They Know as Much as We Do: Knowledge Estimation and Partner Modelling of Artificial Partners

Benjamin R. Cowan (benjamin.cowan@ucd.ie)
School of Information & Communication Studies, University College Dublin
Belfield, Dublin 4

Holly Branigan (holly.branigan@ed.ac.uk)
Department of Psychology, University of Edinburgh
7 George Square, Edinburgh, EH8 9JZ

Habiba Begum (HXB368@student.bham.ac.uk)
HCI Centre, University of Birmingham
Edgbaston Campus, B15 2TT

Lucy McKenna (lucy.mc-kenna@ucdconnect.ie)
School of Information & Communication Studies, University College Dublin
Belfield, Dublin 4

Eva Szekely (eva.szekely@gmail.com)
School of Computer Science, University College Dublin
Belfield, Dublin 4

Abstract

Conversation partners' assumptions about each other's knowledge (their *partner models*) on a subject are important in spoken interaction. However, little is known about what influences our partner models in spoken interactions with artificial partners. In our experiment we asked people to name 15 British landmarks, and estimate their identifiability to a person as well as an automated conversational agent of either British or American origin. Our results show that people's assumptions about what an artificial partner knows are related to their estimates of what other people are likely to know - but they generally estimate artificial partners to have more knowledge in the task than human partners. These findings shed light on the way in which people build partner models of artificial partners. Importantly, they suggest that people use assumptions about what other humans know as a heuristic when assessing an artificial partner's knowledge.

Keywords: knowledge estimation, human-computer interaction, partner modelling, theory of mind, human-computer dialogue

Introduction

Psycholinguistic research on human-human dialogue (HHD) has shown that our language choices are affected by the assumptions we make about our partners as communicative and social beings (i.e. our *partner models*) (Branigan, Pickering, Pearson, McLean, & Brown, 2011): People tend to estimate their conversational partner's knowledge and communicative abilities, and formulate their utterances accordingly. This complex set of judgements is simplified by using a range of heuristics such as accent and social cues (Clark, 1996; Nickerson, 1999) as well as our beliefs about the *social distribution of knowledge*, i.e., assumptions about what information is likely to be known to whom (e.g., students, residents of Dublin, opticians, birdwatchers) (Fussell & Krauss, 1992a).

Such perspective taking is critical to successful communication and is not solely the preserve of HHD. People consistently perceive the flexibility and ability of automated artificial (computer) partners as far lower than those of a human dialogue partner, leading us to categorise them as 'at risk' listeners in dialogue (Oviatt, MacEachern, & Levow, 1998). Moreover, our initial expectations about artificial partner's abilities affect our language choices in Human-Computer Dialogue (HCD) (Branigan et al., 2011; Edlund, Gustafson, Heldner, & Hjalmarsson, 2008). Yet we know little of how people come to have these expectations: What factors impact people's preconceptions and expectations about what an artificial partner is likely to know, before they have even begun to interact with it? In other words, what determines people's *initial partner models* for artificial partners?

Understanding what governs and impacts our partner models when interacting with artificial partners, especially in speech-based interactions, has important theoretical and applied implications (e.g., in developing robust and effective speech-based interfaces). In this paper, we investigate whether our assumptions about what an artificial speech-based interaction partner knows are related to the sense we have of the social distribution of knowledge. We also look at how our beliefs about partner knowledge are influenced by (1) partner type (humans vs. artificial) as well as (2) the partner's signalled nationality.

Perspective-Taking in Dialogue

Imagine that a stranger asks for directions to a local landmark. How do we ensure that the information we include and the language we use to communicate the message is appropriate for them? Research suggests that we use verbal and non-verbal cues to assess our conversational partner's characteristics, e.g. where they are from, their language proficiency, their age, profession etc., and use these cues to construct a partner model to guide our language choices (Nickerson, 1999).

This initial *global partner model* (Brennan, Galati, & Kuhlen, 2010), which is consulted at the stage of initial interaction, is formed through relatively superficial cues (e.g., stereotypes and pre-conceived expectations and assumptions that are in place prior to the dialogue) and assumptions about the social distribution of knowledge (Fussell & Krauss, 1992a). These initial inferences act to give a speaker an initial model of common ground between interlocutors, i.e., a representation of mutual knowledge, assumptions and beliefs shared between the interlocutors in a conversation, crucial to successful and effective communication (Bromme, Rambow, & Nückles, 2001; Clark, 1996). Although our partner models may be subsequently updated by local experiences within the dialogue interaction (e.g. feedback about comprehension, via verbal and non-verbal cues) (Brennan et al., 2010), the global model acts as a guide for our initial interaction, especially before feedback has been gathered from within the dialogue (Fussell & Krauss, 1992a).

Understanding how we develop and form these models is important, as research shows that they guide our language choices. We tend to adjust our language based on our assumptions about our addressees' knowledge. For instance, when people are asked to describe items for their friends, they adapt their descriptions to their friend's knowledge – and these adjustments lead to better communication, i.e., higher accuracy in identification (Fussell & Krauss, 1989). Crucially, studies also show that we are very accurate at assessing others' knowledge and that these assessments guide how we construct our initial message in communication (Fussell & Krauss, 1991, 1992a).

Similar effects of partner models on language choice are thought to drive our dialogue interactions with artificial dialogue partners. People tend to see artificial partners as poorer interlocutors and alter their language choices and speech behaviours as a result (Branigan et al., 2011; Oviatt, Bernard, & Levow, 1998). For example, people are more likely to converge (or *align*) with their partner's choice of referring expression when they believe their partner to be a computer rather than a human. In addition, they adjust their behaviour more in this way when they are led to believe that the artificial partner is a 'basic' interlocutor with restricted capability than a partner with more advanced capability (Branigan et al., 2011). Similarly, people's linguistic choices in a telephone conversation concerning air-fares and timetables change depending on whether they believe their partner to be a human or a computer (Amalberti, Carbonell, & Falzon, 1993). Similar findings have been reported in other work (Bell & Gustafson, 1999; Kennedy, Wilkes, Elder, & Murray, 1988). Compared to HHD, users tend to use simpler grammatical structures, use more words in their descriptions, use fewer pronominal anaphors (e.g. *her/him; he/she*), and use simpler lexical choices (Amalberti et al., 1993; Kennedy et al., 1988). Although such research assumes that people's perceptions and beliefs about their partner's abilities affect their language choices in these contexts, it is not clear what factors determine these beliefs in the first place, and thus what may be driving people's global partner model during their initial interaction with an artificial partner. Our work aims to shed light on this question.

Research on robotic agents has shown that the perceived nationality of the agent, and the content that it is being asked to process, both influence participants' judgements about its abilities (Lee, Lau, Kiesler, & Chiu, 2005). Participants used these cues in a similar way to the way that they are used in HHD: When they were asked to judge the likelihood that a robot 'from New York' or 'from Hong Kong' would know and recognize a set of New York and Hong Kong landmarks, they judged that the robot would be more likely to identify landmarks associated with its perceived nationality (Lee et al., 2005). In this context, accent can play an important role. It acts as a strong signal of identity and a speaker's linguistic background (Ikeno & Hansen, 2007), and allows listeners to identify characteristics such as age, gender and geographic affiliation, as well as stimulating specific stereotypes (Ryan, Giles, & Sebastian, 1982).

Research Aims and Hypotheses

There is currently little understanding of what factors affect people's assumptions about partner knowledge and abilities in HCD contexts. The limited existing research on people's perceptions of artificial dialogue partners tends to focus on affective factors such as interface likeability rather than on assumptions about a computer's knowledge and abilities. Other work in tangential fields such as HRI cannot be assumed to hold more widely as the embodiment of robots tend to facilitate the mapping of human abilities to a robot partner (Kiesler, 2005).

We present a study using a similar method to previous work investigating how people estimate human partners' knowledge (Fussell & Krauss, 1992a), in order to investigate how people estimate artificial partners' knowledge. People are asked to name landmarks and judge the identifiability of those landmarks' names to others. We hypothesise that people will use the same heuristics to estimate partner knowledge for artificial partners as they use for human partners. That is, people will rate both human and

artificial partners as more likely to know the name of those landmarks that are generally more accurately identified by other people (H1). This would be evidence that people have a sense of the spread of knowledge about a topic in the population (i.e., the social distribution of knowledge) with this being related to their assessment of a partners' likely knowledge, including artificial partners. We also expect a strong positive correlation between judgements of humans' and artificial partners' knowledge (H2), giving support to the idea that our judgements of artificial agents are related to our judgements of humans in this context. Based on the intimated difference in partner models between humans and artificial partners in the literature we also hypothesise that there will be a statistically significant difference between people's judgements of how likely a person versus an artificial agent is to know the name of the stimuli (H3). We also hypothesise that people will make different judgements about partner knowledge based on the relation between the system's signalled nationality (UK or US) and the type of content being judged in the experiment (i.e., UK landmarks) (H4).

METHOD

Participants

32 (16 F, 16 M) Native British English speakers with a mean age of 32.0 years (S.D.=12.1) from a UK university took part in the study. The majority (N=26) of participants had previously spoken to an automated system. Those who had used such systems were asked to rate how frequently they used them on a 7 point Likert scale (Very Infrequently-Very Frequently). The mean rating suggests that their level of experience with these types of interfaces was low (M= 2.73, SD= 1.43).

Items

Fifteen UK landmarks were used as the stimuli in the study, selected based on the frequency of accurate naming in a pre-study. This was to ensure that there was variation in the frequency of accurate naming across the items in the experiment.

Conditions

Partner Type All participants were asked to judge both an artificial partner's (i.e. automated agent) and a human partner's (within participants) likely knowledge of the landmark names. The order in which participants were asked to judge the artificial and the human partner was randomised. Participants judged all 15 landmarks in each condition. The display of the 15 landmarks in each partner condition was randomised to reduce potential order effects.

Nationality Participants were asked to judge how likely either an American or British partner (*between participants*) would be to know the landmarks. When in the human partner condition, participants were asked to rate how identifiable the landmarks' names would be to either a British or American person (participants were told that 'identifiable' referred to the likelihood of knowing the landmark name). When in the *artificial partner* condition, participants were told that the researchers were developing a British-based (British nationality condition) or a US-based (US nationality condition) automated agent. They then listened to a sample audio clip taken from the system. Participants listened to a sample audio introduction from the agent (e.g. "*Hello, my name is Laura. How can I help you*?"), simulating the type of content that would guide people's initial partner models in these types of interactions. To further emphasise the nationality, the introductory message from the service was played in either a British or a US accent.

Measures

Participant's ability to name landmarks To identify the spread of knowledge within the sample, all participants were initially asked to name the 15 landmarks used in the study. A 300x250 pixel image of each landmark was displayed along with a textbox. Participants were asked to name the item. They were informed that if they did not know the name of the item they could leave this box blank. The lead author then marked the names given by the participants as either accurate or inaccurate.

Others' knowledge of the landmark names Based on scales used in previous research on perception of others' knowledge in HHD (Fussell & Krauss, 1992b) and human-robot interaction (HRI) (Lee et al., 2005), participants were asked to judge how identifiable they felt the name of each landmark would be to others. This was measured using a 7-point Likert scale from Not Identifiable (1) to Very Identifiable (7).

Procedure

Participants were recruited via email from a British university community. Upon responding to the email participants were sent a link to the online survey. Participants completed the demographic section of the survey. They were then asked to name the 15 landmarks, and subsequently asked to judge how identifiable the name of the landmarks would be to a human (either a British or US person), and then how identifiable the name of the landmark would be to a computer (either British or US accented automated agent). Again, the order of these was randomised. They were then debriefed as to the purpose of the experiment.

RESULTS

Social Distribution of Knowledge

Following previous work on knowledge estimation in HHD (Bromme et al., 2001; Fussell & Krauss, 1992b) we ran analysis on the item level data to test H1 and 2. Using the

item level data means we can see whether landmarks that were more accurately named across the sample were rated as more likely to be known to both human and artificial partners. This would give us a sense of how people's assumptions of knowledge for each item relate to actual levels of knowledge in the group of participants for each item. This type of fine grained insight would not be possible using the participant level data as we would only have a measure of accuracy for each participant, giving us no sense of the spread of knowledge of each item in the sample as a whole.

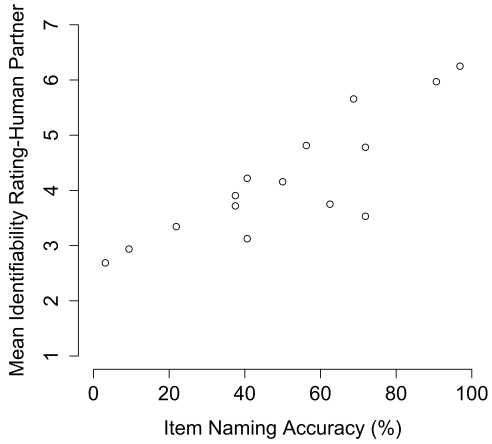

Figure 1: Relationship between percentage accurate item naming and human partner identifiability rating

There was a strong positive correlation between the percentage of accurate responses for an item and participants' mean judgements of other people's [r (13)= .85, p<.001] (Figure 1) as well as an artificial partner's knowledge of its name [r (13)= .86, p<.001] (Figure 2). There was also a strong positive correlation between judgments of other people's knowledge of the names and an artificial partner's knowledge [r (13)= .78, p<.001] (Figure 3).

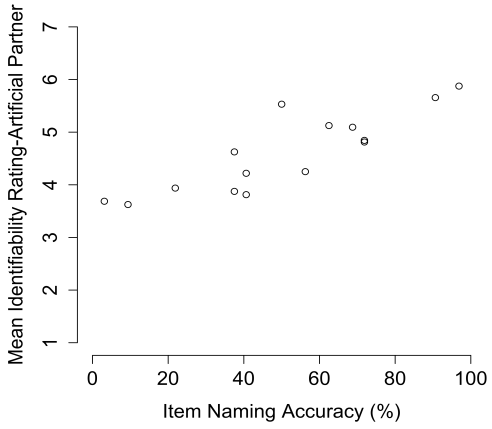

Figure 2: Relationship between percentage accurate naming and artificial partner identifiability rating

These correlations support our hypotheses (H1 and 2). They suggest that people have relatively accurate awareness of the actual distribution of knowledge (with respect to which knowledge is more or less likely to be known) and that this has a strong relationship to their judgements of how likely the name is to be known to a person and an artificial partner.

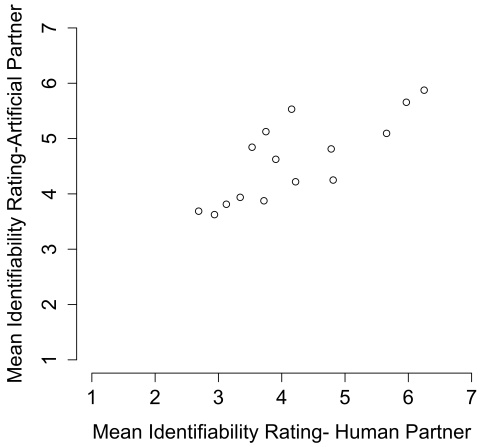

Figure 3: Relationship between human and artificial partner identifiability ratings

Moreover, people's assessment of how identifiable a landmark's name is to an artificial partner seems related to how identifiable they believe it is to a human partner. This supports the idea that people's initial model of an artificial partner's knowledge is related to their initial model of other people's knowledge, with both closely reflecting people's actual rates of accuracy in naming each item.

The Effect of Partner Type & Nationality

To test H3 and H4, we analysed the data at the participant level using a 2x2 Mixed ANOVA looking at the effects of partner type (Human vs. Artificial -within participants) and nationality (US vs. British- between participants) on people's knowledge estimation. We saw a statistically significant main effect of partner type on people's knowledge estimations [F (1, 30)= 6.43, p=.016, η^2_G= 0.058]. People rated item names in general to be more identifiable to an artificial partner (M=4.60, S.D.=1.06) than to a human partner (M=4.19, S.D.=0.74), supporting our hypothesis but contradicting the direction intimated by previous HCD work. There was no statistically significant main effect of nationality [F (1, 30)= 0.31, p=.58, η^2_G=0.007] or interaction effect between partner type and nationality [F(1, 30)=2.94, p=.097, η^2_G=0.028]. Therefore a partner's nationality did not affect people's knowledge judgements of human or artificial partners in relation to the landmarks; H4 was therefore not supported.

DISCUSSION

We found that people have a strong sense of the social distribution of knowledge and this relates to people's

judgements about others' knowledge, irrespective of the *other* being an artificial agent or a human. The number of times each item was named correctly correlated strongly and positively with people's estimations of both artificial and human partners' knowledge of landmark names. We also found that people in general judged the names of the landmarks in the experiment to be more identifiable to a computer than a person. Surprisingly, partner nationality did not have statistically significant effects on knowledge estimation.

Our research highlights that people are relatively accurate at estimating what other people are likely to know based on a sense of the general distribution of that knowledge, similar to previous research (Fussell & Krauss, 1992b; Lau, Chiu, & Hong, 2001). But importantly, these effects also apply to our estimates of artificial partners' knowledge. The actual percentages of correct responses for each item correlated highly and positively with the knowledge estimates for both artificial and human partners. We therefore seem to use our estimates of what other people will know to inform our judgements of what an artificial partner will likely know. That is, people seem to use their perceptions of the social distribution of knowledge among humans to anchor their perceptions of an artificial partner's knowledge.

We also see that people judged an artificial partner as being more likely to know the name of the landmark in the study than a human partner. It is important to note that our finding may reflect users' assumptions about one specific dimension of an artificial partner's abilities in dialogue (i.e., their knowledge of proper names) rather than their communicative capabilities or knowledge as a whole. Participants were asked to judge how identifiable the name of a landmark (e.g., *Stonehenge*) would be. Proper names pick out unique entities in the world. As such, they do not require any complex inferencing, knowledge of ontologies, conceptual relations between categories. They can (usually) be captured by a simple association between the name and a unique object, the kind of data that are prototypically perceived as easy for computer systems to store, index, and retrieve. This may explain why a computer was judged more likely than a human to know the name of the landmarks that we used. Other types of knowledge that involve more complex conceptual relationships, or operations over elements might not show the same pattern. Note however that people did not attribute complete omniscience to the artificial partner; their judgements about its knowledge were strongly related to the social distribution of knowledge.

There is also likely to be a distinction between what we perceive artificial partners to know and what we believe they can do with this knowledge in dialogue, or even whether these names will be recognised effectively in the first place. For instance people may assume that artificial partners know the proper names of landmarks but may not be sufficiently confident that these names will be recognised during speech recognition. Although vast improvements on error rates have been made in speech technology research, there may still be a perception within people's partner models that recognition is poor and inflexible. Hence rather than artificial partners being seen as 'at risk' dialogue actors, people's partner models are likely more nuanced and multi-dimensional, presumably encompass assumptions about both underlying knowledge and processing abilities.

This study focused on how people establish estimates of knowledge in their initial *global* partner models, in the absence of interaction with the system. Within a dialogue, perspective taking is likely to be informed by both the global models we create of our partner (e.g. assumptions of their knowledge and abilities formed by stereotypes and expectations before interaction) and local experiences within the dialogue (e.g. feedback of comprehension via verbal and non verbal cues) (Brennan et al., 2010). Indeed these factors are likely to interact in dialogue interactions. Work on HHD interaction has shown that behaviours within a dialogue that do not match our expected partner models impact our speech (Kuhlen & Brennan, 2010). Research suggests that these models should be considered as being dynamic and adaptable over time (Fussell & Krauss, 1991; Nickerson, 1999). Investigating the dynamism of partner models across the course of an interaction is a critical issue for future research in HCD as it has been in HHD.

In addition, although partner models are assumed to be important in influencing people's language choices and linguistic processing in HCD (Edlund et al., 2008), more research is needed to fully explore the role that they play. This question has received considerable attention in research on HDD, with particular reference to the extent to which our partner models impact processing: Is their influence immediate and pervasive, or delayed and restricted? (see Brennan et al., (2010) for summary of the main theoretical positions). Within HCD research, partner models have been invoked to explain the differences in language use between HHD and HCD (Branigan et al., 2011; Edlund et al., 2008), but recent research has shown that this may not be true in all contexts (Cowan & Branigan, 2015; Cowan, Branigan, Bugis, Obregon, & Beale, 2015). Clearly, partner models affect language choice and processing in both HHD and HCD – but it is not yet clear whether they do so in the same ways and to the same extent. An interesting possibility for future research is that partner models may play a more pervasive and far-reaching role in HCD than in HHD.

Implications & Conclusions

Our research set out to investigate the factors that affect people's expectations about what an artificial partner is likely to know, before they have begun to interact with it. Our findings suggest that we come to interactions with an existing presumption of what an artificial partner is likely to know that is based on assumptions of how knowledge is socially distributed. Moreover we found that under some

circumstances they may have the preconception that an artificial partner knows more than a human partner. These results suggest that models of human-human communication are applicable in important ways to communication with artificial agents. They also have important applied implications for HCD, by casting light on factors that can lead users towards or away from an appropriate mental model of a partner's abilities and intentions, with implications for successful communication (Kiesler, 2005). When designing artificial systems, developers should be aware that people bring with them assumptions about the social distribution of knowledge, which could significantly affect their language use.

References

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representation of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies*, *38*, 547–566.

Bell, L., & Gustafson, J. (1999). Interaction with an animated agent in a spoken dialogue system. In *Proceedings of the Sixth European Conference on Speech Communication and Technology* (pp. 1143–1146). Budapest, Hungary: ISCA.

Branigan, H. P., Pickering, M. J., Pearson, J. M., 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.

Brennan, S., Galati, A., & Kuhlen, A. (2010). Two Minds, One Dialog: Coordinating Speaking and Understanding. In B. Ross (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 53, pp. 301–344). Elsevier.

Bromme, R., Rambow, R., & Nückles, M. (2001). Expertise and estimating what other people know: The influence of professional experience and type of knowledge. *Journal of Experimental Psychology: Applied*, *7*(4), 317–330.

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

Cowan, B. R., & Branigan, H. P. (2015). Does voice anthropomorphism affect lexical alignment in speech-based human- computer dialogue? In *Proceedings of Interspeech 2015*. Dresden, Germany: ISCA.

Cowan, B. R., Branigan, H. P., Bugis, E., Obregon, M., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic alignment in human-computer dialogue. *International Journal of Human-Computer Studies*, *83*, 27–42.

Edlund, J., Gustafson, J., Heldner, M., & Hjalmarsson, A. (2008). Towards human-like spoken dialogue systems. *Speech Communication*, *50*(8–9), 630–645.

Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, *19*, 509–525.

Fussell, S. R., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. *European Journal of Social Psychology*, *21*, 445–454.

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

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

Ikeno, A., & Hansen, J. H. L. (2007). The Effect of Listener Accent Background on Accent Perception and Comprehension. *EURASIP J. Audio Speech Music Process.*, *2007*(3), 4:1–4:8.

Kennedy, A., Wilkes, A., Elder, L., & Murray, W. S. (1988). Dialogue with machines. *Cognition*, *30*(1), 37–72.

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

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating Distracted Addressees: How Speakers' Expectations and Addressees' Feedback Influence Storytelling. *Discourse Processes*, *47*(7), 567–587.

Lau, I. Y.-M., Chiu, C., & Hong, Y. (2001). I Know What You Know: Assumptions About Others' Knowledge and Their Effects on Message Construction. *Social Cognition*, *19*(6), 587–600.

Lee, S., Lau, I., Kiesler, S., & Chiu, C.-Y. (2005). Human Mental Models of Humanoid Robots. *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 2767–2772. https://doi.org/10.1109/ROBOT.2005.1570532

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759.

Oviatt, S., Bernard, J., & Levow, G. A. (1998). Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech*, *41 (Pt 3-4)*, 419–442.

Oviatt, S., MacEachern, M., & Levow, G.-A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication*, *24*(2), 87–110.

Ryan, E. B., Giles, H., & Sebastian, R. J. (1982). An integrative perspective for the study of attitudes towards language variation. In *Attitudes Towards Language: Social and Applied Contexts*. London: Arnold.