Audience design and egocentrism in reference production during human-computer dialogue

Peña, Paola R.; Doyle, Philip; Edwards, Justin; Garaialde, Diego; Rough, Daniel; Bleakley, Anna

Published in:
International Journal of Human Computer Studies

DOI:
10.1016/j.ijhcs.2023.103058

Publication date:
2023

Audience design and egocentrism in reference production during human-computer dialogue

Paola R. Peña [1], Philip Doyle [1], Justin Edwards [1], Diego Garaialde [1], Daniel Rough [2], Anna Bleakley [1], Leigh Clark [4], Anita Tobar Henriquez [3], Holly Branigan [3], Iona Gessinger [1], Benjamin R. Cowan [1]

[1] HCI@UCD, School of Information & Communication Studies, University College Dublin, Ireland

[2] UX'd Group, University of Dundee, Scotland

[3] Department of Psychology, University of Edinburgh, Scotland

[4] Computational Foundry, Swansea University, Wales

Abstract

Our current understanding of the mechanisms that underpin language production in human-computer dialogue (HCD) is sparse. What work there is in the field of human-computer interaction (HCI) supposes that people tend to adapt their language allocentrically, taking into account the perceived limitations of their partners, when talking to computers. Yet, debates in human-human dialogue (HHD) research suggest that people may also act egocentrically when producing language in dialogue. Our research aims to identify whether, similar to HHD, users also produce egocentric language within speech-based HCD interactions and how this behaviour compares to interaction with human dialogue partners. Such knowledge benefits the field of HCI by better understanding the mechanisms present in language production during HCD, which can be used to build more nuanced theories and models of user behaviour to inform research and design of speech interfaces. Through two controlled experiments using an

adapted director-matcher task similar to those used in research on perspective-taking in psycholinguistics, we show that people do take the computer's perspective into account less (i.e. behave more egocentrically) during HCD than in HHD (Experiment 1). However, this egocentric effect is eliminated when computers are framed as separate interlocutors rather than computers integrated in the interactive system and where differences in perspective are made salient, leading to similar levels of perspective-taking as with human partners (Experiment 2). We discuss the findings, emphasising potential explanations for this effect, focusing on how egocentric and allocentric production processes may interact, along with the impact of partner roles and the division of labour in HCD as an underlying explanation for the effects seen.

Keywords: reference production, audience design, egocentrism, human-machine dialogue

Email address: paola.pena@ucdconnect.ie (Paola Peña). UCD School of Information and Communication Studies. Newman Building, first floor, C block, room D107, University College Dublin, Belfield , Dublin 4, Ireland
Preprint so International Journal of Human-Computer Studies August 10, 2022

1. Introduction

With the current growth of conversational interfaces (Dafoe et al., 2021), language interactions with computer dialogue partners are commonplace. Despite this rapid growth, we know little about the psychological and linguistic mechanisms that people employ in speech based human-computer dialogue (HCD). This knowledge is needed to help inform theory building within HCI on what influences user language interactions (Cowan et al., 2023; Peña et al., In press; Shen and Wang, 2023) supporting recent efforts for more formal and computation models of user behaviour in human-machine dialogue (e.g Rothwell et al., 2021). Such insights could also support speech interface development through informing speech technology components (e.g. dialogue management and recognition), whilst informing design decisions so as to take into account their influence on mechanisms for linguistic behaviour (Braunger et al., 2017, Zhao et al., 2022). Our study aims to be a step in identifying how concepts from psycholinguistics in human-human dialogue (HHD), most notably audience design, common ground use and egocentrism in language production, should be considered in how we conceptualise speech based human-computer dialogue.

Our current understanding of the causal mechanisms underlying user language choices in speech-based human-computer dialogue (HCD) interaction is sparse (Clark et al., 2019). What little work there is supposes that we adapt our language choices (e.g. Amalberti et al., 1993; Brennan, 1998; Cowan et al., 2019a; Le Bigot et al., 2007; Luger & Sellen, 2016; Meddeb & Frenz-Belkin, 2010a) based on our perceptions of a computer interlocutor's knowledge and capabilities as a dialogue partner (An et al., 2021; Branigan et al., 2011; Cowan et al., 2019a), akin to the concept of audience design in HHD (Bell, 1984). Yet, recent findings have shown that increased adaptation towards computer partners - compared to human partners - is sometimes absent from language production (Cowan & Branigan, 2015; Cowan et al., 2015), or appears alongside more egocentric language choices (e.g. Dombi et al., 2022), challenging the notion of allocentric processes (e.g. audience design through perspective taking) as the sole driver for

language production in HCD. That is, rather than being driven solely by consideration for the computer's perspective, perceived knowledge, and capabilities (i.e. audience design), the language that we produce when interacting in spoken HCD may also be influenced by a bias towards our own perspective and knowledge (i.e. egocentrism).

Through two studies, our research shows that people tend to take the interlocutor's perspective into account less during HCD than during HHD (Experiment 1). Yet, crucially, this effect is impacted by the role of the computer partner within the dialogue, whereby computer interlocutors that are framed as separate dialogue partners (rather than integrated in the system where interaction takes place) induce similar perspective-taking behaviours as are found in HHD (Experiment 2). We discuss the findings, emphasising the need for HCI theory building in this area to consider how allocentric and egocentric processes may interact, whilst emphasising to designers how the role of a computer partner may influence user language production and user perspective-taking processes in HCD.

1.1 Audience design and egocentrism in human-human dialogue

The audience design approach to HHD (Bell, 1984; Clark & Schaefer, 1987) supposes that communication is a joint activity (Clark, 1992; Clark & Wilkes-Gibbs, 1986), whereby interlocutors are willing to expend effort in the conversation in order for their addressee to understand the intended message and co-ordinate meaning. This coordination relies on a "model of the other person's mind" (Keysar et al., 1998 p. 46), taking into account *common ground* between the interlocutors (Brennan, 1990; Brennan & Metzing, 2004). Common ground is conceptualised as information that is believed to be shared by, and available to, each interlocutor during dialogue (Clark, 1996; Horton & Keysar, 1996; Shintel & Keysar, 2009). This may include information about shared context, assumed knowledge and co-constructed mutual knowledge (e.g. situation-specific information that arises from the environment as well as conceptual pacts on how to refer to specific entities or objects - see Brennan & Clark, 1996; Clark 1996).

This information is thought to be developed and updated collaboratively (Galati & Brennan, 2010), with the aim of maintaining shared representations during dialogue (Bortfield & Brennan, 1997). Such information is important in informing perspective taking across a dialogic interaction (Yoon et al., 2012) by allowing speakers to be more allocentric in their production, optimally designing their utterances to their audience's perceived knowledge state (Horton & Keysar, 1996).

Similar to previous work, we define perspective taking as a speaker's attempt to take into account mutual beliefs as well as what that speaker believes their interlocutor's knowledge state or perspective is pertaining to the dialogue (Keysar & Barr, 2002), which informs allocentric language production processes such as audience design. Although there is strong evidence for audience design by speakers toward their addressees in HHD (e.g. Bortfield & Brennan, 1997; Brennan & Clark, 1996; Clark, 2020; Clark & Wilkes-Gibbs, 1986, Ferreira, 2019; Fussell & Krauss, 1992; Galati & Brennan, 2010; Horton & Gerrig, 2002), the effect is by no means universal. For instance, work on reference production shows that people do not always adequately take listeners' information needs and knowledge state into account when producing utterances (Engelhart et al., 2006; Horton & Keysar, 1996; Lane & Liersch, 2012). Recent work highlights that referent over-specification is common and persistent across dialogue, even though this can lead to increased comprehension difficulty for the addressee (Wu et al., 2013). Speakers sometimes violate Grice's (1975) maxim of quantity by providing more information than necessary (e.g. by naming the only shoe in a grid of objects as the 'blue shoe'). This lack of utterance optimisation is considered to be evidence that, rather than being fundamentally designed for the addressee, language is processed and produced egocentrically by default, only being adjusted when required (i.e. when interpretation leads to errors) or when cognitive resources allow (known as the *monitor and adjust* account of perspective taking (Dell & Brown, 1991; Keysar et al., 2008). This is thought to be because incorporating common ground during dialogue can be cognitively demanding (Keysar et al., 2003), although this claim is debated (see Brennan & Metzing, 2004;

Rubio-Fernández & Jara-Ettinger, 2018). Accordingly, egocentricity in language production is thought to be driven by the speaker wishing to reduce their own effort and to minimise processing demands (Knutsen & Le Bigot, 2014).

Egocentrism in production is largely studied in director/matcher-based paradigms where an asymmetry exists between the knowledge state of the matcher and the director (see Keysar et al., 2000; Wu & Keysar, 2007). A speaker (termed the director) is tasked with asking a listener (the matcher) to move or select objects within a grid. Some of the cells of the grid are covered for one of the interlocutors, thus creating an asymmetry of information between the speaker and the matcher. Previous studies using this methodology (Epley et al., 2004; Keysar et al., 2000) have found that, rather than being influenced by the information that is mutually available to both speaker and listener (i.e. their common ground or shared information), speakers tend to be influenced by the information that is available solely to themselves (i.e. privileged ground). This evidence suggests that egocentric processes are also an important mechanism to consider when investigating language production (Heller et al., 2016; Mozuraitis et al., 2018).

1.2 Application to HCD research

Research on the mechanisms that govern language production in HCD is limited, with recent work calling for further research on this topic (Clark et al., 2019; Peña et al., In Press; Cowan et al., 2023). Most existing literature tends to echo an audience design account, suggesting that language production in HCD is adaptive, being informed by preconceptions of a computer partner's abilities and perceived knowledge (that is, a user's *partner model*; Doyle et al., 2021) (Amalberti et al., 1993; Brennan, 1998; Cowan et al., 2019a; Le Bigot et al., 2007; Meddeb & Frenz-Belkin, 2010a). In comparison to HHD, people tend to use fewer fillers and coherence markers when speaking to a computer (Amalberti et al., 1993), reduce their use of pronominal anaphora, use more basic lexical choices, and make shorter utterances (Kennedy et al., 1988). Similarly, people tend to use simple syntactic structures when interacting linguistically with

animated computer-based agents (Bell & Gustafson, 1999) and also re-use their partner's lexical choices more often with computers than human partners (termed *lexical alignment*; Branigan et al., 2011). Such findings are echoed in more recent literature exploring user language use and speech agent user experience, which highlights that participants adapt their language based on perceived limitations and system capabilities (An et al., 2021), with syntactic and lexical adaptation as well as hyperarticulation after encountering errors in interaction being common (Porcheron et al., 2018; Luger & Sellen, 2016). Consistent with this, people's lexical choices tend to differ when describing tangram pictures (shapes composed of seven simple polygons) for humans or computer partners (Schmader & Horton, 2019). When talking to computers, participants focus more on the geometric features of the images, whereas when interacting with people, participants focus more on the image as a whole. Indeed, recent modelling of alternative explanations for language production in HCD gives strong evidence of audience design being a major driver of language production over other more mechanistic processes such as priming (Rothwell et al., 2021).

This reliance on audience design is thought to be driven by users seeing computers as *at-risk listeners* (Oviatt et al., 1998) or basic dialogue partners (Branigan et al., 2011), leading them to adapt their language to increase the likelihood of communication success based on these perceptions. Even through there has been significant development of more natural capabilities within speech interfaces, recent user studies still demonstrate a category distinction between human and machine conversation (Doyle et al., 2019; Clark et al, 2019; Reeves et al., 2019), with machine partner interaction being perceived as less flexible, having to be learned (Doyle et al., 2019). Others have highlighted that human-like aspects of speech interface design lead to a mismatch between partner models of system capability and the comparatively limited functional capabilities of speech interface, being detrimental to interaction (Luger & Selen, 2016).

Echoing the influence of human-likeness cues, studies have also shown that

phonetic characteristics of speech synthesis and the resulting association with certain speaker groups can impact perceptions of knowledge and capability, leading to language choices that reflect the design of the computer (Cowan et al., 2019a). Recent work looking at the use of US English or Hiberno-English lexicons found that participants were more likely to use US English terms when interacting with a US-accented computer that they were told was US-based compared to an Irish-based and Irish-accented computer (Cowan et al., 2019a).

Yet, other evidence suggests that egocentric processes may also be present in HCD. In Cowan et al.'s (2019) study, although the design parameter of computer accent influenced the likelihood of congruent lexical choices, the majority of lexical items used were still Hiberno-English (i.e. egocentric to the Irish participants taking part in the study), suggesting that participants' lexical choices were influenced by egocentric factors as well as audience design. Moreover, other studies investigating participants' levels of lexical and syntactic choices in HCD have shown no partner effects when comparing human and computer partners (Cowan & Branigan, 2015; Cowan et al., 2015). More recently, a study comparing L2 language learner dialogues with either a fellow human or a computer interlocutor found instances of both audience design and egocentric production in interactions within the computer interlocutor condition (Dombi et al., 2022). Such results suggest that egocentric production processes may also influence interaction in HCD.

1.3 The social role of the computer partner

The Computer Are Social Actors paradigm (CASA; Nass et al.,1994; Nass & Moon, 2000) states that when people interact with computers, they can display social reactions similar to those seen in interactions with human partners. These social responses are easy to generate because computers elicit human-like cues that result in users reacting socially to them (Go & Sundar, 2019; Krämer, 2008; Louwerse et al., 2005; Niewiadomski & Pelachaud, 2010). In the case of speech agents, people use social cues and social signals, such as choice of words, gender of voice, strength of language, etc.,

to guide their social behaviour during conversation (Feine et al., 2019). Such cues and signals elicit social reactions similar to those during human interactions. For example, a social cue of a speech agent's gender of voice may result in the social reaction of applying gender stereotypes to the speech agent.

Research on the CASA paradigm has demonstrated that the role the computer plays during these interactions influences the social reactions users exhibit. In a tutoring task with a computer, users evaluated the performance of the computer either on the same computer where they performed the task, through paper and pencil, or on a different computer (Nass et al., 1999). They found that people evaluated the computer tutor more positively when being evaluated on the same computer rather than in the other two conditions. This suggests that the computer being seen as part of a task or separate from a task may result in different social attributions, norms and responses. Recent work has begun to re-emphasise the importance of partner social roles in the design of conversational user interfaces (Desai & Twidale, 2022; Simpson & Crone, 2022) whereby the norms that come with these roles may influence our interaction. This may influence the way people choose to engage in perspective taking, leading them to either be more or less egocentric in language production if the computer is seen as being part of a task or separate from a task.

1.4. *Study Outline & Motivations*

The research described above suggests that audience design may not be the only determinant of users' language production in HCD, with egocentrism also potentially informing language use. However, current work observing the presence of both allocentric and egocentric behaviour in HCD interactions has focused on the analysis of fragments from multi-turn conversation tasks with spoken dialogue systems (Dombi et al., 2022). Although such research is informative, it does not allow us to causally and systematically assess the differences between egocentric and allocentric language production across HHD and HCD, nor potential variables that may influence these, such

as partner roles. The use of more naturalistic paradigms also complicates the comparison of research findings from HCD with more controlled experimental methods, such as the director-matcher task common in psycholinguistics research on perspective taking in HHD.

The studies presented are designed to identify whether egocentrism occurs in HCD, comparing egocentric language effects across HCD and HHD interactions. Both are motivated by recent calls for more research on mechanisms that govern language production in HCD so as to develop more theoretical understanding of spoken HCD (Clark et al., 2019; Cowan et al., 2023; Peña et al., 2023). Such work is important to inform current efforts to computationally model user language production in HCD (e.g. Rothwell et al., 2021), whilst also highlighting key theoretical issues (e.g. audience design) that may influence and be influenced by speech interface design (Cowan et al., 2019).

To observe this, we had participants describe objects to *a human* or a *computer partner (partner conditions; 2 levels- between participants)* when there was no competing object (object to be described – termed the *target*: small ball; context: one ball picture visible to partner- *One Target* condition), when there was a competing object that their partner could see (target: small ball; context: small ball and large ball, both visible to partner- *Common Ground* condition), or when there was a competing object that their partner could not see (target: small ball; context: small ball visible to partner; large ball not visible to partner- *Privileged Ground* condition) (*perspective conditions: 3 levels- within participants*). Critical to the assessment of whether egocentrism or audience design occurs in these studies is in the participant's use of scalar adjectives (e.g. ‘small’, termed a *scalar modifier*) before the noun when describing target items when competing objects are present and when these are visible (or not) to their partner. Egocentric language production would be said to have occurred when people do not consider the visibility of competitors to their partner (thus more likely to use scalar modifiers whether their partner can see competitors or not). Audience design would be said to have occurred if people

are more likely to consider the visibility of competitors to their partner (i.e. use scalar modifiers more when a competing object is visible to both the partner and the participant). Across both of these experiments we hypothesise that speakers may show egocentrism to different extents depending whether they interact with a human or computer partner. Informed by findings from experiment 1 and by previous work on how a computer's role may influence perceptions of interaction (Nass et al., 1999), we then aimed to explore how the framing of a computer partner as separate from the interaction task, with increased salience in perspective taking as a motivation in interaction, influenced levels of egocentrism and audience design in HCD.

2. Experiment 1

2.1 Hypotheses

The goal of Experiment 1 is to compare the influence of perspective (within participants) on language production when playing a referential communication game (termed the director-matcher task) with either a human or computer partner (between-participants design). The experiment sets out to test whether 1) egocentric language production occurs in HCD and 2) whether this varies significantly depending on whether people interact with a human or computer partner. We hypothesise that there will be a statistically significant difference in the likelihood of using scalar modifiers across the perspective conditions and that this effect will be impacted by partner conditions. More specifically, we predict that people will be more likely to produce scalar modifiers (i.e., more egocentric) when interacting with a computer partner in comparison to a human partner.

2.2 Methods

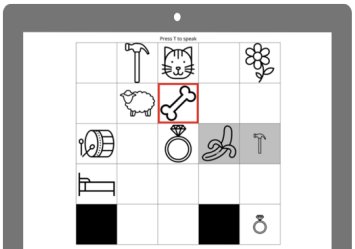
2.2.1 Participants

67 native British-English-speaking participants were recruited using Amazon Mechanical

Turk. Before taking part in the study, participants were asked to confirm that they were native British-English speakers, had normal-to-corrected vision, had normal-to-corrected hearing, and did not suffer from any diagnosed speech or cognitive impairment before taking part in the study. 21 participants were removed from the sample due to being non-native British-English speakers or for inattention during the game (moving to different tabs on their browser five times or more during the study) leaving 46 participants within the sample (15 Female, 28 Male, 1 Non-Binary and 2 Preferred not to say; Mean age=31.26 yrs; SD age=8.25 yrs; Computer Condition- N=23; Human Condition- N=23). The majority of participants held a bachelor's degree (N=20; 43.4%), with 9 (19.5%) holding a master's degree or higher. Four participants (8.6%) held a vocational qualification with the remainder (N=13; 28.2%) holding a secondary- or high school-level qualification. The sample were frequent users of speech agents, with the majority using them either daily (N=17; 36.9%), a few times a week (N=5; 10.8%) or a few times a month (N=10; 21.7%), with the remainder of participants either using them rarely (N=7; 15.2%) or never (N=7; 15.2%). The most commonly used speech agent was Amazon Alexa (N=16; 34.7%) with Apple Siri (N=9; 19.5%), Google Assistant (N=12; 26.08%) and Microsoft Cortana (N=1, 2.1%) also being commonly used. The rest of the participants mentioned they did not use speech agents (N=8, 17.3%). The study was conducted according to British Psychological Society ethics guidelines and was cleared by the University College Dublin ethics procedure for low-risk studies. Participants were given US$6 as an honorarium for taking part.

2.2.2 Design and Materials

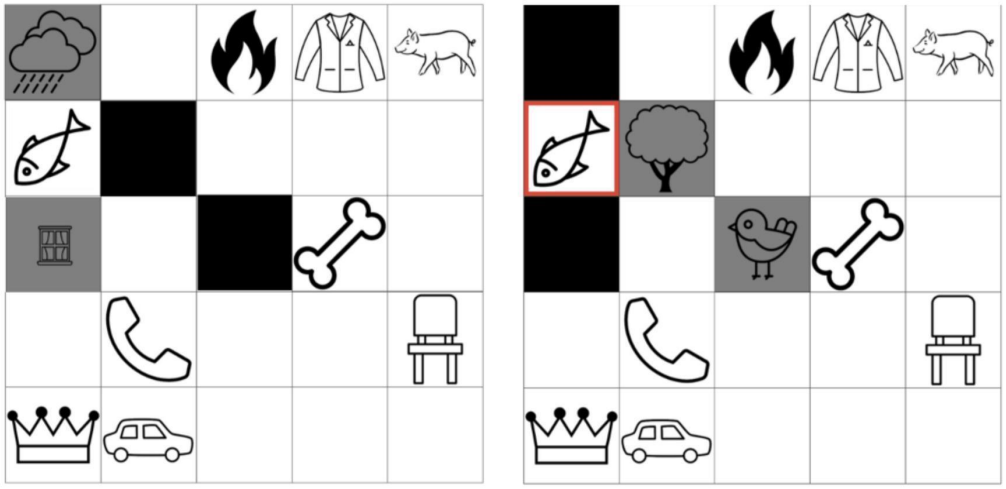
For the experiment, participants were asked to complete an online director-matcher task (see Figure 1) wherein they took turns selecting objects from a grid (i.e. 'matching') based on the description of the partner, and describing objects for the partner to match (i.e. 'naming'). Within each turn, participants were presented with a 5x5 grid where a number of images were displayed.



interface

Fig.1. Schematic representation of the user and the interface they used.

During a matching turn (see Figure 2a), participants heard a description from their partner of an object on the grid, and were tasked with selecting this image as quickly and as accurately as possible. In naming turns (see Figure 2b), participants had to name an object on their grid highlighted by a red box as quickly and as accurately as possible. Importantly, before they commenced the game, participants were informed that their partner's grid differed from their own. How they varied was signalled by the display of images within the grid. Participants were told that images with a white background could be seen by both the participant and their partner (i.e. they were in *common ground*). They were also informed that images with a grey background could only be seen by themself and not their partner (i.e. they were in *privileged ground*). Black squares on the grid represented the placement of images in their partner's privileged ground and were used to further emphasise the differing view of each participant in the game. The pictures used for the game were public domain, monochromatic line drawings of common nouns, selected from lists of common nouns from the Swadesh list (Swadesh, 2017). All materials can be accessed at OSF. Six objects (a *bone*, a *ring*, a *hammer*, a *car*, a *moon*, and a *house*) were used as the experimental objects across the game and seven objects (a *bed*, a *drum*, a *sheep*, a *cat*, a *flower*, a *banana*, and a *plane*) were used as filler objects. For every directing-matching turn pair, eleven objects were displayed on the grid, six of which were filler objects and five of which were experimental objects. Of these, nine objects appeared on a white background (i.e. *common ground*) and two with a grey background (i.e. *privileged ground*). The location of the objects was randomised and the number of turns in which a given object appeared was counterbalanced by condition across all trials.



a) b)

Fig. 2. Example used in the participant instructions to show the difference in what is viewed when being the matcher and the namer for the same trial **Fig. 2a.** demonstrates the screen layout if the participant was the matcher (i.e. In a matching turn). The grey images are those in the participant's privileged ground; those with the white background are in common ground. Black squares represent placement of images in the partner's privileged ground. **Fig. 2b**. demonstrates the same screen layout if the participant was the namer (i.e. in a 'naming' turn). Highlighted square shows the target object to name. The objects displayed do not directly correspond to stimuli used in experiment trials and were used for illustrative purposes only.

2.2.3 Perspective Conditions

Within the naming turns, participants named objects under three different perspective conditions in a within-participants design (six naming turns per condition):

1) *One Target condition:* The target image was the only one of its kind in the grid (see Figure 3a).
2) *Common Ground condition:* The target image had a larger equivalent (i.e. competitor) in the grid, which was visible to both the participant and their partner (see Figure 3b).
3) *Privileged Ground condition:* The target image had a competitor in the grid, but in *privileged* ground, i.e. only visible to the participant (see Figure 3c).

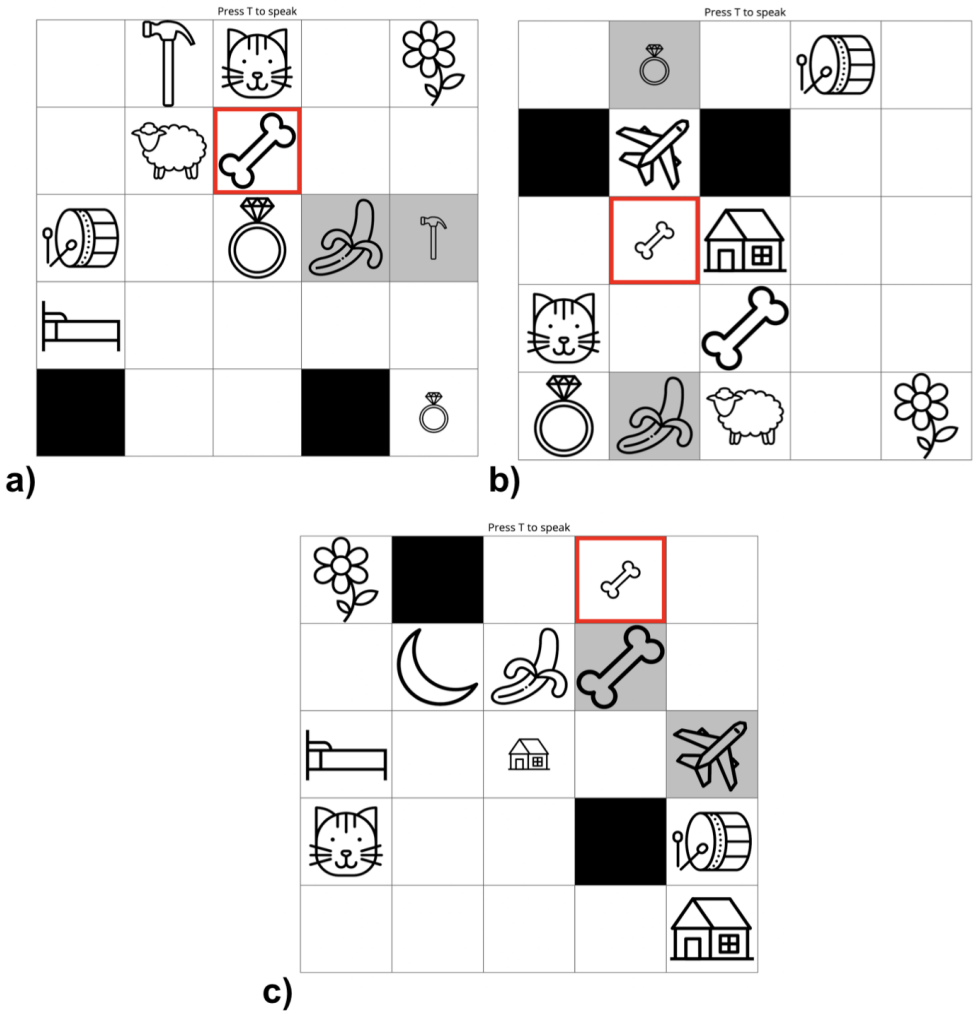


Fig. 3a. Example experiment trial grid for the one-target condition: The target object has no competitor. **Fig. 3b.** Example experiment trial grid for the common ground condition: The target object has a competitor visible to both the participant and the dialogue partner. **Fig. 3c.** Example experiment trial grid for the privileged condition: The target object has a competitor visible only to the participant.

Participants also completed six additional naming turns where they had to name a filler object that had no competitor. The purpose of these additional turns was to hide the aims of the study.

This experimental setup is similar to those employed in previous work on perspective taking research in HHD (e.g. Brown-Schmidt, 2012; Lane et al., 2006; Yoon et al., 2012). In the Common and Privileged Ground conditions, the target images were always small versions of the same experimental object, with a larger equivalent image

acting as a competitor (see Figure 3b and Figure 3c). The order in which the target items and perspective conditions were presented to the participant was pseudo-randomised, ensuring that consecutive naming turns did not use the same target items or perspective conditions.

2.2.4 Partner Conditions

Partners in the director-matcher game were either a *computer partner* (Male Southern British English synthetic voice–Cerevoice William) or *a human partner* (Male Southern British English-accented voice–voiced by one of the authors) in a between-participants design. Similar to previous research (Branigan et al., 2011), both partners were simulated using a reverse Wizard of Oz method. All partners' directing turns were scripted and simulated using pre-recorded audio clips of synthetic speech (for the computer partner) or real recorded audio responses of one of the authors playing the game (for the human partner condition). So as to increase the believability of the human partner, four sessions of one of the research team members playing the game were recorded. One of the four unique audio clips for each relevant item were randomly chosen to be played to the participant to simulate the human partner's turn. Partners only produced utterances of filler objects during their directing turns (e.g. click the bed). During partners' matching turns, participants did not receive verbal or explicit visual feedback about whether the partner selected the correct image, with the progression of the game giving implicit feedback that their partner had made a selection.

2.2.5 Dependent Variable–Scalar Modifier Use

The use of the scalar modifier 'small' within a participant's description of the target item in each of the perspective conditions (termed critical trials) represents the study's binary dependent variable (1= scalar modifier used [e.g. *small bone*]; 0= no scalar modifier [e.g. *bone*] to describe the item). The level of scalar modifier use (e.g. *the small car*) when describing objects in the privileged and common ground conditions captures the influence of shared knowledge–whether an object is in privileged ground or common ground–on the

production of referents. In particular, the use of scalar-modified nouns in the privileged ground condition indicates egocentric language production as it relies on information, the presence of a competitor, that is not known to the partner (i.e., not in common ground).

2.2.6 Procedure

The experiment received ethical clearance from the University's ethics committee, being deemed low risk. Before taking part in the study participants were asked to ensure that they met the requirements of the study in that they must be over 18 years of age, be a native British English speaker with no known cognitive or speech based impediments, have normal to corrected vision or hearing and have speakers, earphones or headphones connected to their computer as a well as a microphone to record their descriptions.

Upon clicking the link to the study, participants were then automatically allocated to either the human or computer condition version of the online experiment. They were then given information about participation, informing them that their picture naming was to be recorded and were asked to give consent. Next, participants were then asked to confirm whether they were a native British-English speaker, whether they had normal-to-corrected vision, normal-to-corrected hearing, or suffered from any diagnosed speech or cognitive impairment. Then, participants were given task instructions. Within these instructions, participants were informed that they would be asked to play a picture naming and matching game with a partner. They were informed that their partner would be *another participant* (human condition) or *a computer* (computer condition) depending on the condition they were allocated. They were told that they would be connected to their partner before they started playing the game. Next, participants were given specific instructions about the grid game. They were told that they and their partner would see grids of images and that each of the player's grids contained the same objects in the same locations. If they saw a red box around an image, they should ask their partner to click on that image (i.e. naming turns). If they did not see a red box around an image,

then their partner would ask them to click on an object (i.e. matching turns). They were told that they would take turns, alternating between clicking their partner's described picture, or telling their partner which picture to click, with the aim of the game being to complete the naming and clicking tasks as quickly and accurately as possible. They were asked to press and hold the T key to speak to their partner during naming turns and, once they named the object, their partner would click on the requested object and advance to the next turn. Participants releasing the T key signalled the end of the naming turn, with the experiment progressing to the next turn. So as to minimise audio recording across the experiment session, recording only commenced at the start of each naming turn and ended once the participant's naming turn had been completed and the experiment had progressed to the next (i.e. matching) turn in the game.

Participants were then informed about the grid layout. They were told that some objects on the grid would be visible only to them and that these objects would be displayed with grey backgrounds for them, but would be displayed as black squares for their partner. They were also told that some objects on the grid would be visible only to their partner and that these objects would be black squares for them but grey-background images for their partner. So as to ensure that they understood the different perspectives, participants were also given a visual example of how the same grid might look to each player (see Figure 2) on a given turn, with a description of what could be seen by each player. After being told that they would first play some practice rounds with their partner, the system then simulated connecting to a partner. This occurred both in the human or the computer condition. This was to enhance the study's realism, and was done by taking five seconds to simulate trying to find the participant's partner, displaying a circular timer with the text "Finding partner" underneath. The text "Partner Found. Practice beginning in a moment" was then displayed on the screen and the game started. The game involved a total of 30 matching-naming turn pairs (6 practice matching-naming turn pairs and 24 experimental matching-naming turn pairs). In the 30 matching turns, participants listened to either the synthetic voice or the natural

voice according to the condition they were allocated (human vs. computer). To avoid priming participants with specific descriptions of the experimental objects, the partner in matching turns only named filler objects without using scalar modifiers (since filler objects were always displayed without competitors).

For practice naming turns, participants named experimental objects without any competitors present so as to familiarise them with the objects before the experimental turns and so as to identify the participant's naming conventions for the objects. After the practice session, participants then completed the 24 experimental naming-matching experimental turn pairs that included 24 naming turns (six naming turns–termed *critical trials*–in each of the three perspective conditions, making a total of 18 critical trials, and six additional filler turns, counterbalanced so that each experimental object was named in each condition) as well as 24 matching turns.

After completing the director-matcher game, participants were asked to complete a demographic questionnaire identifying their highest level of education completed, how often they used speech agents like Alexa, Siri, or Google Assistant, which agents they used most often, as well as asking about their thoughts on the partner they played the game with. Finally, participants were thanked and debriefed as to the aims of the study and how their data will be used and treated. All participants were also debriefed as to the fact that their partner was in fact a computer with all instructions pre-recorded. They were then given a code to claim payment. After completion of the experiment, a message was also sent to all participants thanking them for taking part in the study, taking the opportunity to debrief them again about the purposes of the study, their partner being pre-recorded, what was measured, how data were recorded, confidentiality and de-identification procedures, as well as giving them another opportunity to withdraw from the study if they wished.

2.3 Analysis and Results

2.3.1 Data processing and coding

Target responses (N=828) were collected from 46 participants. 332 responses with descriptions including a scalar modifier were coded as 1 (e.g. *the small ring*), and 465 descriptions not including a scalar modifier were coded as 0 (e.g. *the ring*). Descriptions that did not fit this structure (e.g. '*click the bone under the plane'* or '*click the house in the middle'*), included an alternative scalar modifier (e.g. *the large ring*), or that did not describe the target item accurately like naming the objects with an incorrect referent (i.e. *click the bat*) were classed as Other and coded as NA (N=31). The frequency of scalar modifier use across the three perspective conditions is shown in Table 1.

Table 1
Frequency of scalar modifier use in Human and Computer conditions in Experiment 1

Perspective Condition	Scalar Modifier Use	Computer Condition (N=23)	Human Condition (N=23)
One Target	No scalar modifier	125	134
	Scalar modifier	1	0
	Other	12	4
Privileged Ground	No scalar modifier	56	80
	Scalar modifier	78	56
	Other	4	2
Common Ground	No scalar modifier	41	29
	Scalar modifier	95	102
	Other	2	7

2.3.2 Analysis plan

To test our hypotheses, we used Bayesian generalised linear mixed-effects models (GLMMs), using the *bglmer* function of the *blme* package (Chung et al.,

2013)–an extension of the *lme4* package (Bates et al., 2015)–in RStudio (version 1.2.5033), with a binomial link function, with participants and items as random effects. We attempted to use the maximal random effects structure justified by our design (Barr et al., 2013). In cases where the model would not converge, we simplified the random structure appropriately (e.g. fixing correlations among random effects to zero and/or simplifying slope terms; Bates et al., 2015). Unless otherwise specified, all predictors were contrast coded (-0.5, 0.5). For the GLMM models, we report coefficient estimates (*B*), standard errors (*SE*) and *z* and *p* values for each predictor; 95% confidence intervals (CI) are from the *confint* function (method=Wald).

We conducted two stages of analyses. First, we investigated the presence of egocentrism in participants' tendency to use scalar modifiers, by testing the difference between their propensity to use scalar modifiers in the One Target perspective condition and the other two perspective conditions (Privileged Ground and Common Ground). Second, we investigated the presence of egocentrism in participants' tendency to use scalar modifiers, by testing: i) the difference between Computer Partner condition and the Human Partner condition overall, and ii) how the effect of the Perspective conditions by the Partner condition interacted.

Full model structure (i.e. fixed and random effects) are reported in Tables 2-7 in the Appendix A of the supplementary materials.

2.3.3 Results

Egocentrism and use of scalar modifiers by perspective conditions. Overall, participants were significantly less likely to use scalar modifiers in the One Target perspective condition (M=0.38%[SD=2.5]) than in conditions where there were competing objects displayed (i.e. the Privileged and Common Ground conditions) (M=61%[38]; *B* = 7.58, *SE*= 1.28, *z*= 5.92, *CI*=[5.07, 10.08], *p*<.001), suggesting that participants' tendency to use scalar modifiers was positively affected by the number of objects of the same type that were visually available to them (see Figure 4). Such an

effect is evidence for participants being affected by the visual information available to them in the grids when generating utterances. Moreover, participants were significantly less likely to use scalar modifiers in the One Target perspective condition (M=0.38%[SD=2.5]) than in the Privileged Ground perspective condition (M=50%[SD=38]; *B* = 7.88, *SE*= 1.15, *z*= 6.87, *CI*=[5.63, 10.12], *p*<.001), suggesting that participants' tendency to use scalar modifiers is at least partially driven by an egocentric component. On the other hand, participants were significantly less likely to use scalar modifiers in the Privileged Ground perspective condition (M=61%[38]) than in the Common Ground perspective condition (M=73.78%[34]; *B* = 2.08, *SE*= 0.29, *z*= 7.12, *CI*=[1.51, 2.66], *p*<.001), suggesting that participants' tendency to use scalar modifiers was also affected by audience design. This finding is consistent with previous work evaluating egocentrism and audience design in the research literature on perspective taking. In the presence of two visually available objects (one in common ground and one in privileged ground), there is an interference from one's privileged knowledge (Wu & Keysar, 2007). However, there is a clear sensitivity to the referential context where both objects were mutually available to participant and partner (i.e. common ground) (Yoon et al., 2012).

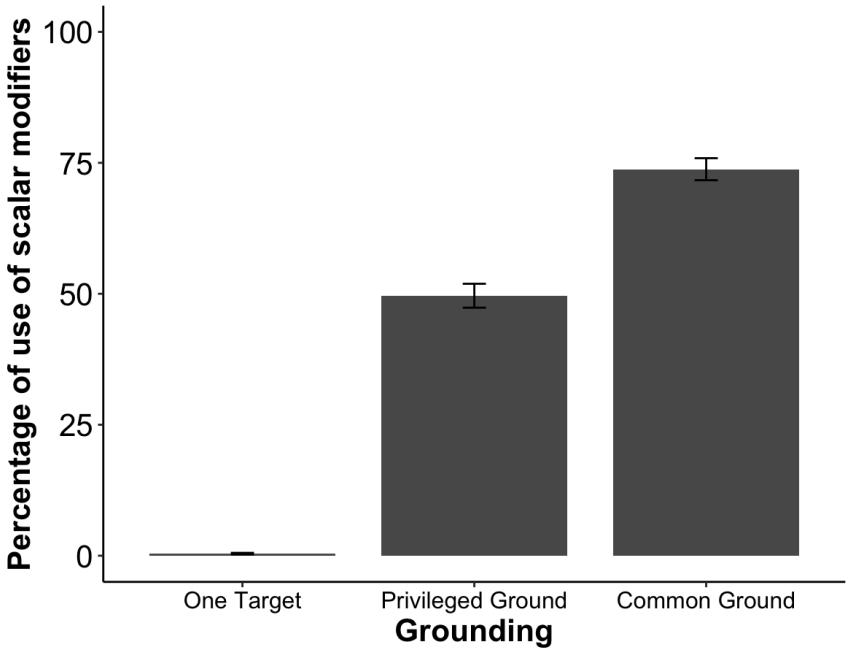


Fig. 4. Mean and standard error of percentage of use of scalar modifiers across Perspective conditions in Experiment 1.

Egocentrism and use of scalar modifiers by partner conditions. Across the perspective conditions, participants used scalar modifiers less often in the Human partner condition (M=39%[41]) than in the Computer partner condition (M=44%[43]), but this difference was not statistically significant overall (B = -1.25, SE= 0.97, z= -1.28, CI=[-3.17,0.67], p=0.20; see Figure 5), suggesting that participants' tendency to use scalar modifiers overall was not generally affected by whether their partner was a human or a computer.

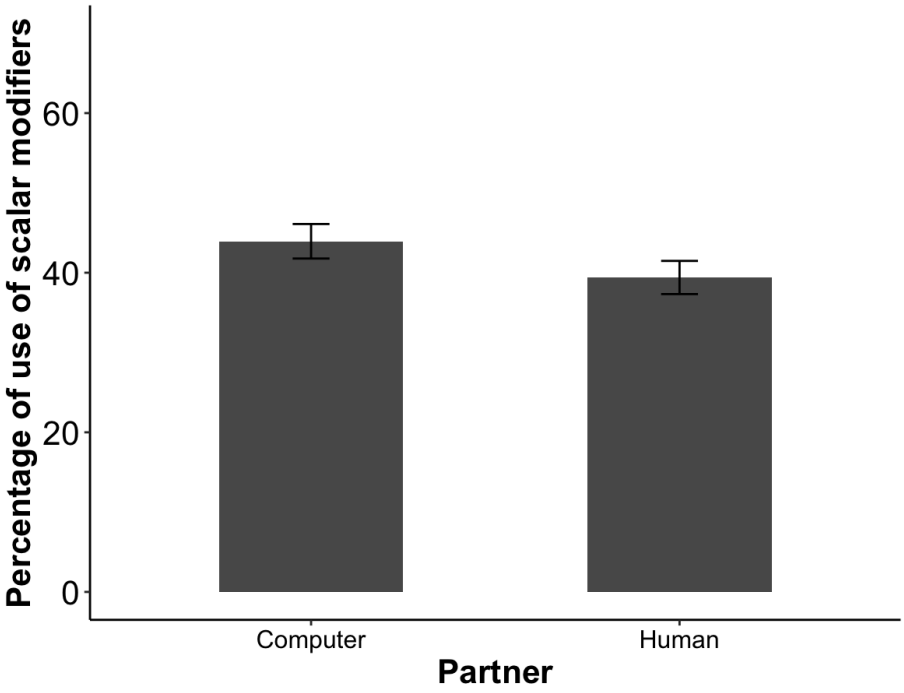


Fig. 5. Mean and standard error of percentage of scalar modifier use across Partner conditions in Experiment 1.

However, a significant interaction between perspective condition and partner condition (B = 1.73, SE= 0.57, z= 3.02, CI=[0.61,2.85], p=0.003) suggests that partner type modulated the effect of the perspective condition on participants' tendency to use scalar modifiers and that this modulation may have differed between perspective conditions (see Figure 6). In particular, participants in the Human partner condition used scalar modifiers on 78% of critical trials in the Common Ground perspective condition versus 41% in the Privileged Ground perspective condition (a difference of 37%; B = 3.67, SE= 1.14, z= 3.22, CI=[1.43,5.89], p<.001). However, participants in the Computer condition used scalar modifiers on 70% of critical trials in the Common Ground versus 58% in the Privileged Ground condition (a difference of 12%; B = 1.92, SE= 0.94, z= 2.04, CI=[0.077,3.77], p=0.04).

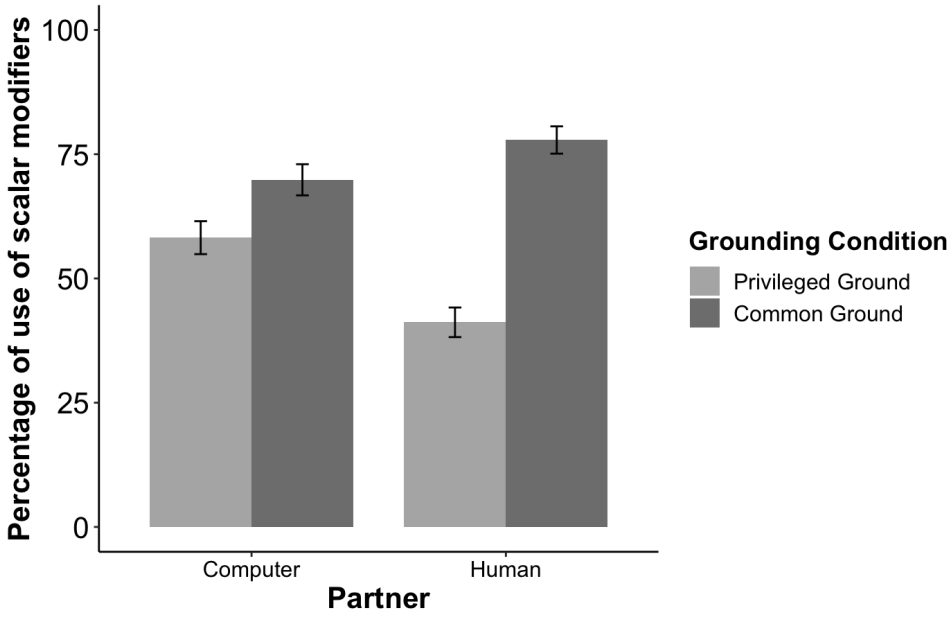


Fig. 6. Mean and standard error of percentage of scalar modifier use across Partner conditions and Perspective conditions in Experiment 1.

2. 4 Discussion

Our results provide evidence about the presence of egocentric behaviour in HCD language production similar to that in HHD (Yoon et al., 2012). However, the stronger presence of scalar modifiers in the computer condition compared to the human condition suggests a stronger bias towards egocentrism when people talk to computer partners. Such egocentrism occurs along with a weaker presence of audience design in the computer condition compared to the human condition. Such difference in the presence of egocentric behaviour and audience design shows that both privileged and common knowledge are processed during dialogue but this occurs to a different extent according to the beliefs of the interacting partner.

The outcome of our experiment supports recent assertions from corpora-based studies that have identified the presence of egocentric biases, along with audience design-related mechanisms when interacting with speech agents (Dombi et al., 2022). Based on previous work that strongly emphasised the role of audience design based processes in HCD production (Amalberti et al., 1993; Bell & Gustaffson, 1999; Branigan

et al., 2011; Cowan et al., 2019a), we might have expected people to exert more effort when communicating with computers and to see a less frequent use of scalar modifiers in the privileged condition, since people tend to hold low expectations of the communicative capabilities of computer partners (Branigan et al., 2011; Luger & Sellen, 2016). Yet, our findings seem to show the opposite in that people tended to be less sensitive to the state of common ground when interacting with computer partners than with human partners, by showing that they consider privileged information in the construction of their utterances to a greater extent in HCD than in HHD.

One possible explanation for this difference might lie in the framing of the computer partner. We suspect that framing the computer partner as a 'computer' can have different interpretations that may result in variation of the results. In addition, the connection to a partner simulation was fast (5s) which may have resulted in people not assuming that a computer was actually connecting to the system independently. This may have given the impression that the computer partner was integrated in the system by which the partner carried out the task, rather than a separate dialogue partner as usually conceived in HHD (i.e. an independent social agent with shared responsibility for mutual comprehension).

The framing of the computer partner as integrated within a specific device or application echoes how speech based IPAs are perceived to function (Clark et al., 2019; Doyle et al., 2019, Cowan et al., 2017). Based on this, it may be that participants felt they could use utterances that were easier for them to produce, without feeling the need to consider a computer partner's perspective. This echoes work that shows the framing of the role of a computer (i.e. whether the computer people use to evaluate an interaction is also the system they conducted the original interaction on vs a separate system used for both) can significantly impact how people evaluate computers (Nass et al., 1999). Indeed, the perception of the social context of the interlocutor (i.e. whether the interlocutor is in a situation where they are unable to collaborate) can also influence how allocentric people are in HHD (Duran et al., 2011).

The current framing of the computer may have primed participants to not consider the computer as a separate partner who has specific information requirements that need to be considered in order to successfully play the game. To investigate this further, we conducted a second experiment where we explored the impact that the computer's role had on the results obtained in Experiment 1. This was implemented by 1) modifying the framing of the role of the computer by telling participants that they would be playing with a virtual voice agent instead of a computer; 2) increasing the time it took for the partners to connect to the game so as to emphasise its status as an communicative partner independent from the platform and by 3) emphasising the differences in perspective, making it more salient that the partners could only see their side of the game. The aim of Experiment 2 was thus to test whether more egocentric behaviour in the computer than the human condition would persist given these manipulations.

3. Experiment 2

3.1 Hypotheses

Similarly to Experiment 1, the aim of the current experiment was to compare the influence of perspective (within participants) on language production when people interact with either a human or computer interlocutor (between participants). Following the same director-matcher paradigm employed in Experiment 1, we again hypothesise that there will be a statistically significant difference in scalar modifier use across the perspective conditions, with evidence for both egocentric and allocentric production. We predict that the effect of the partner conditions on perspective taking seen in Experiment 1 may differ based on the modifications of Experiment 2. More specifically, there will be no difference between the use of scalar modifiers between the computer and human conditions.

3.2 Methods

3.2.1 Participants

100 British Native English-speaking participants were recruited using Prolific. Three participants were removed from the sample due to inattention when completing the study (screen switching 5 times or more during the game), leaving 97 participants within the sample (48 Male, 49 Female; Computer Condition- N=48; Human Condition- N=49) with a mean age of 33.64 yrs (SD = 12.06 yrs). Participants were prescreened to confirm they were British Native English speakers, they had normal-to-corrected vision, normal-to corrected-hearing, did not suffer from any diagnosed speech or cognitive impairment, and had not taken part in a similar study in Amazon Mechanical Turk. The majority of participants held a Bachelor's degree (N=39, 40.20%), followed by Secondary or High School (N=33, 34%), Master's or Higher (N=14, 14.43%), Vocational Qualification (N=10, 10.3%) and No Answer (N=1, 1.03%). Most of the sample were users of speech agents, with the majority using them either daily (N=22, 22.68%), a few times a week (N=14, 14.43%) or a few times a month (N=19, 19.58%), with the remainder of participants either using them rarely (N=22; 22.68%), never (N=19; 19.58%) and No Answer (N=1, 1.03%) The most commonly used speech agent was Google Assistant (N=29, 29.89%), with Amazon Alexa (N=28, 28.86%) and Apple Siri (N=17; 17.52%) also being commonly used. The rest of the participants stated they did not use speech agents (N=23, 23.71%). The experiment was conducted according to British Psychological Society ethics guidelines and was cleared by the first author's university ethics procedure for low-risk studies. Participants were paid £5 for taking part.

3.2.2 Design and materials

The task for participants was identical to the one described in Experiment 1.

3.2.3 Perspective conditions:

Perspective conditions were the same as Experiment 1.

3.2.4 Partner conditions:

The main differences from Experiment 1 lay in the way the computer partner was introduced. The computer partner was introduced as a virtual voice agent rather than solely *as a computer*. Additionally, before the game commenced, participants were explicitly informed that a separate agent had been selected to play the game with them and that they can only see their side of the game, being told that “Voice Agent called ID-X has been selected. “ID-X is a computer voice agent. ID-X can only see its side of the game and can hear what you say when you press the T key. Waiting for the partner to signal they are ready to start”. Similar text was also used for the human condition, displaying a fictitious Prolific ID to increase the believability that another prolific user had connected to the system to play the task. Participants were told that “Partner has been selected: Prolific ID 4AC3O786A5PR2V010AB1350B. 4AC3O786A5PR2V010AB1350B can only see their side of the game and can hear what you say when you press the T key. Waiting for the partner to signal they are ready to start.” All other aspects of the conditions were identical to Experiment 1. For both the human and computer conditions, we also increased the time it took to find a partner on the Finding Partner screen, whereby a circular timer was displayed for 30 seconds instead of 5 seconds of Experiment 1. This was so as to add to the believability of the experiment connecting to a separate partner to play the game with.

3.2.5 Procedure

The main differences between the procedure of Experiment 1 and 2 lay in how the computer partner was framed as a separate dialogue partner from the system where the

game was taking place (see Partner Condition section). All other aspects of the procedure were identical to Experiment 1 including the debriefing of participants. All participants were also debriefed as to the fact that their partner was in fact the recording of a human or a computer with all instructions pre-recorded.

3.3 Analysis and Results

3.3.1 Data processing and coding

Out of the total 97 participants, target responses were collected (N=1746). 773 responses were coded as descriptions including a scalar modifier (1), with 906 coded as not including a scalar modifier (0). As in Experiment 1, descriptions that did not fit this structure, included an alternative scalar modifier or that did not describe the target item accurately were coded as NA (N=67). The frequency of scalar modifiers used are shown in Table 8.

Table 8
Frequency of scalar modifier use in Human and Computer conditions in Experiment 2

Perspective Condition	Scalar Modifier Use	Computer Condition (N=48)	Human Condition (N=49)
One Target	No scalar modifier	264	261
	Scalar modifier	0	1
	Other	24	32
Privileged Ground	No scalar modifier	133	121
	Scalar modifier	150	169
	Other	5	4
Common Ground	No scalar modifier	73	54
	Scalar modifier	213	240
	Other	2	0

3.3.2 Analysis plan

We conducted the same two stage analysis as in Experiment 1. We first tested for the presence of egocentrism in participants' tendency to use scalar modifiers by perspective conditions and then tested for the presence of egocentrism in participants' tendency to use scalar modifiers by partner conditions. Full model structure (i.e. fixed and random effects) are reported in Tables 9-12 in the Appendix B of the supplementary materials.

3.3.3 Results

Egocentrism and use of scalar modifiers by perspective condition. As in Experiment 1, participants were significantly less likely to use scalar modifiers in the One Target perspective condition (M=0.19%[SD=1.9]) than in the other two perspective conditions together (M=67%[37]; *B* = 8.13, *SE*= 1.13, *z*= 7.18, *CI*=[5.91, 10.35], *p*<.001), suggesting that their tendency to use scalar modifiers was positively affected by the number of objects of the same type that were visually available to them (see Figure 7). Moreover, participants were statistically significantly less likely to use scalar modifiers in the One Target perspective condition (M=0.19%[SD=1.9]) than in the Privileged Ground perspective condition (M=56%[SD=38]; *B* = 8.62, *SE*= 1.07, *z*= 8.08, *CI*=[6.5, 10.7], *p*<.001), suggesting that their tendency to use scalar modifiers was again at least partially driven by an egocentric component. Moreover, as in Experiment 1, participants were statistically significantly less likely to use scalar modifiers in the Privileged Ground perspective condition (M=55.7%[38%]) than in the Common Ground perspective condition (M=78.1%[32%]; *B* = 1.95, *SE*= 0.2, *z*= 9.75, *CI*=[1.56, 2.3], *p*<.001), suggesting that the tendency to use scalar modifiers was also generally affected by audience design. This finding is consistent with the results of Experiment 1 and with previous work that emphasise the presence of both egocentric and allocentric processes in dialogue (Wu & Keysar, 2007; Yoon et al., 2012).

Fig. 7. Mean and standard error of percentage of use of scalar modifiers across Perspective conditions in Experiment 2.

Egocentrism and use of scalar modifiers by partner conditions. Participants used scalar modifiers slightly more often in the Human partner condition (M=48% [44%]) than in the Computer partner condition (M=44%[43%]), but this difference was again not statistically significant overall (B = 0.46, SE= 0.71, z= 0.64, CI=[-0.94,1.85], p=0.52; see Figure 8), suggesting that tendency to use scalar modifiers across the experiment was not generally affected by whether participants were interacting with a human or a computer.

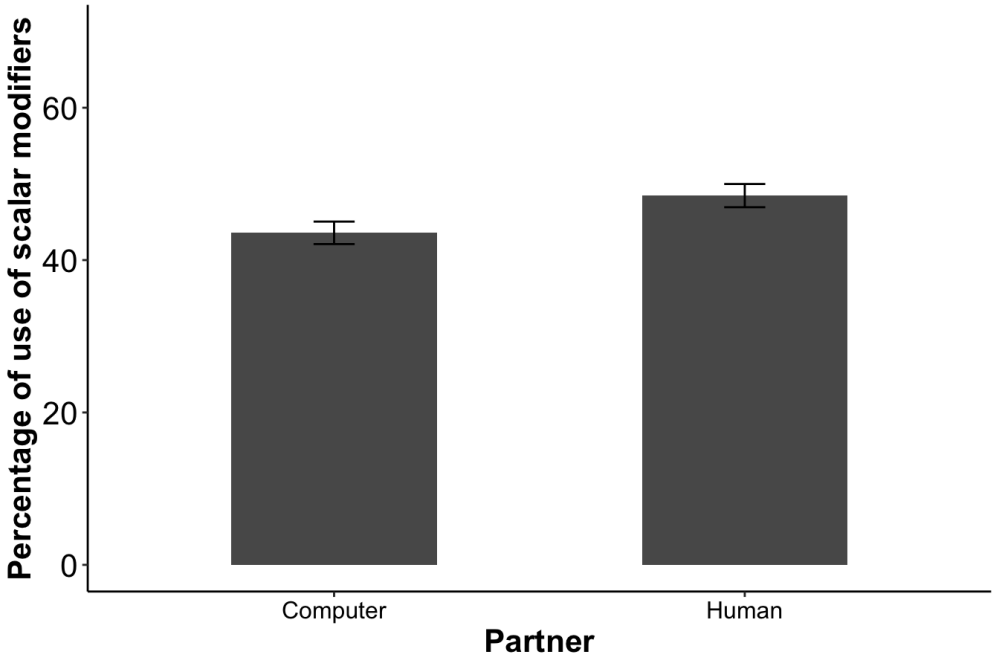


Fig. 8. Mean and standard error of percentage of use of scalar modifiers across Partner conditions in Experiment 2.

In contrast to Experiment 1, the interaction between Partner condition and Perspective condition was not statistically significant (B = 0.88, SE= 0.59, z= 1.5, CI=[-0.27,2.03], p=0.13). This suggests that the partner conditions did not significantly modulate the effect of the perspective condition on participants' tendency to use scalar modifiers (see Figure 9).

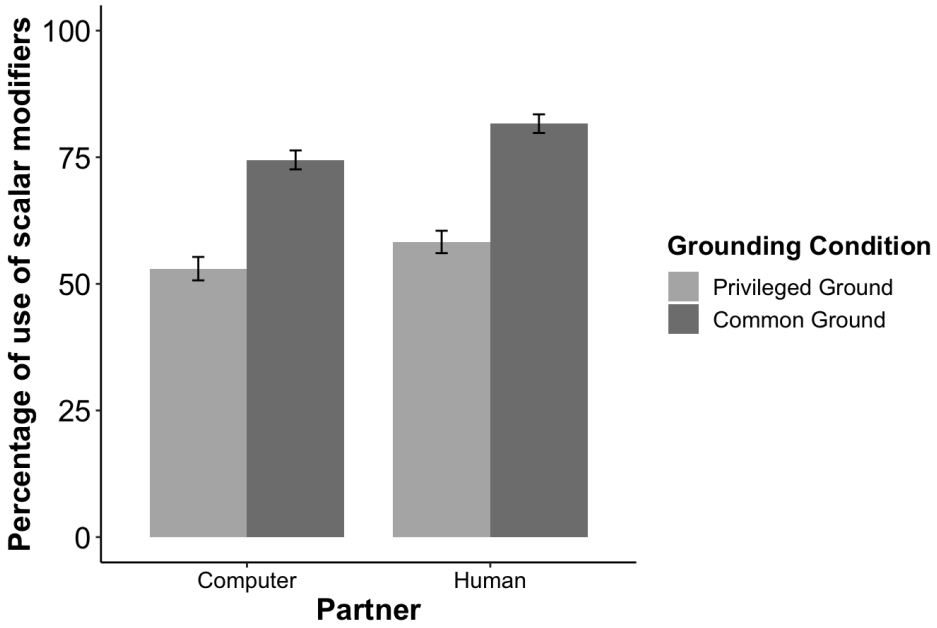


Fig. 9. Mean and standard error of percentage of use of scalar modifiers across Partner conditions and Perspective conditions in Experiment 2.

3. 4 Discussion

Similarly to Experiment 1, the results provide evidence that both privileged and common ground knowledge influence language production (Yoon et al., 2012). However, unlike Experiment 1, the presence of egocentric behaviour and audience design occurs to the same extent in HCD as in HHD. This indicates that the changes implemented in the framing of the computer partner whilst making the perspective differences more salient had an effect on the use of scalar modifiers by participants.

A possible explanation of the opposing results in Experiment 1 and 2 is that of the different processes of perspective-taking: *inferring*, *storing* and *using* (Apperly, 2018; Ferguson et al., 2015). Inferring or being sensitive to the perspective differences does not necessarily mean that this information will be employed in conversation. Rather, this information will be used in language production according to people's motivation to communicate successfully. In this case, in Experiment 1 even if participants were aware of the perspective differences in the task, they might have decided to not use this information to construct their utterances because they had no motivation to do so with the

computer, contrary to what we observed in Experiment 2.

This lack of motivation could be explained by the role the computer partner took in the task. In Experiment 2, where the computer might have been perceived as a separate dialogue partner, there might have been a stronger motivation to engage in perspective-taking to produce optimal utterances for the computer partner compared to Experiment 1. Research in psychology about the social effects in task performance have found that people perform differently when they have a common goal interacting with others (Richardson et al., 2007; Sebanz et al., 2003; Shteynberg & Galinsky, 2011; Spivey, 2007). It may be that in Experiment 2, the emphasis of the computer partner as a separate agent increased the motivation to perceive the computer as more of a collaborator with a common shared goal and a responsibility for the division of labour in dialogue, compared to the computer condition in Experiment 1. Further potential reasons for these differences are discussed in detail in the following section.

4. General Discussion

Language interactions with computer dialogue partners are becoming commonplace (Dafoe et al., 2021). There is currently little understanding of the causal mechanisms that drive user language choices in speech-based HCD (Clark et al., 2019). Yet this knowledge is critical for the ongoing efforts to computationally model user language behaviours in human-computer dialogue within the HCI field (Rothwell et al., 2021), whilst also being important so as to inform technological development and design. The current consensus is that users adapt their language choices based on the perception of the capabilities of computers as dialogue partners (Amalberti et al., 1993; Brennan, 1998; Cowan et al., 2019a; Le Bigot et al., 2007; Luger & Sellen, 2016; Meddeb Frenz-Belkin, 2010a; Rothwell et al., 2021), which can be influenced by design (Cowan et al., 2019), echoing the concept of audience design within human-human dialogue research (Bell, 1984). However, experimental findings comparing language choices with human and computer interlocutors have sometimes found little presence of adaptation

and audience design (Cowan & Branigan, 2015; Cowan et al., 2015), questioning the ubiquity of such an account. Indeed, more recent work observing naturalistic interactions with computer-based language tutors have alluded to more egocentric language production operating alongside audience design (Dombi et al., 2022) in HCD interactions, challenging the notion of audience design as one of the sole drivers for language production in HCD.

Our work aims to build upon recent work (Dombi et al., 2022) by using a controlled experiment paradigm inspired by egocentric language research in human-human dialogue, whether such egocentric language effects occur in HCD. Our experiments advance the field of HCI, in that they are the first to demonstrate experimentally that people do produce language egocentrically when interacting with speech based HCD partners (Experiment 1), and that this tendency is influenced by the saliency of perspective and the role of the computer within the dialogue (Experiment 2). Such an effect therefore needs to be considered when computational modelling user language behaviour in HCD. The findings also emphasise that designers need to consider the role that speech agents take in interaction as well as how shared knowledge is framed and emphasised when designing speech interfaces, as this may influence the mechanisms that are used in user language production Our work also contributes to the HCI field by supplying an online referential communication task paradigm that can be used to advance research on perspective-taking and egocentrism in HCD. Below we discuss the findings, with a particular focus on how allocentric and egocentric processes may work in tandem to influence language production in HCD.

4. 1 The interplay between egocentric and audience design processes in HCD language production

As highlighted in Section 2.3, previous work has emphasised the importance of audience design in HCD language production (Branigan et al., 2011; Oviatt et al., 1998, Rothwell et al., 2021). This aligns with HHD work (Brennan et al., 2010; Clark, 1992,1996; Clark &

Marshall, 1981) that declares dialogue to be a collaborative process where interlocutors adapt their language choices to be more felicitous to the needs of partners in conversation.

Similar to debates in HHD as to the role of audience in dialogue processes (Epley et al., 2004; Keysar et al., 2003), our findings support the notion that audience design may not be a universal mechanism of language production in HCD (Cowan & Branigan, 2015; Cowan et al., 2015; Dombi et al., 2022), with egocentric processes also present in language production. Indeed, this may work in tandem with audience design. Such a dual account echoes HHD work, whereby the role of audience design is seen as probabilistic (Brown-Schmidt & Hanna, 2011) in that the use of perspective when in dialogue is affected by both shared and privileged knowledge. Previous studies have also demonstrated that a number of factors influence whether shared or privileged knowledge is used when producing language, including the communicative scenario or contexts (Heller et al., 2016; Mozuraitis et al., 2018) and time constraints (Horton & Keysar, 1996). Similarly, people may act egocentrically by default, only using perspective to drive language production when they deem it necessary or when they have the cognitive resources to do so (Horton & Keysar, 1996; Keysar et al., 1998). In line with the monitoring and adjustment hypothesis, our findings show experimentally that people act more egocentrically in HCD interactions but that when emphasising the computer partner's independence from the system where the task takes place and making perspective differences more salient (as in Experiment 2), people use perspective knowledge to guide their language production. We note that further studies are necessary to replicate these results. All in all, our findings are an important breakthrough that further our understanding of the causal mechanisms behind user language choices in speech on human-computer interactions (Clark et al., 2019). However, further work is certainly needed to examine how audience design may interleave with more egocentric production processes when in HCD.

4.2 Partner role and the division of labour in human-computer communication

As highlighted above, people may have decided that perspective-taking was more appropriate in HCD interactions when the computer was seen as a separate dialogue partner instead of integrated into the system that was also delivering the game. That is, the differences in framing the partner role may have influenced the level of resources and effort people were willing to give to perspective-taking in the interaction with the computer partner.

Currently little is known about how people decide to allocate perspective taking resources when engaging in conversation with computers. Gricean accounts of language production and language understanding (Clark, 1996; Grice, 1975; Zhang et al., 2006) propose that communication is a collaborative activity between interlocutors. This mutual collaboration aims to minimise joint effort and create a natural division of labour in communication. In HHD, recent work has revealed that this division depends on the expectations the interlocutors have about their addresses' exertion of effort (Hawkins et al. 2021). For example, with uncollaborative partners, people need to allocate more perspective-taking resources or exert more effort to maintain successful communication, whereas with more collaborative partners, people may realise that successful communication can be achieved by exerting less effort. Yet, there is not a clear understanding of how people divide the labour of communication during HCD and how these expectations play a role in the process. In HHD, such division operates because effort is negotiated and executed by both partners (Mey, 2010). In HCD, there is a clear asymmetry in this process, since computers still lack many of the human communicative resources, abilities and systems of knowledge (Doyle et al., 2019; Luger & Sellen, 2016) that permit to divide the labour naturally (Dombi et al., 2022). We call this the *division of labour paradox in HCD* in that it is the human interlocutor alone that negotiates and decides which cooperative strategies to employ during dialogue with a computer as currently computers may have more fixed cooperative capabilities

compared to human interlocutors.

Current work supposes that the expectations of cooperation from the computer's side are low (Branigan et al., 2011; Oviatt et al., 2022) resulting in people exerting more effort to ensure communicative success in HCD compared to HHD. For example, people are more likely to engage in audience design in HCD compared to HHD (Rothwell et al., 2021; Schmader & Horton, 2019) or are more likely to align to computer partners than to human partners (Branigan et al., 2011) because computers are perceived as basic speakers and listeners. However, our results show that the expectations of the users seem to be altered by the role of the computer partner. For instance, in Experiment 1, the similar use of scalar modifiers in the common ground and privileged condition when interacting with computers supposes low collaborative effort from the participant to engage in perspective-taking, shifting the division of labour towards the computer. Given the perception of the computer as integrated in the game, people might not have found motivation to engage in perspective-taking, resulting in doing what was easier for them, echoing studies in egocentrism in HHD (Barr and Keysar, 2002; Keysar et al., 2003).

This differs from the perspective-taking resource allocation in Experiment 2. Since human and computer partners were equally framed as separate from the game, participants seemed to exert the same amount of effort in both the human and computer conditions, accounting for the asymmetries in the grid when producing their utterances. This suggests that people decided to use a similar division of communicative labour when interacting with both partners. This framing of the perspective-taking as a division of labour may be fruitful to explain differences in egocentric and allocentric language use in HCD in that it clearly emphasises the control exerted by the user in driving collaborative effort in perspective taking. It may also give us a mechanism that can explain varying levels of audience design and egocentric production. Future work should aim to further develop this account of perspective taking in HCD interactions. In addition, we encourage future studies to explore how different framings of the computer partner, specifically framing the computer explicitly as a user controlled tool (as mentioned in

speech IPA perceptions- e.g. Doyle et al., 2019), impact division of labour and resultant perspective-taking.

4.3 CASA based explanations for egocentric and allocentric language production in HCD

An alternative explanation for more egocentric language production in Experiment 1 may lie in the perception of computers as social actors. The Computers Are Social Actors (CASA) paradigm (Nass et al., 1994; Nass & Lee, 2001) asserts that, although people know they are interacting with something that may not warrant treatment like a human (i.e. a computer), people still mindlessly apply social heuristics from HHD in HCD, leading them to respond to computer systems as they respond to other people. Work used to support the paradigm highlights that the social categorisation and role of computers can have an effect on how we perceive these systems. For instance, work suggests that we use social categorisation to inform behaviours and judgements of computers in collaborative tasks (Nass & Moon, 1996). Specific to our findings, work on CASA highlights the importance of how the role of a computer within an interaction influences our response to them, with users being more positive (and thus more polite) about a tutoring computer when asked to evaluate it on the computer with which the tutorial was delivered when compared to a separate computer being used for evaluation (Nass et al., 1999). The impact of the computer being seen as part of a task or separate from a task is similar to the dichotomy of the computer partners in our studies. Yet, rather than being more sensitive to the knowledge asymmetries of the computer when it is seen as delivering the task, which may be deemed as more polite socially, we found the opposite in Experiment 1. This is not the case of Experiment 2, where we found similar behaviours between human and computer partners. As mentioned previously, this may be due to the increased emphasis of perspective asymmetries priming more prosocial goals, but also by manipulating the framing of the computer partner in Experiment 2 and making it more explicitly an agent separated from the game system. Future work should

look to specifically test the propositions within the CASA framework against other potential accounts (outlined above) for the behaviours seen in our studies.

4.4. Omniscience and audience design processes

Our results could also be explained by the participants' perceptions of how the role of the computer may have impacted knowledge of the game state. People tend to make appropriate decisions about how to allocate their perspective-taking resources based on the beliefs they have about their partner's knowledge (Clark, 1996; Mainwaring et al., 2003), echoing previous findings (Cowan et al., 2019a). It may be that people may see the computer partner in the first study as omniscient in their knowledge state in situations where it is both collaborating with the user and controlling the information that is being delivered. Previous work has noted that, although anchored to estimates of what they feel others know, people tend to assume more knowledge of computers than humans in object naming tasks (Cowan et al., 2017). These knowledge expectations are critical to people's partner models of computers as dialogue partners (Doyle et al., 2021). Although this may lead to what seems like egocentric based behaviours, the mechanism by which this occurs may in fact be one of audience design based on assuming the knowledge state of the computer is more complete (i.e. they can see both occluded and non-occluded objects) than when interacting with the human partner. Although not tested directly that this was how participant's perceived the partner's knowledge state in our experiment, this explanation could account for the results seen in these studies. In Experiment 2, where the asymmetry in knowledge state was strongly emphasised in the game instructions and the speech agent was made more explicitly independent from the game system, it may have lowered the perception of omniscience. Further work is needed to discount the explanation of perception of knowledge states for our studies.

4.5 Limitations

Our work used a *reverse* Wizard of Oz paradigm (e.g. Branigan et al., 2011) for

both the human and computer conditions, in that we used recordings of human and computer speech respectively rather than a human wizard. Similar to other studies that simulate text-based human partners in dialogue interactions (Hawkins et al., 2021), we aimed to increase the believability of the human partner condition by randomly selecting from a number of recordings of the same partner audio, recorded by one of the authors. We also emphasised whether people were playing with either a partner or computer within the game instructions in addition to simulating connecting to a partner within the study so as to increase believability. That said, people may have still felt that they were not interacting with a real human partner. We attempted to measure this believability within our data through asking participants to self-report whether they thought they played the game with a computer, a person or whether they were unsure. 12 out of 23 participants in the human condition in Experiment 1 and 15 out of 49 participants in Experiment 2 believed they were interacting with another person. The 11 remaining participants in the human condition in Experiment 1 believed to be playing with a computer, whereas in Experiment 2, 27 participants believed to be playing with a computer, 6 were unsure and 1 failed to answer the question. Running our analysis on both the full data and on data that excluded those who did not believe they were interacting with a person led to similar results across both experiments. Indeed our results for the human condition using the full dataset are also similar to those seen in previous human-human dialogue work (Yoon et al., 2012). That said, future work should look at ways to develop this experimental paradigm to improve believability of the human condition.

The present study also did not directly explore how experience may influence audience design and egocentric language processes, and how experience may impact people's preconceptions about a computer partner's communicative abilities. As mentioned in previous work, these perceptions (i.e. user's *partner models*) may be an important component to guiding audience design. Indeed recent definitions of this concept suggest that experiences people have with speech interfaces (Doyle et al.,

2021) may be important to consider. Within our study, past experience data was collected through asking participants how often they used speech agents, by asking them to select whether they use them daily, a few times per week, a few times per month, rarely or never. An exploratory analysis, whereby we grouped participants as experienced (use them daily, a times a week or a few times a month) or inexperienced (used them rarely or never) and compared scalar modifier use within the perspective conditions when interacting with the computer partner showed no statistically significant effect of experience. Such results suggest that past experience with speech agents does not have an effect on the presence of audience design or egocentrism when interacting with a computer partner. However, these findings are exploratory and should be interpreted with caution. Future work should focus on better ways to more directly measure people's preconceptions of computer partner abilities, their relation and dynamics due to experience, along with more systematic control and observation of experience when studying how these have an effect on language production in HCD.

Our work has developed an online referential communication paradigm to observe egocentric processes in dialogue, similar to that used in previous work in HHD (e.g Yoon et al, 2012; Wu and Keysar, 2007). Our paradigm uses a grey background to inform participants when an item is part of theirs but not the addressee's view (i.e. was in privileged ground). This approach is similar to other variations of director-matcher tasks online (Rubio-Ferńandez, 2017). However, it may be that methods used to obscure items in more physical versions of the paradigm are more obvious to participants as physical objects like curtains are used to clearly occlude the competitor from the addressee's view. Although we find similar perspective findings as those seen in HHD work that use more physical forms of the paradigm, the different techniques used to occlude the privileged ground items may indeed vary in how salient the differences in perspective are to participants. As the salience of information is important for the encoding of reference expressions, future studies need to be conducted to compare the effects seen across both physical and online versions of the paradigm proposed.

Experiments could focus on testing mechanisms that make differences in perspective much more salient during tasks along with testing alternative approaches to indicate privileged information in online scenarios such as using images that more clearly communicate occlusion (e.g. curtains).

In addition to these limitations, it is important to acknowledge that the director-matcher task is goal oriented by nature, which could impact the level upon which grounding is used. Previous research has demonstrated that goal-oriented tasks motivates participants to employ more linguistic resources (e.g. grounding and conversational repair) to accomplish the goal (Dideriksen et al., 2020). Given that the game is goal oriented, we may expect that such effects in less goal oriented tasks may differ. Further work should look to investigate the role of task on the effects seen in the work.

This is the first time that this specific experimental paradigm has been employed in the understanding of language production in HCD. Although their use is common in psycholinguistics, previous studies have criticised the employment of the director-matcher task in HHD. For instance, Rubio-Fernández (2017) states that the director task (DT) poses artificial demands on participants' pragmatic abilities, since everyday communication does not impose these types of restrictions. Although the controlled nature does make the dialogue setting less naturalistic and more relevant to task oriented dialogue, these paradigms give the opportunity to control for perspective differences more accurately than more naturalistic approaches. Indeed their task oriented nature is more akin to the types of well defined tasks conducted currently in HCD through speech based devices. We propose that the contribution of both controlled and more flexible setups are essential for understanding this phenomenon.

Finally, our experiment did not provide explicit feedback related to the performance of the partner on their matching turns. Previous work has found that when explicit feedback is available, speakers tend to rely on the cues provided by their partner to assess the optimality of their utterances or employ clarification requests when needed

(Koulouri & Lauria, 2009; Wu et al., 2013). However, when speakers know they cannot rely on feedback, they take more time to plan their utterances. Manipulating the presence or absence of feedback in HCD is therefore a necessary dimension to consider in further work so as to understand how users plan their utterances and allocate perspective taking resources in the presence of feedback during dialogue.

5. Conclusion

Our work demonstrates the existence of both egocentric and allocentric language production in HCD, highlighting that the role that the computer partner has in the dialogue (i.e. whether it is seen as a tool or as a dialogue partner) may influence how allocentric and egocentric people are during interactions. We propose four alternating, but not mutually exclusive, accounts of how this effect occurs focusing on 1) the role of the partner influencing salience of perspective, 2) the division of labour in perspective-taking 3) CASA based explanations as to the perception of the social role of the computer influencing the need for perspective taking and 4) how potential omniscience or overestimation of the computer's knowledge state may have influenced language production within the experiments. Although further work needs to disentangle these accounts and to replicate the findings identified, our work supports recent identification of egocentric production in HCD with experimental evidence. Rather than assuming only audience design based mechanisms for HCD, we suggest that work now needs to explore *when* and *how* people decide to allocate their perspective-taking choices when interacting with computers. The outcomes of such studies would not only drive more theory-driven research in language production and comprehension in HCD, it may also lead us to understand more fully the perspective taking mechanisms within our interactions with speech interfaces.

6. Credit author statement

Paola R. Peña: Methodology, Investigation, Data Curation, Formal Analysis, Writing - Original Draft, Review & Editing - Visualisation, Project Administration.

Philip Doyle: Conceptualisation, Methodology, Writing - Review & Editing.

Justin Edwards: Conceptualisation, Methodology, Experiment coding, Writing - Review & Editing

Diego Garaialde: Writing - Review & Editing

Daniel Rough: Writing - Review & Editing

Anna Bleakly: Conceptualisation, Data Curation

Leigh Clark: Conceptualisation, Writing- Review & Editing

Anita Tobar: Analysis and results

Holly Branigan: Methodology, Writing - Review & Editing

Iona Gessinger: Writing - Review & Editing

Benjamin R Cowan: Conceptualisation, Methodology, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Review & Editing - Visualisation, Funding Acquisition, and Project Administration, and Supervision.

7. Acknowledgements

This project was funded by Science Foundation Ireland ADAPT Centre 405 (13/RC/2106 P2). These organisations had no role in the design, analysis, interpretation, report writing, or choice of publication venue.

8. Data Availability

Code and materials for reproducing the experiment are open and available at https://osf.io/b7htf/?view_only=aba56050356640fc974728d1fc253f99

References

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies*, *38*(4), 547–566. https://doi.org/10.1006/imms.1993.1026

An, S., Moore, R., Liu, E. Y., & Ren, G.-J. (2021). Recipient Design for Conversational Agents: Tailoring Agent's Utterance to User's Knowledge. *CUI 2021 - 3rd Conference on Conversational User Interfaces*, 1–5. https://doi.org/10.1145/3469595.3469625

Apperly, I. (2018). Mindreading and psycholinguistic approaches to perspective taking: Establishing common ground. *Topics in cognitive science*, *10*(1), 133-139.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barr, D. J., & Keysar, B. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, *46*(2), 391–418. https://doi.org/10.1006/jmla.2001.2815

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bell, A. (1984). Language style as audience design*. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Bell, L., & Gustafson, J. (1999). Interaction with an animated agent in a spoken dialogue system. *In Sixth European Conference on Speech Communication and Technology.*

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41–57. https://doi.org/10.1016/j.cognition.2011.05.011

Braunger, P., & Maier, W. (2017, August). Natural language input for in-car spoken dialog systems: How natural is natural?. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue* (pp. 137-146).

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding* (Doctoral dissertation, Stanford University).

Brennan, S. E. (1998). The grounding problem in conversations with and through computers. *Social and cognitive approaches to interpersonal communication*, 201-225.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Chapter 8 - Two Minds, One Dialog: Coordinating Speaking and Understanding. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 53, pp. 301–344). Academic Press. https://doi.org/10.1016/S0079-7421(10)53008-1

Brennan, S. E., & Metzing, C. A. (2004). Two steps forward, one step back: Partner-specific effects in a psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 192–193. https://doi.org/10.1017/S0140525X04240055

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes,* 23(2), 119-147.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62–89. https://doi.org/10.1080/01690965.2010.543363

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2*(1), 11–33. https://doi.org/10.5087/dad.2011.102

Chung, Y., S. Rabe-Hesketh, V. Dorie, A. Gelman, and J. Liu (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. Psychometrika 78(4), 685– 709.

Clark, H. H. (1992). *Arenas of Language Use*. University of Chicago Press.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H. H. (2020). Common Ground. In *The International Encyclopedia of Linguistic Anthropology* (pp. 1–5). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118786093.iela0064

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, England: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, *26*(2), 209–225. https://doi.org/10.1016/0749-596X(87)90124-0

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300705

Cowan, B. R., & Branigan, H. P. (2015). Does voice anthropomorphism affect lexical alignment in speech-based human-computer dialogue? *Interspeech 2015*, 155–159. https://doi.org/10.21437/Interspeech.2015-75

Cowan, B. R., Branigan, H. P., Begum, H., McKenna, L., & Szekely, E. (2017, July). They Know as Much as We Do: Knowledge Estimation and Partner Modelling of Artificial Partners. In *CogSci*.

Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human−computer dialogue. *International Journal of Human-Computer Studies*, *83*, 27–42. https://doi.org/10.1016/j.ijhcs.2015.05.008

Cowan, B. R., Clark, L., Candello, H., & Tsai, J. (2023). Introduction to this special issue: guiding the conversation: new theory and design perspectives for conversational user interfaces. *Human–Computer Interaction*, 1-8.

Cowan, B. R., Doyle, P., Edwards, J., Garaialde, D., Hayes-Brady, A., Branigan, H. P., Cabral, J., & Clark, L. (2019). What's in an accent? The impact of accented synthetic speech on lexical choice in human-machine dialogue. *Proceedings of the 1st International Conference on Conversational User Interfaces*, 1–8. https://doi.org/10.1145/3342775.3342786

Dafoe, A., Bachrach, Y., Hadfield, G., Horvitz, E., Larson, K., & Graepel, T. (2021). Cooperative AI: machines must learn to find common ground. *Nature*, *593*(7857), 33-36.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In: D. J. Napoli, & J. A. Kegl (Eds), *Bridges between psychology and linguistics: A Swarthmore festscrift for Lila Gleitman*. Hillsdale, NJ: Erlbaum.

Desai, S., & Twidale, M. (2022). Is Alexa like a computer? A search engine? A friend? A silly child? Yes. *Proceedings of CUI 2022.*

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2020). *Quantifying the interplay of conversational devices in building mutual understanding*. PsyArXiv. https://doi.org/10.31234/osf.io/a5r74

Dombi, J., Sydorenko, T., & Timpe-Laughlin, V. (2022). Common ground, cooperation, and recipient design in human-computer interactions. *Journal of Pragmatics*, *193*, 4–20. https://doi.org/10.1016/j.pragma.2022.03.001

Doyle, P. R., Clark, L., & Cowan, B. R. (2021, May). What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach. *In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-14).*

Doyle, P. R., Edwards, J., Dumbleton, O., Clark, L., & Cowan, B. R. (2019). Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–12. https://doi.org/10.1145/3338286.3340116

Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, *121*(1), 22–40. https://doi.org/10.1016/j.cognition.2011.06.009

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554–573. https://doi.org/10.1016/j.jml.2005.12.009

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective Taking as Egocentric Anchoring and Adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138-161.

Ferguson, H. J., Apperly, I., Ahmad, J., Bindemann, M., & Cane, J. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition*, *139*, 50–70. https://doi.org/10.1016/j.cognition.2015.02.010

Ferreira, V. S. (2019). A Mechanistic Framework for Explaining Audience Design in Language Production. *Annual Review of Psychology*, *70*(1), 29–51. https://doi.org/10.1146/annurev-psych-122216-011653

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*(3), 378–391. https://doi.org/10.1037/0022-3514.62.3.378

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51. https://doi.org/10.1016/j.jml.2009.09.002

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304-316.

Grice, H. P. (1975). Logic and Conversation. *Speech Acts*, 41–58. https://doi.org/10.1163/9789004368811_003

Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The Division of Labor in Communication: Speakers Help Listeners Account for Asymmetries in Visual Perspective. *Cognitive Science*, *45*(3), e12926. https://doi.org/10.1111/cogs.12926

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120. https://doi.org/10.1016/j.cognition.2015.12.008

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589–606. https://doi.org/10.1016/S0749-596X(02)00019-0

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Kennedy, A., Wilkes, A., Elder, L., & Murray, W. S. (1988). Dialogue with machines. *Cognition*, *30*(1), 37–72. https://doi.org/10.1016/0010-0277(88)90003-0

Keysar, B., & Barr, D. J. (2002). 8. Self-Anchoring in Conversation: Why Language. Heuristics and biases: The psychology of intuitive judgement, 150.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The Egocentric Basis of Language Use: Insights From a Processing Approach. *Current Directions in Psychological Science*, *7*(2), 46–49. https://doi.org/10.1111/1467-8721.ep13175613

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science,* 11(1), 32-38.

Keysar, B., Converse, B. A., Wang, J., & Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science,* 19(12), 1280-1286.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Koulouri, T., & Lauria, S. (2009). Exploring Miscommunication and Collaborative Behaviour in Human-Robot Interaction. *Proceedings of the SIGDIAL 2009 Conference*, 111–119. https://aclanthology.org/W09-3915

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, *21*(6), 1590–1599. https://doi.org/10.3758/s13423-014-0620-7

Krämer, N. C. (2008, September). Social effects of virtual assistants. A review of empirical results with regard to communication. In *International Workshop on Intelligent Virtual Agents* (pp. 507-508). Springer, Berlin, Heidelberg.

Lane, L. W., & Liersch, M. J. (2012). Can you keep a secret? Increasing speakers' motivation to keep information confidential yields poorer outcomes. *Language and Cognitive Processes*, *27*(3), 462–473. https://doi.org/10.1080/01690965.2011.556348

Lane, L. W., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological science,* 17(4), 273-277.

Le Bigot, L., Terrier, P., Amiel, V., Poulain, G., Jamet, E., & Rouet, J.-F. (2007). Effect of modality on collaboration with a dialogue system. *International Journal of Human-Computer Studies*, *65*(12), 983–991. https://doi.org/10.1016/j.ijhcs.2007.07.002

Louwerse, M. M., Graesser, A. C., Lu, S., & Mitchell, H. H. (2005). Social cues in animated conversational agents. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(6), 693-704.

Luger, E., & Sellen, A. (2016). 'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Mainwaring, S. D., Tversky, B., Ohgishi, M., & Schiano, D. J. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial cognition and computation,* 3(1), 3-42.

Meddeb, E. J., & Frenz-Belkin, P. (2010a). What? I Didn't Say THAT!: Linguistic strategies when speaking to write. *Journal of Pragmatics*, *42*(9), 2415–2429. https://doi.org/10.1016/j.pragma.2009.12.022

Mey, J. L. (2010). Reference and the pragmeme. *Journal of Pragmatics*, *42*(11), 2882–2888. https://doi.org/10.1016/j.pragma.2010.06.009

Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling reference production as the probabilistic combination of multiple perspectives. *Cognitive science*, *42*, 974-1008.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology: applied*, *7*(3), 171.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., Moon, Y., & Carney, P. (1999). Are People Polite to Computers? Responses to Computer-Based Interviewing Systems1. *Journal of Applied Social Psychology*, *29*(5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78).

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.

Oviatt, S. (2022). Multimodal Interaction, Interfaces, and Analytics. In *Handbook of Human Computer Interaction* (pp. 1-29). Cham: Springer International Publishing.

Oviatt, S., MacEachern, M., & Levow, G.-A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication*, *24*(2), 87–110. https://doi.org/10.1016/S0167-6393(98)00005-3

Peña, P. R., Doyle, P. R., Ip, E.Y., DI Liberto, G., Higgins, D., McDonnell, R., Branigan, H., Gustafson, J., McMillan, D., Moore, R.J., & Cowan, B. R. (In press). A Special Interest Group on Developing Theories of Language Use in Interaction with Conversational User Interfaces. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*.

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3174214

Reeves, S., Fischer, J. E., Porcheron, M., & Sikveland, R. (2019). Learning how to talk: Co-producing action with and around voice agents. *Mensch und Computer 2019-Workshopband*.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination. *Psychological Science*, *18*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2007.01914.x

Rothwell, C. D., Shalin, V. L., & Romigh, G. D. (2021). Comparison of Common Ground Models for Human--Computer Dialogue: Evidence for Audience Design. *ACM Transactions on Computer-Human Interaction*, *28*(2), 9:1-9:35. https://doi.org/10.1145/3410876

Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention?. *Psychonomic bulletin & review,* 24(4), 1121-1128.

Rubio-Fernández, P., & Jara-Ettinger, J. (2018). Joint inferences of speakers' beliefs and referents based on how they speak. In *CogSci*.

Schmader, C., & Horton, W. S. (2019). Conceptual effects of audience design in human–computer and human–human dialogue. *Discourse Processes,* 56(2), 170-190.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, *88*(3), B11–B21. https://doi.org/10.1016/S0010-0277(03)00043-X

Shen, H., & Wang, M. (2023). Effects of social skills on lexical alignment in human-human interaction and human-computer interaction. *Computers in Human Behavior*, *143*, 107718.

Shintel, H., & Keysar, B. (2009). Less Is More: A Minimalist Account of Joint Action in Communication. *Topics in Cognitive Science*, *1*(2), 260–273. https://doi.org/10.1111/j.1756-8765.2009.01018.x

Shteynberg, G., & Galinsky, A. D. (2011). Implicit coordination: Sharing goals with similar others intensifies goal pursuit. *Journal of Experimental Social Psychology*, *47*(6), 1291–1294. https://doi.org/10.1016/j.jesp.2011.04.012

Simpson, J. & Crone, C. (2022). Should Alexa be a Police Officer, a Doctor, or a Priest? *Proceedings of CUI 2022.*

Spivey, M. (2008). *The Continuity of Mind*. Oxford University Press.

Swadesh, M. (2017). *The origin and diversification of language*. Routledge.

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review, 19(4), 699-707.*

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological science*, *18*(7), 600-606.

Wu, S., Barr, D., Gann, T., & Keysar, B. (2013). How culture influences perspective taking: Differences in correction, not integration. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/article/10.3389/fnhum.2013.00822

Zhang, T., Hasegawa-Johnson, M., & Levinson, S. E. (2006). Cognitive state classification in a spoken tutorial dialogue system. *Speech Communication*, *48*(6), 616–632. https://doi.org/10.1016/j.specom.2005.09.006

Zhao, X., & Malle, B. F. (2022). Spontaneous perspective taking toward robots: The unique impact of humanlike appearance. *Cognition*, *224*, 105076.