA helps attendees find sessions and people to meet based on their interests (Bishop, 2018; Cassell, 2019). We analyzed the social cues of SARA based on identified videos, papers, and news articles. Consequently, the investigation is non-exhaustive and primarily serves demonstration purposes. In general, SARA exhibits a wide range of social cues as it is a comprehensive and fully embodied CA. Consequently, the taxonomy enabled us to systematically identify four categories and 9 subcategories of social cues (i.e., all except haptic cues) that SARA may exhibit. First, SARA uses several verbal cues such as greetings and farewells, express a name (e.g., "*Hi, I am SARA*"), self-disclosure (e.g., "*I've been asked to play matchmaker by helping attendees find sessions to attend and people to meet*" or "*I certainly find it difficult to remember information without noting it down*"), praise (e.g., "*I've never met someone like you before. It's refreshing*"), and reference to the past. Its verbal style can be considered as rather formal (e.g., "*May I ask your name?*" or "*I can send a message on your behalf*"), rather complex, and with high lexical diversity. In addition, many visual cues were identified in SARA's design. For example, SARA has a comic-like, 3D visual appearance of a female person with black hair, glasses, and rather formal clothing. SARA stands behind a desk with a screen showing the logo of the World Economic Forum behind it. Moreover, SARA uses arm and hand gestures (e.g., touching its head), head movement (e.g., nodding), eye movement (e.g., gaze shift, blinking, eyebrow lifting), facial expressions, such as smiling (e.g., when taking a selfie with a conference attendee), and shifts its posture. Additionally, SARA exhibits auditory cues. It has a female voice and varies its voice quality (Cassell, 2019). Based on the information available to us (e.g., videos, papers, news articles), we could not identify any vocalizations. Finally, SARA also uses chronemic cues. For example, there is a pause of a few seconds, when it searches for recommendations. In summary, we could exemplary identify a set of 24 social cues that were (either intentionally or unintentionally) implemented in SARA's design. In contrast to the other two examples, SARA exhibits a larger number of visual cues due to its realistic, animated 3D-representation. In summary, the taxonomy enabled us to

Table 6
Narrative literature review and analysis about social cues of ECAs and their corresponding classification according to the taxonomy.

Source	Investigated cue	Social response	Social cue category	Social cue subcategory
Cassell (2000b)	Greeting and farewell, eye movement, facial expression, head movement, posture shifts, arm and hand gestures, degree of human-likeness, vocal segregate, response-time.	Impact on CA's collaboration, cooperativeness, natural language capabilities, and benefits on task.	Verbal, Visual, Auditory, Invisible	Content, Kinesics, Appearance, Vocalization, Chronemics.
Cassell and Thorisson (1999)	Acknowledgement, confused expression, strength of language, eye movement, head movements, facial expression, posture shift, facial feature.	Change of user's speech patterns, hesitations, frustrations, rating of lifelikeness, and fluidity.	Verbal, Visual	Verbal, Style, Kinesics, Appearance.
Rosis et al. (2003)	Facial feature, facial expression, head movement, eye movement.	Impact on believability and persuasion of CA.	Visual	Kinesics, Appearance.
Bickmore and Cassell (2005)	Smalltalk, facial expression, hand and arm gesture, head movement, posture shifts, eye movement, facial feature, vocal segregates, pitch range.	Impact on knowing, liking, feeling close, feeling comfortable, perceived friendliness, warmth, information, and knowledge.	Verbal, Visual, Auditory	Content, Kinesics, Appearance, Vocalizations.
Thiebaux et al. (2008)	Posture shifts, head movement, eye movement, response time.	*No user reactions reported*	Visual, Invisible	Kinesics, Chronemics.
Kopp and Wachsmuth (2004)	Arm and hand gestures, posture shift, head movement, facial expression, response time.	*No user reactions reported*	Visual	Kinesics.
Cassell (2001)	Eye movement, posture shift, arm and hand gestures, facial expression, pitch range, vocal segregates.	Impact on user's trust and perceived system skills.	Visual	Kinesics, Appearance, Voice qualities.
Ryokai et al. (2003)	Decontextualized language, temporal expressions, spatial expressions, sentence complexity, head movement, eye movement, facial expression, background, facial feature, greeting, age, gender.	Facilitates peer interactions, improves children's quoted speech, and temporal and spatial expressions.	Verbal, Visual	Content, Style, Appearance, Kinesics, Proxemics.
Carolis et al. (2004)	Facial expression, eye movement, head movement, facial feature.	*No user reaction reported*	Visual	Kinesics, Appearance.
Bailenson and Yee (2005)	Eye movements, head movements, facial feature, degree of human-likeness, gender.	Changes in user's self-report, and in cognitive and behavioral measures.	Visual	Kinesics, Appearance.

Note: The analysis is non-exhaustive and primarily serves to demonstrate how the taxonomy can be used to classify social cues of CAs.

systematically identify a wide variety of different social cues implemented in the three real-world CAs (see Table 7).

5.3. Limitations and future research

Although we followed established guidelines and aimed to ensure a high rigor in the research project to build a taxonomy of social cues for CAs, there are limitations that should be considered.

First, the search strategy might have missed relevant publications. As with any literature review, the identified and selected publications have an impact on the social cue identification process. Thus, we acknowledge that a different search strategy and selection process might have resulted in a different list of identified social cues. Therefore, we do not argue that the list of identified social cues of CAs is exhaustive and represents all investigated social cues in the extensive body of existing knowledge. Particularly, many researchers may not have framed their study as an investigation of social cues of CAs. Thus, our search strategy and search term might have missed other relevant publications and their corresponding social cues. Particularly, other search terms could have been included to reveal additional social cues (e.g., dialogue systems, spoken dialogue systems, interactive voice response (IVR) systems). However, we argue that the set of 48 social cues identified in 92 relevant publications represents a sufficient foundation to provide researchers and practitioners with an initial overview of different social cues of CAs. Moreover, we argue that the initial list of social cues is suitable as a starting point for our iterative taxonomy development process as we did not only follow an empirical-to-conceptual taxonomy development, but also derived all categories of the taxonomy by closely following a conceptual-to-empirical taxonomy development process (Nickerson et al., 2013).

Second, the level of abstraction of the identified social cues is the result of the authors' coding process and our conceptualization of social cues. Thus, all cues need to be design features of a CA salient to the user that presents a source of information but do not account for the underlying meaning they are supposed to convey (i.e., their social signal). However, drawing a clear line between a social cue and a social signal might be difficult sometimes. Thus, we abstracted the investigated cues at the level to that they are perceived by the user and can be designed by the researcher or practitioner (e.g., tempo, gesture). However, we did not break them down into their different design characteristics (tempo: fast or slow, volume: loud or quiet) or the communicative functions for specific user, tasks, and contexts of an interaction (e.g., emblems, illustrators, affect displays, regulators, adaptors, see Ekman, 1973). Thus, we acknowledge that the level of abstraction of social cues can also be further broken down. For example, tune (as a form of melody) can be operationalized through several identified social cues (e.g., pitch range, tempo) and then by itself constitutes an own meaningful social cue that perceived by a human can transform in a meaningful social signal. Therefore, future research can extend the hierarchical structure of the taxonomy by integrating additional social cue sub-category layers that capture additional levels of abstraction.

Third, although the categories of taxonomy were derived from interpersonal communication theory, the final classification of the identified social cues is influenced by the authors' subjective assessment. Therefore, we closely followed the established interpersonal communication theory and applied the method by Nickerson et al. (2013) as objectively and rigorously as possible. We discussed deviations among the authors extensively, reviewed relevant interpersonal communication theory, and resolved them by mutual agreement. Finally, we argue that the ten subcategories of the taxonomy are mutually exclusive, but we do not argue that they are collectively exhaustive as a new category may be added (e.g., olfactory). However, we argue that the social cue categories (i.e., verbal, visual, auditory, invisible) are mutually exclusive and collectively exhaustive as they are based on the well-established categories from existing classifications in interpersonal communication (Leathers, 1976). However, not all categories of the taxonomy will be equally important for all researchers. Therefore, future studies could extend the taxonomy, by including additional social cues and developing new sub-categories of other technologies such as physical robots (e.g., Hegel et al., 2011; Wiltshire et al., 2014). This would verify whether the taxonomy is generalizable and applicable to other, not yet identified social cues and to other, not yet investigated contexts and types of CAs. In particular, since the initial set of identified social cues is non-exhaustive, future work can investigate the generalizability of the taxonomy to other contexts (e.g., Cronbach, 1972). This could expand the applicability of this taxonomy beyond CAs and could create a more complete classification of social cues.

Fourth, the assignment of social cues to only one of the four communication systems (i.e., verbal, visual, auditory, invisible) should be reflected critically. Several interpersonal communication researchers point out that all communication systems transfer meaning by interacting, reinforcing, and conflicting with the other systems and thus, never act on their own (Burgoon et al., 2010; Knapp et al., 2013; Leathers, 1976; Leathers and Eaves, 2015). As a consequence, several researchers investigate how different social signals are created through the co-occurring, temporal arrangement, multimodal realization, and reciprocal adaptation of social cues of CAs (Bevacqua et al., 2010; Chollet et al., 2014; Kopp et al., 2006; Pelachaud, 2005). Although social cues usually do not occur isolated from each other, the distinction is commonly practiced to understand the relevant elements (Burgoon et al., 2010). However, researchers and practitioners should be aware of potential interrelations between two or more social cues and thus, should apply the taxonomy with care. Particularly, as meaningful social signals include the complex constellation of several social cues and the context of the interaction (Vinciarelli et al., 2012), future work could further investigate the co-occurring, temporal, multimodal, and reciprocal relationships of social cues in experiments in order to investigate outcomes of a specific social cue design (i.e., what functions and meanings they convey), how an outcome can be operationalized in various contexts (i.e., technical and multimodal realization), and in which temporal and sequential order the social cue design should be displayed. To achieve this, future research could leverage ontological models in order to store effects of individual and multimodal social cue realizations and provide tool support for a meaningful social cue design (Feine et al., 2019b).

Fifth, we only evaluated the taxonomy with potential users. However, according to Nickerson et al. (2013), a taxonomy needs to be applied by real users to thoroughly assess its usefulness. Although the taxonomy meets all formal criteria (Nickerson et al., 2013) and we evaluated the categories and definitions with potential users from both research and practice (Moore and Benbasat, 1991), further evaluation with real users in a real-contexts should be carried out at a later stage. To facilitate this process, we provide researchers and practitioners with a taxonomy web application that eases access to the study findings and helps to further accumulate the existing body of knowledge about social cues of CAs.

6. Conclusion

In this article, we developed and evaluated a comprehensive taxonomy of social cues for CAs that extends existing classifications. To demonstrate its usefulness, we applied the taxonomy to classify and analyze existing research and to identify social cues in the design of three real-world CAs. Our work contributes to the body of knowledge on designing CAs. It provides guidance for researchers to systematically classify research on social cues of CAs from different research fields and supports practitioners, such as CA designers, in identifying, implementing, and testing possible types of social cues. Thus, both practitioners and researchers can use the taxonomy as a starting point for further, interdisciplinary research and design in order to avoid reinventing the wheel in the design of CAs.

Table 7
Exemplary analysis of social cues implemented in three real-world CAs.

	Example Type of CA	**Poncho Text-based CA**	**Alexa Voice-based CA**	**SARA Embodied CA**
Identified Social Cues	**Verbal**	• **Content**: greeting and farewells, refer to past, joke, small talk • **Style**: formality (informal), sentence complexity (low)	• **Content:** greeting and farewells, joke, self-disclosure, self-focused questions, small talk, express name (Alexa) • **Style:** formality (informal), sentence complexity (low)	• **Content:** greeting and farewells, self-disclosure, praise, refer to past, opinion conformity, express name (SARA) • **Style:** formality (formal and polite), sentence complexity (complex sentences), lexical diversity (high)
	Visual	• **Appearance:** name tag (Poncho), facial features (smile), photorealism (comic-like), • **CMC:** typeface (neutral)	-	• **Appearance:** 3D visualization, gender (female), photorealism (comic-like), facial features (black hair, glasses), clothing (formal) • **Proxemics:** background (desk and screen) • **Kinesics:** arm and hand gestures, head movement, eye movement, facial expressions, posture shift
	Auditory	-	• **Voice qualities:** gender of voice (female), volume, pitch range, voice tempo • **Vocalizations:** whisper, laughing, vocal segregates ("Hmm")	• **Voice qualities:** gender of voice (female), pitch range (varying)
	Invisible	• **Chronemic:** response time	• **Chronemic:** response time	• **Chronemic:** response time
No. of identified social cues		11	16	24

Note: The analysis is non-exhaustive and primarily serves to demonstrate how the taxonomy can be used to identify implemented social cues of existing CAs.

Declaration of Competing Interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix

Acknowledgment

We thank the associate editor and the anonymous reviewers for their excellent comments and constructive feedback that have significantly improved the quality of this article. The authors also thank all participants of the evaluation for their help in evaluating and improving the taxonomy.

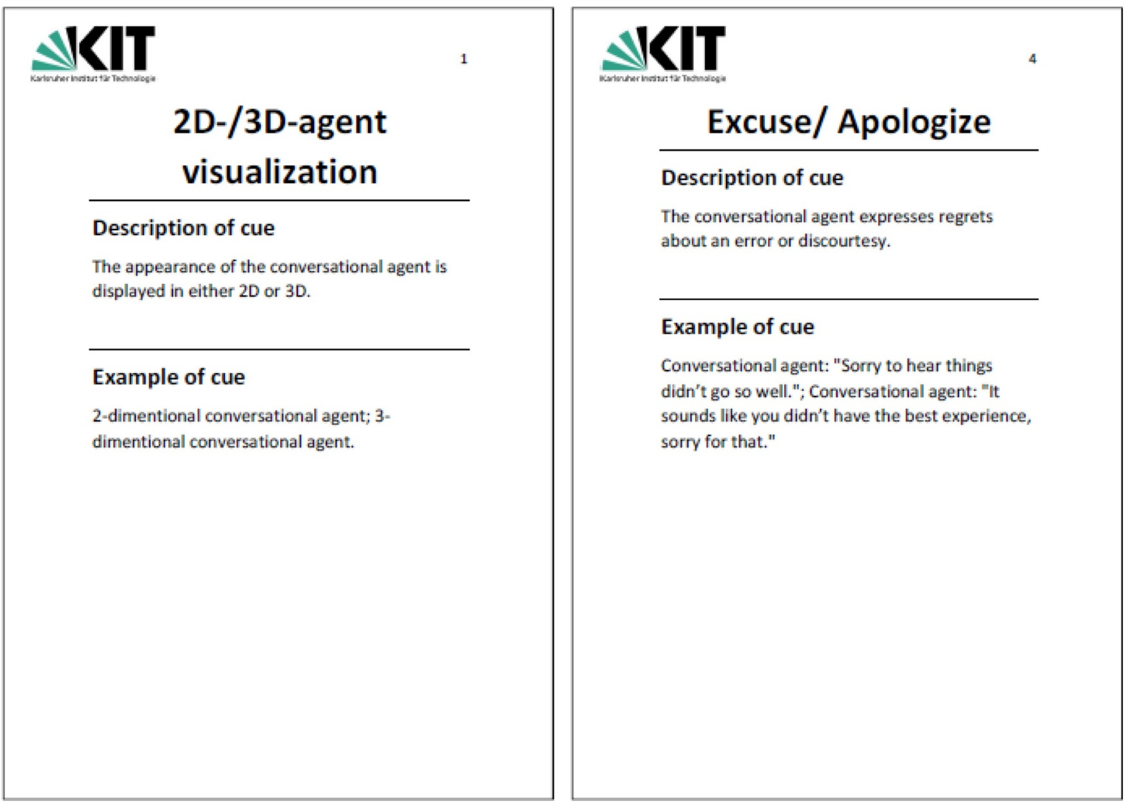


Fig. A1. Example cards of card sorting rounds.

Table A1
List, description, and examples of identified social cues.

ID	Social cue	Description	Examples
[SC 1]	2D-/3D-agent visualization	The appearance of the CA is displayed in either 2D or 3D.	2-dimensional CA, 3-dimensional CA.
[SC 2]	Abbreviations	The CA uses shortened forms of a word or phrase.	CA: “No. of times in the last year having a feeling of guilt?” instead of “Number of times in the last year having a feeling of guilt”.
[SC 3]	Age	The biological age of the appearance of the CA.	Old agent visualization (e.g., grey hair), young agent visualization (e.g., shiny skin, full hair).
[SC 4]	Excuse/ Apologize	The CA expresses regrets about an error or discourtesy.	CA: “Sorry to hear things didn't go so well”, CA: “It sounds like you didn't have the best experience, sorry for that”
[SC 5]	Arm and hand gesture	The CA moves its arm or hand.	Point the finger at something, raise an arm to express an idea or feeling.
[SC 6]	Ask to start/ pursue dialog	The CA requests the user's permission to start, continue, or end the conversation.	CA: “Do you want to start the session?”, “now we are finished. Do you want to continue with the conversation?”
[SC 7]	Attractiveness	The visual characteristics of the CA are perceived as being pleasing or beautiful.	Symmetrical face of the CA.
[SC 8]	Background	The picture, scene, or design around the CA forms a setting for the interaction.	CA being in front of a hospital, CA being in front of a police station.
[SC 9]	Clothing	The CA wears items, such as clothes and trousers, to cover, protect, or decorate its body.	CA wears a pullover or a suit.
[SC 10]	Color of agent	The overall color of the appearance of the CA.	White, black, or pink color of CA.

(*continued on next page*)

Table A1 (*continued*)

ID	Social cue	Description	Examples
[SC 11]	Conversational distance	The spatial separation between the CA and its environment during a conversation.	CA appearing far away when conversing or CA is really close to user.
[SC 12]	Degree of human-likeness	The appearance of the CA ranges from a natural human form to the shape of a cartoon or artificial object.	CA has a human-like, object-like, or robot-like appearance.
[SC 13]	Emoticons	The CA sends textual or pictorial depictions of facial expressions.	:-) ;-) ^_^
[SC 14]	Eye movement	The CA moves its eyes to intentionally or unintentionally fixate or track objects.	CA is looking in the direction of the other person's face.
[SC 15]	Facial expression	The CA expresses a gesture by executing one or more motions with his facial muscles.	CA smiles or moves its eyebrows.
[SC 16]	Facial feature	The static and non-changing characteristics of the CA's face.	Static smile of CA.
[SC 17]	First turn	The CA starts the conversation and is the first to say something about a topic.	CA goes first during an interaction.
[SC 18]	Formality	The CA expresses words and sentences that are either compliant or incompliant with conventional rules.	CA addresses user with last name instead of first name (e.g., “Hello Mr. Tree” instead of “Hello Peter”).
[SC 19]	Gender	The CA belongs to either one of the two sexes (male, female) or it is ambiguous.	Male, female, ambiguous CA appearance.
[SC 20]	Gender of voice	The voice of the CA belongs to either one of the two sexes (male, female) or it is ambiguous.	Female voice, male voice.
[SC 21]	Greetings and Farewells	The CA expresses a word of welcome or marks someone's departure.	CA: “Nice to meet you”, “Welcome”, “Goodbye”.
[SC 22]	Grunts and moans	The CA makes a low sound.	Grunts and moan to express emotions.
[SC 23]	Head movement	The CA moves its head.	Head nodding, head turning.
[SC 24]	Joke	The CA expresses phrases that cause amusement or laughter.	CA: “I am very good at sleeping. I can do it with my eyes closed”.
[SC 25]	Laughing	The CA makes sounds that indicate its amusement.	Sounds of laughing (e.g., “ahahaha”)
[SC 26]	Lexical diversity	The CA expresses many different unique words.	CA uses various synonyms such as “also”, “further”, “likewise”, “besides”, etc.
[SC 27]	Name tag	A tag or badge with the name of the CA.	Name of the agent is displayed on the agent or within the interface (e.g., Anna).
[SC 28]	Opinion conformity	The CA shares the same opinion as the user regarding some topic.	CA: “I agree, this painting is beautiful”.
[SC 29]	Photorealism	The appearance of the CA ranges from an extremely photorealistic appearance to a comic appearance.	Real photo of a human being, comic figure.
[SC 30]	Pitch range	The degree of variation from the CA's average pitch.	Small pitch range is a very monotone voice, whereas a high pitch range voice sounds really animate.
[SC 31]	Posture shift	The CA moves its upper or lower body.	CA turns its body.
[SC 32]	Praise	The CA expresses approval, gratitude, or admiration for the user.	CA: “You did a great job”, “You are really clever”.
[SC 33]	Refer to past	The CA refers to content from past conversations.	CA: “Hello Peter! Welcome back. Did you enjoy the recipe that I told you last Tuesday?”.
[SC 34]	Response time	The amount of time it takes for the CA to respond to the user's input.	Immediate response (e.g., 0 s delay), slow response (e.g., 2 s delay).
[SC 35]	Self-disclosure	The CA reveals intimate information about itself.	CA: “People lie. I tell lies too if I have to”.
[SC 36]	Self-focused questions	The CA asks questions about itself.	CA: “How good did I support you?”, “Did you like my performance?”
[SC 37]	Sentence complexity	The complexity of the CA's sentences in terms of number of used words, grammatical quality, and length of chosen words.	Well-constructed sentences with the use of sophisticated words or grammatically wrong sentences with simple short words.
[SC 38]	Small talk	The CA engages in casual conversations.	CA: “How are you today?”, “What about the weather?”.
[SC 39]	Strength of language	The messages of the CA range from strong and assertive statements to submissive and equivocal statements.	Strong and assertive language (e.g., “you should definitely do this”, “you have to do this”) or submissive and equivocal language (e.g., “perhaps you should do this”).
[SC 40]	Tactile touch	User perceives a tactile sensation initiated by the CA.	User gets touched.
[SC 41]	Temperature	The user perceives a warm or cold surface while interacting with a CA.	Cold keyboard, warm keyboard.
[SC 42]	Thanking	The CA expresses thankfulness to the user.	CA: “Thanks for playing!”, “Thank you for your time”.
[SC 43]	Tips and advice	The CA provides help on a specific task that the user needs to solve.	CA: “I will help guide you through the screening process”, “I guess the answer is B”.
[SC 44]	Typeface	The design of letters and symbols used by the CA in the chat window.	Times new roman, comic sans.
[SC 45]	Vocal segregates	The CA makes sounds that get in the way of fluent speech.	CA: "uhs", "ums", stuttering.
[SC 46]	Voice tempo	The pace of the CA's voice.	Slow pace to emphasize certain ideas. Quicker pace to show excitement or humor.
[SC 47]	Volume	The loudness of the CA's voice.	Extreme loud voice, medium loud voice, quiet voice.
[SC 48]	Yawn	The CA makes a sound of inhaling deeply to express tiredness or boredom.	Yawning due to tiredness or boredom.

Table A2
Social cue placement ratios for the first sorting round.

		Actual characteristics											Total	Hit ratio
		Content	Style	Kinesics	Proxemics	Agent app.	CMC	Voice set	Voice qual.	Vocalizat.	Chronem.	Haptics		
Assigned characteristics	Content	59	1			1					1		62	0.95
	Style		24										24	1.00
	Kinesics			22									22	1.00
	Proxemics				9								9	1.00
	Agent app.			3	1	47							51	0.92
	CMC					2	10						12	0.83
	Voice set							5	4				9	0.56
	Voice quality								11				11	1.00
	Vocalizations									20			20	1.00
	Chronemics	1									9		10	0.90
	Haptics											10	10	1.00
		Total placements: 240				Hits: 226			Overall hit ratio: 0.94					

Table A3
Social cue placement ratios for the second sorting round.

		Actual characteristics											Total	Hit ratio
		Content	Style	Kinesics	Proxemics	Agent app.	CMC	Voice set	Voice qual.	Vocalizat.	Chronem.	Haptics		
Assigned characteristics	Content	58	1										59	0.98
	Style	2	24				1						27	0.89
	Kinesics			23									23	1.00
	Proxemics				10								10	1.00
	Agent app.			2		48							50	0.96
	CMC					2	9						11	0.82
	Voice set							5	2				7	0.71
	Voice quality								13				13	1.00
	Vocalizations									20			20	1.00
	Chronemics										10		10	1.00
	Haptics											10	10	1.00
		Total placements: 240				Hits: 230			Overall hit ratio: 0.96					

Table A4
Social cue placement ratios for the third sorting round.

		Actual characteristics										Total	Hit ratio
		Content	Style	Kinesics	Proxemics	Agent app.	CMC	Voice qual.	Vocalizat.	Chronem.	Haptics		
Assigned characteristics	Content	60										60	1.00
	Style		25				2					27	0.93
	Kinesics			25								25	1.00
	Proxemics				10							10	1.00
	Agent app.					50	1					51	0.98
	CMC						7					7	1.00
	Voice quality							19				19	1.00
	Vocalizations							1	20			21	0.95
	Chronemics									10		10	1.00
	Haptics										10	10	1.00
		Total placements: 240				Hits: 236		Overall hit ratio: 0.98					

Table A5
Assignment of publications to taxonomy.

Publication	Category									
	Verbal		Visual				Auditory		Invisible	
	Content	Style	Kinesics	Proxemics	Agent appearance	CMC	Voice qualities	Vocalizations	Chronemics	Haptics
Andonov et al. (2016)	X									
Appel et al. (2012)			X						X	
Becker et al. (2004)			X				X	X		
Becker et al. (2005)			X					X		
Bee et al. (2009)			X							
Beer et al. (2015)			X		X					
Beldad et al. (2016)					X					
Beun et al. (2003)					X					
Bickmore and Picard (2005)	X	X	X	X						
Bickmore et al. (2010)	X		X				X			X
Bickmore and Cassell (2001)	X		X				X			

(*continued on next page*)

Table A5 (*continued*)

Publication	Category Verbal		Visual				Auditory		Invisible	
	Content	Style	Kinesics	Proxemics	Agent appearance	CMC	Voice qualities	Vocalizations	Chronemics	Haptics
Bonito et al. (1999)			X							
Brahnam and Angeli (2012)					X					
Brandão et al. (2013)						X				
Braslavski et al. (2018)		X								
Cafaro et al. (2016)	X		X	X						
Campano et al. (2015)	X									
Candello et al. (2017)						X				
Cassell and Bickmore (2000)	X		X							
Cassell et al. (1999)			X					X		
Catrambone et al. (2002)	X				X					
Chae et al. (2016)					X					
Chue et al. (2018)										X
Cowell and Stanney (2005)			X		X		X	X		
Danilava et al. (2013)									X	
Derrick and Ligon (2014)	X									
Ding et al. (2014)			X					X		
Dybala et al. (2009)		X								
Endrass et al. (2010)	X									
Fogg (2002)	X	X			X				X	
Fogg and Nass (1997)	X									
Forlizzi et al. (2007)					X					
Gebhard et al. (2014)	X		X						X	
Guo et al. (2016)	X		X							
Hastie et al. (2016)	X									
Hayashi (2016)			X		X					
Hess et al. (2005)			X				X			
Hoffmann et al. (2009)	X									
Hone (2006)	X				X					
Huisman et al. (2014)										X
Isbister and Nass (2000)		X	X	X						
Kang and Gratch (2011)	X									
Kang and Fort Morie (2013)	X									
Keeling et al. (2004)					X					
Kim et al. (2016)					X					
Klein et al. (2002)	X									
Knijnenburg and Willemsen (2016)					X					
Kraemer et al. (2016)			X		X					
Krämer et al. (2013)			X							
Lee and Nass (2003)		X					X			
Lee and Choi (2017)	X									
Li and Graesser (2017)		X								
Li et al. (2017)	X	X			X	X				
Lisetti et al. (2013)	X		X							
Lortie and Guitton (2011)	X	X								
Louwerse et al. (2005)					X		X			
Mayer et al. (2006)		X								
McBreen (2002)					X					
McBreen and Jack (2001)			X		X					
Mersiol et al. (2002)			X		X		X			
Moon (2000)	X									
Moon and Nass (1996)		X							X	
Moon and Nass (1998)		X							X	
Morkes et al. (1999)		X							X	
Wuenderlich and Paluch (2017)					X					
Nass and Moon (2000)	X	X			X		X		X	
Nass et al. (1999)	X									
Nass et al. (1997)							X			
Nass et al. (1995)		X			X				X	
Nass et al. (1994)	X						X			
Niculescu et al. (2010)					X					
Niewiadomski et al. (2013)			X					X		
Nowak (2004)					X					
Nunamaker et al. (2011)			X		X					
Pertaub et al. (2001)			X							
Pecune et al. (2015)			X					X		
Pfeifer and Bickmore (2009)								X		
Puetten et al. (2010)			X							
Puzakova et al. (2013)		X								
Richards and Bransky (2014)	X									
Rickenberg and Reeves (2000)			X							
Rossen et al. (2008)					X					
Sah and Peng (2015)		X			X					
Shiban et al. (2015)					X					

(*continued on next page*)

Table A5 (*continued*)

Publication	Category Verbal Content	Style	Visual Kinesics	Proxemics	Agent appearance	CMC	Auditory Voice qualities	Vocalizations	Invisible Chronemics	Haptics
Skantze and Hjalmarsson (2013)								X		
Trovato et al. (2015)					X					
Verhagen et al. (2014)	X				X					
Visser et al. (2016)	X				X					
Pütten et al. (2009)			X							
Wu et al. (2017)					X					
Yuksel et al. (2017)					X		X			
Zhang et al. (2017)				X	X					

References

Akhtar, Z., Falk, T., 2017. Visual nonverbal behavior analysis: the path forward. IEEE Multimed. 25 (2), 47–60. https://doi.org/10.1109/MMUL.2017.265091158.

Allison, D., 2012. Chatbots in the library: is it time? Libr. Hi Tech. 30 (1), 95–107. https://doi.org/10.1108/07378831211213238.

Amazon, 2019. Speech Synthesis Markup Language (SSML) Reference. https://developer.amazon.com/de/docs/custom-skills/speech-synthesis-markup-language-ssml-reference.html Accessed 1 March 2019.

Amazon, 2019. Speechcon Reference (Interjections): English (US). https://developer.amazon.com/de/docs/custom-skills/speechcon-reference-interjections-english-us.html Accessed 1 March 2019.

Amazon, 2019. Writing for Voice. https://developer.amazon.com/de/docs/alexa-auto/writing-for-voice.html Accessed 1 March 2019.

Andonov, A., Schmidt, M., Niehues, J., Waibel, A. (Eds.), 2016. Using tweets as "ice-breaking" sentences in a social dialog system. Speech Communication 12 ITG Symposium.

Andonova, E., Taylor, H., 2012. Nodding in dis/agreement: a tale of two cultures. Cogn. Process. 13 (Suppl 1). https://doi.org/10.1007/s10339-012-0472-x.

Antaki, C., Billig, M., Edwards, D., Potter, J., 2003. Discourse analysis means doing analysis: a critique of six analytic shortcomings.

Antaki, C., 2008. Discourse analysis and conversation analysis. In: Alasuutari, P., Bickman, L., Brannen, J. (Eds.), The SAGE Handbook of Social Research Methods, pp. 431–447.

Appel, J., von der Pütten, A., Krämer, N.C., Gratch, J., 2012. Does humanity matter?: analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. Adv. Hum.-Comput. Interact. 2012 (2), 1–10. https://doi.org/10.1155/2012/324694.

Araujo, T., 2018. Living up to the chatbot hype: the influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput. Hum. Behav. 85, 183–189. https://doi.org/10.1016/j.chb.2018.03.051.

Bailenson, J.N., Yee, N., 2005. Digital chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments. Psychol. Sci. 16 (10), 814–819. https://doi.org/10.1111/j.1467-9280.2005.01619.x.

Bailey, K., 1994. Typologies and Taxonomies. SAGE Publications, Thousand Oaks, CA, USA.

Baur, T., Mehlmann, G., Damian, I., Lingenfelser, F., Wagner, J., Lugrin, B., André, E., Gebhard, P., 2015. Context-aware automated analysis and annotation of social human-agent interactions. ACM Trans. Interact. Intell. Syst. 5 (2). https://doi.org/10.1145/2764921. 11:1208;11:33.

Becker, C., Kopp, S., Wachsmuth, I., 2004. Simulating the emotion dynamics of a multimodal conversational agent. In: Affective Dialogue Systems, Proceedings, pp. 154–165.

Becker, C., Prendinger, H., Ishizuka, M., Wachsmuth, I., 2005. Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game. Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 466–473.

Bee, N., André, E., Tober, S., 2009. Breaking the ice in human-agent communication: eye-gaze based initiation of contact with an embodied conversational agent. Intelligent Virtual Agents. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 229–242.

Beer, J.M., Smarr, C.-A., Fisk, A.D., Rogers, W.A., 2015. Younger and older users' recognition of virtual agent facial expressions. Int. J. Hum.-Comput. St 75, 1–20. https://doi.org/10.1016/j.ijhcs.2014.11.005.

Beldad, A., Hegner, S., Hoppen, J., 2016. The effect of virtual sales agent (VSA) gender – product gender congruence on product advice credibility, trust in VSA and online vendor, and purchase intention. Comput. Hum. Behav. 60, 62–72. https://doi.org/10.1016/j.chb.2016.02.046.

Bell, A., 1997. Language style as audience design. In: Coupland, N., Jaworski, A. (Eds.), Sociolinguistics: A Reader. Macmillan Education UK, London, pp. 240–250.

Beun, R.J., Vos de, E., Witteman, C., 2003. Embodied conversational agents: effects on memory performance and anthropomorphisation. Intelligent Virtual Agents. pp. 315–319.

Bevacqua, E., Pammi, S., Hyniewska, S.J., Schröder, M., Pelachaud, C., 2010. Multimodal backchannels for embodied conversational agents. Intelligent Virtual Agents. Springer, Berlin Heidelberg, pp. 194–200.

Bickmore, T., Cassell, J., 2001. Relational agents: a model and implementation of building user trust. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. The SIGCHI Conference, Seattle, Washington, United States. New York, NY. ACM, pp. 396–403.

Bickmore, T., Cassell, J., 2005. Social dialogue with embodied conversational agents. In: Kuppevelt, J.C.J., Bernsen, N.O., Dybkjær, L. (Eds.), Advances in Natural Multimodal Dialogue Systems, vol. 30. Springer, Dordrecht, pp. 23–54.

Bickmore, T., Gruber, A., 2010. Relational agents in clinical psychiatry. Harv. Rev. Psychiatry 18 (2), 119–130. https://doi.org/10.3109/10673221003707538.

Bickmore, T.W., Picard, R.W., 2005. Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput. Hum. Interact. 12 (2), 293–327. https://doi.org/10.1145/1067860.1067867.

Bickmore, T.W., Fernando, R., Ring, L., Schulman, D., 2010. Empathic touch by relational agents. IEEE Trans. Affect. Comput. 1 (1), 60–71. https://doi.org/10.1109/T-AFFC.2010.4.

Bishop, T., 2018. What happened to Yahoo's $10 M Alliance with CMU, and How it Could Help AI Restore Our Humanity. https://www.geekwire.com/2018/happened-yahoos-10m-alliance-cmu-help-ai-restore-humanity/ Accessed 1 March 2019.

Bonito, J.A., Burgoon, J.K., Bengtsson, B., 1999. The role of expectations in human-computer interaction. In: Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work. New York, NY, USA. ACM, pp. 229–238.

Brahnam, S., Angeli, A.de, 2012. Gender affordances of conversational agents. Interact. Comput. 24 (3), 139–153. https://doi.org/10.1016/j.intcom.2012.05.001.

Brandão, C., Reis, L.P., Rocha, A.P., 2013. Evaluation of embodied conversational agents. In: 8th Iberian Conference on Information Systems and Technologies (CISTI). Lisboa. pp. 1–6.

Brandtzaeg, P.B., Følstad, A., 2018. Chatbots: changing user needs and motivations. Interactions 25 (5), 38–43. https://doi.org/10.1145/3236669.

Braslavski, P., Blinov, V., Bolotova, V., Pertsova, K., 2018. How to evaluate humorous response generation, seriously? In: Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR '18. the 2018 Conference, New Brunswick, NJ, USA. 11.03.2018 - 15.03.2018. New York, New York, USA. ACM Press, pp. 225–228.

Burgoon, J.K., Guerrero, L.K., Floyd, K., 2010. Nonverbal Communication. Routledge, NY, USA.

Burgoon, J.K., Guerrero, L.K., Manusov, V., 2011. Nonverbal signals. In: Knapp, M.L., Daly, J.A. (Eds.), The SAGE Handbook of Interpersonal Communication. SAGE Publications, Thousand Oaks, CA, USA.

Cafaro, A., Vilhjalmsson, H.H., Bickmore, T., 2016. First impressions in human-agent virtual encounters. ACM Trans. Comput.-Hum. Interact. 23 (4). https://doi.org/10.1145/2940325.

Campano, S., Langlet, C., Glas, N., Clavel, C., Pelachaud, C., 2015. An ECA expressing appreciations. In: International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 962–967.

Candello, H., Pinhanez, C., Figueiredo, F., 2017. Typefaces and the perception of humanness in natural language chatbots. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. New York, NY, USA. ACM, pp. 3476–3487.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., Pelachaud, C., 2007. Virtual agent multimodal mimicry of humans. Lang. Resour. Eval. 41 (3–4), 367–388. https://doi.org/10.1007/s10579-007-9057-1.

Carolis, B., de Pelachaud, C., Poggi, I., Steedman, M., 2004. APML, a markup language for believable behavior generation. In: Prendinger, H., Ishizuka, M. (Eds.), Life-Like Characters: Tools, Affective Functions, and Applications. Springer, Berlin Heidelberg, pp. 65–85.

Cassell, J., Bickmore, T., 2000. External manifestations of trustworthiness in the interface. Commun. ACM 43 (12), 50–56. https://doi.org/10.1145/355112.355123.

Cassell, J., Thorisson, K.R., 1999. The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. Appl. Artif. Intell. 13 (4–5), 519–538. https://doi.org/10.1080/088395199117360.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M., 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In: Proceedings of the 21st annual Conference on Computer Graphics and Interactive Techniques. ACM.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., Yan, H., 1999. Embodiment in conversational interfaces. In: The CHI is the Limit. CHI 99 Conference Proceeding. the SIGCHI Conference, Pittsburgh, Pennsylvania, United States. 5/15/1999 - 5/20/1999. New York. ACM, pp. 520–527.

Cassell, J., Sullivan, J., Churchill, E., Prevost, S., 2000. Embodied Conversational Agents.

MIT Press, Cambridge, MA, USA.
Cassell, J., 2000a. Embodied Conversational Agents. MIT Press, Cambridge, MA, USA, pp. 430.
Cassell, J., 2000b. Embodied conversational interface agents. Commun. ACM 43 (4), 70–78. https://doi.org/10.1145/332051.332075.
Cassell, J., 2001. Embodied conversational agents - representation and intelligence in user interfaces. AI Mag. 22 (4), 67–83.
Cassell, J., 2019. SARA: The Socially Aware Robot Assistant. http://articulab.hcii.cs.cmu.edu/projects/sara/ Accessed 1 March 2019.
Catrambone, R., Stasko, J., Xiao, J., 2002. Anthropomorphic agents as a user interface paradigm: experimental findings and a framework for research. In: Proceedings of the Cognitive Science Society. 24.
Chae, S.W., Lee, K.C., Seo, Y.W., 2016. Exploring the effect of avatar trust on learners' perceived participation intentions in an e-learning environment. Int. J. Hum. Comput. Interact. 32 (5), 373–393. https://doi.org/10.1080/10447318.2016.1150643.
Chakrabarti, C., Luger, G.F., 2015. Artificial conversations for customer service chatter bots: architecture, algorithms, and evaluation metrics. Expert Syst. Appl. 42 (20), 6878–6897. https://doi.org/10.1016/j.eswa.2015.04.067.
Chokshi, N., 2018. Amazon Knows Why Alexa was Laughing at its Customers. https://www.nytimes.com/2018/03/08/business/alexa-laugh-amazon-echo.html Accessed 1 March 2019.
Chollet, M., Ochs, M., Pelachaud, C., 2014. From non-verbal signals sequence mining to Bayesian networks for interpersonal attitudes expression. Intelligent Virtual Agents. Springer International Publishing, Cham, pp. 120–133.
Chue, V., Rios, J.S., Wohn, D.Y., 2018. A hot bot: testing effect of temperature on feelings of closeness when interacting with a chatbot. In: Proceedings of the 2018 ACM Conference on Supporting Groupwork. ACM, New York, NY, USA, pp. 136–139.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46. https://doi.org/10.1177/001316446002000104.
Collier, G., 2014. Emotional Expression. Psychology Press, New York, NY, USA.
Cowell, A.J., Stanney, K.M., 2005. Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. Int. J. Hum. Comput. Stud. 62 (2), 281–306. https://doi.org/10.1016/j.ijhcs.2004.11.008.
Cronbach, L.J., 1972. The Dependability of Behavioral Measurements: Theory of Generalizability For Scores and Profiles. Wiley, New York, pp. 410.
Crystal, D., 1969. Prosodic systems and intonation in English. CUP Archive. University of Cambridge.
D'Arcy, A., 2016. It's Lonely on Top: Why and How Poncho Became the Best Bot. https://medium.com/@hiponcho/its-lonely-on-top-why-and-how-poncho-became-the-best-bot-222d42d9c858 Accessed 30 August 2018.
Dale, R., 2016. The return of the chatbots. Nat. Lang. Eng. 22 (5), 811–817. https://doi.org/10.1017/S1351324916000243.
Danilava, S., Busemann, S., Schommer, C., Ziegler, G., 2013. Why are you silent?-Towards responsiveness in chatbots. Avec le Temps! Time, Tempo, and Turns in Human-Computer Interaction. Workshop at CHI 2013, Paris, France.
de Visser, E.J., Monfort, S.S., McKendrick, R., Smith, M.A.B., McKnight, P.E., Krueger, F., Parasuraman, R., 2016. Almost human: anthropomorphism increases trust resilience in cognitive agents. J. Exp. Psychol. Appl. 22 (3), 331–349. https://doi.org/10.1037/xap0000092.
Demeure, V., Niewiadomski, R., Pelachaud, C., 2011. How is believability of a virtual agent related to Warmth, Competence, Personification, and Embodiment? Presence 20 (5), 431–448.
Derrick, D.C., Ligon, G.S., 2014. The affective outcomes of using influence tactics in embodied conversational agents. Comput. Hum. Behav. 33, 39–48. https://doi.org/10.1016/j.chb.2013.12.027.
DeVito, J.A., 2013. The Interpersonal Communication Book, thirteen ed. Pearson, Boston, MA, USA.
Ding, Y., Prepin, K., Huang, J., Pelachaud, C., Artières, T., 2014. Laughter animation synthesis. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp. 773–780.
Donath, J., 2007. Signals, cues and meaning. February Draft For Signals, Truth and Design. MIT Press.
Dybala, P., Ptaszynski, M., Rzepka, R., Araki, K., 2009. Activating humans with humor - a dialogue system that users want to interact with. IEICE Transactions on information and systems E92D (12), 2394–2401. https://doi.org/10.1587/transinf.E92.D.2394.
Ekman, P., 1973. Darwin and Facial Expression: A Century of Research in Review. Academic Press, Oxford, England, pp. 273 xi.
Endrass, B., Rehm, M., André, E., 2010. Planning small talk behavior with cultural influences for multiagent systems. Comput. Speech Lang. 25 (2), 158–174.
Følstad, A., Brandtzæg, P.B., 2017. Chatbots and the new world of HCI. Interactions 24 (4), 38–42. https://doi.org/10.1145/3085558.
Feine, J., Morana, S., Gnewuch, U., 2019a. Measuring service encounter satisfaction with customer service chatbots using sentiment analysis. In: 14. Internationale Tagung Wirtschaftsinformatik (WI2019).
Feine, J., Morana, S., Maedche, A., 2019b. Leveraging machine-executable descriptive knowledge in design science research - the case of designing socially-adaptive chatbots. Extending the Boundaries of Design Science Theory and Practice. Springer International Publishing, Cham, pp. 76–91.
Fernández-Dols, J.-.M., 2013. Nonverbal communication: origins, adaptation, and functionality. In: Hall, J.A., Knapp, M.L. (Eds.), Nonverbal Communication. De Gruyter Mouton, Berlin/Boston, pp. 69–92.
Fiore, S.M., Wiltshire, T.J., Lobato, E.J.C., Jentsch, F.G., Huang, W.H., Axelrod, B., 2013. Toward understanding social cues and signals in human-robot interaction: effects of robot gaze and proxemic behavior. Front. Psychol. 4, 859. https://doi.org/10.3389/fpsyg.2013.00859.
Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76 (5), 378. https://doi.org/10.1037/h0031619.
Fogg, B.J., Nass, C., 1997. How users reciprocate to computers, in: CHI '97 Extended Abstracts on Human Factors in Computing Systems. CHI '97 extended abstracts, Atlanta, Georgia. ACM, New York, NY, pp. 331.
Fogg, B.J., Nass, C., 1997. How users reciprocate to computers. In: CHI '97 Extended Abstracts on Human Factors in Computing Systems. CHI '97 Extended Abstracts, Atlanta, Georgia. New York, NY. ACM, pp. 331.
Fogg, B.J., 2002. Computers as persuasive social actors. Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann Publishers, San Francisco, CA, USA, pp. 89–120.
Forbes, 2018. Amazon Echo, Google Home Installed Base Hits 50 Million; Apple has 6% Market Share, Report Says. https://www.forbes.com/sites/johnkoetsier/2018/08/02/amazon-echo-google-home-installed-base-hits-50-million-apple-has-6-market-share-report-says/#3c8d669a769c Accessed 4 March 2019.
Forlizzi, J., Zimmerman, J., Mancuso, V., Kwak, S., 2007. How Interface agents affect interaction between humans and computers. In: Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces. New York, NY, USA. ACM, pp. 209–221.
Frommert, C., Häfner, A., Friedrich, J., Zinke, C. (Eds.), 2018. Using Chatbots to Assist Communication in Collaborative Networks. Springer, pp. 257–265.
Gamble, T.K., Gamble, M., 2014. Interpersonal Communication: Building Connections Together. SAGE, Los Angeles, CA, USA, pp. 464.
Gao, Y., Pan, Z., Wang, H., Chen, G., 2018. Alexa, my love: analyzing reviews of amazon echo. In: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing. Internet of People and Smart City Innovation.
Garrido, P., Barrachina, J., Martinez, F.J., Seron, F.J., 2017. Smart tourist information points by combining agents, semantics and AI techniques. Comput. Sci. Inf. Syst. 14 (1), 1–23. https://doi.org/10.2298/csis150410029g.
Garrison, A., Remley, D., Thomas, P., Wierszewski, E., 2011. Conventional faces: emoticons in instant messaging discourse. Comput. Compos. 28 (2), 112–125. https://doi.org/10.1016/j.compcom.2011.04.001.
Gartner, 2017. Top Trends in the Gartner Hype Cycle for Emerging Technologies. https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/ Accessed 30 August 2018.
Gartner, 2018. Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020. https://www.gartner.com/newsroom/id/3858564 Accessed 22 November 2018.
Gebhard, P., Baur, T., Damian, I., Mehlmann, G., Wagner, J., André, E., 2014. Exploring interaction strategies for virtual characters to induce stress in simulated job interviews. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems. Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems, pp. 661–668.
Gerber, A., Baskerville, R., van der Merwe, A., 2017. A taxonomy of classification approaches in IS research. In: Proceedings of the Twenty-Third Americas Conference on Information Systems (AMCIS). Boston, CA, USA. AISel.
Ghazali, A.S., Ham, J., Barakova, E., Markopoulos, P., 2018. The influence of social cues in persuasive social robots on psychological reactance and compliance. Comput. Hum. Behav. 87, 58–65. https://doi.org/10.1016/j.chb.2018.05.016.
Gnewuch, U., Morana, S., Maedche, A., 2017. Towards designing cooperative and social conversational agents for customer service. In: Proceedings of the 38th International Conference on Information Systems (ICIS). AISel, Seoul.
Gnewuch, U., Morana, S., Adam, M., Maedche, A., 2018a. Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. In: Proceedings of the 26th European Conference on Information Systems (ECIS). Portsmouth, United Kingdom. pp. 23–28.
Gnewuch, U., Morana, S., Heckmann, C., Maedche, A., 2018b. Designing conversational agents for energy feedback. In: International Conference on Design Science Research in Information Systems and Technology (DESRIST 2018).
Go, E., Sundar, S.S., 2019. Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. Comput. Hum. Behav. 97, 304–316. https://doi.org/10.1016/j.chb.2019.01.020.
Gorin, A.L., Riccardi, G., Wright, J.H., 1997. How may I help you? Speech Commun. 23 (1–2), 113–127.
Gregor, S., 2006. The nature of theory in information systems. MIS Quart. 30 (3), 611–642. https://doi.org/10.2307/25148742.
Guerrero, L.K., DeVito, J.A., Hecht, M.L., 1999. The Nonverbal Communication Reader. Waveland Press Lone Grove, IL.
Guo, Y.R., Goh, D.H.-L., Muhamad, H.B.H., Ong, B.K., Lei, Z., 2016. Experimental evaluation of affective embodied agents in an information literacy game. In: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries. New York, NY, USA. ACM, pp. 119–128.
Hall, E.T., 1990. The Silent Language. Doubleday, Garden City, N.Y., USA.
Hastie, H., Lim, M.Y., Janarthanam, S., Deshmukh, A., Aylett, R., Foster, M.E., Hall, L., 2016. I remember you! Interaction with memory for an empathic virtual robotic tutor. In: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems, pp. 931–939.
Hauser, M.D., 1996. The Evolution of Communication. MIT Press, Cambridge, Mass., USA.
Hayashi, Y., 2016. Lexical network analysis on an online explanation task: effects of affect and embodiment of a pedagogical agent. IEICE Trans. Inf. Syst. 99 (6), 1455–1461. https://doi.org/10.1587/transinf.2015CBP0005.

Heath, A., 2018. Meet Poncho, the Weather Bot in Facebook that had Everyone Talking this Week. http://www.businessinsider.com/poncho-weather-bot-in-facebook-messenger-2016-4?IR=T Accessed 30 August 2018.
Hegel, F., Gieselmann, S., Peters, A., Holthaus, P., Wrede, B., 2011. Towards a typology of meaningful signals and cues in social robotics. In: Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication. Atlanta. pp. 72–78.
Hermann, E., 2019. Prosody Samples. https://www.amazon.com/Enno-Hermann-Prosody-Samples/dp/B071FV257B Accessed 1 March 2019.
Hess, T.J., Fuller, M.A., Mathew, J., 2005. Involvement and decision-making performance with a decision aid: the influence of social multimedia, gender, and playfulness. J. Manag. Inf. Syst. 22 (3), 15–54. https://doi.org/10.2753/MIS0742-1222220302.
Hill, J., Randolph Ford, W., Farreras, I.G., 2015. Real conversations with artificial intelligence: a comparison between human–human online conversations and human–chatbot conversations. Comput. Hum. Behav. 49, 245–250. https://doi.org/10.1016/j.chb.2015.02.026.
Hoffmann, L., Krämer, N.C., Lam-chi, A., Kopp, S., 2009. Media equation revisited: do users show polite reactions towards an embodied agent? In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (Eds.), Intelligent Virtual Agents: 9th International Conference, IVA 2009 Amsterdam, The Netherlands, September 14-16, 2009 Proceedings. Berlin, Heidelberg. Springer Berlin Heidelberg, pp. 159–165.
Hone, K., 2006. Empathic agents to reduce user frustration: the effects of varying agent characteristics. Interact. Comput. 18 (2), 227–245. https://doi.org/10.1016/j.intcom.2005.05.003.
Huisman, G., Bruijnes, M., Kolkmeier, J., Jung, M., Darriba Frederiks, A., Rybarczyk, Y., 2014. Touching virtual agents: embodiment and mind. Innovative and Creative Developments in Multimodal Interaction Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 114–138.
Isbister, K., Nass, C., 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. Int. J. Hum.-Comput. St 53 (2), 251–267. https://doi.org/10.1006/ijhc.2000.0368.
James, A., 2017. Prosody and paralanguage in speech and the social media: the vocal and graphic realisation of affective meaning. Linguistica 57 (1), 137–149. https://doi.org/10.4312/linguistica.57.1.137-149.
Johnson, K., 2017. Microsoft Bot Framework is Now Used by Over 130,000 Developers. https://venturebeat.com/2017/05/10/microsoft-bot-framework-is-now-used-by-over-130000-developers/ Accessed 30 August 2018.
Kalman, Y.M., Gergle, D.R., 2010. CMC cues enrich lean online communication: the case of letter and punctuation mark repetitions. In: Proceedings of the Fifth Mediterranean Conference on Information Systems. Tel-Aviv.
Kalman, Y., Gergle, D., 2014. Letter repetitions in computer-mediated communication: a unique link between spoken and online language. Comput. Hum. Behav. 34, 187–193. https://doi.org/10.1016/j.chb.2014.01.047.
Kang, S.-H., Fort Morie, J., 2013. Users' socially desirable responding with computer interviewers. CHI '13 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 229–234.
Kang, S.-H., Gratch, J., 2011. People like virtual counselors that highly-disclose about themselves 167 Stud. Health Technol. Inform.
Keeling, K., Beatty, S., McGoldrick, P., Macaulay, L., 2004. Face value? Customer views of appropriate formats for embodied conversational agents (ECAs) in online retailing. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences. pp. 1–10.
Kerly, A., Hall, P., Bull, S., 2007. Bringing chatbots into education: towards natural language negotiation of open learner models. Knowl.-Based Syst. 20 (2), 177–185. https://doi.org/10.1016/j.knosys.2006.11.014.
Kim, S., Chen, R.P., Zhang, K., 2016. Anthropomorphized helpers undermine autonomy and enjoyment in computer games. J. Consum. Res. 43 (2), 282–302. https://doi.org/10.1093/jcr/ucw016.
Kitchenham, B., 2004. Procedures for Performing Systematic Reviews. Keele University Technical Report TR/SE-0401.
Klein, J., Moon, Y., Picard, R.W., 2002. This computer responds to user frustration: theory, design, and results. Interact. Comput. 14 (2), 119–140.
Klopfenstein, L.C., Delpriori, S., Malatini, S., Bogliolo, A., 2017. The rise of bots: a survey of conversational interfaces, patterns, and paradigms. In: Proceedings of the 2017 Conference on Designing Interactive Systems. New York, NY, USA. ACM, pp. 555–565.
Knapp, M.L., Daly, J.A. (Eds.), 2011. The SAGE Handbook of Interpersonal Communication. SAGE Publications, Thousand Oaks, CA, USA.
Knapp, M.L., Hall, J.A., Horgan, T.G., 2013. Nonverbal Communication in Human Interaction. Wadsworth, Cengage Learning, Boston, MA, USA.
Knijnenburg, B.P., Willemsen, M.C., 2016. Inferring capabilities of intelligent agents from their external traits. ACM Trans. Interact. Intell. Syst. 6 (4). https://doi.org/10.1145/2963106. 28:1208;28:25.
Kopp, S., Wachsmuth, I., 2004. Synthesizing multimodal utterances for conversational agents. Comput. Anim. Virt. Worlds 15 (1), 39–52.
Kopp, S., Krenn, B., Marsella, S., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H., 2006. Towards a common framework for multimodal generation: the behavior markup language. In: Gratch, J. (Ed.), Intelligent Virtual Agents. 6th International Working Conference, IVA 2006: Marina del Rey, CA, USA, August 21-23, 2006: Proceedings. Berlin. vol. 4133. Springer, pp. 205–217.
Krämer, N.C., 2005. Social communicative effects of a virtual program guide. Intelligent Virtual Agents. Springer, Berlin Heidelberg, pp. 442–453.
Krämer, N., 2008a. Soziale Wirkungen Virtueller Helfer: Gestaltung und Evaluation Von Mensch-Computer-Interaktionen, first ed. Stuttgart, Kohlhammer, pp. 283 S.
Krämer, N.C., 2008b. Social effects of virtual assistants. A review of empirical results with regard to communication. In: Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings. Berlin, Heidelberg. Springer, pp. 507–508.
Krämer, N., Kopp, S., Becker-Asano, C., Sommer, N., 2013. Smile and the world will smile with you-the effects of a virtual agent's smile on users' evaluation and behavior. Int. J. Hum.-Comput. Stud. 71 (3), 335–349. https://doi.org/10.1016/j.ijhcs.2012.09.006.
Kraemer, N.C., Karacora, B., Lucas, G., Dehghani, M., Ruether, G., Gratch, J., 2016. Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. Comput. Educ. 99, 1–13. https://doi.org/10.1016/j.compedu.2016.04.002.
Lakoff, G., 1987. Women, Fire, and Dangerous Things: What categories Reveal About the Mind. University of Chicago Press, Chicago, IL, USA.
Lamolle, M., Mancini, M., Pelachaud, C., Abrilian, S., Martin, J.C., Devillers, L., 2005. Contextual factors and adaptative multimodal human-computer interaction: multi-level specification of emotion and expressivity in embodied conversational agents. In: Dey, A., Kokinov, B., Leake, D., Turner, R. (Eds.), Modeling and Using Context, Proceedings. vol. 3554. pp. 225–239.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174. https://doi.org/10.2307/2529310.
Larivière, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C., Wünderlich, N.V., Keyser, A.de, 2017. "Service encounter 2.0": an investigation into the roles of technology, employees and customers. J. Bus. Res. 79, 238–246. https://doi.org/10.1016/j.jbusres.2017.03.008.
Laver, J., 1980. The Phonetic Description of Voice quality: Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.
Leathers, D.G., Eaves, M., 2015. Successful Nonverbal Communication: Principles and Applications. Pearson/Allyn and Bacon, Boston, MA, USA.
Leathers, D.G., 1976. Nonverbal Communication Systems. Allyn & Bacon, Boston, MA, USA.
LeBreton, J.M., Senter, J.L., 2008. Answers to 20 questions about interrater reliability and interrater agreement. Organ. Res. Methods 11 (4), 815–852. https://doi.org/10.1177/1094428106296642.
Lee, S., Choi, J., 2017. Enhancing user experience with conversational agent for movie recommendation: effects of self-disclosure and reciprocity. Int. J. Hum.-Comput. Stud. 103, 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005.
Lee, K.M., Nass, C., 2003. Designing social presence of social actors in human computer interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA. ACM, pp. 289–296.
Li, H., Graesser, A., 2017. Impact of pedagogical agents' conversational formality on learning and engagement. Artificial Intelligence in Education. Springer International Publishing, Cham, pp. 188–200.
Li, J., Zhou, M.X., Yang, H., Mark, G., 2017. Confiding in and listening to virtual agents. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI, Limassol, Cyprus. 13.03.2017 - 16.03.2017. ACM Press, pp. 275–286.
Liebman, N., Gergle, D., 2016. It's (Not) simply a matter of time: the relationship between CMC cues and interpersonal affinity. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York, NY, USA. ACM, pp. 570–581.
Lisetti, C., Amini, R., Yasavur, U., Rishe, N., 2013. I can help you change! An empathic virtual agent delivers behavior change health interventions. ACM Trans. Manag. Inf. Syst. 4 (4). https://doi.org/10.1145/2544103. 19:1208;19:28.
Lobato, E.J.C., Warta, S.F., Wiltshire, T.J., Fiore, S.M., 2015. Varying social cue constellations results in different attributed social signals in a simulated surveillance task. In: FLAIRS Conference, pp. 61–66.
Lortie, C.L., Guitton, M.J., 2011. Judgment of the humanness of an interlocutor is in the eye of the beholder. Plos One 6 (9), 1–7. https://doi.org/10.1371/journal.pone.0025085.
Louwerse, M.M., Graesser, A.C., Lu, S.L., Mitchell, H.H., 2005. Social cues in animated conversational agents. Appl. Cogn. Psychol. 19 (6), 693–704. https://doi.org/10.1002/acp.1117.
Maedche, A., Morana, S., Schacht, S., Werth, D., Krumeich, J., 2016. Advanced user assistance systems. Bus. Inf. Syst. Eng. 58 (5), 367–370. https://doi.org/10.1007/s12599-016-0444-2.
Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., Söllner, M., 2019. AI-based digital assistants. Bus. Inf. Syst. Eng. 61 (4), 535–544. https://doi.org/10.1007/s12599-019-00600-8.
Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K., 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. J. Artif. Intell. Res. 30, 457–500.
Mayer, R.E., Johnson, W.L., Shaw, E., Sandhu, S., 2006. Constructing computer-based tutors that are socially sensitive: politeness in educational software. Int. J. Hum. Comput. Stud. 64 (1), 36–42. https://doi.org/10.1016/j.ijhcs.2005.07.001.
McBreen, H.M., Jack, M.A., 2001. Evaluating humanoid synthetic agents in e-retail applications. IEEE Trans. Syst. Man Cybern. Part A: Systems and Humans 31 (5), 394–405. https://doi.org/10.1109/3468.952714.
McBreen, H., 2002. Embodied conversational agents in e-commerce applications. Socially Intelligent Agents: Creating Relationships with Computers and Robots. Springer US, Boston, MA, pp. 267–274.
McTear, M., Callejas, Z., Griol, D., 2016. The Conversational Interface: Talking to Smart Devices, first ed. Springer International Publishing, Switzerland.
McTear, M.F., 2017. The rise of the conversational Interface: a new kid on the block? Future and Emerging Trends in Language Technology. Machine Learning and Big Data. Springer International Publishing, Cham, pp. 38–49.
Mersiol, M., Chateau, N., Maffiolo, V. (Eds.), 2002. Talking heads: which matching between faces and synthetic voices? Proceedings. Fourth IEEE International Conference on Multimodal Interfaces.
Mimoun, M.S.B., Poncin, I., Garnier, M., 2012. Case study—embodied virtual agents: an

analysis on reasons for failure. J. Retail. Consum. Serv. 19 (6), 605–612. https://doi.org/10.1016/j.jretconser.2012.07.006.
Moon, Y., 2000. Intimate exchanges: using computers to elicit self-disclosure from consumers. J. Consum. Res. 26 (4), 323–339. https://doi.org/10.1086/209566.
Moon, Y., Nass, C., 1996. How "Real" are computer personalities? Commun. Res. 23 (6), 651–674. https://doi.org/10.1177/009365096023006002.
Moon, Y., Nass, C., 1998. Are computers scapegoats?: Attributions of responsibility in human–computer interaction. Int. J. Hum.-Comput. Stud. 49 (1), 79–94. https://doi.org/10.1006/ijhc.1998.0199.
Moore, G.C., Benbasat, I., 1991. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf. Syst. Res. 2 (3), 192–222. https://doi.org/10.1287/isre.2.3.192.
Moore, R.K., 2013. Spoken language processing: where do we go from here? In: Trappl, R. (Ed.), Your Virtual Butler: The Making-Of. Springer, Berlin Heidelberg, pp. 119–133 Berlin, Heidelberg.
Morkes, J., Kernal, H., Nass, C., 1999. Effects of humor in task-oriented human-computer interaction and computer-mediated communication: a direct test of SRCT theory. Hum.-Comput. Interact. 14 (4), 395–435. https://doi.org/10.1207/S15327051HCI1404_2.
Myers, L., 2017. New SSML Features Give Alexa a Wider Range of Natural Expression. https://developer.amazon.com/de/blogs/alexa/post/5c631c3c-0d35-483f-b226-83dd98def117/new-ssml-features-give-alexa-a-wider-range-of-natural-expression Accessed 1 March 2019.
Nöth, W., 1995. Handbook of Semiotics. Indiana University Press, Bloomington, USA.
Nass, C., Moon, Y., 2000. Machines and mindlessness: social responses to computers. J. Soc. Issues 56 (1), 81–103. https://doi.org/10.1111/0022-4537.00153.
Nass, C., Steuer, J., Tauber, E.R., 1994. Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA. ACM, pp. 72–78.
Nass, C., Moon, Y., Fogg, B.J., Reeves, B., Dryer, D.C., 1995. Can computer personalities be human personalities? Int. J. Hum. Comput. Stud. 43 (2), 223–239. https://doi.org/10.1006/ijhc.1995.1042.
Nass, C., Fogg, B.J., Moon, Y., 1996. Can computers be teammates? Int. J. Hum.-Comput. Stud. 45 (6), 669–678. https://doi.org/10.1006/ijhc.1996.0073.
Nass, C., Moon, Y., Green, N., 1997. Are machines gender neutral?: gender-stereotypic responses to computers with voices. J. Appl. Soc. Pyschol. 27 (10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x.
Nass, C., Moon, Y., Carney, P., 1999. Are people polite to computers? Responses to computer-based interviewing systems. J. Appl. Soc. Pyschol. 29 (5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x.
Nickerson, R.C., Varshney, U., Muntermann, J., 2013. A method for taxonomy development and its application in information systems. Eur. J. Inf. Syst. 22 (3), 336–359. https://doi.org/10.1057/ejis.2012.26.
Niculescu, A., Hofs, D., van Dijk, B., Nijholt, A., 2010. How the agent's gender influence users' evaluation of a QA system. In: International Conference on User Science and Engineering (i-USEr).
Niewiadomski, R., Pelachaud, C., 2010. Affect expression in ECAs: application to politeness displays. Int. J. Hum. Comput. Stud. 68 (11), 851–871. https://doi.org/10.1016/j.ijhcs.2010.07.004.
Niewiadomski, R., Hofmann, J., Urbain, J., Platt, T., Wagner, J., Piot, B., Cakmak, H., Pammi, S., Baur, T., Dupont, S., Geist, M., Lingenfelser, F., McKeown, G., Pietquin, O., Ruch, W., 2013. Laugh-aware virtual agent and its impact on user amusement. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems. Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems, pp. 619–626.
Nowak, K., 2004. The influence of anthropomorphism and agency on social judgment in virtual environments. J. Comput.-Mediat. Commun. 9. https://doi.org/10.1111/j.1083-6101.2004.tb00284.x.
Nunamaker, J.E., Derrick, D.C., Elkins, A.C., Burgoon, J.K., Patton, M.W., 2011. Embodied conversational agent-based kiosk for automated interviewing. J. Manag. Inf. Syst. 28 (1), 17–48. https://doi.org/10.2753/mis0742-1222280102.
Ochs, M., Pelachaud, C., McKeown, G., 2017. A user perception-based approach to create smiling embodied conversational agents. ACM Trans. Interact. Intell. Syst. 7 (1). https://doi.org/10.1145/2925993. 4:1208;4:33.
von der Pütten, A., Krämer, N., Gratch, J., 2009. Who's there? Can a virtual agent really elicit social presence? In: The 12th Annual International Workshop on Presence.
Pantic, M., Cowie, R., D'Errico, F., Heylen, D., Mehu, M., Pelachaud, C., Poggi, I., Schroeder, M., Vinciarelli, A., 2011. Social signal processing: the research agenda. In: Moeslund, T.B., Hilton, A., Krüger, V., Sigal, L. (Eds.), Visual Analysis of Humans: Looking at People. Springer, London, London, pp. 511–538.
Paré, G., Trudel, M.-.C., Jaana, M., Kitsiou, S., 2015. Synthesizing information systems knowledge: a typology of literature reviews. Inf. Manag. 52 (2), 183–199. https://doi.org/10.1016/j.im.2014.08.008.
Pecune, F., Mancini, M., Biancardi, B., Varni, G., Ding, Y., Pelachaud, C., Volpe, G., Camurri, A., 2015. Laughing with a virtual agent. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems. Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems, pp. 1817–1818.
Pelachaud, C., Bilvi, M., 2003. Computational model of believable conversational agents. In: Communication In Multiagent Systems: Agent Communication Languages And Conversation Policies, pp. 300–317.
Pelachaud, C., 2005. Multimodal expressive embodied conversational agents. In: Proceedings of the 13th Annual ACM International Conference on Multimedia. New York, NY, USA. ACM, pp. 683–689.
Pelachaud, C., 2009a. Modelling multimodal expression of emotion in a virtual agent. Philos. Trans. R. Soc. B 364 (1535), 3539–3548. https://doi.org/10.1098/rstb.2009.0186.
Pelachaud, C., 2009b. Studies on gesture expressivity for a virtual agent. Speech. Commun. 51 (7, SI), 630–639. https://doi.org/10.1016/j.specom.2008.04.009.
Pelachaud, C., 2017. Greta: a conversing socio-emotional agent. In: Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents. The 1st ACM SIGCHI International Workshop, Glasgow, UK. 11/13/2017 - 11/13/2017. New York, NY. ACM, pp. 9–10.
Perez, S., 2017. Alexa Learns to Talk Like a Human With Whispers, Pauses & Emotion. https://techcrunch.com/2017/04/28/alexa-learns-to-talk-like-a-human-with-whispers-pauses-emotion/?guccounter=1 Accessed 1 March 2019.
Pertaub, D.P., Slater, M., Barker, C., 2001. An experiment on fear of public speaking in virtual reality. Stud. Health Technol. Inform. 81.
Pfeifer, L.M., Bickmore, T., 2009. Should agents speak like, um, humans? The use of conversational fillers by virtual agents. In: Proceedings of the 9th International Conference on Intelligent Virtual Agents. Berlin, Heidelberg. Springer-Verlag, pp. 460–466.
Poyatos, F., 1991. Paralinguistic qualifiers: our many voices. Lang. Commun. 11 (3), 181–195. https://doi.org/10.1016/0271-5309(91)90005-G.
Prat, N., Comyn-Wattiau, I., Akoka, J., 2015. A taxonomy of evaluation methods for information systems artifacts. J. Manag. Inf. Syst. 32 (3), 229–267. https://doi.org/10.1080/07421222.2015.1099390.
Prepin, K., Ochs, M., Pelachaud, C. (Eds.), 2013. Beyond Backchannels: Co-Construction of Dyadic Stancce by Reciprocal Reinforcement of Smiles Between Virtual Agents.
Puetten, A.M., von der Kraemer, N.C., Gratch, J., Kang, S.-.H., 2010. It doesn't matter what you are!" Explaining social effects of agents and avatars. Comput. Hum. Behav. 26 (6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012.
Puzakova, M., Rocereto, J.F., Kwak, H., 2013. Ads are watching me. Int. J. Advert. 32 (4), 513–538. https://doi.org/10.2501/IJA-32-4-513-538.
Recanati, F., 2001. What is said. Synthese 128 (1/2), 75–91. https://doi.org/10.1023/A:1010383405105.
Reeves, S., 2017. Some conversational challenges of talking with machines. In: Talking with Conversational Agents in Collaborative Action, Workshop at the 20th ACM conference on Computer-Supported Cooperative Work and Social Computing (CSCW '17). Portland, USA.
Reidsma, D., Katayose, H., Nijholt, A., Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssafidou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., Jones, H., Ochs, M., Pelachaud, C., Porayska-Pomsta, K., Rizzo, P., Sabouret, N. (Eds.), 2013. The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews: Advances in Computer Entertainment. Springer International Publishing, pp. 476–491.
Rezabek, L., Cochenour, J., 1998. Visual cues in computer-mediated communication: supplementing text with emoticons. J. Vis. Lit. 18 (2), 201–215. https://doi.org/10.1080/23796529.1998.11674539.
Richards, D., Bransky, K., 2014. ForgetMeNot: What and how users expect intelligent virtual agents to recall and forget personal conversational content. Int. J. Hum.-Comput. Stud. 72 (5), 460–476. https://doi.org/10.1016/j.ijhcs.2014.01.005.
Rickenberg, R., Reeves, B., 2000. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA. ACM, pp. 49–56.
Rosis, F.de, Pelachaud, C., Poggi, I., Carofiglio, V., Carolis, B.de, 2003. From greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. Int. J. Hum.-Comput. Stud. 59 (1–2), 81–118. https://doi.org/10.1016/S1071-5819(03)00020-X.
Rossen, B., Johnsen, K., Deladisma, A., Lind, S., Lok, B., 2008. Virtual humans elicit skin-tone bias consistent with real-world skin-tone biases. In: Prendinger, H., Lester, J., Ishizuka, M. (Eds.), Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings. Berlin, Heidelberg. Springer Berlin Heidelberg, pp. 237–244.
Rugg, G., McGeorge, P., 2005. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. Expert Syst. 22 (3), 94–107. https://doi.org/10.1111/j.1468-0394.2005.00300.x.
Ryokai, K., Vaucelle, C., Cassell, J., 2003. Virtual peers as partners in storytelling and literacy learning. J. Comput. Assist. Learn. 19 (2), 195–208. https://doi.org/10.1046/j.0266-4909.2003.00020.x.
Sah, Y.J., Peng, W., 2015. Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. Comput. Hum. Behav. 45, 392–401. https://doi.org/10.1016/j.chb.2014.12.055.
Schötz, S., 2002. Linguistic & Paralinguistic Phonetic Variation in Speaker Recognition & Text-to-Speech Synthesis. Speech Technology.
Searle, J.R., Kiefer, F., Bierwisch, M., 1980. Speech Act Theory and Pragmatics. Springer, Dordrecht, pp. 336.
Selting, M., 2009. Communicative style. In: D'hondt, S., Östman, J.O., Verschueren, J. (Eds.), The Pragmatics of Interaction. Handbook of Pragmatics Highlights. John Benjamins Publishing Company, Amsterdam.
Shechtman, N., Horowitz, L.M., 2003. Media inequality in conversation: how people behave differently when interacting with computers and people. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Ft. Lauderdale, FL, USA. pp. 281–288.
Shiban, Y., Schelhorn, I., Jobst, V., Hoernlein, A., Puppe, F., Pauli, P., Muehlberger, A., 2015. The appearance effect: Influences of virtual agent features on performance and motivation. Comput. Hum. Behav. 49, 5–11. https://doi.org/10.1016/j.chb.2015.01.077.
Skantze, G., Hjalmarsson, A., 2013. Towards incremental speech generation in conversational systems. Comput. Speech Lang. 27 (1), 243–262.
Smith, J.M., Harper, D., 2003. Animal Signals. Oxford University Press, Oxford, UK.

Tannen, D., 1984. Conversational Style: Analyzing Talk Among Friends. Ablex Publishing Corporation, Norwood, NJ, USA.

Thiebaux, M., Marsella, S., Marshall, A.N., Kallmann, M., 2008. Smartbody: behavior realization for embodied conversational agents. In: International Foundation for Autonomous Agents and Multiagent Systems.

Thomas, P., Czerwinski, M., McDuff, D., Craswell, N., Mark, G., 2018. Style and alignment in information-seeking conversation. In: Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR '18. the 2018 Conference, New Brunswick, NJ, USA. 11.03.2018 - 15.03.2018. New York, New York, USA. ACM Press, pp. 42–51.

Titscher, S., Meyer, M., Wodak, R., Vetter, E., 2000. Methods of Text and Discourse Analysis: In Search of Meaning. SAGE.

Trager, G.L., 1958. Paralanguage: a first approximation. Stud. Linguist. 1958 (13), 1–12.

Trenholm, S., Jensen, A., 2011. Interpersonal Communication. Oxford University Press, Oxford, UK.

Trovato, G., Ramos, J.G., Azevedo, H., Moroni, A., Magossi, S., Ishii, H., Simmons, R., Takanishi, A., 2015. "Olá, my name is Ana": a study on Brazilians interacting with a receptionist robot. In: 2015 International Conference on Advanced Robotics (ICAR), in: 2015 International Conference on Advanced Robotics (ICAR).

Verhagen, T., van Nes, J., Feldberg, F., van Dolen, W., 2014. Virtual customer service agents: using social presence and personalization to shape online service encounters. J. Comput.-Med. Commun. 19 (3), 529–545. https://doi.org/10.1111/jcc4.12066.

Vinciarelli, A., Pantic, M., Bourlard, H., 2009. Social signal processing: survey of an emerging domain. Image Vis. Comput. 27 (12), 1743–1759. https://doi.org/10.1016/j.imavis.2008.11.007.

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., Schroeder, M., 2012. Bridging the gap between social animal and unsocial machine: a survey of social signal processing. IEEE Trans. Affect. Comput. 3 (1), 69–87. https://doi.org/10.1109/T-AFFC.2011.27.

Visser, E.J.de, Monfort, S.S., McKendrick, R., Smith, M.A.B., McKnight, P.E., Krueger, F., Parasuraman, R., 2016. Almost human: anthropomorphism increases trust resilience in cognitive agents. J. Exp. Psychol. Appl. 22 (3), 331–349. https://doi.org/10.1037/xap0000092.

Wallis, P., Norling, E., 2005. The trouble with chatbots: social skills in a social world. In: AISB 2005 Convention: Proceedings of the Joint Symposium on Virtual Social Agents: Social Presence Cues for Virtual Humanoids Empathic Interaction with Synthetic Characters Mind Minding Agents, pp. 29–36.

Walther, J.B., Tidwell, L.C., 1995. Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. J. Organ. Comput. 5 (4), 355–378.

Walther, J.B., 1992. Interpersonal effects in computer-mediated interaction: a relational perspective. Commun. Res. 19 (1), 52–90. https://doi.org/10.1177/009365092019001003.

Walther, J.B., 2006. Nonverbal dynamics in computer-mediated communication, or :(and the net :)'s with you, :) and you :) alone. In: Manusov, V., Patterson, M.L. (Eds.), The SAGE Handbook of Nonverbal Communication. SAGE Publications, Thousand Oaks, CA, USA.

Walther, J.B., 2008. Social information processing theory: impressions and relationship development online. In: Baxter, L.A., Braithwaite, D.O. (Eds.), Engaging Theories in Interpersonal Communication. Multiple Perspectives. SAGE Publications.

Webster, J., Watson, R.T., 2002. Analyzing the past to prepare for the future: writing a literature review. MIS Quart. 26 (2), xiii–xxiii. https://doi.org/10.2307/4132319.

Weizenbaum, J., 1966. ELIZA - a computer program for the study of natural language communication between man and machine. Commun. ACM 9 (1), 36–45.

Wiltshire, T.J., Lobato, E.J.C., Velez, J., Jentsch, F.G., Fiore, S.M., 2014. An interdisciplinary taxonomy of social cues and signals in the service of engineering robotic social intelligence. In: Unmanned Systems Technology XVI, 90840F.

Wolfswinkel, J.F., Furtmueller, E., Wilderom, C.P.M., 2013. Using grounded theory as a method for rigorously reviewing literature. Eur. J. Inf. Syst. 22 (1), 45–55. https://doi.org/10.1057/ejis.2011.51.

Wu, Z., Lin, T., Xu, H., Tang, N., 2017. Predicting personality associations evoked by multicolored appearance of virtual agents: an exploratory study. Color Res. Appl. 42 (6), 753–763. https://doi.org/10.1002/col.22154.

Wuenderlich, N.V., Paluch, S., 2017. A nice and friendly chat with a bot: user perceptions of AI-based service agents. In: Proceedings of the 38th International Conference on Information Systems (ICIS). AISel, Seoul.

Youssef, A.B., Chollet, M., Jones, H., Sabouret, N., Pelachaud, C., Ochs, M., 2015. Towards a socially adaptive virtual agent. Intelligent Virtual Agents. Springer International Publishing, Cham, pp. 3–16.

Yuksel, B.F., Collisson, P., Czerwinski, M., 2017. Brains or beauty: how to engender trust in user-agent interactions. ACM Trans. Internet Technol. 17 (1). https://doi.org/10.1145/2998572. 2:1208;2:20.

Zhang, Z., Bickmore, T.W., Paasche-Orlow, M.K., 2017. Perceived organizational affiliation and its effects on patient trust: role modeling with embodied conversational agents. Patient Educ. Couns. 100 (9), 1730–1737. https://doi.org/10.1016/j.pec.2017.03.017.