EXPLORING VAGUE LANGUAGE USE AND VOICE VARIATION IN HUMAN-AGENT INTERACTION

LEIGH MICHAEL HARRY CLARK, BA (Hons.)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

October 2016

Contents

Abstract

This thesis addresses the linguistic phenomenon of vague language (VL) and its effect on the creation of identity in the emerging and developing field of human-agent interaction (HAI). Current research on VL has focused on human interaction, while similar existing literature on language in HAI has focused on politeness theory and facework. This thesis brings the two research fields together and uses them as a focusing lens to investigate the issue of identity in agents – software with varying degrees of autonomy and intelligence.

Agents are increasingly common in our everyday lives, particularly in the role of an instructor. Intelligent personal assistants are a frequent feature on smartphones, automated checkout systems pervade supermarkets both large and small, and satellite navigation systems have been a mainstay for over a decade now. Despite their frequency, there is relatively little research into the communication challenges surrounding HAI. Much like other people, the language and voice of agents have the ability to affect our perceptions and of them, and shape the way in which we create their identities. Instruction giving, amongst other facets of talk, in human communication can be mitigated through the use of VL. This can reduce the imposition we have on interaction partners, pay respect to a listener's face, and establish and maintain a positive rapport with our interlocutors. This can have a profound effect on the desire to interact with someone again. Furthermore, agents that use speech to communicate are assigned one of two varieties of voice – synthesised or pre-recorded human speech, both of which have documented benefits and drawbacks. Given the rise of agents in the modern world, it is in the best interest of all parties to understand the salient variables that affect our perceptions of agents, and what effect VL and other variables such as voice in language and voice may have in our interactions with them.

This thesis provides a novel approach to investigating both VL and voice in HAI. A general framework is presented with the use of a specific VL model to apply in the interactions, which is designed around verbal agents giving people instructions on how to construct Lego models. The first study compares the effects of a vague and non-vague verbal agent in this context, while the second study focuses on the comparative use of synthesised text-to-speech voices and professional human recordings in the same context.

The results from the investigation reveal key findings regarding the use of VL in a verbal agent instructive context. The first study indicated that a synthesised agent voice is better suited to using non-vague instructions, while the second study revealed that a professional voice actor is a preferable candidate for using VL in comparison to two different synthesised voices. These findings discuss the issue of

identities in HAI. They reveal that, when an agent instructor is perceived to have a voice that is non-human and machinelike, it is more likely that its use of VL will be received less positively. This is often because the combination of voice and language do not mix, but is also a result of a clash of perceived group identities between agent and human speech. As agents are typically direct, the use of "humanlike" VL can create a large disparity between a person's expectations of agent speech and the reality of the interaction. Similarly, if an agent's voice has more of a humanlike feel to it, then its use of VL will create less disparity and has the potential to bridge the gap between these two group identities.

This poses discussions on the nature of agent identity and how it compares to those in humans. The thesis concludes with reflection on the findings in light of existing linguistic theories, and how further research into this field may assist agent designers, researchers, and agent users alike. A suggestion of employing a corpus linguistics approach to HAI is proposed, which may pave the way for future success in this area.

Acknowledgements

I owe immense gratitude to those who have helped me in the years working on this thesis. My PhD supervisors Svenja Adolphs and Tom Todden provided me with both academic and non-academic guidance, and without Khaled Bachour mentoring me the first couple of years would have been a lot harder.

My fellow PhD students across English, Horizon, and the MRL have helped immensely. I owe great thanks to Adbulmalik Ofemile for being an attentive research colleague and friend through these past years – from going through dozens upon dozens of research designs to co-presenting my first conference paper in Tsukuba. The final months were also made a lot easier with the friendship and support of Annie Quandt, and I look forward to returning the favour in kind.

I am thankful too for the people and friends I have met outside of my PhD during my time in Nottingham. The many people at Melton Hall made the first year a memorable one. So too those I have lived with other the years – Paul, Laura, Ma'ie, Hannah, Max, & Avril. Recounting those days would be a thesis in itself. Thanks also to Pete & Rob for offering their guidance when I most needed it.

I am tremendously grateful for all of the friends, in and outside of Nottingham, that have supported me. Those that I have known prior to my studies in Huddersfield and York have seen less of me than perhaps any of us would like.

Finally, my thanks go to all of my family for their belief, encouragement, and support they have provided me all these years. Without them this would not have been possible.

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of variance statistical test
CASA	Computers as social actors paradigm
CL	Cepstral Lawrence: one of the synthesised agent voices
CP	CereProc Giles: one of the synthesised agent voices
CLAN	Software used for playing video data and transcribing within the same program
FTA	Face-threatening act
HAI	Human-agent interaction
HCI	Human-computer interaction
HHI	Human-human interaction
HRI	Human-Robot Interaction
Nvivo	Qualitative data analysis software used for multimodal analysis
VA	Voice Actor: the agent created with using a voice actor
VL	Vague Language

1. Introduction

1.1 Initial Overview

Much of our daily lives increasingly involve digital interaction of a wide variety. Computers have moved far from the stationary desks of the home and office alike, and digital devices now permeate a wider space. Smartphones, tablets and other devices have saturated the world of personal mobile computing, and our interactions with them having become ever more sophisticated. A large part of this sophistication comes from the research and development into agents – software that, amongst other features, displays degrees of autonomy, social capabilities and sometimes humanlike characteristics (Wooldridge and Jennings, 1995). Our collaboration with these intelligent agents is known as human-agent interaction or HAI. This is essentially a sub-field of human-computer interaction or HCI, which encompasses digital interactions with agents, computers and other machines. Many of the theories and literature discussed throughout this thesis will often discuss HCI specifically, but there is extensive crossover between these two fields. Because of this many of the theories, hypotheses and ideas are transferrable from one to the other. Moreover, there is also some crossover in the specific types of agents being discussed and how the results presented may influence research into them. This agent is focused on speech as a modality, but others may have multimodal capabilities. Any fundamental findings then can be considered for these other agents too, but their multimodality and any other differentiating features also have to be considered.

With the greater prevalence of human-agent interaction comes the need to address challenges in how our relationships with agents will develop as their sophistication increases as they are given more autonomy, responsibility and varied roles in collaboration (Jennings et al., 2014). One of the most salient of these roles is that of the agent instructor. Already there are a wide variety of agents that instruct us every day. Map based applications in smartphones and satellite navigation systems in cars direct people across cities and the country, automated self-checkout use in supermarkets has boomed (Orel and Kara, 2014), and telephone based spoken dialogue systems have been a mainstay in society for over a decade now (Nass and Moon, 2000).

Because most interactions in HAI involve the agent instructing the human, successful communication requires that humans be open to being directed (Sukthankar et al., 2012), and able to engage with agents at a peer level (Maes, 1994). Agents are capable of dealing with some types of information in quantities and complexities that would overwhelm humans (Ball and Callaghan, 2011), and it is in these situations that they are ideally suited for a role as instructor, making

quick decisions with vast amounts of data. Moreover, agents provide a cheaper alternative than employing human beings. A lot of work has been done on the role agents can play in the management of complex and information rich situations such as emergencies (Schaafstal et al., 2001) and damage control (Bulitko and Wilkins, 1999). Agents have also been shown to be able to hold more advisory roles such as a personal tutor (Heylen et al., 2003) or by assisting patients and medical staff in diagnoses (Doswell and Harmeyer, 2007; Chan et al., 2008). Evidently, there is a vast area in agent instructors that are both currently deployed and in development in the laboratory.

Many of these agents use speech as a key mode of interaction with its users and signify a shift towards a greater use of natural language in agent interfaces – i.e. using language in a more *natural* form as it appears in conversations and interactions between humans (Cowan et al., 2015) . This presents unique challenges in understanding the effects of spoken discourse in human-agent interaction, as speech contains a wealth of interactional complexities that build and maintain the way people see each other in terms of power, identity and personality (Goffman, 1967; Goffman, 2002; Cameron, 2001; Coulthard, 2013). Identity here is seen as the "social positioning of *self* and other" (Bucholtz and Hall, 2005: p.586). With the frequency in which we interact with such agents and its impending rise, it is important to know how changes in this spoken discourse affect the human-agent dynamic for both users and developers alike.

How human-agent interactions differ from their human-human counterparts is particularly important. Our interactions with agents, computers and other media have been shown to be similar to that of other humans in that the same social rules underpinning both are instinctive in nature, and draw from the same social resources (Nass et al., 1994; Nass et al., 1995; Nass et al., 1996). This allows researchers to take inspiration from existing theories in linguistics, psychology, sociology and communication theory amongst others, and apply them to interactions with agents and other technologies. This includes adorning agents with humanlike features seen in human communication and interaction and seeing how users perceive them and interact with them in laboratory and real world contexts. Aspects of this human likeness include in the specific area of verbal agents include manipulating language (Clark et al., 2014; Strait et al., 2014; Rosé et al., 2008) and vocal capabilities such as the specific voice and prosodic properties being used to convey information to users (Dahlbäck and Jonsson, 2010; Tamagawa et al., 2011; Grichkovtsova et al., 2012). Despite having this wealth of human interaction to draw from, there is no guarantee that human-agent interaction will be the same when these are used to design verbal agent instructors.

Instruction giving in humans can be a delicate process. Being the social actors that we are, there is a desire to not infringe upon the personal

space and rights of others by asking them to do something, and not to presents ourselves in a negative light in the same breath (Goffman, 2002). We often desire to save face in these situations. To help mitigate these and attempt to build and maintain a rapport with interlocutors, there are a number of linguistic strategies used to manipulate the potential adverse effects of giving instructions, such as those outlined in politeness theory (Brown and Levinson, 1987). There have been attempts to research both the general effect verbal agent instructors have on their users in a game setting (Moran et al., 2013), as well as the effects of mitigated instructions in both pedagogical tutoring (Wang et al., 2008; Wang et al., 2010) and task-based scenarios (Torrey, 2009; Torrey et al., 2013; Strait et al., 2014). Although they are sometimes used successfully, for example in the pedagogical setting, but in task-based scenarios have received mixed results. At times, this type of linguistic strategy in agents makes them seem more considerate, kind and likeable, whereas in others it as deemed as inappropriate for the interaction.

The approaches used in this type of research have provided interesting results but there is not always a consistency in how the agent's communications have been designed and implemented. Politeness strategies for example are broad and many and can include greetings and praise, as well as face saving (Brown and Levinson, 1987). Without explicit description of the language being used it, nor a consistency, the results can be hard to put into context as to which linguistic features are causing the specific positive and negative points of discussion that arise from the data.

This thesis presents initial steps in creating a linguistic framework of implementation based on a different phenomenon – vague language (VL). VL refers to types of language that are inherently imprecise and used to achieve a variety of interactional and social goals (Channell, 1994). The categories of VL that Channell refers to VL for example are vague approximators such as *like, about, a bit of*; vague category identifiers such as *and so on, and stuff*; and placeholder words including *thingy* and *thingamy*. Different authors have described different categorisations of VL and these are discussed further in Chapter 2. This type of language is different from *vagueness* that arises from genuine uncertainty. VL can be used as a mark of social cohesion (Cutting, 2007), and can help towards creating an informal and less direct atmosphere (Channell, 1994; Cheng and O'Keeffe, 2014).

There are several reasons why VL was employed in this thesis as a potentially useful linguistic strategy to be employed by verbal agent instructors. It is a common feature of spoken interaction in particular, although it does also appear in writing (Channell, 1994; Jucker et al., 2003; Cheng and O'Keeffe, 2014). Given the agents described in this thesis are primarily of a spoken nature, and that there are a growing number of such agents that people currently interact with as outlined

previously, VL represents a good candidate for investigation in this type of spoken interaction. Although it does not always communicate effectively and may sometimes leads to miscommunications in interaction (Cutting, 2007; Jucker et al., 2003), VL is also neither necessarily good or bad. Rather, it is either appropriate or inappropriate for the context in which it is used (Channell, 1994). How appropriate VL and other language use is within interactions, and the extent facework and politeness functions, can be affected by a variety of factors, including gender, social status, and cultural background. One of the key research aims of this thesis is to ascertain from initial investigations as to whether or not VL can be appropriate or not for the use in a verbal agent instructor.

The focus in this research is to first create an explicit VL model that originates from the lexical level – that is it starts with individual words and phrases and not broad strategies as seen in some politeness research in HAI. Drawing on previous literature and attempt to categorise VL, this model presents a description of the categories deemed appropriate for this research context and what lexical items these contain. Previous literature is discussed in both Chapters 2 and 3, while the specifics of the model are outlined in Chapter 3 alone. Chapter 3 also includes the general approach to applying VL to a HAI context for the research studies in Chapters 4 and 5.

Primarily this research presents initial steps in creating linguistic analysis frameworks for understanding human-agent interaction, as well as a linguistic focused methodology for investigating the comparative effects of vague and non-vague language, as well as synthesised and human recorded voices, on participants' perceptions and attitudes towards a verbal agent instructor. There are also benefits for the designers of these. The results presented here provide another initial step into some of the preferences users display towards these two variables, which may be taken into consideration for future developments of voice agent technology. Investigating the use of VL in particular in these interactions allows for insights into whether this is a viable option to improve user experience when interacting with verbal agents, and in a broader sense assess the effects language may have on the way in which users perceive agents and project identities onto them, as well as any effects this might have on their task performance in a specific model assembly context.

This thesis also deals with the issue of voice in human agent interaction. Agents typically have either a synthesised voice or a human recording, and there are benefits and drawbacks to both. Synthesised voices can use human recordings to create a database of natural speech, from which the appropriate sub-word features can be used to output virtually any utterance (Black, 2002). A text-to-speech (TTS) system can accomplish this, where textual output is turned into synthesised speech output. Human recordings, on the other hand,

provide output from a finite list of pre-recorded utterances. The latter represents another shift towards human likeness, though synthesised voices are becoming more advanced and getting closer to the same positive perceptions as human recordings receive (Forbes-Riley et al., 2006; Georgila et al., 2012). Both are used in verbal agents and have their respective benefits and drawbacks. Although human recordings are usually of a higher quality and perceived more positively, they are more expensive and require much more time in preparing. Text-to-speech systems on the other hand are fairly cheap and can produce the same speech output in a fraction of the time. While comparisons have been made between the two there has not been an assessment as to how it affects a linguistic phenomenon such as VL that is so ingrained in human communication but not in HAI. Analysing both synthesised and human speech covers one variable present in speech technology that already has prior research, though not on the effects of VL use. Any benefits and drawbacks on the use of either type of speech in verbal agent instructors can further contribute to this research, and provide recommendations on some combinations of voice and VL use in such contexts.

Both the language used and how it is produced can influence how one perceives a speaker, and how they in turn create different identities for them as a result. Comparing synthesised and human voices with VL allows for the data to inform future agent designers who may explore further atypical agent speech.

It can also be seen whether this a viable option to improve the user experience when interacting with verbal agents. This means we can see whether or not paying a greater attention to language can influence the way in which users perceive agents, whether this matters in regards for their future interaction with them and if there is any effect on task performance. The latter is likely only useful in situations where human-agent collaboration is non-leisurely, but is a possibility if not a probability in the future. Similarly, people will be interacting with agents that can do more and interact with them for longer. This is in essence another continuum from early computers to human interaction. If something as simple as manipulating language to make it vague can improve perception of particular attributes, rapport, user experience, efficiency, clarity etc. then it would be wise to incorporate it. This all has basis in human communication and is not merely an arbitrary inclusion of some imprecise language into human-agent interaction.

1.2 Research Aims and Objectives

The core aim of this thesis is to explore the use of VL, a lexical strategy and linguistic phenomenon of human interaction, in the continuously developing area of human-agent interaction. Specifically this research looks at the use of VL by verbal agent instructors in context of them

guiding human users to complete model assembly tasks. This also includes analysing the effects of synthesised and human voices on the VL used by these agents. This is achieved first by creating a model of VL that can be implemented into a HAI context. The implementation into a Lego model assembly task is analysed using a mixed methods approach. This approach allows for the both the analysis of attitudes towards the vague agents and the analysis of descriptive accounts of participants interacting with them. Combining quantitative and qualitative approaches, this goes towards developing linguistic analysis frameworks that can account for how agents and humans interact in one space in a specific context of interaction. The specific aims of this thesis are as follows:

> *1) How do users perceive and project identities towards verbal agent instructors that use VL and what contrasts can be seen with human communication?*

User perception is a fairly general term, but this refers to the mixed methods approach and analysis of quantitative questionnaires and qualitative interviews. VL exists in abundance in human communication but not in HAI, so this leaves a large gap in the knowledge of how successful and appropriate VL can be in this interaction space. This is despite research into similar linguistic areas, though there is little information to be found in combining this with theories of identity. The contrasts with human communication focus on whether users are able to identify the use of VL as appropriate or not, in light of the different linguistic and social capabilities that may be attribute to agents.

> *2) Are there any differences in these identities when comparing vague agents to non-vague agents?*

Study One in Chapter 4 focuses on the comparisons between vague and non-vague agent interactions. It is expected that there will be marked differences between the two agent types in both the quantitative and qualitative data analyses.

> *3) Are there any differences in these identities when vague agents use synthesised and human voices?*

Given the two types of voices that are used in verbal agents, it is important to test the use of VL in both. Although human recordings in agents are often preferred to the synthesised alternatives, it is unknown how users will react when both are using VL. Study Two in Chapter 5 focuses on the comparisons between vague agents using synthesised and human voices. Different agent voices can affect people's perceptions of agent interfaces, though the relation to voice and language in these HAI contexts is not fully understood. It is

expected that there will also be marked differences in the different agents in this study.

> *4) Does the use of VL in an instruction based task in HAI affect a user's ability to conduct a task?*

As well as accounting for users' created identities and attitudes towards the agents, this thesis wishes to understand the effects on their ability to accomplish the task of model assembly. This occurs both in the comparative analysis vague and non-vague agents, as well as comparing vague agents using synthesised and human voices. This is done through analysing metrics such as the time taken to complete tasks and how many times participants request a repeat of the information provided to them. Agents giving instructions often require tasks to be completed and sometimes these occur within time constraints. Analysing the appropriate use of language and voices in these contexts becomes even more important as a result. Both studies in Chapters 4 and 5 address the issue of task performance.

1.3 Thesis Outline

This thesis consists of seven chapters in total. Chapter 2 looks at the related work in the field, starting with the discussion of agents in modern society and the social reactions people display towards them. This chapter then progresses onto the concept of identity – how it is defined, its relation to language use, and the notion of identity in agents. The last point focuses on how salient linguistic variables in agent design can affect the identities users create for agents. This chapter also includes discussions on the linguistics theories of VL, politeness and face, and how they may be incorporated into HAI. Chapter 3 builds on the discussion of VL and creates a bespoke VL model to be implemented in the two experiments in subsequent chapters. This includes details of the individual lexis and their uses in context of the agent's instructions. This chapter also includes the general approach to designing the two studies in the later chapters, including designing the verbal agent interface, the assembly task, and the instructions the agent provides to participants.

Chapter 4 describes Study One from the implementation of the VL model, the specific methodology of agent and task design, and then presentation and discussion of the results. Its focus lies on the comparison between a vague and non-vague agent instructor. Chapter 5 builds upon the first study and presents Study Two. This includes improvements to the methodology and data analysis, with a focus on the comparisons between synthesised and human voices in a vague agent instructor.

Chapter 6 addresses the findings of Study One and Two and reflects on their contribution towards the understanding of linguistic theories of

identity, politeness, face, relational work, and VL. Discussions subsequently include the nature of agent identities and how it relates to human identities depending on the linguistic variables it possesses. Finally, Chapter 7 provides a summary of the thesis and all its individual chapters, and frames the findings of the thesis around understanding identities in vague verbal agents and building approaches to understanding them better in the future. The limitations of this thesis are also discussed alongside the avenues of investigation for future work. This concludes with a brief discussion on the future use of a corpus linguistics approach in HAI.

2. Literature Review

2.1 Introduction

This thesis addresses how a verbal agent instructor using vague language is perceived by its users, and to what extent the voice used by the agent to deliver this language affects these perceptions. This chapter begins with discussion on the modern rise of agents in society. This is followed by discussions of the Computers as Social Actors (CASA) paradigm and Media Equation theory, and the similarities between our social reactions towards computers and those towards other people. This chapter then moves towards discussing the concept of identity – one of the fundamental concepts in this thesis – including how it can be defined and how it relates to language. Identity in human-agent interaction is then discussed, and what salient linguistic variables of voice, language, and prosody (i.e. the modification of speech in its speed, intensity, stress etc.), can affect how users created identities for agents. The linguistic theories of politeness and face are also discussed in the context of identity and linguistic variables. This thesis concludes with by discussing the concept of VL, its functions, and the contexts in which it is used. These then form the foundation for the discussion in Chapter 3.

It should be noted that the literature discussed in this chapter is not always wholly based on human-agent interaction (HAI) and crosses boundaries with both human-robot interaction (HRI) and human-computer interaction (HCI). Arguably, the latter came before the other two and sometimes theories and research blur lines in these areas depending on the granularity it wishes to delve into. Nevertheless, clarification as to the particular fields being discussed is included when relevant, particularly when concerning the study of robots. The distinction between agents and computers is perhaps less obvious, though agents often have some degree of autonomy when interacting with their users (see Wooldridge and Jennings 1995).

2.2 The Modern Rise of Agents

We live in a world in which our interactions with digital media and devices are part of the fabric of our everyday lives. Computers have moved away from the desktop and are now built into the world around us (Jennings et al., 2014). These appear under various forms such as networked computers, tablets, smartphones, personal devices and wearable technologies. One particular type of system that is growing in frequency is that of the agent – computer systems that have varying capabilities to act autonomously, intelligently, socially, and sometimes with humanlike characteristics (Wooldridge and Jennings, 1995). They pervade our daily lives and research suggests they can also be used for scenarios such as citizen science and disaster response (Jennings et al., 2014).

Although agents may interact with users in a variety of modalities, it is systems that use speech that perhaps face the most interesting linguistic and social challenges when it comes to interacting with their users. There has been a rapid rise in spoken interactions with agents. This includes both the user using speech to command and query, and the agent using natural language to reply to users (Cowan et al., 2015). Examples of these agents include smart televisions; satellite navigation systems; automated checkouts in supermarkets, and telephony systems. Intelligent personal assistants such as Apple's Siri, Google Now, Microsoft's Cortana and Amazon Echo are also growing in prevalence (Jiang et al., 2015; Kiseleva et al., 2016). As exchanges with these speech interfaces or *verbal agents* continue to increase, it becomes more important to understand what salient features of these agents affect their interactions with human users. Investigating these features may help us better inform agent designers, and create a better body of knowledge of frameworks and paradigms in linguistics, psychology, and sociology, as our lives move ever towards the digital.

Understanding the effect of language use and communication and their effects on rapport and human perception are some of the key aims of researchers and designers working with systems such as Embodied Conversational Agents or ECAs (Smith et al., 2011). These agents often focus on companionship and sociability. Other agents, such as the ones listed in the previous paragraph, may have more of a combination of companionship and practicality, each with varying degrees, depending on the specific agent. Something these spoken agents do often have in common is that they instruct, command, or request things from their users using natural speech, i.e. that which mimics human speech and, being speech, carries with it social information as a result (Cowan et al., 2011). This social information may not always be present in other modalities such as text. Agents using speech bring with them the nuances that do not appear in non-spoken interaction. These include voice quality - how a voice sounds in terms of age, gender, and class,for example, and prosody, both of which allow them to tap into the vast riches of spoken paralinguistic features[1] that do not necessarily appear in other forms of interaction. This also includes accent – patterns of pronunciation often linked to a particular group of people, such as in a geographical region (e.g. a Yorkshire accent). As agent technology develops, we may observe a shift towards agents whom their users can have conversations with, and vice versa, as well as form bonds with them. Discussion of one type of agents is included in Chapter 6 (6.3.3).

[1] Paralinguistic features refer to "phenomena that are modulated onto or embedded into the verbal message, be this in acoustics (vocal, non-verbal phenomena) or in linguistics (connotations of single units or of bunches of units)" cited in Schuller et al. (2013: p.5)

There is also an indication that users are more accustomed to being instructed by agents and not taking control of interactions (Moran et al., 2013). As a result, we have an increase in verbal agent interfaces that may instruct their users either as a primary or supplementary form of its language use. For the sake of clarity, these will be referred to as *verbal agent instructors*. Some can speak to their users, and with others their users can speak to them. Some of these agent instructors have both capabilities. There are questions, however, as to how agents should talk to users and to what extent natural language and humanlike characteristics should be used as part of their design, if at all.

2.3 Computers as Social Actors

Designing a verbal agent instructor without any natural language or humanlike qualities would be nigh on impossible. Despite this, there is research to suggest that humans respond to social cues such as language use and voice quality in human-agent interaction, much as they do in human interaction. This research is originally concerned with human-computer interaction but these also extend to HAI.

It is fairly established in the field of HCI that people often treat computers as social actors. The Computers as Social Actors (CASA) paradigm (Nass et al., 1994) and The Media Equation (Reeves and Nass, 1996) pioneered this theory – that in interaction people treat computers as they would do other people. In analysing the results of five experiments the authors noted several key findings regarding HCI. One of the key findings that is of particular interest in this thesis are that HCI is profoundly social and as a result many of the theories from research in psychology, sociology, and other similar fields of study are also relevant in HCI. As a result, this makes the research of such theories in this alternative interaction context a viable means of investigation. This includes theories found in linguistics.

In one experiment, the authors found that aspects of identity can be applied to computer voices as well as people. It was found that notions of both "self" and "other" applied in HCI even when the computer interface was not particularly advanced. This indicates that even small changes in computers can evoke a "wide range of social responses" – something also seen with social robots that have limited capabilities (de Graaf et al., 2015). Similarly, the experiments showed that some people apply politeness norms to computers as they would with other people. One caveat, however, is that these experiments were not focused on the linguistic notions of politeness strategies. They did, however, contribute to the overall notion that interaction with such technologies can often be a social one. This can even contradict a person's own perceptions of an interaction, in that they may insist they do not treat a computer system like they would a person, despite the data suggesting otherwise. Results such as these indicate that there is

no switch that people can turn on and off when interacting with computers instead of people, and that these social behaviours continue to manifest in these types of interactions. As such, they argue that the social rules guiding interactions with people can apply equally to HCI.

Research following the initial CASA paradigm was expanded upon to include suggestions that computers can be perceived by users to have personalities similar to humans (Nass et al., 1995). This study introduced the idea that personalities can be perceived by users even when the computer does not possess advanced features (i.e. a more basic computer can still be perceived to have personality). Because of this, a computer may only need a small and limited language output in order for personalities to be perceived in interactions with them, and even small changes in creating these perceived personalities can elicit social behaviours from their users (Lee, 2010). The notion of users treating computers like humans even though they are aware the computer does not possess human motivations or an actual self as such, was also reinforced.

One of the fundamental findings of this research was that because these social rules apply in HCI, one can take a theory from a field such as linguistics, sociology or psychology for example that discusses human-human interaction (HHI) and apply it to HCI (Nass et al., 1994). Observations can then be made on how the HCI context differs from the evidence produced in HHI. Such investigations have found that people can identify themselves as a "teammate" of computers (Nass et al., 1996), and can be flattered by them much like they would be with other people (Fogg and Nass, 1997).

Further research in theories of human interaction in HCI contexts has included the use of applying politeness strategies (Wang et al., 2008; Scheutz et al., 2011, Torrey et al., 2013; Clark et al., 2014; Strait et al., 2014; de Graaf et al., 2015; Mayer et al., 2006; Nass and Moon, 2000). Similarly, research has been conducted regarding the effects of an interface's voice on interaction and on users' perceptions of verbal interfaces (Lee et al., 2000; Nass and Lee, 2001; Dahlbäck et al., 2007; Jonsson and Dahlbäck, 2011; Tamagawa et al., 2011; Grichkovtsova et al., 2012). Further details of these studies are discussed at relevant points throughout this chapter.

There are numerous studies on human likeness in agents and other technologies. These have provided mixed results, and the different focus that each study takes can make development of wide-reaching theories problematic. Experiments looking at aspects of human likeness in voice and language of an agent, for example, may not display the same results when analysing an agent's appearance. Negative responses to human likeness, however, often refer back to the idea of the "uncanny valley", which focused on appearance (Mori et

al., 2012)[2]. The uncanny valley suggests that as a non-human character becomes more human like in appearance, we often find it more comfortable to interact with. Furthermore, as the perceived human likeness increase, so too can the feelings of comfort and rapport (Mitchell et al., 2011b). This linearity reaches a peak; however, when the rapport suddenly dips as the nuances of non-human characteristics become too apparent, causing a dip or "valley". This creates "eeriness" as a result of a character that looks imperfectly human. These imperfections and flaws are more noticeable as it moves ever closer to being humanlike. An example of this in fiction, which is also discussed in 2.5.2, can be seen in the television series *Star Trek: The Next Generation* (Roddenberry, et al., 2007). The series features a humanlike android "Data", one of the commanding officers aboard a space exploration vessel, whose humanlike capabilities are highly sophisticated. However, this also brings light to the anomalies he possesses, such as physical moment and sentence production. These are recurring themes in episodes spanning the series.

There is then some debate as to what is appropriate for a non-human entity to both look and sound like. The context of this thesis focuses on human participants receiving instructions from agents and is limited to minimal tactile input with a computer interface and no spoken participation by the participants. Creating an agent with more humanlike social cues, as described in social agency theory, may improve the way it is perceived (Mitchell et al., 2011a). These cues include those vocal features such as pitch (perception of physical sound corresponding to a physical frequency[3]), and prosodic features such as speech rate,[4] that can both affect perception in human interaction. This also extends to human-computer interaction. Human voices, for example, are often preferred to machinelike synthesised counterparts (Lee, 2010; Mayer et al., 2003). As Mitchell et al. (2011) state, however, this preference may be overstated by their users. There is no research on how these different voices can impact on a user's perception of agents when they use VL. Similarly, there is no comparative study on the effects of using vague and non-VL with a synthesised agent voice, and how this may inform the design of verbal agents. These two gaps are addressed in the two study chapters (4 and 5), as well as the specific discussion of VL and its application in these studies (Chapter 3).

---

[2] The original work was in Japanese and has not been listed here but the discussion can nevertheless be found in this English reference and other references linked to the uncanny valley that have been mentioned in this and other sections in the thesis.

[3] See Matthew (2007) The concise Oxford dictionary of linguistics.

[4] Arguably, both features can be seen as being part of spoken prosody, though this distinction or lack thereof is dependent on the individual.

With humanlike language this becomes more difficult to analyse. There has been success with computers and agents using various politeness strategies (Wang et al., 2008; Torrey et al., 2013). Humans have been shown to display politeness towards them in turn, but direct comparisons with human interaction are somewhat lacking, as are the relative extents to which these effects are observed (Mitchell et al., 2011a). The uncanny valley also is often discussed in terms of robot and machine appearances (Mori et al., 2012), but how this may or may not extend to VL has not been covered by previous research.

The CASA Paradigm and Media Equation were first discussed almost two decades prior to the research undertaken as part of this thesis – before smartphones, before tablets and before the boom in computers, agents, and other media alike. Although the Media Equation, for example, is still cited in the current day (Strait et al., 2015).

It is debatable as to the extent to which interactions with technologies are different or similar to human interactions, and how people will perceive increasingly complex media, especially those that aim to mimic humanlike interaction. The number of interactions that one may typically have with such media on an everyday basis is higher or at least more complex than it was twenty years earlier. This has not been instantaneous, however, and certain generations have even grown up with agents as a familiar type of interaction. People who have not grown up with the technology, but have gradually been exposed to it, may have grown accustomed to these interactions. The perceptions, as a result, may have grown more positive over time. This is reflected in a study with social robots (de Graaf et al., 2015).

The research described in this section shows that HCI, while containing the same social rules of interaction from the user's viewpoint, can have noticeably different outcomes when variables of voice and language of a computer or agent are altered. Linguistic factors, including voice, prosody, and language, can also affect a user's preference for future interactions with these technologies. Affecting perceptions of these technologies can in turn alter the perceived identities that users create for these agents, which will be discussed further in 2.5.

2.4 Understanding Identity

This thesis is interested in the way a verbal agent instructor using VL is perceived by its users. It investigates to what extent the voice used by the agent to deliver this language affects these perceptions and how the participants construct the identities of the agents. The concept of identity can be abstract in nature and difficult to provide a single definition for (Tajfel, 2010). The following section aims to outline some of the definitions of identity that can be transposed to HCI and

HAI, as well as some of the salient features of identity creation in these contexts that have been highlighted in previous literature.

2.4.1 Defining Identity

The historical development in how identity has been defined and redefined over the years has perhaps contributed to a loss in its meaning (Gleason, 1983). However, it is noted that in simple terms that identity can be used to mean different things depending on context. One such example may be to describe personal characteristics that an individual possesses. Another may be to describe social groups to which an individual belongs. Fearon (1999) provides a historical account of the definitions and how they have developed over time. However, he also attempts to provide some simple definitions. Firstly, that identity may relate to social groups and expected characteristics associated with them. Secondly, it may also relate to, either combined with the first definition or on its own merit, characteristics of a person that are "socially distinguishing" and that one may take pride in and is a result of social interaction. Despite their different historical accounts provided by these two authors, their conclusions draw some similarities, in that identity exists both with regards to individuals and as well as to groups. Not only this, but identity can be seen as pluralistic, as opposed to singular, so a person can have different individual identities and different identities belonging to various social groups.

The notion of *identities* rather than identity is one supported by many scholars, particularly in the field of sociolinguistics. When thinking of identity it is tempting to imagine a static, unchanging collection of personal characteristics that one is and that one displays to others – that there exists deep down within us a stable and fixed self. However, identity is perhaps better understood as a process steeped in social activity and history (Hall, 1996; Hall, 2013), something that changes in response to the contexts in which we interact (Llamas and Watt, 2010) and arguably only fully exists when in interaction (Joseph, 2010). Rather than there being an "absolute self" independent of interaction, identity is a product of interaction and discourse, rather than a precursor to it (Benwell and Stokoe, 2006). De Fina et al. (2006) provide a thorough account of the different types of identities. However, for the purposes of this research a more simplistic view will be taken and is perhaps best summarised by one definition of identity (De Fina, 2006: 265):

> "Identity can been (sic) seen and defined as a property of the individual or as something that emerges through social interaction; it can be regarded as residing in the mind or in concrete social behaviour; it can be anchored to the individual or to the group"

This definition consists of several key points. Firstly, identity can be seen as properties that an individual has, similar to their personality, as well as emerging within social interaction. Secondly, because of this, there are both individual identities that a person can have as well as social identities that they belong to as a part of a group, and it can be both a mental and social construct. If we consider identity as also being the "social positioning of self and other" (Bucholtz and Hall, 2005: p.586) the identity of the *other* is created during interaction and developed in *my own mind* from the information I am receiving from them as the interaction(s) develop. As such, other people's identities are created in our minds by the way in which we perceive others. The way they speak, dress, behave, and gesture – all of these and more require perception from our own point of view and these perceptions drive the identities we create for them, even if they differ from a person's own self-identity. Another person's behaviour, for example, may be perceived as very aggressive when we process the information of their speech and non-verbal behaviour. In turn, the identity we create for them may be negative, aggressive, or threatening, for instance. Even though that person may not be intedning that information by perceived in such a way, our perceptions are key to identity creation.

While De Fina (2006) provides a relatively comfortable definition of identity, there still exists the question as to how it is created. As mentioned, there is not one fixed identity but a series of dynamic identities that are performed. In interaction, however, even with this lack of stability, the factors that determine these performances must come from somewhere. Hall (2013) discusses two aspects of identity creation. The first are the aspects of society we are born into e.g. ethnicity, social class, and geographical location. The second are those we choose to be a part of or are otherwise brought into by others, be they family, friends, colleagues or a result of community practices. These two aspects of society see us perform a wide variety of roles. Though they may not always directly determine how our language will change from role to role, they have a strong influence in the outcome of its production (Hall, 2013).

Language and identity are strongly linked and can be seen as a key aspect of human nature (Llamas and Watt, 2010). Not only can language be used to describe the identity of ourselves and of others, but perceptions of someone's language use also allows us to identify them as both an individual and as belonging to particular social groups and communities. We are capable of associating patterns of language use to an individual, which we may describe as being part of their idiolect. Similarly, we are also able to identify patterns related to different social groups. We are able to differentiate between various dialects – patterns of language that we associate with particular geographical regions or social groups (Hughes et al., 2013).

Language use consists of much more than simply communicating a message. How a speaker interacts with a listener can have a negative or positive impact on how that speaker is perceived. Using an example from (Joseph, 2010), the simple act of notifying another person that their shoes are untied, carries with it not just the factual information but an indication of concern for another being, and in turn creates a social bond. However, the way in which a speaker chooses to address the listener in such a situation will not only reflect their intention, but also have an effect on how others identify them. This holds true for not only the choice of words (such as vague choices discussed in 2.7 and Chapter 3) but also for how the words are said. Changes in prosody may shift the speaker's intention from concern to ridicule or humiliation and consequently change the way in which the listener perceives the speaker.

2.4.2 Linguistics and Identity

The strong link between language use and identity has prompted a focus on developing a linguistic framework around analysing identity. This thesis draws influence on one such framework developed by Bucholtz and Hall (2005), who devised five principles on understanding identity in data from a linguistic perspective. This section will summarise their contribution, before discussing the concept of identity in human-agent interaction in 2.4.3 onwards. Bucholtz and Hall (2005) focus on creating a framework that pertains to human interaction. This leads to the creation of five principles that result in the framework.

Their first principle discusses that identities are a product of linguistic practices and emerge through interaction, rather than wholly being a fundamentally and pre-existing phenomenon. The temporary roles interlocutors take up in interactions will play a part in the emergence of these identities, as discussed in their second principle. For a user taking the role of an instruction follower, and the agent assuming the role of the instructor, these already set in place a starting point in which emergence will take place. Previous interactions with humans and agents alike with both of these roles will likely influence how identities will emerge in future interactions. Bucholtz and Hall go on to discuss that emerging identities are perhaps best identified in those cases where "speakers' language use does not conform with the social category to which they are normatively assigned" (p. 588). This is of particular interest in regards to how an agent's identities are perceived by users. Users may categorise agents as being typically direct in their language use and possessing a particular type of voice. As a result, they may not be receptive to agents that use language with a greater focus on relational and social goals, rather than just transactional. This may also be affected by other factors such as context.The realities of agents conforming or not conforming to a

user's prescribed set of norms are both interesting in regards to how they position agents socially.

This notion of positioning is also discussed in their third principle in which they discuss linguistic forms and structures being associated with particular identities. They refer to this process as "indexicality" which accrues through interactions. This means that dialect, lexis and phonology can all be associated with particular repertoires, social stances, and identities. For agents, this may mean that they have both macro and micro levels of language associated as part of their linguistic repertoires. Using the specific case of VL in this research, it can be argued that users may have some form of indexed identity categories related to the use of VL. How this combines with and indexed categories of agent instructors may inform to what extent this style of language in this particular context is useful and appropriate for the user.

Their fourth principle discusses relationality – the notion that identities require social actors and social meaning, and that they are linked with other categories of identity. This is often seen in the similarities and differences between interlocutors. Often there are distinctions in identities that arise a result of language use. These may mark others as being part of a particular social category, for example, and displaying similarities and differences between speaker and listener will both affect the identities being projected by either party. To clarify, the agent being typically direct in its language use will be used again as an example. A user may associate an agent with direct language use, and through social interaction with the agents used in this research encounters language that is more similar to something they use themselves, particularly for various social reasons[5], i.e. the VL. This may be at odds with their indexed identity categories of agent language use and, depending on the user, may have a negative or positive effect on their on-going and emerging projections of its identities. This VL use may be similar enough to the user's own language use, or their identity categories they are familiar with, and result in the agent being able to display itself as more likeable for example, and attempting to engage in rapport as well as instruction giving. Conversely, it may be too much of a distinction from expected norms in relation to expected identity categories that it is not perceived as such, and may adversely affect the social positioning it is given as discussed previously.

The final principle in this framework is that of partialness, and posits the following (p. 606):

[5] At least in regards to a particular style of language they may employ in different contexts of human interaction.

> "Any given construction of identity may be in part deliberate and intentional, in part habitual and hence often less than fully conscious…"

This suggests that, consciously or not, parts of the identities social actors wish to construct and project onto others are somewhat deliberate. For agents, this is slightly different. Currently, no agents possess consciousness. Verbal agents can be programmed, however, to converse in particular ways based either on pre-written statements or as a result of algorithms and progressively accrued knowledge from on-going interactions. Perhaps more so with pre-programmed statements, this means that language use is deliberate and, if an agent designer so wishes, the attempt to project a particular identity also deliberate. With the vague agent, there is an attempt to assess whether it can project a softer, less controlling identity through its use of face saving strategies arising from lexical choice.

Table 1: Summary of linguistic principles of identity based on Bucholtz and Hall (2005) with additions and modifications of the descriptions.

Principles of Identity in Linguistic Interaction	
Emergence Principle	Identity emerges as product of linguistic practices through interaction; not wholly pre-existing
Positionality Principle	Local identity categories & temporary roles, along with macro categories (e.g. age, gender) contribute to emerging identities
Indexicality Principle	Linguistics forms & structures are associated with specific identities e.g. dialect (macro); idiolect (micro)
Relationality Principle	Identities emerge in relation to one another; categories are linked and not independent
Partialness Principle	Identity construction can be partly habitual; intentional; part of others' perceptions; constantly shifting

This linguistic approach to identity, summarised in Table 1, provides an explicit framework not only for human interaction but also for human-agent interaction. Sometimes these principles apply fairly equally to agents and in other principles there are obvious differences. In short, this can be used to form a basis on which to analyse how agent identity through language use is formed through socially emerging, relational, and sometimes deliberate linguistic interaction, both on a micro and macro level. This is also affected by the context in which the interaction occurs and the roles in which agents and users take on. For a verbal agent instructor using VL, it is unclear how users will relate their indexed identities towards a vague agent, and their expectations of how an agent may typically interact with them. This framework also justifies the use of combining the micro level lexical

choices of VL in the model outlined in Chapter 3, and the macro social and interactional functions of this lexis.

2.4.3 Exploring Agent Identity

The CASA paradigm discussed in 2.1 suggests that voices are social actors and the notions of "self" and "other" apply to them (Nass et al., 1994). Computers using speech have also been shown to have personality (Nass and Lee, 2001; Lee et al., 2006). Using Bucholtz and Hall's (2005: 586) broad definition of identity, "the social positioning of self and other," it can be argued that verbal agents can also have identities that emerge in interaction. There are differences, however, in how the principles of language and identity are realised in HAI when compared to HHI. Agents, for example, are not born into aspects of society such as ethnicity and class, rather they are programmed to belong to these entities or either provides the user options with which to customise these. These customisable options include features such as voice, gender of the voice, and language. Users are not always provided with a choice regarding these factors and so may make assumptions as to the identity of the agent.

As for aspects of society, agents being brought into them are determined primarily by their designers and the intended user demographics. Satellite navigation systems will unlikely pervade many other aspects of society or interaction than that which they have been designed for – providing directions in a vehicle. This may differ when using agents that do not have a single-track purpose, however (e.g. automated checkout). Intelligent personal assistants for example could be used as a personal means of retrieving information and performing tasks, though when used with two or more people can provide a means of entertainment by way of exploring its capabilities and limitations.

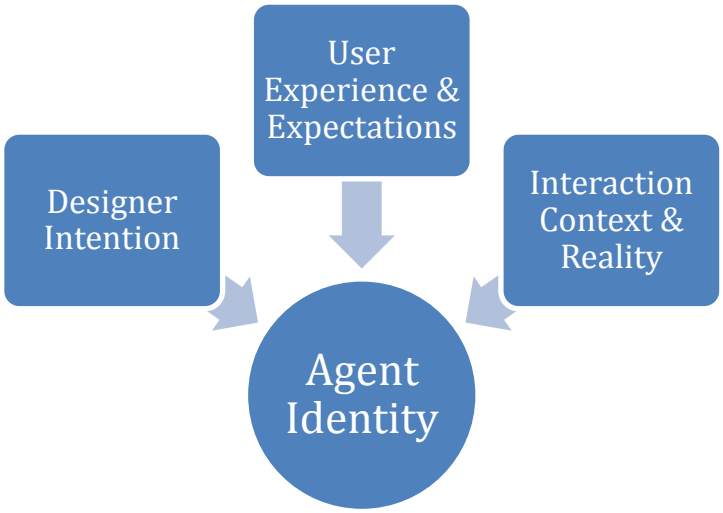

Figure 1: Formation of Agent Identity.

The diagram above is a representation of how the users, designers, and interaction contexts can coincide and affect agent identity. There are similarities shared with this diagram and the principles of identity summarised in Table 1. While identity emerges through interaction and is not wholly pre-existing, according to the emergence principle, there are elements that are pre-existing for agents that differ from humans. Those who design an agent will have decided the features and characteristics that an agent will consist of, and the linguistic variables that it will use to communicate with. This overlaps with the partialness principle. The creation of agent identity will be partly intentional, from the agent designer, and partly because of the perception of others, which comes from the users. These perceptions will be fuelled by previous experiences too, as discussed in the relationality principle, because identity categories are linked to one another, rather than being independent. This means that an interaction with one agent, such as a sat nav, may influence other sat nav interactions that the user has in the future. Similarly, it may influence other interactions with verbal agents that are not sat navs. In the case of the latter, they are related to previously indexed categories of identity (indexicality principle). Finally, there is also the positionality principle. Part of this principle suggests that the roles people assume in interaction are temporary, though with agents this may not always be the case, at least from the viewpoint of the designer. The designer may intend for an agent to always be an instructor, for example, though this does not guarantee all users will perceive the agent in the same manner all of the time. Figure 1 demonstrates that the designer of the agent takes on some of the responsibility of how agents perform their identity towards their users, and in turn, how users create identity for the agents.

The contribution of a user's expectations towards an agent may take several forms. This could be what the user expects the agent to be able to accomplish. Examples include its functional purposes, the medium in which it communicates, and the linguistic variables it uses to

communicate with. The latter is particularly salient for verbal agents as this includes the voice it uses. To this end user expectation relates to the theory of affordances (Gibson, 1977). Gibson's theory describes that how an object's possibilities are perceived can affect the actor's relationship with that object. That is to say what someone perceives something they interact with can do can have an effect on the way their relationship with that object will transpire. If a user perceives an agent to afford certain actions, they will have certain expectations about that agent and about their relationship and interaction with it. This could include a single linguistic variable, or combination of them, that they relate to other indexed categories of agent identity and, in turn, an agent's affordances.

In regards to HCI, there exists both real and perceived affordances – possibilities of what a system is actually capable of, and what its users think it is capable of (Norman, 2013). A glass, for example, affords seeing through, holding a volume of liquid, and being fragile. A verbal agent may afford interaction through speech in either a passive (one-way) or active (two-way) discourse. There are also social and cultural aspects of the human interaction to consider for affordances in HCI (Kaptelinin and Nardi, 2012). The users' expectations in regards to the capabilities and characteristics of an agent (such as in voice and language as discussed later in this section) may not actually coincide with the reality of an interaction. As shown in Figure 2, this can occur before, during and after interaction with an agent, and prior experience may affect these perceived affordances as much as an on-going interaction. These social, cultural and individual features of each user's own mind and experiences may also affect these perceived affordances.

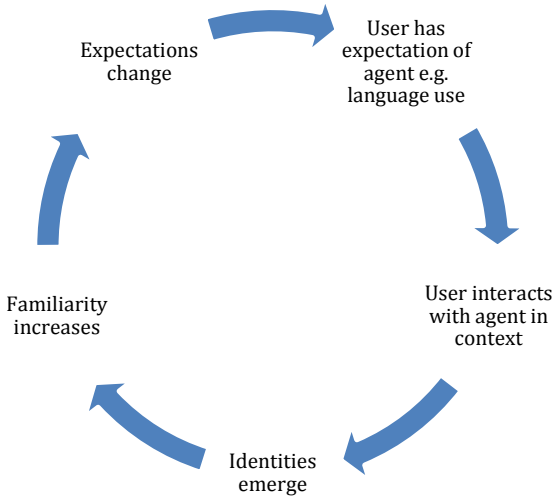

Figure 2: An example of how a user's expectations help shape the identities they create for an agent.

The figure above is a representation of how a user's expectation may develop during an interaction with an agent. As they interact with the

agent, the perceived affordances of the agent generated by the user may or may not emerge, which can influence the way in which they prescribe an identity. This occurs both during and in between interactions (Bucholtz and Hall, 2005). This is also similar to how people frame expectations of other people (Tannen, 1993). It may provide them with a sense of familiarity as the interaction goes on, thus affecting their opinion towards the affordances of the agent. Similarly, their opinions as to particular features of the agent will form both early on and as the interaction progresses. There are no guarantees, however, that the perceived affordances will align with the real affordances.

An agent that operates at purely an instructor level will differ in affordances from another that operates at a peer level. Agents involved in other aspects of communication, often appearing in research and development, include those that aim to achieve rapport with its users (Bickmore and Cassell, 2001; de Graaf et al., 2015) that take on roles such as carer, advice giver and, to an extent, friend. The actual and perceived roles of an agent link strongly again to the theory of affordances in HCI. Not only may users' perception of agents change in regards to what affordances they perceive, but their behaviour towards them may also change depending on what they believe its capabilities are and what expertise it projects (Pearson et al., 2006).

2.5 Identity in HAI and the Effect of Linguistic Variables

In generating a better understanding of what identity means in HAI, I will discuss a set of examples taken from previous research in which agents have been designed with specific variables of prosody, voice, and language. In these examples, the researchers evaluate the agents in interactions with their users, using both quantitative and qualitative approaches. How people evaluate these different agents in different scenarios allows for more insight as to what place agents have in this world in terms of personality, functionality and purpose. Also, insight can be gained into how this may compare to humans in similar contexts. In turn, this can provide insight into how voice and language can affect the identities that users create for these agents.

In gaining these insights, we may unveil indications as to how different identities emerge in human-agent interaction and what factors may lead to the creation of different identities. This allows for generation of a bigger picture of agents, piece-by-piece. This includes where they fit into this world, how much of an "us versus them" relationship is occurring and what they are primarily seen as e.g. tools, machines, assistants, or perhaps friends. Another way of describing this outlook towards agent identity is assessing how agents are perceived, what metrics can be used to recognise this (e.g. quantitative surveys or

qualitative content analysis) and how these factors will be used as a means of understanding identity.

2.5.1 Applying Humanlike Variables to Agent Communication

Humans react socially to agents, and even those which possess lower-end synthesised voices can create the perception of personality (Nass and Lee, 2000). However, it is unclear whether or not there is linearity in the social responses people have to computers as human likeness increases. The uncanny valley hypothesis, discussed in 2.1, states that reactions to human likeness in robots are usually positive until reaching the cusp of human likeness, without quite reaching it. A recent review of research into this hypothesis, however, provides an alternative guideline – increasing human likeness generally leads to more positive reactions in users, although it is conceded that exceptions to this rule may exist (Kätsyri et al., 2015). The greatest negative affinity was also discovered to be when there was a perceptual mismatch for the user i.e. that certain features of the respective artificial entities were not perceived to match up. For example, a mismatch in human and synthesised faces and voices resulted in higher sense of eeriness for the users (Mitchell et al., 2011a)[6]. It remains unknown, however, whether a sense of eeriness, discomfort, or otherwise negative feeling will arise if often used conversational human strategies appear in something that is very non-human.

With vocal and linguistic qualities, this challenge of similarity with human likeness in agents is less clear. There are so many differences in human interaction alone that humans experience on a daily basis – regional accents and dialects, idiolects and different languages are some of the differences in language quality that are commonly faced. Similarly, differences in speech cadence, pitch and frequency vary tremendously (Hughes et al., 2013). All of these allow us to distinguish between individuals and social groups. While agents are traditionally direct and can often possess synthesised voices, this is not always the case, as agents become more widespread and diverse, and include the use of human recordings as their voices.

2.5.3 Prosody

The following sections discuss some of the literature on the features of of verbal agents that are pertinent to this research – namely language use, voice and prosodic capabilities. The various effects these agents and their varying degrees of ability in incorporating these phenomena into their interactions with humans are discussed to provide a context

[6] Although this paper presents a small sample size of papers that are also fairly recent rather than spread throughout years of research.

for the two studies discussed in subsequent chapters. They focus both on how variations of these aspects affect interaction and perception, both in general changes and in regards to being similar or dissimilar to their human counterparts.

In human interaction an interlocutor who is similar may be preferred to someone who is different (Montoya et al., 2008). This is sometimes referred to as similarity-attraction theory (Nass and Lee, 2000; Nass and Lee, 2001). Preference for similarity has also been documented in HCI. Designing agents with the aim of creating a sense of similarity and in-groupness may be a valid and useful design approach (Häring et al., 2014) and may make interactions more effective (von Scheve, 2013). More extreme examples of similarity preference may be observed that have underlying racist or xenophobic inclinations, though this is not something that is addressed within this thesis.

One study focusing on similarity in HCI and similarity has focused on prosodic features (Mitchell et al., 2011a)[7]. Vocal cues in agents such as pitch and intensity, and speech cues such as speech rate and nonfluencies, influence the perceptions people have on speakers Users may also change the loudness of their speech and duration of their pauses to match those of computer-generated speech (Suzuki and Katagiri, 2007). Prosodic alignment may also be affected by voice in that a more humanlike voice may be seen as more competent, potentially changing the levels in which users align themselves to an agent using a human voice as opposed to a synthesised one (Cowan et al., 2015). Similarly, prosodic alignment by a computer may create a positive response from the user (Branigan et al., 2010). Another study revealed that users might display a preference towards an agent that shares similar prosodic features in intensity of speech, fundamental frequency, frequency range, and speech rate (Nass et al., 2001). This study also suggested that setting parameters on paralinguistic features might help to maximise the perceptions of likeability and trust a user perceives when interacting with a computer voice.

Accent and pitch can also both affect the evaluation of a system, as seen in a comparative study on the use of British and Singaporean accents in voice user interfaces (Niculescu, 2011). This was conducted on native Singaporeans, and it was shown that a British accent might be seen as more polite and having a higher voice quality and ease of dialogue than a Singaporean accent – a result of it culturally having more esteem – and a higher pitch might be seen as more enjoyable to interact with and rank higher in voice appeal, personality and behaviour.

[7] A further in depth discussion of the literature on this topic can be found in the reference cited here.

The findings from studies on prosody in HCI indicate that similarity in an agent's prosody can induce positive responses from their users, and that there are also sociocultural factors involved in how computer voices are perceived. While this thesis is not strongly focused on prosody in isolation, its features contribute towards the voice as a whole, which relevant to both studies in Chapter 4 and 5.

2.5.4 Voice

Voice as well as prosody can provoke different responses in users. Voice can be a powerful tool in assessing identity, even if it is with someone one cannot see, and allows us to make assumptions on their age and gender, and allows us to prescribe an identity towards that voice (Watt, 2010; Latinus and Belin, 2011). Given the similarities between human interaction and HCI one can assume that an identity is also prescribed towards a humans and non-human voices in agents.

Agents can either produce verbal output using synthesised speech technology or pre-recorded human speech (Georgila et al., 2012). The responses that different voices in agents can create are particularly evident when comparing synthesised and human voices. One study assessed the difference between a machine and human voice, on speaker rating for users receiving an educational narrated animation (Mayer, et al., 2003). The results revealed that the machine voice group rated the speaker as "less dynamic, less attractive, and less superior" than the human voice group. Lee (2010) shows that human voices in an anthropomorphic interface are rated more positively. He does present one caveat with the effect of personality type on the results, as these only occurred in participants who were "less analytical or more intuition-driven." Having a human voice may also create the impression of an agent that has greater communicative abilities, and may be rated as more advanced and capable than a synthesised voice (Cowan et al., 2015). In a similar vein to Lee's (2010) study, Cowan et al. (2015) highlight that further investigation might be needed in understand the characteristics of the users, such as cognitive styles, in the understanding of human-machine interactions.

The differences between synthesised and human voices can change when comparing lower-end and higher-end synthesised voices. Using human speech to create an "advanced voice" for example, can result in greater interaction satisfaction than a "basic" computer voice (Cowan et al., 2012). However, this study also reported no significant statistical difference between the advanced voice and an actual human partner. Similar effects may be observed when comparing synthesised voices to amateur and human recordings (Georgila et al., 2012). In this study, voice actor recordings were rated as more likeable, conversational, and natural than both amateur human recordings and synthesised voices. However, it was also discovered that a "high-quality general-purpose voice or a good limited-domain voice" can outperform amateur human recordings (p.8). Georgila et al. (2012) argue that

synthesised voices have reached a point in quality where there may be room for synthesised voices to replace amateur human recordings, particularly in regards when there is a large disparity in performance versus cost. The voice actor, however, remains the most preferable. These studies indicate that the distinction between synthesised voices and human recordings can be seen as a cline – a gradated continuum in which there are "certain focal points where phenomena may cluster" (Hopper & Traugott, 2003: p.6). In such a cline, a cluster of voices that are perceived to sound humanlike, such as those recorded by professional voice actors may populate one end. At the other, lower-end synthesised voices that do not sound humanlike may cluster together.

Despite the findings discussed above, this research was measuring a series of individual utterances given to participants, rather than a prolonged interaction of any kind. Moreover, the interface that was being used in testing synthesised and human voices is designed for use in assisting military personnel and family members to seek help for conditions such as post-traumatic stress disorder. While a useful contribution, the lack of a continuous interaction that may more resemble a real human interaction leaves a gap to be filled in comparing synthesised and human voices. Similarly, there is more a focus on the voice quality than any investigation into language use. While it does provide evidence in the potential benefits of using a voice actor, there are no guarantees that this extends to communicative phenomena such as VL. The gaps in this research are some of the focal points addressed in Chapter 4 and 5.

It is quite apparent that agent voices are not neutral, and with them come cues such as gender, age and personality, which in turn affect the identities user project onto that voice (Dahlbäck and Jonsson, 2010). This makes it an important consideration designers of these verbal interfaces, as the research discussed in this section has shown that an agent's voice can affect the perceptions the user will create of that agent. Moreover, the relationship between VL in agent instructors and different voices is not clear. Given that both synthesised and human voices are the two common options in designing verbal agents, and are both options that are available for designers, it is an important variable in the overall perceptions that users develop. VL is introduced as another variable in this thesis and, in Chapter 5, is combined with the approach of comparing synthesised and human voices seen in similar research. In doing so, the existence of similar drawbacks and benefits of these two voice types discussed in previous literature discussed above can be investigated.

2.5.5 Language

Language and identity can be strongly related (Jaworski and Coupland, 2014)[8] and seen as a performance that is affected by "private and institutional discourses"[9]. This was discussing dynamic interaction, i.e. two-way interaction; however this is not something that all agents are capable of. The agents used in this thesis (Chapters 4 and 5) do not have this dynamic capability and instead the agent can only speak to the user and not be spoken to. For agents that are dynamic, one may assume that a designer will be to some extent aware of the discourse context in which their system is being deployed, the demographics targeted, or at least the style in which they wish their agent to communicate. In using this awareness, an agent's lexical outpute may be used to affect how it is perceived by its users (Fong et al., 2003).

In agent language, there is often a tendency for it to be direct and (Clark et al., 2014). This could arguably be seen as a common style and strategy that designers employ for use in human-agent interaction. There are indications that modifying machine language to be more humanlike (i.e. moving away from direct) can be beneficial in creating a positive identity, such as in the use of politeness strategies of face saving, greetings and personal pronoun use (Wang, et al., 2008; Torrey, 2009; Torrey et al., 2013). Making an agent's language more humanlike may then be seen as straying from the perceived status quo and group identity of direct language in HAI and moving towards natural language use of human interaction – language that encompasses styles, strategies and a concern of face and rapport maintenance. Using natural language may also help towards amplifying the Media Equation effect in agents (de Graaf et al., 2015), though it is uncertain whether this would have a positive or negative impact on the interaction.

It should also be noted that linguistic strategies are not always perceived positively in interactions with machines (Strait et al., 2014; Strait et al., 2015). In Strait et al. (2014), for example, the interaction distance between the robot helper and user was an important contributing factor in the overall perceptions. In other research on politeness, direct interactions were not actually observed, which

[8] Further literature highlighted in this reference give a more detailed discussion. Also, the second edition is referenced here.

[9] These discourses referred explicitly to subjectivities of the self and how they alter these subjectivities. This is opposed to the notion of the self being natural.

somewhat limits the scope of the findings (Torrey, 2009, Torrey et al., 2013). This research focused on observations of an interaction, with the observers giving the positive feedback, rather than any direct interaction participants[10]. The example studies in this paragraph are focused on task-scenarios, as are the studies in this thesis, though there are gaps that the investigations in Chapter 4 and 5 aim to address. This is discussed further in 2.4 and 2.5.

The idea of moving towards agents using natural language may also come into contention with the familiarity of their user base with that agent's language. Familiarity can be seen as having prior knowledge of something that informs expectation of future interaction, or defines an intimacy with something or somebody (Turner, 2008). An example in discourse would be that of common ground (Clark, 1996) in which two or more speakers have a mutual assumption that those in an interaction are familiar with particular items of language and references to them. Prior knowledge, experiences and uses of language are then built upon as the interactions emerge and develop, and can inform opinions their prior experience will inform people's opinions in the present (Fong et al., 2003). It is difficult to be fully aware of which agents people have experienced and they are familiar with, and in what contexts these interactions occurred. While prior linguistic research in human communication can provide indications of how something such as politeness or VL may work, for example, it is difficult to predict their effects in human-agent interaction. Given the often direct nature of agent speech (Clark, et al., 2014), we can postulate that using strategies such as VL may create perceptions of more humanlike language, though in an unfamiliar context.

The increasing use of natural language in agents may shift their language use away from direct and towards atypically indirect, and more similar to "human language". This creates an interesting discussion as to how this may shift individual and group agent identities, if at all. Discussions also concern the extent people are comfortable with smaller distances between agent and human speech, particularly when using language for both social as well as functional goals (see 2.4 and 2.5).

2.5.6 Identity and Adaptability

As agent identities can be individual to the user, there is also the matter of the extent to which the individual is catered for, and how any changes in agent speech and other behaviour is instigated (e.g. by agent or user). Giving a user the choice to modify elements of the agent, such as the voice, is one possibility. In a study on in-car interfaces, this ability to choose can be positive, even though it may not be something the user prefers in the long run (Lee et al., 2011). The

[10] A point also highlighted by Strait et al. (2014, 2015).

study also reported that choice of even several voices in such an interface may null the effect of similarity-attraction, and argues that having a user's own voice or a voice designed by them may be a goal that interface designers could aim for. Changes in agent behaviour may also be instigated by the agent itself, rather than the user. Social robots for use with the elderly, for example, were received positively when they adapted to personal needs (de Graaf et al., 2015). Wider cultural aspects of behaviour may also have to be considered when designing socially intelligent agents (Mascarenhas, et al., 2015). The extent to which an agent caters for an individual or a group may impact upon its user's perceptions, both of which may either be instigated by the user or the agent. In understanding the use of VL in agents, it may be useful to first understand the impact it has on user perception and its correlation with an agent's voice, before instigating the discussion on modifying its behaviour.

2.6 Language, Identity & Mitigation

One of the significant issues explored in this thesis is the issue of identity in instruction giving agents. In human interaction instructions can be given on a direct-indirect spectrum. This in turn can affect the way in which hearers perceive the identity of speakers. This section discusses some of the issues in and around indirectness in human communication, notably politeness theory and facework. Following this, previous research into the notion of politeness in HAI contexts is discussed, including an example of relevant linguistic changes in a real world HAI environment.

2.6.1 Mitigation in Human Interaction

The use of mitigated communication as a means of not imposing oneself on their interlocutors is one of the key features of Politeness Theory (Brown and Levinson, 1987) and is one of the primary modes of polite communication (Carter, 2004). When being polite, speakers intend to maintain a social equilibrium between themselves and their interlocutors and maintain an amicable interaction (Leech, 1983).

Politeness Theory builds on the work of Erving Goffman and the notion of face. Face is the public social image we project to others during interactions (Goffman, 1967, Goffman, 2002, Goffman, 2012). It is usually in the interest of all people in an interaction to protect this self-image, known as saving face. This is done through on-going mutual understanding of what is acceptable and unacceptable in any given interaction, with the knowledge that there is a potential mutual damaging of face should there be diversions away from acting on this understanding. A person's face changes not only between different interactions, but can also change and develop throughout an interaction. Through experience of interactions, however, there is some sense of face that lies within the persona of an individual as well as in the emerging interaction (Spencer-Oatey, 2007). In these social

interactions the aims are to establish, understand, and maintain the face of all parties, as well as to prevent any negative emotions and feelings that may arising from losing face. Engaging in an interaction that creates an imbalance of power may have threats of face loss, such as giving directives – a term for when a speaker wishes to instruct one or more recipients to do something (Clark, 1996). Face saving strategies can thus be employed through the use of indirect language.

Politeness Theory split the idea of face into negative and positive categories (Brown and Levinson, 1987). Negative face refers to the right of a person not to be imposed upon by others, and to their freedom of action without obstruction. Positive face refers to showing approval of the public self-image of others in interaction. In order to prevent imposition and to be approved by others, speakers can use a variety of what Brown and Levinson refer to as politeness strategies. Speakers can therefore be indirect or imprecise to mitigate potentially Face Threatening Actions or FTAs. FTAs can depend on variables such as the social distance between speaker and listener, the different in power, and the relative impact an utterance can have in any given culture. Politeness strategies to prevent FTAs are again split into positive and negative categories.

Positive politeness strategies to prevent FTAs include acknowledging that both speaker and listener are part of the same in-group, and recognising the wants of the listener so they may reciprocate in kind. Negative politeness strategies focus on recognising the listener's negative face and their right not to be imposed upon. This includes using indirect language to avoid imposition and potential face threats. It should be noted also that although the aim of these strategies is to prevent FTAs, they could also be used to soften FTAs, as some of them may not be preventable.

Brown and Levinson (1987) discussed the use of a formula for calculating the weightiness of an FTA. This is essentially the impact a face-threatening act in an interaction. This formula was written as follows:

$$W_x = D(S,H) + P(H,S) + R_x$$

The individual symbols refer to the following: W = weightiness, or the impact; D = social distance between speaker and hearer, often described as being a symmetrical relationship; P = distance of power between hearer and speaker, often as an asymmetric relationship; R = rankings of FTAs in a particular culture i.e. the degree of imposition that it causes; x = the specific face threat being assessed. While it is possible to attribute numbers to all the values in the FTA equation, it can also be thought of as a representation of how threats to face are impacted by social and contextual variables.

The social distance between speaker and hearer can refer to the frequency of interaction that they have, as well as the exchange of material and non-material goods between them. This includes things such as physical items (money, items) and non-physical (face). According to Brown and Levinson this reflects social closeness, in that the more S and H interact with one another, the more they will begin to show keenness towards respecting the wants of each other, as well as sharing their own. Power can also have physical and non-physical elements, and is said to be authorised (money, physical force) and non-authorised (control over actions of others). It is also described as referring to the extent to which H can impose their own plans and face at the expense of those belonging to S. The ranking of an FTA in a culture, R, can change both between and within cultures and is reliant on the context in which the FTA takes place. Context, in this case, refers to any variable that may affect the weightiness of the FTA. This can include the physical environment that S and H inhabit and the rights either S or H has of conducting a particular FTA. Given the wide-reaching definition of context here, this could be extended to include the personalities and identities of S and H at a given time and space.

There are some criticisms towards Brown and Levinson's approach towards politeness. Attempts to classify universal theories of politeness ignore the various social, cultural and individual differences in what is polite and impolite. Many scholars have differentiated between two types of politeness as a result. The first-wave politeness or Politeness$_1$ refers to those folk or lay interpretations that are social, individual and evaluative; involve collaborative negotiation with a view towards acknowledge face and preventing its loss; and depend on whether or not the listener categorises an utterances as polite or not (Watts, 2003, Group, 2011, Eelen, 2014). Politeness$_2$ or second-wave politeness is the wider theory of politeness and what academics discuss as being polite in interaction. The reason for the distinction is that Politeness$_2$ does not always take into account what interlocutors in any given interaction perceive as being polite. Politeness strategies outlined in the original theory, for example, may be impolite in some interactions. Similarly, language that may be classed as impolite in some social groups and cultures may be polite in others. Politeness$_2$ also discusses (im)politeness – the whole spectrum of both polite and impolite talk, though often from an outsider's interpretation and perspective, whereas Politeness$_1$ focuses on user evaluations of what is polite in interaction, with a focus on the polite end of the spectrum. Watts (2003) goes on to discuss that it is impossible to create a predictive theory of politeness given the need to evaluate on-going interactions, and that linguistics items are not inherently polite or impolite as a result (e.g. Culpeper, 1996).

Watts also criticises the FTA formula discussed on the previous page, arguing that it implies the variables are equal, when in reality they may not be (Watts, 2003: 93). The factor of the power difference,

between S and H, for example, may have more impact towards the weightiness of the FTA than the social distance, or the absolute ranking of the FTA in a particular culture. This criticism of the equation is not wholly unwarranted. Mathematical formulae do have certain rules in their form, and displaying a way of calculating the weightiness of an FTA in such a manner can bind them to the same rules. Nevertheless, if the formula is not taken too strictly and, if seen as a representation as mentioned previously, it does present itself as a useful analytical tool.

Although speakers do use politeness strategies to prevent the loss of face in interaction, it is nigh on impossible to create a universal description of politeness, as it would ignore too many sociocultural and individual variables. However, that is not to say we cannot provide descriptive accounts of politeness strategies and attempts to mitigate face threats. It is also possible to categorise descriptive accounts on a micro and macro level, but with the proviso that if used to analyse future interactions there is no guarantee any descriptions and their functions will appear in the same way. In this sense, it comes to somewhat of a full circle: a politeness$_1$ approach can be used to describe and categorise, to then inform politeness$_2$ approaches which can be used in turn for future analyses. This approach is useful for HAI contexts as language can often be controlled, unlike genuine and spontaneous human interactions, at least from the observer's point of view. As a result, agents can be equipped with specific linguistic output, and other variables such as its voice can be controlled, which makes for an ideal and almost isolated testing environment[11]. The politeness1 patterns that emerge from describing the interactions can then inform the politeness2 approaches with specific mentions of salient variables that affect a user's perception of that agent. It should be stressed again that this does not provide guarantees, but can be used to inform future testing environments.

In regards to the notion of politeness, the following three chapters will be focusing mostly on the negative politeness strategies that are used in facework. The application of politeness for the two studies is done via the implementation of a specific VL model that is discussed in Chapter 3. This chapter describes VL and its different functions, while also providing examples of it in context of the instructions that are used by the agents in Chapters 4 and 5. For this research context, the primary focus of using VL is on the mitigation of imperative language and conducting facework, however this is note the sole function of VL. This approach towards using VL is revised somewhat in Chapter 6 (6.2.5) in light of the results of the two studies.

[11] Isolated in the sense of being different to other agent interactions that possess differences in their linguistic variables.

2.6.2 Politeness in Human-Agent Interaction

As well as the CASA paradigm and related research showing that humans display politeness norms towards computers (see 2.1), there has been some work on incorporating politeness strategies into agent communication. For example, when used in the classroom, polite agents using both text and speech were seen to improve learning outcomes (Wang et al., 2008). Students who received polite agents scored better than those who received a direct alternative. Similarly, when used in advice giving robots using speech for a baking task it was shown to make them appear more likeable, considerate and less controlling (Torrey et al., 2013). This study explored the successful use of hedges and discourse markers as polite communication. It did, however, analyse the opinions of participants watching others interact with a robot, as opposed to analysing the opinions of those who actually interacted with it. In robot-instructed drawing tasks, politeness strategies did not always help to improve agent perception (Strait et al., 2014). Although there were similar improvements in likeability and considerateness as in Torrey et al. (2013), they argue that speech efficiency and human likeness is of more importance. Also, these effects were only noticed in third-person interactions, which were similar to the observations of interactions in Torrey et al. (2013). While the results of Strait et al. (2014) did not fully correlate with those of Torrey et al. (2013), the latter provided some guidelines on how politeness may function in an HRI context, which were applied and modified by the former (Strait et al., 2014).

In conditions where actual human-robot interaction was present i.e. the robot was sharing the same space as the participant, these positive effects were less prominent. The study by Strait et al. (2014) did take into account the limitations of having a specific help-giving robot operating within defined parameters. However, the measures used had a heavy focus on statistical analysis of questionnaire data and the use of functional near infrared spectroscopy (fNIRS). While providing further insight into the previous work on politeness in this area, there was no insight into the participants' experiences in their own words. This leaves a gap of analysing the opinions of the participants, which could supplement these quantitative measures with qualitative data, and use a mixed-methods approach to further bolster data analysis.

There are also examples of polite agent communication outside of laboratory studies. In 2015, media outlets reported that the automated checkout systems at Tesco supermarkets were being updated to change both the voice and the language used to guide their customers through checking out their own purchases. Reports stated that Tesco were shifting towards using "softer phrases" (Collins, 2015). Examples 1 and 2 are taken from this article and discussed below:

1) *Old – "Unexpected item in bagging area. Remove this item before continuing"*
   New – "This can now be placed in your bag"

This changes the directive of *remove* to a representative *this can*. The function of the utterances remains the same, in that the system alerts a customer of an item not being placed in the bagging area as expected. The illocutionary force, however, changes, and the imperative changes to an alternative, which implies the user has power over the act of placing the item where it is required.

2) *Old – "Please take your items"*
   New – "Thank you for shopping at Tesco"

The old phrase here did include the polite use of *please*, which somewhat mitigates the imposition of the directive *take your items*. However, the new phrase is now a positive politeness strategy of thanking (Brown and Levinson, 1987) while still indicating that the interaction with the automated checkout has reached an end.

3) *Old – "Approval needed"*
   New – "We just need to approve this"

Example 3 is taken from a report on the same changes by Robarts (2015). This new phrasing hedges the fact that an item needs to be approved by a member of staff before the interaction can continue. The phrase *just* is used here to do this, making the phrase a negative politeness strategy. This is often seen as a hedge, though is described as a minimiser in this thesis, as discussed in 3.3. In this sense, the use of *just* here attempts to reduce the implied impact of needing approval for an item, perhaps indicating that it will not take a long time for the customer to wait for this to happen.

As well as the language, the voice of this system was also changed. The discussion in 2.3.4 highlighted that voice can have a strong impact on the way agents are perceived. In the case of this automated checkout system's voice, this was changed along with the language (Collins, 2015; Robarts, 2015). The company reported that customers had referred to the old voice as "shouty" and "irritating", while reporting that the new voice is "friendlier, more helpful and less talkative"(Robarts, 2015). In a televised interview segment with the voiceover artist for a local news broadcast, the artist referred to the guidelines he was given in producing the voice. He discussed the company mentioning the following regard the quality of the voice (BhamUrbanNewsUK, 2015):

"warm and friendly and I think one of the phrases they used was above all human."

The artist's comments on being "above all human" suggest an effort to making the voice sound less machinelike, and changing it to appear "warm and friendly" indicates that the previous voice may have been more imposing towards its users. In speech, both the voice of the speaker and the language used can affect the identities a listener projects onto their interlocutor. We do, however, possess the ability to control parts of both[12]. This ability can be used purposefully to foster particular identities arising in our interaction partners, and this is also true of agents that have pre-programmed verbal outputs. In manipulating both in this example, Tesco appear to have the aim in creating a more positive interaction between their checkout systems and their customers, presumably with the intent of promoting subsequent future interactions and increasing their sales.

2.7 Vague Language

The example of the Tesco checkout is one way of creating a more indirect communication between agent and customer. Another way in which indirectness can be conducted is through the use of vague language (VL). VL can be used as a politeness strategy and to conduct facework, though these are not the only functions that VL is used for. However, these particular functions are a central theme within this thesis. This section will discuss some of the definitions and functions of VL, which explored further in Chapter 3 in as part of the framework of investigation for the studies in Chapter 4 and 5.

2.7.1 Defining Vague Language

Unlike some politeness research, it can often begin with a micro lexical account before its effects on the wider interaction and discourse. VL is deliberately imprecise language that is used to achieve a wide range of both functional and interpersonal goals, often simultaneously. The description of VL has origins Channell's seminal work (Channell, 1994). She describes it as having the following characteristics: it can be contrasted with another word or expression which appears to render the same proposition; it is "purposely and unabashedly vague"; its meaning arises from the "intrinsic uncertainty" referred to by Peirce (1902)[13]. This uncertainty refers to indefinite "habits of language" that a speaker uses, rather than ignorance on their part.

---

[12] Discussion of voice in particular can be found in Watt, D. (2010). The identification of the individual through speech. Language and Identities. C. Llamas and D. Watt. Edinburgh, Edinburgh University Press**:** 76-85.

[13] The reference for this as found in Channell (1994) is as follows - PIERCE, C. 1902. Vagueness. *Dictionary of Philosophy and Psychology II*., though is referred to as "Peirce" in the author's text and "Pierce" in the citation used here.

To provide some examples of VL, Channell described different categories, which are also described in Cutting (2007). These are: vague additives such as *about* and *approximately*; vagueness through lexical choice including *thing* and *whatsit*; *vagueness* through implicature such as *Sam is six feet tall (when Sam is actually six and quarter feet tall);* and *tags (or something, and so on)*. There are numerous and varied categorisations of VL, often in regards to particular contexts surrounding individual areas of research. This can be seen to taking both an evaluative approach and theoretical approach seen in Politeness$_1$ and Politeness$_2$ – describing the language being used and merging it with the general theories of VL. Trappes-Lomax (2007: 122) adopts a Politeness$_1$ type approach to the definition of VL, referring to it as:

> "any purposive choice of language designed to make the degree of accuracy, preciseness, certainty or clarity with which a referent or situation (event, state, process) is described less than it might have been."

This provides a more evaluative account of VL, but the categorisations cannot be discounted as they appear throughout various sources of literature. There are common hedging functions across the literature for example (Cutting, 2012).

There can often be little consensus when describing the terminology of VL (Cotterill, 2007). In describing the different functions of VL, authors have either adopted or sometimes modified Channell's VL framework described above, or made an effort to create a different framework. Cheng (2007) adopted Channell's aforementioned framework, as did Koester (2007). Adolphs et al. (2007) followed suit but included hedges – namely approximators (*somewhat*) and shields (*I think*). Rowland (2007) also considers the use of hedges and their subcategories, and includes plausibility shields (*I think*), attribution shields (*according to X…*), adaptors (*a little bit*) and rounders (*around*).

Cheng and Warren (2003) adopted a slightly different framework in their description of "vagueness". They differentiated between "vagueness additives to numbers" (*around, about*), "vagueness by choice of vague words" (*and things like that*), and "vagueness by scalar implicature" (*a little bit, some*). Wang (2005) also approached their framework differently, which can be seen below as cited from Cotterill (2007: p.99):

- 'Impression' indicators: vague quantifiers ('a lot', 'many') and approximators ('approximately', 'about', 'roughly')
- 'Unspecificity' indicators: ('after 10 o'clock', 'at six-ish')
- 'Fuzziness' indicators: ('sort of', 'kind of')
- 'Etcetera' indicators: additives ('and so', 'and things like that')
- 'Uncertainty' indicators: vague adverbs ('maybe', 'probably')

Cotterill (ibid.) herself used this guide to focus primarily on fuzziness and etcetera "markers".

There are clearly differences in the approaches to describing and categorising VL, and a consensus is rarely established. However, despite this lack of consensus, many similarities can be seem when comparing these different approaches and frameworks. Often it is the vague lexis being discussed that actually remains the same, albeit under different guises. Channell's (1994) tags, for example, are similar to Wang's (2005) additives in 'etcetera' indicators. This is not necessarily a negative thing, as scholars have the right to categorise VL in a means that is sufficient for their research purposes. Outlining and defining the VL framework sufficiently is more important than achieving a consensus in nomenclature. The subsequent third Chapter is another example of a bespoke model being created in addressing specific research goals.

2.7.2 Contexts and Functions of Vague Language

VL has been reported to appear in a wide variety of contexts. Rowland (2007) reports the use of VL in a mathematics classroom. For example, when a student is answering a mathematics questions in classroom and responds with, "but it's around 50 basically?" In this example, the speaker conducts the functional goal of answering a question given by a teacher, while also fulfilling the relational goal of protecting oneself from full commitment to the answer and potential error by being imprecise using "around" and "basically", thus saving face. Adolphs et al. (2007) report the use of VL in healthcare contexts, for example when reducing the markedness of particular phrases and reducing the distress patients may be exposed to e.g. "meningitis type symptoms" where *meningitis* is reduced by the following word *type* to offer up further possible conditions other than meningitis alone. The authors also reported the use of VL in chaplaincy-patient interactions, where the chaplain is a professional who can offer spiritual and pastoral care for patients in hospitals. They may discuss a variety of matters with patients – medical, social, religious, and emotional, for example. Chaplains in these contexts can use VL to elicit patient discourse while maintaining rapport and reducing perceived face threats.

Koester (2007) observed the use of VL in North American and UK offices. Part of her research revealed the use of VL in interactions where there is an imbalance of power, for example with a manager and employee or a customer and supplier, where the risk of conducting face-threatening acts is higher. With customer-supplier interactions and service encounters in particular, the risk of threatening a customer's face can also threaten potential business and sales performance. The vague tag, "an' things," for example, was used and

Koester argued this contributes the informality of a customer-supplier interaction (2007: p.50). This type of interaction is similar to that between the automated self-checkout machine and customer discussed in 2.6.2. The use of VL and "warmer" language allows Tesco to conduct a hopefully less face-threatening encounter, which will in turn be profitable for the company through future interactions.

VL can also appear as tension management devices in academic conferences (Trappes-Lomax, 2007), by using minimisers such as "just" or approximators such as "partly" to downplay research findings. Each chapter in Cutting's (2007) edited book on VL highlights a different area of interaction where VL is present. It is likely that there are few contexts in which some forms of VL do not exist, as it is a very common feature of language, particularly in speech[14] (Channell, 1994, Cheng and O'Keeffe, 2014) which has long been known to contain numerous accounts of non-specific language (Brown and Yule, 1983). Humans also have no trouble in decoding VL and its various meanings in context, allowing for its frequent use. Again, the mention of context here is important as, in a similar vein Watts' (2003) definitions of $Politeness_1$ it is not inherently good or bad, but dependent on the context in which it occurs. The speakers in context of an interaction will decide progressively if it is appropriate or inappropriate. In HCI contexts and verbal instruction giving in particular, one may expect there to be a more direct approach to language use. Given that interactions with verbal agents are a relatively new context, there is very little evidence of VL being used or studied. It may run counter to our expectations of instructional agents being focused on transactional rather than interpersonal goals. It is unknown whether in these contexts these two goals can be brought together using VL. This is one of the research gaps this thesis aims to address.

2.7.3 General Functions of Vague Language

As highlighted in 2.7.2, speakers can use VL in a variety of specific contexts to achieve specific goals. Generally, VL can be used to perform a wide number of different functions. This section describes some of the more general functions of VL with examples. These descriptions are mostly drawn from both Channell (1994) and Cutting (2007). They both describe a wide array of these functions in their respective work. VL may be used simply because a speaker is able to be precise enough with the vague lexical items they use as the context requires. This allows a speaker to be efficient, for example when using vague nouns or placeholders such as *thing* instead of the full referent to the noun or noun phrase, which may not only be irrelevant but also repetitive and time consuming. This shortening of nouns also provides greater clarity for a speaker, especially when engaging in a lengthy period of talking.

[14] As opposed to other modalities of communication such as written and textual forms.

Koester (2007) argues the speed and accuracy that VL may provide is worthy of future research. It could be argued that this would be of particular interest in areas of talk that are limited to short amounts of time, or linked to wider social or functional goals that are similarly limited in length.

These same vague nouns may be used for a genuine lapse in lexis (Channell, 1994) For example, *thing* may be used perhaps either to replace a forgotten word, or substitute something that the speaker believes the listener may not understand or that the speaker decides is not worth mentioning. This may, however, along with other variations of VL, result in a miscommunication between both parties (Cutting, 2007). The need for a mutual and assumed common ground between interlocutors may also require more processing for the listener and as such the meaning may not also be derived as the speaker intended (Jucker et al., 2003). Again, the decision to use VL is thus something for the speakers to decide. If there is a desire to communicative effectively then some contexts that individuals may deem more appropriate than others.

Sometimes there may be a desire, however, for a speaker to use VL as a means of deliberately miscommunicating or omitting information from others. Although VL can be used to claim in-group membership, it may also be used as a tool to assert or as a means of excluding others from identifying meaning in an interaction (Cutting, 2007). This could include non-L1 English speakers (e.g. L2, L3), for example, as the use of implicature and assumption rather than explicitness may create communicative barriers. Other social groups in which certain VL items are unfamiliar to some of the participants may also feature in this function. This may or may not be intentional, and in some cases may be impolite, depending on the context. There may also be the need to withhold certain information from listeners depending on the audience. Vague nouns such as *thing* or *stuff* can thus be useful to speakers if this is the case.

When using VL it to claim in-group membership, however, the modification or omission of words can work towards "reducing social distance" and establish "interpersonal solidarity" (Terraschke and Holmes, 2007). It can also correlate with and indicate intimacy between interlocutors (Cutting, 2007). VL can therefore be used to both conduct facework with strangers through the use of items such as hedges (*just, you know*) and also be an indicator of a close relationship through interlocutor reliance on common ground (McCarthy and Carter, 2006).

VL has other social functions. Some of these often tend to lean towards the polite end of the (im)politeness spectrum (Channell, 1994)[15], Expressions can be softened to reduce the imposition on a listener, as well as reduce the impression of a speaker being too authoritative or direct (McCarthy and Carter, 2006). Channell (1994) also offers the example of giving a listener a choice. This can be achieved through the use of exemplar + tag phrases e.g. "*would you like a drink or something?*" with the tag *or something* providing a listener with the option of an alternative to a drink. Even if the speaker has no intention of providing an alternative, they can still balance and level the power in an interaction by giving the linguistic choice (Carter, 1998). VL is also discussed as "a marker of social cohesion" (Cutting, 2007) and an interactional strategy (Jucker et al., 2003) that helps contribute towards making a conversation natural[16]. Making a conversation more natural is but one of the functions that can make VL a worthwhile endeavour of exploration in human-agent interaction. VL, then, can serve many different functions in speech. While the primary focus of VL in this thesis concerns mitigation and reducing imposition on face, there are many functions that fall outside of these categories. This is similar to the discussion by Trappes-Lomax (2007: p.123), who dictated that, "some but not all VL has avoidance (defensive/protective) purposes, and some but not all avoidance behaviour is expressed through VL." Similarly, not all hedges are necessarily vague, and not all VL types are necessarily hedges. These are agreed upon in-interaction and in context.

With regards to natural language, verbal agents are already using natural language and will likely continue to do so, and in greater numbers. Given VL's frequency in speech, it may become more common as part of an agent's linguistic repertoire. There are research gaps regarding how VL is perceived by an agent's users, which this thesis aims to address. Similarly, the impact of an agent's voice on its use of VL will also be explored. Deciding which aspects of VL to use and how to apply it to a HAI context for investigation is the focus of the next chapter.

2.8 Summary of Literature

This chapter provided discussion on the increasing prevalence of agents in our modern world, and how our interactions with them are driven by the same social rules that underpin human interaction. The

[15] Channell (1994) refers to politeness as a function of VL, rather than discussing the spectrum of (im)politeness explicitly.

[16] Cutting (2007) here references McCarthy (1998) in the discussion of naturalness in conversation. Both of these references are included in the bibliography. The citation for Jucker et al. discusses the notion of vague language being an interactional strategy.

concept of identity was also discussed, first in relation to human interaction and language, followed by identity in human-agent interaction (HAI). The relation between identity and linguistic variables in an agent, such as prosody, voice, and language was also covered. This chapter also reflected on the linguistic concepts of politeness theory, face, and vague language (VL). VL in particular was defined and its functions explained. Politeness theory in previous HAI research was also discussed.

It is emphasised in this chapter that there are gaps on the understanding of how VL use in a verbal agent instructor affects its users. Similarly, there are gaps in the methods used in studies on politeness in HAI that this thesis aims to improve upon in looking at VL, which are addressed in Chapters 3, 4 and 5. Gaps also exist regarding the effect an agent's voice has on the identities users create for it. Chapters 4 and 5 touch upon the issue of identity in light of the study findings, while Chapter 6 discusses this concept in further detail.

3. Assessing Vague Language in Human-Agent Interaction: Creating a Framework

3.1 Introduction

Having covered some of the literature on vague language (VL) more directly in 2.5, this chapter presents a framework for applying and assessing VL use in a human-agent interaction context. In this context, the agents take on the role of a verbal instructor. In applying VL to the verbal agent instructors, first the initial design of an instruction based task using the assembly of Lego models is discussed, followed by the creation of appropriate verbal instructions for these assembly tasks. The creation of a VL model is then presented. This builds upon the definitions and functions of VL discussed in Chapter 2, while specifying the relevant categories of VL that underpins the rest of this thesis, including examples of the model being applied to the task instructions. In the final sections of this chapter, the general design approaches towards the agents and the task interactions are discussed.

3.2 Initial Task Design

This thesis is concerned with investigating the variables of VL and voice used by verbal agent instructors, as opposed to other genre types. As a result, the tasks chosen for the studies in Chapters 4 and 5 followed a similar approach to existing research. The two studies in this thesis follow similar themes of recent human-robot interaction (HRI) literature, in using an instruction-based task to assess the effect of language use in interactions with machines (Torrey, 2009, Torrey et al., 2013, Strait et al., 2014). As discussed in 2.4.2, Torrey et al. (2013), for example, looked at the effects that polite advice giving robots had on people observing such interactions, with the task at hand being cupcake making. Similarly, Strait et al. (2014) analysed observations and direct interactions in drawing tasks. This paper emphasised the need for researching direct interactions with robots, rather than researching observers' opinions of an interaction. This direct interaction approach is used here, and the design of the agent and the two studies in this thesis focuses on having a co-located interaction with a verbal agent interface.

In a similar setting to the advice giving tasks discussed in the previous paragraph, the studies here use Lego model assembly tasks. This type of task has a precedent in human interaction research, for example in investigating collaborative assembly and investigating participants speaking whilst monitoring others for understanding (Clark and Krych, 2004). The Lego in this study was used as they could not be described as "familiar objects such as bridges, animals, or buildings." The Lego models used in this study are similar, in that many pieces are not familiar, and describing them is often difficult and lacks

consensus[17]. This allows for the liberal use and testing of vague nouns as discussed later in 3.1.4. Lego is also not overly complex, with the age range specified as 6-12 years[18]. Focusing on the use of verbal instructions, rather than the traditional Lego pictorial instruction booklets, however, made the tasks more difficult but still achievable. Pre-task discussions indicated that although most participants had some familiarity with Lego in general, the specific Hero Factory[19] models being used here were largely unfamiliar. Three models from this range were chosen – "*Aquagon*," "*Nex*,' and "*Stringer*". These are all similar types of model that have a humanoid form and a similar number of pieces[20]. All three were used in Study One, while only the first two were used in Study Two.

3.3 Instruction Design

Having established the models that were used, the instructions were subsequently designed for the agent to verbalise. The non-vague instructions were designed first, as the vague instructions would be a modification of them, as is discussed further in this chapter. As Lego models come equipped with pictorial instruction booklets, all instructions had to be written without any assistance from the manufacturer on naming conventions for any pieces or assembly processes. Only four sets of instructions were required in total (non-vague and vague for *Aquagon* and *Nex*), as the model *Stringer* was only used as a practice task in Study One, using the original pictorial booklet for the instructions (see 4.3).

The manufacturer's booklets were followed for the other two models to establish an appropriate assembly order. Although no words are included in these booklets, there were three distinct categories in each step of the assembly process, which can be seen below in Figure 3.

[17] This was identified initially during preliminary testing prior to this study, as well as in some of the post-task interviews in both studies.

[18] As seen for one of the models, "*Aquagon*", for example, in the following link http://shop.lego.com/en-US/AQUAGON-44013?p=44013

[19] This Lego range has since been discontinued.

[20] *Aquagon* has 41 pieces, *Nex* 39 pieces, and Stringer 42 pieces.

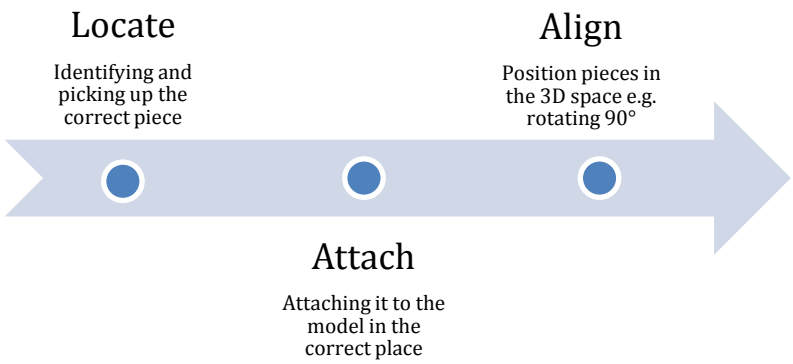

Figure 3: The process of a step in assembly of Lego models.

Repeating this process for every piece was seen to be potentially tedious, and consequently some steps in the instructions contained multiple stages of the process. A shorter step, such as the example below from the *Nex* instructions, may contain one instance of stage in the process:

> *Step 11: Locate the largest black piece with seven ball joints. This is the body.*

Step 11 only involves the location stage of the process, while also providing information as to what the piece is. The description of the body piece was used in all sets of instructions so participants would find the task somewhat easier, and in turn be exposed to more of each agent's speech. Other steps, such as in the example below from the *Aquagon* instructions, contained multiple instances of one stage in the process, in this case three instances of the location stage:

> *Step 14: Take a black cylinder, a grey cylinder and a small light grey piece with a curved fin.*

Another example from the *Aquagon* instructions provides an example of two different stages being used within a single step, with Step 31 includes instances of both the attachment and alignment stages.

> *Step 31: Attach the yellow socket to the top black joint of the body so the piece points forwards and is in line with the feet.*

Highlighting these examples shows that the steps in the instructions were varied and included various iterations of the three stages in the assembly process. This was to counteract some of the predictably that the agent may have otherwise projected, again with a view to maintaining a greater focus on the agent's speech.

Instructions were also designed so that pieces were described in simple terms, with the aim that participants would be able to understand what pieces are being referred to. There were several pieces in each model that were similar to one another, and included some of the same descriptive phrases. For example there were many

mentions of *socket, ball joint* and *cylinder*. The latter is a common three-dimensional shape, and the first two used are in both anatomical descriptions (such as the bones of the shoulder) and mechanical joints. Other pieces proved more difficult to describe, such as any references to pieces of *armour*. Nevertheless words such as *armour, spikes* and *fins* were used in attempts to communicate the instructions effectively.

In the first iteration of the instructions for Study One, there were forty-seven steps for *Aquagon* and forty-eight for *Nex*. While the difference in one instruction was not observed to have a significant impact on the results, these were altered slightly in Study Two so that both models had forty-seven steps in total. While there were more individual steps than pieces, this was due to some of the steps not including a stage of locating a piece. As they also had to be attached and aligned, the steps were spread in to avoid overloading participants with too much information at once, and detracting from the agent's speech. Moreover, this prevented participants from proceeding too quickly in either of their verbal tasks.

3.4 Creating a Model of Vague Language

Once the non-vague instructions had been completed, there was a need to design a model of VL that could be applied to create vague instructions, which is discussed in detail in this section. Before this, two experimental sessions were undertaken with two participants, who both instructed the researcher on how to construct one of the Lego models in their own individual sessions. This was in order to get a better understanding of what types of VL might be used in this context and in what position of individual steps they appear. It also provided an opportunity to compare how other people verbalised the visual information with the non-vague instructions. One participant was a 24-year-old female while the other was a 25-year-old male. Both participants were L1 English speakers. Each participant was provided with the manufacturer's booklet of one of the Lego models and asked to verbalise the visual instructions to the researcher. The researcher then assembled the model under the guidance of the participants until they were satisfied that the model was complete. This process took approximately 12 minutes for each participant, and the audio of each interaction was recorded on a MacBook Pro 10.2. No monetary incentive was provided to participant, though they were expressed with extreme gratitude.

The audio recordings of these interactions were observed multiple times. Despite the small sample size, they revealed several patterns of language use. Discourse markers (*so, now*) were commonly used when starting a new stage of location or attachment in the assembly process, and were introduced at the initial stages of utterances when included in the participants' speech. Hedges and minimisers such as "twist it a bit" and "just sort of attach it" were also frequently used. So too were vague nouns, particularly "thing" for describing various pieces, often

with a reference to their size, shape, and colour. In almost every instruction, some form of VL was used.

These preliminary investigations gave some guidance on the type of vague categories that would and would not be relevant in an instruction-giving context. The research context is one of the fundamental concepts in this chapter. As discussed in 2.7.1, previous research provides examples of existing categorisations of VL and these are often based on data from human interaction contexts[21]. Cutting (2007: 5) discusses Channell's (1994) VL categories as follows: vague additives (*around ten*); vague implicature and vague quantifiers (*15,000 died*); vague placeholders (*thing, whatsit*) and tags (*or something, and so on*). However, despite her seminal work, her approach to categorising VL is not always emulated. This does not come necessarily as a criticism of Channell's work, rather out of necessity of the research context and the data gathered by the researchers from their respective investigations. The differences in labelling aren't necessarily important so much as long as the framework for the context is set out clearly, which is the goal this section – to describe the VL model in the context of a verbal agent instructing users on assembling Lego models. As the functions of facework and conducting politeness are a common theme in this thesis, the selection of the lexis in the VL model is focused on attenuation and mitigation of instructions. This can be seen particularly in the first three categories described in the subsequent paragraphs.

3.4.1 Hedges: Adaptors

Hedges are lexical items that alter the truth condition of a statement by attributing "fuzziness" to it i.e. utterances are made less definite and precise (Lakoff, 1973) and (Zhang, 1998). They have different functions depending on the type of hedge being used. Prince *et al.* (Prince et al., 1982) describe two categories of hedges: *shields* and *approximators*. Shields themselves are divided into *plausibility shields* and *attribution shields*. The first are phrases that a speaker uses to declare a degree of uncertainty to a statement they are making (e.g. *I think, possibly, as far as I can tell*). The second is a shield that attributes responsibility of a statement to someone or something other than the speaker (*it has been said that, according to X*) (Fraser, 2010). Speakers pass the buck of the blame as it were in order to distance themselves from the statement.

[21] Further insight into the varied categorisation of vague language can be found extensively in the collection of edited contributions in J. CUTTING, *Vague Language Explored*, Palgrave Macmillan, 2007.

Approximators, the other class of hedges, are also subdivided into two categories: *rounders* and *adaptors*. Rounders provide estimations, usually of measurement, and convey a range of values (*approximately fifty metres, about here, around half past ten*). Adaptors create imprecision through the reduction of class membership (*somewhat, sort of, kind of, a little bit*) as opposed to using a definite alternative. An example of this can be seen in Figure 4 below.

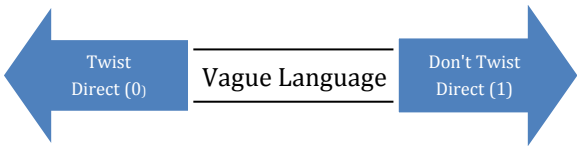

Figure 4: VL can occupy the fuzzy space between direct alternatives.

In the context of the Lego assembly instructions used in this study the directive to *twist* a component appears several times. For the sake of clarity take the example "give this piece a little bit of a twist". Though twist may arguably be a vague term in itself, it implies some form of rotation of an object from its current position in 3-D space. The adaptor phrase *a little bit* modifies the class membership of the verb twist. The benefit of this is twofold. Firstly, the adaptor helps to mitigate the impact of the instruction, which may otherwise appear assertive and abrupt. Secondly, it opens up a wider set of possibilities for the listener that exist between the direct alternatives which has shown to be of benefit for the user for their peace of mind and agent for its perception (Clark et al., 2014).

Of the categories of hedges discussed thus far, it is adaptors that are included in the VL model. Rounders are excluded from the VL model, as are both classes of shields. Rounders provide estimations of measurements and value, and it is believed their inclusion in the agents' instructions may project a lack of expertise and possibly make the task more confusing or difficult than is intended. Similarly, the use of plausibility shields, such as "*I think,*" or attribution shields, such as "*It has been said that*…" may also project a lack of expertise in the agents. A lack of expertise may lead to participants ignoring the agent's instructions or potentially distract them. The method described in this chapter aims to allow participants to follow the agent's instructions and go some way to finishing the tasks, if not completing them outright.

Having an agent instructor guessing the amount of pieces to pick up for example does not suggest competency in its ability to collaborate

with a human in an assembly-building context. It is assumed users will have expectations regarding the expertise of the agent using shields; however this is not a focal point of this study.

3.4.2 Discourse Markers

Discourse markers are words or phrases that function primarily as a structuring unit of spoken language (Fraser, 1990, Jucker et al., 2003) and despite containing no grammatical information are common in natural speech (Laserna et al., 2014). Examples of discourse markers include *now, well, so* and *actually*. Structurally, they can be used as a bridge from one section of information to another, as well as to indicate a change in topic. The ability to structure different turns of information already makes them ideal for an agent delivering assembly instructions, where there are various stages and sub-stages of building involved. In a humanoid Lego model, for example, the lower body may be seen as one stage, with the feet a sub-stage leading up to it. Using discourse markers can help group these together and alert users to a shift in the assembly process. While part of the VL model, the phrase *so* does also appear in the non-vague instructions as well, as a bridge between the attachment and alignment stages in the assembly process. In the non-vague instructions, these only appear as conjunctions between these stages, such as in the step below from *Nex*:

> *Step 38: Attach the grey piece to the orange holes so the wider end is closest to the head.*

Discourse markers are not a feature usually discussed in VL, though Quaglio (2009) does refer to the discourse markers "I mean" and "you know" in a discussion of VL in television dialogue and normal conversation. However, structuring talk is not their only feature. They can also operate as a hedging device by reducing markedness of phrases that may have an effect on a listener, indicates loose or non-literal utterances and lessen the assertiveness of a speech act (Adolphs et al., 2007, Andersen, 1998, Fleischman and Yaguello, 2004). Drave (2001: p.27) also discusses some qualities that support the use of discourse markers here. He points out that VL helps in "maintaining an atmosphere of friendliness, informality or deference", "emphasising (and de-emphasising certain information)," and contributing "to the overall tenor of the conversation" in achieving interpersonal goals. De-emphasising certain information is similar to the markedness-reducing effects discussed earlier in the paragraph. By including discourse markers at the beginning of instructions (e.g. *take the body piece vs. so take the body piece*), the aim is to reduce the focus the participants have on the imperatives, which can then in turn reduce the perceived assertiveness of the instruction, and contribute to the face-protecting interpersonal goals of the interaction.
Given the qualities of discourse markers and VL listed above, those used in the VL model can be said to having hedge-like effects and

functioning almost like a hedge does. However, although they have potential attenuating qualities, they also retain the function of structuring the instructions from one segment to another. In this way they are different from the adaptors and minimisers – they have a structuring function as well as an attenuation function.

3.4.3 Minimisers

Although discourse markers can have hedge-like effects (Jucker and Ziv, 1998) and be associated with informality (Brinton, 1996), there are some that do this more than others and blur the lines between themselves and hedges. These are described here as minimisers, a term borrowed partly from describing the use of 'just' as a tension management device in academic presentations (Trappes-Lomax, 2007). In this thesis they take three different forms – *like, basically* and *just*. While discourse markers such as *so* and *now* operate primarily at the beginning of information structures, minimisers appear both at the beginning and mid-sentence. Not only this, they also seek to reduce the simultaneously reduce the assertiveness of an instruction whilst lessening the apparently difficulty of the task associated with that instruction. Take the following examples from the Lego instructions:

> *Step 12: Now locate the largest black piece that has seven ball joints. This is the body.*
>
> *Step 37: So now locate the yellow face piece.*

Here, *now* and *so now* are operating at the start of a new turn. Step 12 introduces the body and Step 37 the face. Compare these to the next examples:

> *Step 13: Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs.*
>
> *Step 22: Just connect the yellow joints to the socket of each fist.*

In Step 13, the wording is such that the task of finding the narrow end and attaching the ball joints to the leg sockets are not challenging, showing belief in the user's capabilities and minimising the imposition of the instruction. Step 22 is similar and places the subsequent phrases in more positive light. Let us analyse one final example to highlight the differences between these categories:

> *Step 9: So keep these black pieces vertical and just twist each one a little bit 90 degrees or so to the right. These are the legs.*

This instruction opens with the discourse marker *so* but also includes the minimiser *just* in the middle of the sentence. If these were to be interchanged it would look like this:

> *Step 9: Just keep these black pieces vertical and so twist each one a little bit 90 degrees or so to the right. These are the legs.*

This orientation lacks the mid-sentence flow provided by *just*, yet there is nothing out of place about it being used at the beginning of a sentence. Minimisers can be understood as discourse markers that can function strongly as hedges and vice versa, whereas there are hedges and discourse markers that do share this ability, at least not to the same degree.

3.4.4 Vague Nouns

The finally category of the VL model is vague nouns. Vague nouns substitute the full description of a noun with a concise alternative, which include phrases such as *thing*, *thingamy* or *whatsit*. They may also be referred to as general or dummy nouns (Halliday and Hasan, 2014), as well as placeholders (Channell, 1994), though the latter my refer to nouns that contain vague lexemes such as *thingummy*.

Words such as *piece*, *end* and *thing* that are used in the instructions can be said to be vague nouns in that they operate in a similar manner to *thing* in human interaction. While *piece* and *thing* can be used to one of the constituent parts of the Lego models, *thing* may arguably represent a more open set of potential nouns. It may be the case that most participants equally attribute them both to parts of the model, however it is believed *piece* is better attributed to the parts of a model assembly. Because of this, it is included in the majority of steps throughout the instructions, whereas *thing* is restricted[22]. *End* is also used in each model's instructions. This noun may not be as vague as the others, as *end* refers to an extremity of one of the constituent model parts (e.g. two points the greatest length from one another on a cylinder). Nevertheless, in isolation without additional language describing the *end* it is perhaps just as vague. For the purposes of this research, the abovementioned nouns are considered vague nouns, in a similar vein to Channell's (1994) vague category identifiers that refer to a set of things that share characteristics and equivalency[23].

---

[22] *Thing* appears twice in *Aquagon* and once in *Nex*. The discussion of the differences between the sets of instructions for the two models are discussed further in this chapter.

[23] In sharing a collective characteristic, the vague nouns are all constituent parts of the model. For *end* specifically, the shared characteristic amongst *ends* is the referent to an extremity of one of these parts.

As well as contributing towards rapport management and facework, vague nouns also serve a strong functional purpose in the instructions. Rather than having to refer to potentially unfamiliar model parts by repeating full noun phrases, they can be replaced by vague nouns such as *thing* or *piece*. This prevents the interaction from quickly becoming tedious, which was observed when assembly instructions were designed with too many diagrams (Agrawala et al., 2003). Similarly, in HCI, it may be the case that having more information than is required can be undesirable (Niculescu, 2011). Furthermore, it is another measure to ensure maximum potential exposure of the agent to the participants.

One point of distinction for vague nouns is that they are also included in the non-vague instructions, rather than exclusively used in the vague instructions. One of the reasons behind this design choice is the functional nature of vague nouns in this context. The other categories have a stronger relational purpose, whereas in for Lego model assembly vague nouns are used to describe unfamiliar model parts. This does not discount their relational purposes, however. Attempts to not use vague nouns in the non-vague instructions also created a strong imbalance in the amount of language that either agent was speaking, and resulted in a lot of repetition. Furthermore, testing the inclusion vs. exclusion of vague nouns goes beyond the focus of this thesis. Despite the inclusion of vague nouns in both sets of instructions, non-vague in this thesis refers to those instructions that contain vague nouns, but no other lexical items from the other categories in the VL model[24]. A summary of the VL model can be seen in Table 2.

[24] This is apart from *so,* used as a conjunction (3.3.2).

Table 2: The vague language (VL) model.

Category	VL items used	Function	Example in context
Adaptors	*more or less; a little bit; sort of; a bit; a little; pretty much; or so; somewhat*	Hedging instructions; reduce assertiveness; minimise imposition; reduce face threats	So keep these black pieces vertical and just twist each one a <u>little bit of a twist</u> 90 degrees <u>or so</u> to the right
Discourse Markers	*so; now*	Structure new turns at talk; some hedging (see above)	<u>Now</u> pick up the two black pieces with ball joints
Minimisers	*just; like; basically*	Structure talk; hedging; reduce perceived task difficulty	<u>Basically</u>, find the end that is a bit more narrow than the other one and <u>just</u> attach the side ball joints to the sockets on the legs
Vague Nouns	*thing; piece; end*	Improve language efficiency	Just place the big spikes into the holes that are closest to the edge of each <u>piece</u>

This table shows the VL lexis being used, their intended functions, and an example of the lexis in context of the instructions. 3.4.1 identifies some of the VL categories that are not being used in the instructions, namely rounders, plausibility shields, and attribution shields. There are also other categories of VL that are not presented in this model. Tags (*or anything*), or "etcetera indicators" as Wang (2005) describes them, are not included in this VL model. While useful to hedge utterances, for example in healthcare discourse (Adolphs et al., 2007), there is a similar concern regarding the lack of expertise these may project. This is similar to the three other types of hedges excluding adaptors. This is also true for what Wang (2005) calls "uncertainty indicators" (*maybe, probably*).

3.5 Applying the VL Model and Refining Assembly Instructions

The vague instructions were designed so that each step would contain at least one vague item, in a similar manner to the preliminary human instructor tests discussed at the beginning of this section. Longer steps within the instructions included more vague items. This was a deliberate attempt to include enough VL so that it would be noticeable in the instructions[25] though this approach admittedly leaves room for oversaturation. Noticeable in this context refers to participants being aware of the VL used in the instructions, to the extent that they are able to comment on it during the interviews. This is essential in understanding VL use in this specific HCI context, and using too few vague items may hinder this process, and result in little to no comments on the VL. Having comments at this stage is a positive, as it provides a basis from which similar research can feed off. In this thesis alone, the comments of the first study provided the foundation for the second. Negative, positive, and neutral responses from participants all provide useful information on how to progress on the VL use in verbal agent instructors, regardless of the ratio in which they appear.

Once the instructions were written, they were tested on three individuals to assess whether they could understand the instructions. This was done without an agent interface, and involved the researcher verbally instructing the pilot participants. This approach provided valuable input towards the design of the agent interface.

It was observed that following the instruction process that is outlined in the visual instructions, but with the verbal instructions, was not well received and often confusing. This was a result of the instruction booklet switching between various sections of the model, such as the head and the legs, rather than focusing on one at a time. Following feedback from these three pilot participants, it was decided that a bottom-up approach for the assembly was a better alternative i.e. from the feet upwards to the head. Although the task is designed to be somewhat of a challenge, one of the main aims is to provide participants with the change to interact with the agent as much as possible, rather than have them unable to complete the tasks.

Further feedback provided suggestions to alter some of the phrases used to describe certain pieces. Of particular importance was the need to more clearly differentiate between some of them. Some pieces were referred to as a *cylinder* for example, whereas feedback suggested longer pieces may be best called a *tube*, while keeping the smaller ones as *cylinders*. Further detail in describing the colours of pieces was also requested, in regards to the light and dark shades of grey that were present in the *Aquagon* model. The vague and non-vague instructions were both changed with the feedback in mind before being used in

[25] A full account of the instructions can be seen in Appendix A.

further experiments.

Following the development of the VL model, there was a need to design a methodology for its application in a human-agent interaction context. This section addresses the general approach to the design of the studies discussed in Chapter 4 and 5. The specific methodologies tailored to each study are discussed in their respective chapters.

3.5.1 Designing the Agent

Following the creation of the VL model and the instructions, a verbal agent instructor was designed. In doing so, it did not seem necessary to create an agent in the sense that it had the features of autonomy or intelligence in the traditional definition of the term (Wooldridge and Jennings, 1995). Instead it seemed an easier and more viable approach to design a simpler interface that would provide instructions from a library of pre-recorded sound files. To this end, this agent is perhaps better described as a *virtual* or *simulated agent* in that it does not possess abilities of agency as such, but aims to create a believable agent-like interaction[26]. This is similar to the notion of *intersubjectivity* (Cassell and Tartaro, 2007). This refers to creating a believable interaction as opposed to a believable life-like agent, and moving the benchmark of interaction towards something that creates a sense of sameness in HAI that appears in human counterparts. Cassell and Tartaro (2007) describe this sameness as humans displaying the same conscious and unconscious behaviour in reactions towards agents as they would other humans, much like the theories of the CASA paradigm and Media Equation (see 2.1).

The interfaces that provided the instructions to participants were designed to be minimalistic so as to keep the focus on its spoken information. Each interface consisted of an HTML file first, which was written in Java and created in the Eclipse program (https://eclipse.org). An example of one of the interfaces using the Lego model *Aquagon* can be seen in Figure 5. Each interface was linked to its own library of sound files corresponding to the specific voice and Lego model. These were numbered individually and contained within a subfolder called “sounds”. This subfolder and the interface file were contained within one folder. Further details regarding the voices and how they were implemented into the interfaces are discussed in specific methods in Chapters 4 and 5. The interface itself could be opened in any Internet browser, though for consistency Google Chrome was used throughout this thesis.

[26] For the sake of clarity and consistency, they are described as agent, interface, or agent interface throughout the thesis, rather than *virtual* or *simulated* agents.

Four different types of information were displayed to the participants. Two of these were non-interactive. First at the top of the interface was the name of the current model they were assembling – either *Aquagon* or Nex. Below this the specific step[27] they were on was shown and the relation to the current step and the total number of steps e.g. "Step 4 of 47". Below these were the two buttons that participants could actively interact with. The first was *Next Instruction* (or *Start* for the initial instructions) that moved participants onto the next step of the instructions. The second was *Repeat (or Finish for the last step)*, which repeated the most recently played instruction. When one of the steps was being played, either for the first time or repeated, the interactive buttons were greyed out. This approach ensured participants could not play more than one step at a time, nor they could skip any instructions.

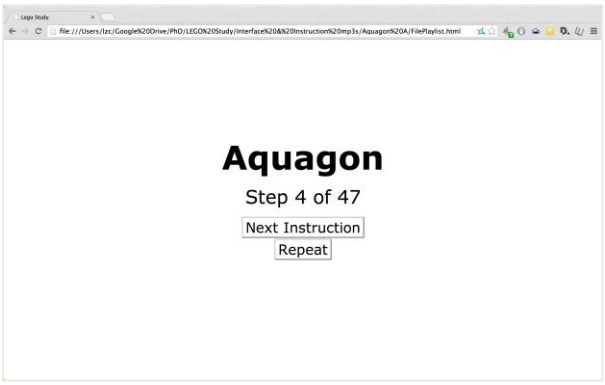

Figure 5: One of the interfaces for the model *Aquagon* in Study One.

There were other parts of the interface hidden from the participants. Below the visible information several logs were kept for tracking information regarding each task. These logs recorded the time taken for each step of the task, the number of times participants requested repeats, and the specific steps that had been repeated.

3.5.2 Designing the Interactions

Following the completion of the instructions and the agent interface, the application of them into an actual interaction was designed. In doing so, the approach was to emulate the direct interactions used in some of the interactions seen in Strait et al. (2014), as opposed to observations of interactions in Torrey et al. (2013). Given that direct interactions with agents are increasingly common, opting for observations did not present itself as a preferable alternative. The direct interactions in this thesis refer to participants interacting with the agent interface via a laptop (MacBook Pro 10.2), where there tactile input provides verbal agent output. Despite there being a lack of

[27] Both "step" and "instruction" are used to describe the forty-seven different stages of the model assembly given by the agent, and are essentially synonymous.

a two-way dynamic interaction in terms of speech, with only the agent interface speaking, the method outlined in this chapter still provides an interaction. While the participants are not afforded the ability to communicate verbally to the agent, they provide their own input through their tactile manipulation of the interface and are reciprocated with information. Participants request instructions from the agent through their own volition and decision making, and in turn are provided with the next step or a repeated instruction. After the agent presents the requested instruction, the participants then act upon them, as they deem fit, and then subsequently engage with the model assembly or the interface again.

There were other approaches to the interaction design that aimed to address some of the drawbacks of those used in the abovementioned studies. The duration of interactions, for example, are not made in Strait et al. (2014). While they are not included in Torrey, et al. (2013), previous postgraduate research by Torrey (2009) indicates that the length of interactions that participants observed was approximately three minutes, and contained five steps of a baking recipe. Strait, et al. (2014) refer to the pacing of the task being set by the cues given by the participants, which is also the case here, though without any explicit indication of the time allowed for participants to complete the task. While Torrey et al. (2009) describes the time and number of steps in each of the "video vignettes", the use of only five steps in three minutes appears somewhat short.

Addressing the time explicitly may benefit future researchers for replicating studies, as well as frame the interaction in greater detail. Similarly, there may be differences when researching the effect of agent instructors and robot helpers over longer periods of time within a single interaction. The approach used in this thesis uses two different duration types. These are discussed in further detail in the following two chapters, like the agent's voices.. In summary, Study One included both tasks in which there was no time-limit, followed by those in which there were time-limits. Study Two consisted of tasks with a consistent time limit throughout every session. Participants could complete the task by either reaching the time limit, or by completing all of the required steps.

3.5.3 Data Collection

With regards to the data collection, it was found that the studies referenced in 3.4.2 could benefit from a more rounded mixed methods approach to analysis. The use of post-task questionnaires, rating specific characteristics of the agents, is continued in this thesis in attempts to replicate results, though this is also supplemented with thorough qualitative analyses. Following each task in both studies, semi-guided interviews are introduced in order to generate a richer understanding of the participants' perceptions of the agent, with a

view also to bolstering and giving explanation to the quantitative measures. This post-task data collection approach allowed for people to reflect on their interactions with the agents and then describe it in detail relevant to the research questions. Survey measures have been used in similar studies involving polite robots (Torrey et al., 2013; Strait et al., 2014), though often lacked a substantive qualitative approach. This is remedied by created the mixed methods approach described above, and provides the opportunity for a grounded theory approach in the analysis of the qualitative data.

In regards to alternative approaches to linguistic analysis, many of these were excluded as an option for several reasons. Conversation analysis, interactional sociolinguistics, exchange theory, and even the cooperative principle were all excluded here. While these approaches are well defined in linguistics, they often focus more on a two-way spoken interaction, as are the studies of service encounters, for example. Although the participants are still engaging in an interaction with the agents in the studies in Chapters 4 and 5, the speech is only coming from the agent as the participant saying anything towards the agent will not cause it to react, perform an action, or otherwise adapt. As such, it is perhaps not the same type of interaction that is typically associated with these linguistic approaches. Essentially, this thesis takes the approach of understanding people's attitudes towards the interactions rather than the interactions themselves, which contributes to the exclusion of the linguistic approaches discussed above.

3.5.4 Population

The choice of population for collecting data was also an important aspect of the methodology to consider. L1 English speakers were chosen for both studies in this thesis, as it was believed English VL would be more familiar to them than L2 or L3 English speakers. The term "native speaker" is avoided here, as it may have different connotations in HAI and HCI contexts that are yet to be fully explored. It's possible to link the concept of a native speaker to the degree of familiarity a user has with an interface, as well as the extent to which the interfaces is similar to the user, for example in its voice and its language use. Familiarity and similarity are both explored further in Chapter 6.

3.6 Summary

This chapter has provided an overview of the framework used in Chapters 4 and 5, on how VL in a HAI context will be investigated. The initial design of the context including Lego assembly tasks and the instructions used within them were explored, as was the creation of a VL model to apply to these tasks. This chapter also provides details on the general approach to applying the VL model in the context of the

Lego assembly tasks with verbal agent instructors. The design of the agent and interaction are also explored, along with some of the research gaps in methodology that this thesis aims to address. This chapter provides the foundation to the more specific approaches that are discussed in Chapters 4 and 5, which both provide further detail on the nuances of each study.

4. Study One: Comparing Vague and Non-Vague Verbal Agents in Lego Assembly Tasks

4.1 Introduction

This chapter discusses the implementation of the vague language (VL) model outlined in Chapter 3, in the form of a research investigation referred to as Study One. The specific questions this chapter aims to address are first discussed, along with their respective hypotheses. The specific variations of the general framework in Chapter 3 are then discussed, followed by the results and discussion of the data.

Ultimately this data will be used to assess participants' performance in the task with vague and non-vague agents in both stress (with a time limit) and no-stress (without a time limit) tasks, their perception of the agent's language use, and thoughts on the interaction as a whole. This will be discussed in the latter stages of this chapter, along with observations as to the identities that users construct towards the vague and non-vague agents. Finally, concluding remarks on the contributions and limitations of this study are discussed.

4.2 Aims and Objectives

The review of previous literature in Chapter 2 discussed several research gaps in regards to agents using VL. VL explicitly has not been investigated in a human-agent interaction (HAI) context, although the theories of politeness and facework have both featured in previous research (Wang, et al., 2008; Torrey et al., 2013; Strait et al., 2014). The results of these studies are mixed, but some of the findings indicate that agents using politeness strategies can be beneficial in increasing the positive perceptions of them by their users.

The central aim of this study is to compare responses to vague and non-vague verbal agent instructors in Lego assembly tasks. The first point of comparison is in task performance – the extent to which participants can accomplish the tasks, and what effects the agents have on these outcomes. The second consists of the perceptions towards the agent. This includes how they rate it on a predefined set of characteristics, how they describe their interaction preferences for either agent, and what identities they appear to project onto them during post-task interviews.

The duration of the agent interactions was also considered in this study, as in previous literature this was not always explicitly defined (3.5.2). As a result, this study includes a further condition based on the duration of the tasks. The aim of this condition is to assess whether tasks with a time limit affect perceptions of either agent, in comparison to tasks where there is no time limit.

4.3 Experimental Questions and Hypotheses

To achieve the aims outlined above, this study focuses on addressing four experimental questions, along with their respective hypotheses. This section discusses these specifically.

> ***EQ1-1****: Is there a difference in how vague and non-vague agents are rated in regards to specific characteristics of the agent i.e. friendly, authoritative, trustworthy, likeable, controlling, sociable, clear and direct?*

This takes a similar approach to the studies on polite agents discussed previously and aims to assess whether or not VL can provide any similar patterns in the results. Some of these characteristics are the same as those used in these studies, and further ones have been added.

> ***EQ1-2****: Are there differences in performance of the tasks for the vague and non-vague agent i.e. time taken to complete the task; number of repeated instructions requested?*

The focus here is on measuring the participants' ability to successfully assemble the model based on their completion time and the number of times they requested information to be repeated. This tests both their ability to construct the model and their comprehension of instructions.

> ***EQ1-3****: Does the addition of a time limit affect a participants' ability to perform the task and does it affect their perception of the agents?*

An additional variable was introduced in this study, which was named *stress*. This referred to whether the task being conducted was under a time limit (stress) or not (no-stress). As some interactions with verbal agents are conducted under time critical conditions, such as those with satellite navigation systems, this variable was introduced to understand whether it contributes to any effects seen in the first two sets of measures.

> ***EQ1-4****: How are identities towards the vague and non-vague agents presented by participants and what contrasts and similarities are observed?*

Finally, there are questions as to how participants will present their perceived identities of the agents, particularly in regards to the notions of identity discussed in Chapter 2. These are mostly related to the discussions that occur after each task in the semi-structured interviews. This includes, for example, discussions as to the appropriateness of the VL, how this reflects on the wider social implications of HAI, and attitudes towards any other salient features of the agent and the interaction.

With these four research questions come hypotheses based on previous literature surrounding both human-agent interaction and VL in human communication. Although results are difficult to predict in a somewhat novel research context, the prior work on politeness strategies and indirect speech in particular allows for some predictions to be made. The hypotheses are discussed below.

> ***EH1-1****: Users will rate the vague agent as more likeable, friendly, trustworthy and sociable than the non-vague agent.*

Although mixed results have been observed with politeness strategies and indirect speech, it is believed that the social benefits of using VL in establishing and maintain a positive rapport with interlocutors will also appear in HAI. It is thought the social levelling effects of VL should provide a less authoritative vague agent, and the frequency in which they occur in human speech should create a more comfortable and familiar space of interaction.

> ***EH1-2****: Non-vague agents will be rated as more controlling, authoritative, clear and direct than the vague agents.*

Similarly with the social leveling effect of VL (Carter, 1998) not being present in the non-vague agent, it is thought that this will give users an impression of a more direct and authoritative persona in that it does not attempt to hedge or soften its instructions. Given the lack of VL, the non-vague agent is expected to achieve a higher rating for clarity from the participants.

> ***EH1-3****: Users will display a better task performance in the vague agent rather than the non-vague conditions.*

The non-vague agent is expected to be more direct and imposing, though it is also hypothesised to provide more clarity. However, because of the assumed naturalness of the interaction that should be evoked in participants as a result of the VL, it is thought that this would create a less stressful task environment. As a result, it is expected that participants will request less repeats from the vague agent and complete the tasks in less time. This may seem somewhat counterintuitive, given that there will be more overall time spent by the agent instructing the participants. As highlighted in Figure 4 in 3.4.1, however, by occupying the "fuzzy space" between direct alternatives, this in turn may create a greater number of options as to what constitutes successful execution of an instruction. This is for both in where pieces ought to be placed on the model and in what orientation they should be positioned. The non-vague agent is believed to require of participant a more exact placement and orientation.

> ***EH1-4****: Any significant differences observed in the above hypotheses will be reduced in the stress condition when compared*

to the no-stress condition.

As the stress condition tasks will include a time limit, it is expected that participants will want to finish within this limit and as such repeat less and complete the tasks quicker (EH1-3). Similarly, with the focus expected to be greater on completing the task in time it is also expected that less attention will be paid to the language of the agents, and that perceived characteristics in the vague agent will contrast less with those of the non-vague agent (EH1-1, EH1-2).

4.3 Method

The approach to investigating the experimental questions and hypotheses builds upon the general framework outlined in Chapter 3. This section discusses the specific approach to investigating the comparative responses between vague and non-vague verbal agent instructors.

4.3.1 Agent Design

With the focus on the two types of agents in this study (vague and non-vague), using two different Lego models (*Aquagon* and *Nex*), a total of four different agent interfaces were created. The interfaces were identical in design, apart from the name of the Lego model being displayed and the library of sound files being presented verbally as the instructions.

As discussed in Chapter 3, the design of the agent's voice is discussed specifically in each chapter. While the instructions had already been completed, a voice had to be chosen to present them verbally to the participants. A synthesised voice was deemed to be appropriate, as the issue of the VL being used was seen as the most salient aspect of investigation before the agent's voice. Furthermore, synthesised voices are more readily available for change than amateur or human voice recordings, and represent an inexpensive option. The synthesised voice Cepstral Lawrence was chosen for this study (http://www.cepstral.com).

To apply the instructions to this voice, the program Text2SpeechPro was used (http://www.hewbo.com). This is a text-to-speech program that allows for textual input to be outputted as audio files in a specific voice. In this program, the individual instructions were inputted and then exported as .mp3 files. These were then organised into the libraries for their respective agent and model type, along with the HTML files. This completed all four interfaces.

4.3.2 Participants

Thirty L1 English speakers studying at the University of Nottingham were recruited for this study and reimbursed with a £10 Amazon

voucher for their participation. L1 English speakers were used as the nuances of the potential social and interpersonal effects of the vague language may not be universal, in a similar manner to politeness strategies (Brown and Levinson, 1987). Of these participants nineteen students were male (63.3%) and eleven were female (36.7%). Five were postgraduates and twenty-five were undergraduates. The ages of the participants ranged from 18-30 years old. They were recruited through email advertisements of the study.

4.3.3 Procedure

Prior to the first participant session, it was decided that the first twelve would be part of the no-stress condition. This was to establish an appropriate time limit for each model in the stress condition, as opposed to using an arbitrary number. The mean averages of these no-stress times were calculated and then an additional two minutes and thirty seconds added. Although this may appear generous, this was to ensure that all participants would have a fair chance at being exposed to all the instructions in any given task. To counter the notion that a task time limit appeared generous, these time limits were not revealed to participants in the stress condition. Instead, they were only informed that a time limit was in place without the revealing the specifics of its duration. This was done with the aim that this would instill the opinion that the task could end at any time, creating the desired stress.

Conversely, those in the no-stress condition were informed they had as much time as they needed to complete each model. The stress condition tasks also consisted of a further twelve participants, again with alternating agent conditions as a counterbalancing measure. The final six participants, however, had the same agent but with alternating stress conditions. This was to analyse whether any effects could be observed in participants that had the opposite alternative conditions to the first twenty-four participants.

With regards to counterbalancing, the first group of participants (N = 24) were balanced so that the order of both the models being used, as well as the agent conditions, were presented in an equal number across the group. Similarly, the second group of participants (N = 6) was counterbalanced in a similar manner with the model order and stress conditions, however the aim of full counterbalancing was incomplete was the second group did not meet the required number (8).

After all prior measures were conducted the experiments were ready to be conducted. All participants were given the same procedure. First each participant was briefed on the session process, which was supplemented with an information sheet on the task. This described the background information on the study, as well as what the session

would entail for the participants[28]. A consent form was also provided, describing the recording of data, its future uses, and notification that they are able to withdraw at any point in time[29]. The consent form had to be signed and dated by the participants before any tasks were to begin.

Before interacting with the agent interfaces participants were given a practice task to get them accustomed with the nuances of this particular type of Lego[30]. This was conducted with the manufacturer's booklet rather than any verbal instructions, which also allowed for any comparisons to be made during the semi-structure interviews later on in the sessions. Like the other tasks, these were done in both stress and no-stress conditions. Following this model they were notified that the following two tasks would be recorded, as outlined in the briefing. The aim of the practice model was to accustom the participants to the particular tasks they would be engaging in. As discussed in 5.2, however, this was a process that was eventually discarded as unnecessary.

Each session was filmed from two angles. The native camera on a MacBook Pro 10.2, which provided the interfaces for each task, was set to record the entire session to capture the front facing angle of each task. This allowed for the recording of participant facial gestures. A Panasonic HDC-SD80 camera was also set up to record each session from the side to allow for a more detailed view of the model assembly (Figure 6).

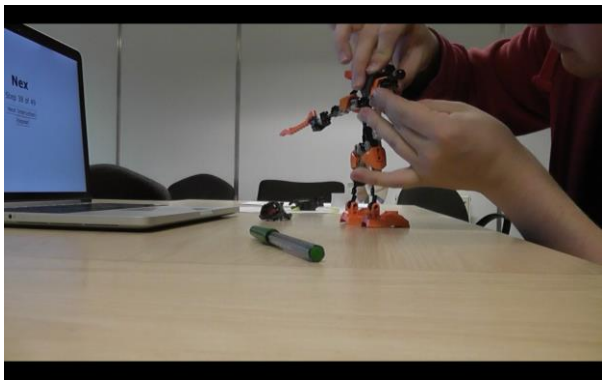

Figure 6: A side angle view of a participant engaged in one of the tasks.

For all tasks including the practice model, participants were informed of whether or not they had a time limit set. Those who did not were informed they had as much time as they needed to complete the model, whereas those who did were informed of the time limit but not

[28] See Appendix B for a copy of the information task sheet.

[29] See Appendix B for a copy of the consent form.

[30] This was the third Hero Factory model called "*Stringer*".

notified of what it was. The researcher kept track of the time once they clicked the "Start" button on the interface for all tasks. The timer was stopped was once participants clicked "Finish" on the interface. The stress tasks were completed either when "Finish" was selected or the time limit was reached. Once each task with the verbal agent was complete, participants were asked to complete a questionnaire held on SurveyMonkey[31]. This was then followed by a semi-structured interview. Interviews were recorded using the same cameras that filmed the tasks.

4.3.4 Measures

This study uses a combination of both quantitative and qualitative measures. The first of the quantitative measures is task performance. Essentially this means how well a user performs in the task in view of a series of metrics. These metrics were influenced by similar research into task-oriented contexts. The time in which it took participants to complete the task was used as one of the performance metrics, as seen in similar tasks such as bicycle repair and Lego building (Kraut et al., 2003, Clark and Krych, 2004). Although this initially could be seen as unfair towards those in the timed sessions, the timings were intentionally long so as to give the majority of participants enough time to complete each assembly. The second metric used was the number of times in which participants requested an instruction be repeated via the interface. This measure was fairly novel in that such a method of requesting repeated information is lacking in previous literature. The interface allows this measurement to be fulfilled by the means of the "Repeat" button. Both of these metrics were logged via the interface. The completion time was measured from the moment participants clicked "Start" until they clicked "Finish" and measured with a timer by the researcher, depending on whether the participant was in the stress or no stress condition. The number of requested repeats was also logged on the interface automatically. This also logged which specific steps had been repeated and on how many occasions this occurred.

The other set of quantitative measures used were found on the questionnaire given following each task. On these, amongst other questions, were eight five-point Likert scale questions (1= strongly agree; 5 = strongly disagree)[32]. These asked participants to rate the instructor on eight different characteristics: likeable; friendly; trustworthy; sociable; controlling; authoritative; clear and direct. These were modified from an existing voice attribute scale used in previous literature surrounding synthesised voices, accents and user

[31] Found at https://www.surveymonkey.com/.

[32] Full scale was as follows: 1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, 5 = strongly disagree.

perception of robots (Tamagawa et al., 2011). VL often promotes a level social environment, hence the inclusion of the first four characteristics. The latter for were based on the hypothesis that the non-vague agent would display characteristics associated with direct language.

Qualitative data first consisted of open-ended questions in the post-task questionnaires[33], and semi-guided interviews. The interviews in particular sought to achieve a greater understanding of the participants' experience, as discussed in 3.5.3.

4.4 Results

The results presented in this section first discuss the quantitative data. As most of the results concern the differences in the vague and non-vague agent interactions, the first twenty-four participants' data is included and the last six excluded. This is because these six participants interacted with a vague or non-vague agent twice, though in different stress conditions. The other twenty-four, however, interacted with both agents and the order of the voices was counterbalanced between participants. The qualitative data, however, does include comments from all thirty participants, as it is not concerned with statistical analysis.

In this chapter, the analyses of task performance data are discussed first, followed by the measures of agent perception used in the questionnaires. This consists of statistical analysis of the questionnaire responses in regards to the perceived characteristics of the data, and comparisons between the vague and non-vague agents in the stress and no-stress task variables. SPSS was used as the software to conduct all the analyses. The qualitative data then follows with common themes in the data discussed, and presented along with extracts from the interviews.

4.4.1 Task Performance

A one-way ANOVA between the two agent conditions (Table 3) revealed that, although there were noticeable differences in both the time taken to complete tasks and the number of repeated instructions requested, the results were not statistically significant, $F(1, 58) = 1.94$, $p = .17$[34]. Non-vague tasks took a mean average of 749s (SD = to complete and vague tasks 820s. The overall mean was 787s. Non-significant differences were also found in the number of repeats requested in each agent condition, $F(1, 58) = .18$, $p = .67$. Vague agent

[33] See Appendix B for a copy of the questionnaire.

[34] Although participants took part in two tasks, each one was inputted into SPSS as a different data point, hence the high degrees of freedom figure.

tasks saw an average of 7.03 repeats and non-vague an average of 7.64.

Table 3: ANOVA on task performance between agents.

Agent	Time (seconds)		Repeats	
	M	SD	M	SD
Non-Vague	749	184	7.64	6.37
Vague	820	207	7.03	4.82
TOTAL	787	198	7.32	5.55

Another one-way ANOVA was run for the two stress conditions. There was no significant difference in regards to the time between stress and no-stress conditions, $F(1, 58) = 2.32$, $p = .13$. There was, however, a significant difference between the number of repeats requested, $F(1, 58) = 5.97$, $p = < .05$.

Table 4: Comparing task performance in no-stress and stress conditions.

Stress Condition	Time (seconds)		Repeats	
	M	SD	M	SD
No-stress	825	228	9.00	6.09
Stress	748	158	75.63	4.45
TOTAL	787	198	7.32	5.55

Data was also then grouped into four categories representing the permutations of both agent and stress conditions: vague stress; vague no-stress; non-vague stress and non-vague no-stress. A one-way ANOVA was conducted on these four conditions (Table 5). No significant differences were found for time, $F(3, 56) = 1.47$, $p = .23$, or repeats, $F(3, 56) = 2.1$, $p = .11$, indicating that although there was a significant effect in stress there were no significant effects when combining stress and agent type. Findings were similarly non-significant when assessing gender as a variable (Table 6) for both time, $F = (1, 58) = 2.39$, $p = .13$, and repeats, $F = (1, 58) = 2.95$, $p = .09$.

Table 5: Comparing task performance in stress and agent type combined.

Agent	Time (seconds)		Repeats	
	M	SD	M	SD
Non-vague no-stress	777	221	8.93	7.19
Non-vague stress	720	234	6.36	5.37
Vague no-stress	867	140	9.06	5.18
Vague stress	772	173	5.00	3.52
TOTAL	786	198	7.32	5.55

Table 6: Comparing task performance between female and male participants.

Gender	Time (seconds)		Repeats	
	M	SD	M	SD
Female	838	181	8.91	5.61
Male	757	220	6.39	5.39
TOTAL	787	198	7.32	5.55

4.4.2 Survey Measures

The questionnaire of agent characteristics also underwent similar statistical analysis, both in the tests used and the variables being tested. A one-way ANOVA was conducted to compare the mean values of each attribute against both the vague and non-vague agents (Table 7 on the following page). The ANOVA revealed the non-vague agent was rated as significantly more authoritative than the vague agent, F (2, 58) = 6.143, p = .016, as well as significantly more direct, F (2, 58) = 10.345, p = .002. No other significant differences were observed.

Table 7: Comparing attributes between non-vague and vague agents.

Agent	Authoritative*		Direct*		Friendly		Trustworthy		Likeable		Sociable		Controlling		Clear	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Non-Vague	2.25	.585	1.82	.612	2.79	.917	2.46	.838	3.00	.981	3.57	.92	2.75	.70	2.5	.923
Vague	2.75	.96	2.56	1.08	2.41	.798	2.84	1.05	3.09	.995	3.16	1.05	2.94	.982	2.84	1.14
TOTAL	2.52	.813	2.21	.958	2.58	.869	2.66	.968	3.05	.982	3.35	1.06	2.85	.86	2.68	1.05

*p values: * = $p < .05$*

The mean averages (M) and standard deviations (SD) are included. Lower mean scores indicate a higher rating for that characteristic.

A separate one-way ANOVA was run comparing the same values between the stress and no-stress conditions, however no significance was observed (Table 8). In combining the stress and agent conditions to make four separate groups again, there was a significant result in the direct attribute, F (3, 56) = 4.95, p = < .01. A Post-hoc Tukey's HSD test showed that both non-vague non-stress and non-vague stress tasks were rated as significantly more direct than the vague no-stress tasks. No other significant interactions were observed.

Table 8: Comparing authoritative and direct attributes between combined agent and stress agent types.

Agent	Authoritative		Direct	
	M	SD	M	SD
Non-vague no-stress	2,36	.633	1.79	.579
Non-vague stress	2,14	.535	1.86	.663
Vague no-stress	2,81	.981	2.88	1.20
Vague stress	2.69	.873	2.25	.857
TOTAL	2.52	.813	2.21	9.58

The mean averages (M) and standard deviations (SD) are included. Lower mean scores indicate a higher rating for that characteristic.

Contrasts in gender were also seen (Table 9). Another one-way ANOVA revealed that the female group perceived the agents overall as significantly more authoritative (M = 2.23, SD = 0.62) than the male group (M = 2.68, SD = 0.87). The agents were also collectively seen as more direct in the female group (M = 1.91, SD = 0.13) than the male group (0.18). This result, although close, was not statistically significant.

Table 9: Comparing authoritative and direct ratings between female and male participants.

Gender	Authoritative		Direct	
	M	SD	M	SD
Female	2.23	.612	1.91	.610
Male	2.68	.873	2.39	1.08
TOTAL	2.52	.813	2.21	.959

The mean averages (M) and standard deviations (SD) are included. Lower mean scores indicate a higher rating for that characteristic.

In assessing the survey measures against other independent variables and combinations of them, no further significant results were observed.

4.4.3 Interaction Preferences

The questionnaire that contained the survey measures also included open-ended questions on the participants' preferences for agent interaction. This section reviews the results of the three relevant questions that pertained to these preferences. Only the first 24 participants' responses are included here, as the remaining 6 did not interact with both agent types. The first of these questions asked the following:

> *What were your feelings towards the agent's voice? Would you have preferred being instructed by a human voice instead?*

The results here focus on the second part of the question and the responses were coded and placed into four categories – *yes, maybe/no preference, no,* and *not clear*. The categories of *yes* and *no* were definitive and in these responses participants specifically declared their preference for either. For the answers categorised as *maybe/no preference*, participants indicated there were possibilities that they may choose a preference but did not definitely specify one. Moreover, this includes responses that said the participants specifically had no preference. Finally, the few responses in *not clear* contained responses in which participants had not clearly answered the question, and as such could not be placed into any of other categories. The results of these categories can be seen in Figure 7.

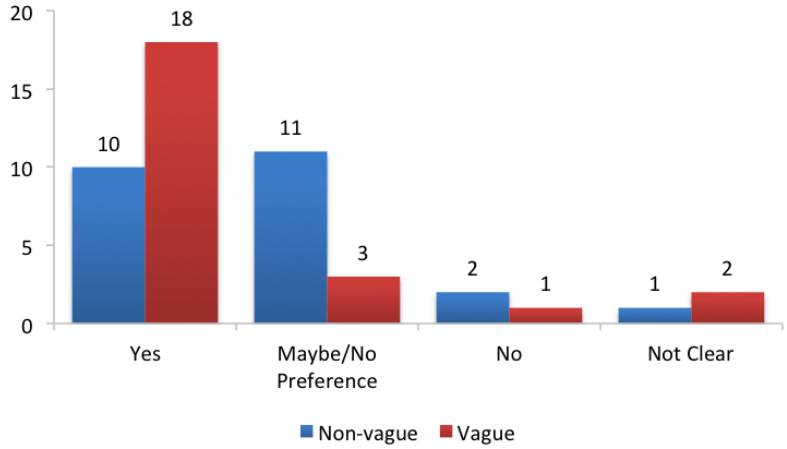

Figure 7: Human voice preferences in vague and non-vague interactions.

As this figure shows, in the vague agent interactions more participants indicated a preference for interacting with a human voice (75%) than those in the non-vague agent interactions (42%). Similarly, less ambiguity was observed, with an almost even number of *yes* and *maybe/no preference* responses emerging from the non-vague agent interactions. A breakdown of the responses (see Appendix C) shows that of those who preferred a human voice in the vague agent interactions, 33% of participants indicated this was because of a lack of clarity. A further 12.5% signified this was because of the agent's language. In the interviews that are reported in 4.4.4, the responses indicate that 63% of participants preferred interacting with the non-vague agent, while only 8% preferred the vague agent. The remaining 29% were undecided or did not display a preference.

The results above suggest that while there are preferences for a human voice in both agents, the lack of clarity partly due to the use of the language, had a significant impact on the vague agent interactions affecting the participants' choice of voice in a future interaction. The second question regarding interaction preferences was as follows:

Would you be happy to interact with the agent again?

Again, these responses were coded as *yes, no, maybe/no preference,* and *not clear*. As Table 10 shows, these were broken down into the categories of reasoning that the participants provided in their answers. Only the *yes* category was broken down as the other categories did not have clear or elaborated reasons behind the answers. There were notable differences between the two agents, with more participants preferring to interact with the non-vague (75% vs. 46%) agent again, and more participants indicating directly that they would not interaction with the vague agent again (33% vs. 8%). Furthermore, there were more caveats to in those responses that had elaboration for interacting with the vague agent. In contrast, there was only one caveat for the *yes* responses in the non-vague interactions.

Table 10: Comparison of responses for interacting with the non-vague and vague agents again.

	Non-vague	Vague	TOTAL
Yes: little/no elaboration	19	11	30
Yes: if I had to	0	3	3
Yes: for a short time	0	1	1
Yes: for this task	1	0	1
No	2	8	10
Maybe/No Preference	1	1	2
Not Clear	1	0	1
TOTAL	24	24	48

The final question regarding interaction preferences asked the following:

> *Would you be happy to have the same voice for personal devices e.g. smartphone, sat nav?*

This question builds on the previous one, though instead specifies using the voice on a personal device, and the responses were coded in the same way as the previous questions.

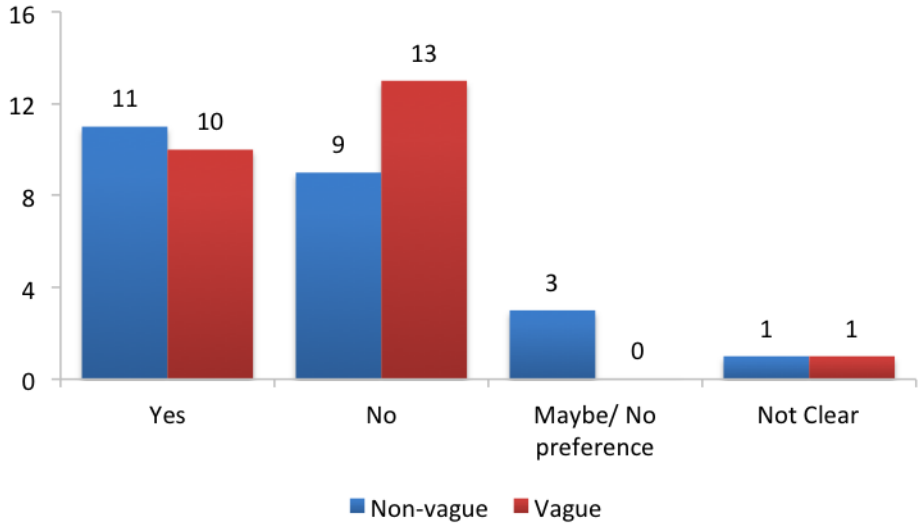

Figure 8: Preferences for the same voice in personal devices.

The results in Figure 8 show that the responses in the *yes* category were very similar, and that there was more ambiguity this time with the vague agent interactions, while there was none in the non-vague agent interactions. Still, there were more responses in *no* for the vague agent (54%) than the non-vague agent (38%). Furthermore, a

breakdown of these categories (see Appendix C) reveals that in the *no* responses, there were an equal number of participants who provided the reasoning that a human or humanlike voice would be better, that the current iteration of the vague agent would be annoying, and that they would not use it in a personal device because of the language (all 12.5%). Overall, the numbers in the *yes* category lower for both the non-vague and vague agent interactions in the use of a personal device than the responses in the second question.

Results from all three questions indicate that the language can impact upon the participants' desire to interact with an agent again, and that the preference for a human voice is higher in the vague agent interactions, while still being a noticeable caveat for the non-vague interactions.

4.4.4 Qualitative Analysis

The qualitative analysis presented in this section is derived from the interviews conducted after each task. These were condensed down into noticeable patterns of data or *emerging themes*. Most of these themes discuss the differences between the vague and non-vague agent, with a particular focus on the VL itself, as well as wider reflections on human-agent interaction outside of this study. Extracts from the interviews are marked with the agent type the participant is referring to (non-vague = NV; vague = V).

4.4.4.1 Voice Clarity and Expectations

One of the most frequent talking points in the vague agent interviews was regarding the quality of the Cepstral Lawrence voice and how its use of VL affected the participants' perceptions of the. Given the non-human nature of the voice, however, it sometimes proved difficult to understand even in the non-vague tasks:

> *P12: It was just some of the words like sometimes the pace changes or how it says the word and that makes it harder… it's just the fact that some words are said a bit weird you know? (NV)*

> *P13: A few times I couldn't actually understand what it was saying. The text to speech thing was a bit off. (NV)*

Both P12 and P13 highlight there being something almost ineffable about the way in which Lawrence produced particular words in the instructions. This meant that some words became hard to understand, and sometimes impossible to follow. When dealing with a voice that can be difficult to understand, the introduction of language that is not typical of such an agent could be a source of further confusion and miscommunication. In other cases it was not miscommunication that was an issue, but simply disliking the quality of the voice:

P21: It's fine but like the voice makes it less because it's computerised it's like less of a personal experience so it makes it very robotic. (NV)

In this example a computerised voice is deemed as a less personal experience, suggesting that a non-computerised alternative would elicit the opposite response. This may be due to the lack of vocal similarity between themselves and the agents, perhaps lacking the closeness that a human voice could achieve. Other contexts were suggested as being suitable for the voice:

P23: I guess for mobile devices it would be better to have a computerised voice because it's commands and one lines. (NV)

P23 proposes that mobile devices are better suited with synthesised voices, particularly due to the way in which they communicate, which they believe involves simple sentences and commands. However, further on in the interview they suggest a "realistic" voice being more appropriate for a context such as satellite navigation, highlighting a disparity between voices and contexts.

In other cases where the voice was not a cause for, the non-vague tasks were often well received: "It was fine. I wouldn't really change anything"; "It's kind of what you expect from a computerised voice". When expectations were matched the response was often positive, or the agent was at least deemed appropriate for the interaction.

4.4.4.2 Voice and Language Disparity

As discussed at the beginning of this section, one of the more prominent issues participants had in the non-vague tasks was the use of the VL in combination with the voice:

P23: It's just the combination of the voice and the script didn't work. (V)

Sometimes this was quite general: "It sounded too forced." It appeared that there was some disconnect between the words that the agent was using and the ways in which they were being pronounced. The majority of participants suggested that improving the quality of the voice would improve their perception of the agent and how the language would be received: "It would sound better with a more natural voice"; "The voice is holding it back." Other feedback suggested that the vague agent was unable to execute the VL as would be seen in human interaction: "It felt like it was insincere"; "It seemed fake…it was trying too hard", indicating a certain lack of success in creating the positive social effects that were intended.

Some more specific recommendations included references to the prosodic capabilities of the agent: "Change the speed since it's quicker in human speech," as opposed to just referencing the general quality of the voice. Comparing the verbal agent to human speech indicates that the atypical nature of VL in this context, and that if such language is going to be used then having a more humanlike voice would be the appropriate step forward.

As well as general discussions of the language, specific lexical items from the VL model were also discussed. The most commonly identified items from the VL model were *basically, like* and *just*. These were followed by *kind of, sort of* and *more or less*, and to a less extent *should, so* and *now*. For the non-vague agent there were explicit mentions of any of its language use in regards to specific lexical items. Although there were mentions of the VL agent in the majority of interviews, there were still some cases where participants did not notice anything in regards to language use. When they did it was often in varying degrees of negativity: "It emphasised phrases like *more or less* strangely." The phrase *basically* was not talked about in any positive manner, with it being described as "inappropriate and somewhat demeaning," "annoying and too vague," and contributing towards creating a "condescending tone" (i.e. a condescending way of speaking through language choice). Similar responses were seen with just and sometimes created purposefully humorous and hyperbolic responses: "If I had this in my sat nav I would probably crash my car"; "It made me want to kill myself."

Unlike basically, however, the use of *just* also received positive feedback on its suitability with instructions that matched the procedure: "It was more consistent with the step…like it was just a little twist" and in creating a better impression of the agent: "I thought it was friendlier…it sounded like a natural thing for him to be saying". This suggested that to some extent the face saving attempts of this minimiser were used successfully, and even contributed to creating something more "natural" which is assumingly synonymous with humanlike. Similarly, having *just* used in tandem with an instruction that required a minimal movement such as "*just* a little twist" was sometimes seen as a positive addition as there was a truth behind it being a small movement, and created further attempts at minimising this instruction.

There were occasions in the interviews where participants did not notice any of the VL being used:

> *P10: Erm no. It's like your genuine computer program. It was concise and to the point so I guess it was efficient in terms of communication…You don't expect like politeness from a generated voice. (V)*

Here P10 likens it to "your genuine computer program," which assumingly is a reference to this agent being both similar to others that P10 has interacted with previously, and also an agent that had no attempts to pay attention to the user's face. This point is reinforced by this participant saying politeness is not expected when dealing with a synthesised voice, referred to as "generated" here, which is somewhat ironic given they were interacting with the vague agent. It should also be noted that this participant also interacted with the non-vague agent and found little to no contrast between the two.

4.4.4.3 Agent Preferences

When participants were able to compare both agents there was a strong preference for the direct alternative due to its lack of VL: "I liked the lack of fluffy words". Its simplicity in language was also praised: "It just said what was needed"; "It was much more straightforward so you just do what it tells you", indicating participants sometimes found the VL superfluous to the instructions as a whole, and instead preferred just the information pertinent to assembling the models. Sometimes the non-vague agent was praised when comparisons were made to agents participants had previously interacted with:

> *P27: No, no actually it was surprisingly kind of like accessible. It was talking to you in the right kind of language…. No and I am one of those people who actually wants to shout back at one of those kind of like automatic checkout things so actually no it was fin. (NV)*

P27 indicates how accessible the language was especially in comparison to an automated checkout, despite them having a tendency to not like them. This is perhaps a matter of being overly familiar with the checkout systems and this being a novel interaction. It may also be the case of interact with an agent in a task that they enjoy, with their own time purposefully set aside to complete it.

There were mixed response to the vague agent, though the findings leaned towards a theme of negative perceptions. One of the reasons is the atypical use of VL that did not always correlate with previous agent interactions. There are also strong indications that the synthesised voice being used did not combine well with this atypical language, and two solutions are often proposed: either that the VL is changed to be more direct, or that the voice is improved to be of a higher quality.

Removing the VL completely is not always seen as necessary, as much as simply reducing frequency in which it occurs within the instructions. Doing so may improve how the vague agent is perceived: "I wouldn't mind it if they didn't say like so much"; "It's okay but think it says just too much"; "It's used too heavily", also pointing to a lack of

variety in the language and some items being used too much in comparison with others.

> *P13*: *So yeah less frequently and maybe different ones. I don't think like is a good word to be using when explaining stuff. (NV)*

The example from P13 here suggests that removing the word *like* would be a good starting point, but also references the point that the language being used in the vague agent is not context appropriate i.e. unsuitable for "explaining stuff". The instructive nature of the task was often a talking point when concerning the VL use in the agent. It appeared to VL. It appeared to interfere with task performance and in turn the attitudes towards the agent: "I just wanted to get the job done"; "The extra information was meant to help but it ended up being confusing". Conversely, the non-vague agent received praise on the appropriateness of its instructions: "I think direct is just the way it has to be when getting instructions".

4.4.4.4 Effect of Instructive Context

There were suggestions that the context of interaction was an obstacle in how appropriate users perceived the vague agent' appropriate use of VL:

> *P11: Erm no. Well though having said that maybe if it wasn't instructions it would be better, say if you weren't trying to get something done because it kind of felt like it was just adding words when you just want the instruction to do it, where if it was something else it would be okay, yeah. It's like if it was just for listening to that would be better. (V)*
>
> *P22: I think it's [non-vague] just the way it has to be if you are getting instructions.(NV)*
>
> *P28: I think with that kind of thing it's probably better to have something that's a bit more friendly. (NV)*

P11 here brings up the point that the VL was just "adding words" and given the fact it was amongst the instructions, seemed unnecessary. A different context that did not involve instruction giving would perhaps be a better environment for the vague agent, although P11 was lukewarm in how this may turn out, suggesting it would be "okay" and "better". P28 suggests that attempts to create a friendlier identity are perhaps more important in conversational contexts, referred to as "that kind of thing," including with intelligent personal assistants such as Siri. Although they acknowledge that direct language is the preferable option in this interaction, they see the value in having VL that may create a friendly agent in contexts that a more casual and less task focused. P22 was similar in their response to the vague agent.

Following a vague agent with a non-vague agent task, they suggested that direct is the "way it has to be" for when someone is receiving instructions. Other responses were similar and some were more positive: "It's fine so long as it doesn't impact on what needs to be done"; "If it was for something that wasn't so precise - as instructions - then it'd work". So long as it attends the needs of the task at hand and does not interfere in this, it would appear that the VL could be considered appropriate. However, understanding the specific needs for each task is likely an individualistic matter.

As well as other contexts were there some proposals that other demographics may benefit from the vague agent, rather than the relatively young user base seen here:

> *P23: If it was like an old person interacting and they've got all day to sit down and converse with a computer then maybe but for people who just want to use computers to get things done then no. (V)*

This is similar to research suggesting the use of social robots can improve the quality of life for elderly people, by the means of keeping them independent, fit and in company rather than in isolation (de Graaf et al., 2015). If one of the aims in this area of research is to provide some form of company, then being direct may not always be the best option. Adopting the use of some types of VL may provide a solution. As well as other demographics, P23 does also put forth another contextual improvement regarding how much time one has to interact with an agent, in that if there is more time to engage with them then perhaps using VL is more appropriate.

As well as the voice and the language sometimes causing difficult and negative perceptions, the modality in which instructions were given also appeared to have its drawbacks. Given that the practice model used the visual booklet, participants already had a comparison between this and the speech: "They're easier to relate to"; "Visual is easier for locating the right piece". Previous experience, which many had, is also to have been a likely contribution in these comparisons. As there was only speech here and no visual information, it was often less well received in some aspects of the task:

> *P11: I didn't like not being able to see it as well because I like seeing stuff. (V)*

> *P12*: *I think it's easier to see if it shows you the piece or something or shows you a picture of what you are doing. (V)*

> *P13: Yeah. I mean I don't mind the voice thing just like the occasional picture just for like the difficult ones. (NV)*

All three of these participants found the verbal instructions troublesome to some extent, both in the locating and assembling of pieces. Although the visual was often described as easier, the verbal agent was sometimes better received than the visual instructions: "Verbal was easier for the actual assembly"; "It was easier to navigate around the 3D space with the spoken". Although some participants did suggest that the voice be replaced entirely, the supplementation of it with visual aids appeared to be another acceptable solution: "A visual supplement would make things easier"; "A mixture of both would be nice". This suggests that an agent with speech as its primary modality agent may not be the best for instructions and similar tasks, and that including visual information as well as verbal may improve a user's ability to conduct the task. It is not known whether this would improve the perception of other aspects of the agent such as language, though this remains a possible avenue of further research as discussed later in 4.5.

4.4.4.5 Agents in Society

A lot of interview examples discussed in this section are in regards to the specific features of the agent such as the voice, the language and the modality of interaction. There are also examples of participants discussing the wider implications of the effect on the identities users create for agents. This includes reflections on the general social acceptance of agents, classifications of agent likeness as a group identity, and what contrasts and comparisons exist with human likeness.

As mentioned above there were problems with how users perceived the combination of the Cepstral Lawrence voice and the VL. At times this was simply a matter of prosodic deficiencies – the voice was not able to pronounce words in a manner people were used to hearing. This did not also coincide to the expectations of all participants:

> *P11: It tried to be like friendly like basically do this but you don't expect it from a computer so you don't like it. I rather it be what you expect. I rather it be a computer voice and speak like you expect it to. (V)*

In this extract P11 begins with an interesting point that using VL such as *basically* was an attempt at being friendly, and so recognises and verbalises one of the uses of VL in human communication. However, because it is not expected from a computer, it does not achieve this aim of portraying a friendly identity. This is somewhat separate from the second point regarding it being a "computer voice" and as such should speak in an expected manner that does not include attempts at being friendly. This again provides further substance to the notion that having a non-computerised voice, but still used in a verbal agent interface, could successfully portray a friendly identity by using VL.

This extract provides credence to the idea that individuals expected machines to speak within certain expected limitations, and this includes the use of VL.

There were further concerns regarding the vague agent's attempts at imitating human speech, but not successfully executing these efforts:

> *P13: It just felt like it was trying to be human but it was kind of forcing it…kind of felt like it was making fun of how people explain things. (V)*

> *P27: If it's just trying to pretend to be human then it's just like putting a mask on and it's actually getting in the way of its ability to communicate with you and yeah that is frustrating… it's trying to be chummy with you and you know you can't be chummy with a machine. (V)*

P13 raises an issue as to how the agent appears to be forcing attempts to sound like a human through the use of VL, implying that this did not create a positive identity. They also indicate that the agent using VL in turn makes fun of human instruction giving and explanations, rather than successfully using the language itself. P27 follows on with a point that using the VL is like wearing a mask and inhibits its communicative abilities in regards to instruction giving. Interestingly the choice of referring to this as a mask suggests that the interface is in some way covering up its true agent likeness. Both participants indicate this VL is atypical of agent speech and inappropriate in the interaction. P27 shows awareness of VL able to create rapport between humans, but that its use in a verbal agent interface is redundant as this same rapport cannot exist between human and machine. A further extract from them captures articulately some of the points that they and other participants were discussing:

> *P27*: *It feels really sinister… I think we like machines to actually know their place as machines and crossing that boundary is a bit strange for me… I think we've got kind of acceptable terms in which we interact with computers and sort of machines in general and once you start blurring that boundary it does get creepy. (V)*

This continues the theme of separation between agents and humans, with P27 arguing that there are boundaries separating the two that an agent should not cross. The point regarding blurring these boundaries being creepy is similar to the uncanny valley discussion from Chapter 2. Attempting to achieve a humanlike sense of rapport and lack of imposition through VL created a "sinister" identity for the vague agent here. Another participant made a similar type of distinction:

P23: *If it was real time maybe but computers think a lot faster than humans so it probably wouldn't be needed. (V)*

In discussing other contexts, P23 posits that because of the speed at which a computer thinks, adding VL would not be necessary, especially when a user is just aiming to complete a particular task or accomplish something soon[35]. Again, there is a categorisation of the differences between human and agent qualities, even if only on a small scale. Other comparisons to human speech were observed in the interviews, though these were not necessarily concerning such the same social matters:

P21: It just sounded like a mate trying to describe like tell you how to do it down the phone like an actual instruction. (V)

P29: It is weird to have him saying things like, but at the same time I know a human put that information there so it does make sense because it's how you would have said it if you said it yourself. (V)

P21 here likens the vague instructions to how a friend would describe it to them, even though they discuss further in their interview that the VL would be best left out of the interaction. This may indicate the effect of the VL being a social leveller (McCarthy and Carter, 2006) and that these effects can still be acknowledge in agents, even if not always seen as appropriate. P29 indicates that VL on first thought does not make sense in agent speech, but understands that a human designer has input the information in the agent first. As they see words such as *like* as sounding more like a human instructor, the connection from designer to agent output then affords a greater acceptance for VL in this interaction. This then comes back to the notion of similarity in that P29 can relate to the designer of the vague agent and the use of VL in their instructions, creating some form of similarity-attraction. This disagrees somewhat with some thoughts in the CASA paradigm that suggest users react to a computer system and not the designer behind it. It may be the case that both entities are involved in the reality of the interaction. This suggests perhaps that sometimes users do have the agent designer in mind rather than just the agent itself. This may also relate to a feeling of similarity in that they create an identity that is related to the human designers, and that humans use this type of language.

4.5 Discussion

The aim of this study was to investigate how users react to vague and non-vague verbal agents, with a focus on the contrasts that emerge from the vague agent interactions. This was conducted by running

[35] This was discussed elsewhere in the interview.

participants through Lego assembly tasks with a verbal agent interface instructing them on how to complete each model. A mixed methods approach to data capture and analysis was used to gather results. First, participants were asked to complete a Likert scale questionnaire regarding different characteristics of the agent they had just interacted with. This was followed by a semi-structured interview in which these attitudes were further investigated. Statistical and content analysis was used to analysis the quantitative and qualitative data respectively.

4.5.1 Agent Characteristics and Task Performance

In analysing the quantitative data the results show that the questionnaire was not as definitive as thought prior to the study. Despite some significance in the results, the others were largely non-significant even if differences between the vague and non-vague agents were still observed. Firstly, EH1-1 was not observed in these results. The vague agent was not seen as significantly more likeable, sociable, friendly, and trustworthy than the non-vague agent. It was hypothesised that the face saving strategies of the VL would create a less imposing and more likeable agent, however this did not transpire. This contrasts with the results of two similar studies in politeness strategies used by robots (Torrey et al., 2013, Strait et al., 2014).

EH1-2 was partly observed, in that the non-vague agent was rated as significantly more authoritative and direct than the vague agent, though no significance was seen for the characteristics of *clear* and *controlling*. In regards to task performance, EH1-3 hypothesised that participants would perform better in the vague tasks. Although there was a notable difference in how often participants would repeat instructions, with the vague agent having less repeats, this was not significant. This was also the outcome for the time in which it took them to complete the tasks. EH1-4 posited that the effect of the stress condition would see a reduction in the differences between the two agents in the significant results, though there was little effect observed when comparing the stress and no-stress tasks, although generally there were less repeats in the stress tasks than the no-stress tasks when it came to task performance. There was a difference in how direct the agents were perceived to be in the different task conditions, where both the non-vague no-stress and non-vague stress tasks were significantly more direct than the vague stress task.

Responses from participants regarding the attributes of both agents were less definitive than in research studying advice-giving interactions, with significant differences found only in the direct and authoritative characteristics. The results of these two attributes were perhaps the least surprising, as the lack of VL in the non-vague agent naturally creates a direct tone, which is typical of agent speech. User performance did not vary significantly across the two agent conditions, though less repeats were used in the stress condition. This is likely a

result of the unknown time limit creating a sense of urgency, leaving less time to check instructions again and perhaps forcing participants to employ a greater focus on speed during those tasks.

4.5.2 Qualitative Contributions

Though the hypotheses were not observed as expected, the qualitative data yielded insights as to why this may have occurred. Although some of the participants verbally expressed their understanding of VL in human communication and some of its uses, they noted the vague agent failed to execute the VL successfully. Often this was due to the quality of the synthesised voice not coinciding to their familiarity of hearing VL. The interaction preferences discussed in 4.4.3 show a strong preference for both the non-vague agent and a preference for having a human voice instead, particularly for the vague agent.

VL items such as *basically* were often discussed, and analysis of the instructions containing this word indicated a tendency for Cepstral Lawrence to prolong the pronunciation of the last syllable, more so than the preceding ones, creating a noticeably stressed hang on the *"-ly"* phoneme. The non-human quality of voice and prosody combined to be, for the most part, a distraction or an annoyance. Participants, as discussed in multiple subsections of 4.4.4, also described their expectations of what an agent should sound like, though often while describing it under the umbrella term of "computer" or "machine". The vague agent again did not live up to their expectations or correlate with familiar experiences of other agent interactions. Comments included participants' expectations of computers speaking in a particular and often direct manner, rather than using any (or at least as much) of the VL used here. This suggests that these participants associate this type of agent voice, and perhaps this type of interface and system as a whole, with identities of directness and lacking the social nuances and capabilities that humans possess.

Further qualitative analysis in 4.4.4 revealed participants who had particular expectations about the language practices of an agent seemed less willing to accept the vague agent, and often commended the non-vague agent for its precision and clarity. However, there were some who did commend the vague agent too for its appropriate use of language in the on-going context of the interaction. When the minimiser *just* for example, coincided with a minimalistic assembly step, its minimising effects were sometimes appreciated. It indicated to some participants that the current instruction did not require much effort, while maintaining a lack of imposition. Moreover, the lexical items that made up the VL model were not recognised by participants equally, nor perceived equally. Phrases such as *more or less* were not frequently observed compared to *just* and *basically*, while *so* and *now* were only commented upon by a few of the participants. The

discussion of vague nouns was very rarely touched upon and remained relatively insignificant in comparison.

On the one hand, some of the positive reactions towards the VL such as the minimisers showed that positive identities did emerge and develop within the interaction itself. On the other hand, there were those who would seem to have had preconceived notions of identity for agents and their use of voice and language, and so encountered a disconnection between their expectations of agent speech and the reality of the interaction.

Participants who did describe the vague agent negatively in regards to their expectations may have indexed VL as coming from humans only, and so regard agent VL as a difficult concept to initially accept. This same indexing may have agents and synthesised voices associated with direct language too. As well as perceptions of this specific agent, there also emerged the wider social discussion of how agents should interact with humans. The vague agent was often derided in its efforts in trying to sound human by using VL and in turn being unsuccessful in its attempts to be friendly. This may have contributed towards the large majority of vague interactions resulting in 75% of participants preferring to interact with a human voice.

This was also met with some discussion in 4.4.4 that agents cannot be friendly with humans, and that there are boundaries between agent and human likeness that an agent should not cross. Other comments included that the use of VL contradicted the nature of being an agent somewhat, in that it created a sense of imprecision in a typically precise and knowledgeable entity. Though does appear to have merit, as one may expect any number of agents discussed in the previous chapters to function with the precision and knowledge for the purposes they are designed for. In the vague interactions, the VL, in this particular voice, has the potential to obfuscate these.

In this sense there appears to be, for a selection of the participants, two group categories of identities that should be distinct from one another, and that the relationality between them creates some incongruence. That is to say a verbal agent instructor with a synthesised voice may belong to a group in which directness and precision is expected, whereas the use of VL may be contained within groups where a human voice is expected to be speaking.

While there are interesting patterns that emerge from this study, the problems of individual vs. group design do arise here. Although there are patterns as to what may constitute this agent likeness, such as being direct and not attempting to engage socially with human users, it cannot be ignored that for some participants the VL was at times positive. For others, it was just a part of the agent that warranted no discussion or was not even noticed. This provides a viable avenue of

investigation into whether the voice is a fundamental obstacle in the accept of VL or whether it is a larger issue of the agent itself i.e. an interface being displayed on a laptop without any avatar or human embodiment, or indeed a combination of these two.

As discussed in the qualitative results it was the voice that was often a point of negative feedback. Sometimes this was in the non-vague tasks as well as the vague, though the use of the VL in combination with the synthesised voice was not often discussed positively. This raises further questions in regards to identity, particularly when concerning the framework discussed in Chapter 2 by Bucholtz and Hall (2005). One of the principles discussed in this framework was that of similarity and difference. As noted previously, participants may index VL as belonging only to human speech, and see the agent's use of it as an encroachment of sorts upon the human social space. However, this may a result of the language being used being similar to what they encounter in human interaction, but the voice being very alien from this interaction space. The mixed methods results discussed in 4.4 suggest that if the quality of the voice were to be improved, be more natural, or made more humanlike, then the use of the VL would be received more positively. Often it was the combination of the voice and the language together that created the negative perceptions, which perhaps would go some way towards explaining why the non-vague agent was fairly well received with the same voice. This lack of symmetry between voice and language was so apparent in some tasks that participants would laugh at the agent during the assembly, and on some occasions comment negatively on what they were hearing. It may be the case that they were not reacting socially to the agent and were instead indicating their perceptions to both the researcher and the camera. Nevertheless, it is a firm indication of how strongly some participants felt towards the agent "trying to be human".

Another reason for the lack of success with the vague agent perception may be a lack of familiarity and exposure to such agents. As discussed already, participants often had some notion of what an agent is and what its speech should be like. However, with the growing number of verbal agents and increasing diversity of their interactions, these types of agents may become more familiar to people in the future (such as relational agents discussed later in 6.2.3). An increasing frequency of these types of agent interactions may alter people's expectations of what they should sound like, and in turn push the boundaries of what agent identities can be. When it comes to VL in particular, perhaps increasing the amount of exposure a participant has with a vague agent would contribute to a more familiar and more positive perception of their interaction.

The issue of context is also an important factor to consider when analysing the results of this study. As the interaction centred on instructions, participants would often highlight that VL was an

obstacle in them receiving and understanding the necessary information to conduct the task. This was opposed to the agent attempting to create a friendly atmosphere or identify as an unimposing instructor. Although the agent was taking on the role of the instructor, it appeared that direct language was seen as acceptable because the information it provides is all that is required to assemble the models. While this may not be true for human instructors for the same participants, there is again the link between familiarity, expectations, interaction realities, and identities. Participants who are familiar with being instructed by agents, likely in a direct style of language, may have had some expectations of the same directness here with this verbal agent. The interview data showed that there was also some support for using the vague agent in an alternative context to instruction giving, both from those who were positive towards the VL and those who were not. Conversational and leisure based contexts appeared to be more appropriate, such as when interacting with intelligent personal assistants. Given that social agents and robots are increasingly common in research, and in contexts that aim to build rapport over time (again see 6.2.3 for discussion on relational agents), VL may be a useful linguistic tool to achieve these aims.

4.5.3 Limitations and Moving Forward

Despite the findings there are still some limitations to consider for this study. Some points worth considering are those regarding the methodology. Firstly, the stress and no-stress task conditions did not show great significance; even though there was some effect on the number of repeats and how direct agents were perceived to be. Even though there were numerous suggestions that a vague agent would perhaps be better utilised in a context where there is ample time for interaction, it is likely that instructional contexts have some inherent need for accomplishing tasks within a certain timeframe. Given this, it is assumed that a time limit may be best used as a constant time for all participants, or removed entirely. There are also limitations to consider regarding the sample size of participants (N = 30). Although there were consistent patterns in the qualitative data in particular, this may have contributed somewhat to the mixed results seen overall. The separation between those who took part in a session with both the vague and non-vague agent was also greater (N = 24), than those who interacted with just one of the agents (N = 6). Having the remaining six participants interact with both agents may have strengthened some of the findings that are presented in this chapter. Some of the question wording in the quantitative measures may benefit from being refined and more specific, and in turn provide a greater clarity in participants' responses. The third question on interaction preferences in 4.4.3, for example, referred to interacting with the voice again on a personal device. While participants did note that language and its clarity were still obstacles in wanting to interact again in these contexts, framing the question around the voice rather than the agent as a whole

(including the language) may have influenced the answers in a different direction than intended. This may account for some of the disparity between responses in the third question and the other two questions.

There was also some imbalance in the gender of participants, with more males than females represented here. This may have had an effect on the overall patterns seen throughout the data. So too was there an imbalance in the educational background of the participants. The population consisted mostly of computer science students and did not reflect as diverse a background that it could have.

Overall, despite some mixed responses and limitations, the data reveals that there may be potential uses for a vague agent giving verbal instructions; however there are numerous obstacles to first consider. There appears to be a clear imbalance between the voice used in the agent and the VL it was using for the instructions, in that there was rarely praise for the synergy of the two. The numerous calls for a higher quality of voice if VL was to be used were a strong indication that synthesised voices that are deemed to be of a significantly lower quality than human speech are not best equipped to start handling language that has numerous levels of social uses. Although other factors such as the instruction based context cannot be ignored, it appears the logical next step would be to analyse the effects of a similar task with a higher quality voice. This would provide an investigation as to whether it is indeed the voice that is the biggest obstacle in the use of VL for verbal agent instructors, whether it is the context of interaction, or perhaps the wider social implications of the blurring of lines between agent and human likeness.

It is clear that the non-vague agent was relatively successful in its interactions. Participants often expected the direct language and the synthesised voice appeared to be better matched up with the direct language than the vague, particularly with the instructive context of model assembly. This provides some valuable data that if a lower quality voice is to be used in a verbal agent instructor, for whatever reason, then direct language is probably the style to opt for.

For both the vague and non-vague agents participants projected various identities onto both which often saw some consistencies. The non-vague agent was often expected and seen as familiar, which gave some sense of continuous identity for participants as their expectations had been fulfilled. For the vague agent, there was often an identity of a machine trying to be human and not succeeding. Although some of the VL was seen as positive, for example when minimisers would effectively minimise an instruction, often there were not expectations that such an agent would be using such a humanlike style of language. With the short time in which participants interacted there

was little time to become accustomed to the nuances of the vague agent, and it was often received negatively.

4.6 Summary

This chapter discussed the specific methodology and findings of the first study. Using the general approach described in Chapter 3, this chapter presented the nuances of the methods used in Study One, along with the experimental research questions and hypotheses. The specific procedure of the study is discussed, including details of the participants. The results are then presented. First, the quantitative data of agent characteristics, task performance, and interaction preferences were provided. These were followed by the qualitative analysis of the participants' interviews. Finally, the discussion of the results and limitations of the study were presented.

5. Study Two: Comparing Synthesised and Human Voices in Vague Verbal Agents

5.1 Introduction

This chapter focuses on the second investigation into verbal agents using vague language (VL). First, the findings of the previous study are discussed and how these can be used as a foundation to build the second study. This includes not only using the results to drive the next research questions and hypotheses, but also to refine methods on methodology and data analysis so that an improved study can be performed. This in turn aims to improve both the scope and depth of the data analysis. One of the main themes from the previous study that is focused on in this chapter is that of voice quality. As mentioned in the previous chapter this was the most salient point that came out of the qualitative data and one that participants were often critical of, particularly in regards to the disconnect between the voice and VL. This chapter will begin with a discussion on bridging the gap between the two studies before going on to briefly review relevant literature around voice quality in similar contexts. This will be followed by a description of a new approach towards exploring voices in this context alongside new research questions and hypotheses. The rest of the chapter will continue much like the previous one. First a report on the explicit methodology used is discussed. This focuses on increasing the amount of voices used and focusing on comparisons between synthesised and recorded human voices, followed by how this is implemented in a further series of Lego assembly tasks. Results and discussions of the data gathered from these tasks are then discussed, followed by views on its limitations, future work, and concluding remarks.

5.2 Reflections on Study One and Related Work

The first study revealed that there were a number of significant and frequently occurring obstacles in the interaction with the vague agent that hindered the successful use of VL, although some of these were also present in the non-vague interactions. The most pertinent of these was the quality of the agent's voice and the apparent disconnection between the language the vague agent was using and the quality of speech that was being used to produce this language. This was less apparent in the non-vague agent, although there were still comments regarding misunderstandings for the tasks where it was the instructor. There were also contextual and social implications of the vague agent that were obvious in the participant interviews. Given that the interest still remains the VL, and that there are qualities of the vague agent that can easily be manipulated to tackle some of the challenges that arose, the non-vague agent is no longer of interest in this second study. The first already showed that a non-vague agent with a synthesised voice, often of lower quality, is usually expected to use direct language and so

it succeeded in being appropriate for the instructive context of interaction. Although it was not significantly more likeable or friendly than the vague agent it did not need to be, and was still often discussed as being preferable option for instruction giving. Results from this study will also go towards reinforcing or diverging from the notion that non-vague verbal agents are the best choice for instruction giving.

This section will begin with a short summary continuing from the need to assess the emerging themes from the previous study, the most common of which was apparently the quality of the voice. This study aims to assess whether it is indeed the quality of the voice that impedes the successful use of VL in the agent or whether the voice, no matter how good it is, can not cover up for the inappropriate use of VL in this context, as well as if it again impedes upon the boundaries of being humanlike.

5.2.1 Voice Quality in Human-Agent Interaction

Before discussing the experimental questions addressed in this study, this section will readdress some of the discussion points on voice in human-agent interaction. Chapter 2 discussed some of the features that contribute towards the construction of identity in human-agent interaction. These included language, voice and prosody (2.5). The latter two are of particular interest in this study, which aims to compare the perception synthesised and human recorded voices as used by verbal agent instructors in assembly tasks. In total three voices will be compared. Two of these are synthesised and provided by two different commercially available text-to-speech (TTS) systems. A professional voice actor provides the final voice. A voice actor's recordings represent one of the best qualities of voice that can be achieved. Voice actors are professionals and may provide a consistency that might not be achieved with amateur recordings. The lack of consistency was sometimes a drawback in the participants' perceptions of the agents in Study One. The VL item *basically* was discussed more than any others. Cepstral Lawrence produced an inconsistent final phoneme that was not consistent with the rest of the word (4.5.2).

Discussions regarding voices in HAI in Chapter 2 looked at previous research in comparing human and synthesised voices (2.5.4). While there is progress with some synthesised voices, a human voice is often the preferred option (e.g. Cowan et al., 2012; Georgila et al., 2012). Georgila et al. (2012: 8) noted that a "high-quality general-purpose voice or a good limited-domain voice can perform better than amateur human recordings." They also argued that although the voice actor remains more preferable, the gap between amateur human recordings and synthesised voices has reached a point where the two may perform equally. This study did focuses on sentences, however. Although some interactions with agents will only consist of a few

sentences, others will have a prolonged interaction. Comparing the two lengths with both styles of voice would provide important information on if and how the two interactions differ.

The wider use of voice actors can have its drawbacks. If new output is required from the actor it will mean new recording sessions. This is also true if an amateur human recording is used. This is where one advantage of using synthesised speech can be found. With synthesised voices any utterance required could be inputted, likely with a text-to-speech interface, and be ready for use in a fraction of the time. Although synthesised voices will have voice actors used to gather the speech necessary to develop it, it does not require a new recording each time a new utterance is needed. Despite human recordings possessing the obvious benefit of sounding more natural (i.e. humanlike), synthesised voices in TTS systems allow for an arguably infinite number of utterances. All that is required is to type in the information that is needed. This is perhaps the biggest benefit and indeed reason behind using synthesised voices, particularly if human likeness in any aspect is a goal. If HAI contexts are to become more complex, dynamic and useful then it becomes less viable to use voice actors that would require more and more hours of recordings, and extremely beneficial to have a humanlike text-to-speech alternative.

Another study looking at low-cost TTS systems, and whether high quality voice systems are needed for tutoring, found that there was little difference between a low-end TTS system and a pre-recorded human speaker (Forbes-Riley et al., 2006). The main difference that did occur was in learning gains, specifically within their ITSPOKE system. Although there may have been little difference in this study, the results in the previous chapter suggest that there will be a greater difference when these voices are using VL.

There appears to be a closing in of the gaps between synthesised voices and human recordings, though the benefits and drawbacks of both are still apparent. There is no known indication of previous studies observing the differences in VL between these two types of voices, and whether this can inform future agent design. Even in previous studies looking at politeness strategies in HCI and HRI (Torrey et al., 2013, Strait et al., 2014) did not account for the quality of the voice being a contributing factor to its appropriate use and success in interaction. Given that verbal agent instructors may use either style of voice, it is important to understand the differences when concerning more sophisticated styles of communication. Voice is a strong indicator of identity and can influence the way in which we perceive a speaker (Latinus and Belin, 2011) and understanding the effects of similarity and social indexing of identity is important. Analysing these contrasts and similarities can provide a greater understanding of user preferences and inform future agent design.

5.2.2 Experimental Questions and Hypotheses

There are new experimental questions and hypotheses that this chapter aims to address. These build upon those addressed in Chapter 4, though instead they focus on comparing effects between three levels of the same variable (three voices) as opposed to two levels of two variables (vague vs. non-vague; stress vs. no-stress).

> ***EQ2-1****: Is there a difference in how synthesised and human voices are rated in regards to specific characteristics of the vague agent?*

This first question takes a similar approach to the first used in Study One, but expands the amount of characteristics being assessed. While eight characteristics were used previously, here there are nineteen. These form part of three groups of characteristics:

Group One: *likeable, would want to interact with the voice again, annoying*.
Group Two: *precise, coherent, intelligible, comprehensible*.
Group Three: *rude, imposing*
Group Four: *assertive, controlling*
Group Five: *anxious, enabled completion of the task, apprehensive interacting with similar systems*
Group Six: *humanlike*

There is some overlap between the groups and they are not necessarily being grouped together in the statistical analysis, but the groups represent the similar themes between the characteristics. In the actual questionnaire some of the questions are worded negatively so that there is some balance e.g. *likeable, intelligible* vs. *incomprehensible, imprecise*. Direct and authoritative were removed as they were for the analysis of the non-vague agent that is not being used for this study. Group Six is separate from the other eighteen and a group as such, but including *humanlike* will test the differences in quality between the three voices.

> ***EQ2-2****: Are there differences in performance of the tasks for the synthesised and human voiced vague agents i.e. time taken to complete the task; number of repeated instructions requested?*

This is the same research question from the previous study focusing on several performance metrics. This will again test the comprehension of instructions.

> ***EQ2-3:*** *Is the VL accepted more in the human voice or synthesised voice agents?*

Voice quality was such a prominent feature in the interview data last time, so this is one of the most important questions to answer. This will test whether or not participants accept the VL more when the voice is another human, despite them being unaware of it, and whether or not the differences between agent and humankind still emerge.

> ***EQ2-4****: How are identities towards the synthesised and human voices in vague agents presented by participants and what contrasts and similarities are observed?*

This is similar to the acceptance of the VL by the participants, but extends to include the agent as a whole. Again this is done mostly through analysis of the interview data and will again be looking at aspects of how aspects such as voice and language are described, as well as wider social implications of these.

Again, along with these research questions are hypotheses. These are grounded in previous literature and results from Study One:

> ***EH2-1:*** *Users will rate the voice actor agent higher in some if not all of the positive characteristics than the synthesised voice agents.*

By positive characteristics H1 refers to those that are positively worded and contain positive connotation, such as *likeable, coherent, intelligible, would want to interact with the agent again*. Although not all of these may be seen as positive, and similarly other characteristics that are not positively worded as negative, this is simply a means of grouping some of the characteristics together. These are the attributes that it is assumed a high quality voice will achieve, and those that are attributed to a positive identity based on the similarity between the agent and the participant.

> ***EH2-2:*** *Users will display a better task performance in some or all of the metrics in the voice actor tasks compared to the synthesised voice tasks.*

Because of the higher quality provided by the voice actor and the increased naturalness of the interaction, it is thought participants will be more comfortable in completing the task. They should also comprehend the instructions better in the voice agent tasks, requiring them to repeat less and complete the tasks in less time.

> ***EH2-3:*** *Acceptance of the VL should appear higher in the voice actor tasks than the synthesised voice tasks.*

Even though some degree of unfamiliarity and disconnection between the voice actor agent and its use of VL is expected, the gulf in quality should be so apparent that participants are more accepting of it using

VL. This will be analysed in the interviews where it is expected participants should describe these views.

> ***EH2-4:*** *The voice actor agent should have more positive identities projected onto it than the synthesised agents.*

Because of the acceptance in VL being higher in the voice agent tasks, the identities being projected onto it and described in the interviews should reflect a more positive collection of identities. It is thought the synthesised voices will be described in a similar manner to the previous study, whereas the voice actor should be seen as more natural and less encroaching upon human likeness.

Studying the differences between the synthesised and voice actor agents allows for the investigation as to the uses of either when dealing with VL, and hopefully wider linguistic phenomenon as a result. Synthesised voices have already shown to be a good option for when direct language is being used, but it remains to be seen whether a voice actor can provide the same for a vague alternative. Given the uses for both pre-recorded agents and text-to-speech systems, this study aims to highlight some of the contrasts between the two.

5.3 Method

Before discussing the specifics of the methodology for the second study, there are some elements of the design of the previous study that are addressed first. Firstly, the practice model was removed as this appeared to do little but increase the time of each session and the learning effect for each participant. Secondly, the approach to data collection and transcription was below par in the first study. Some of the video data was of poor quality due to camera positioning, which has been changed here, and the laptop camera has been replaced with a higher quality high-definition camcorder. With regards to transcription, the first study used salient quotes and extracts from the video data but not all of the interviews were transcribed in full. This has been addressed in Study Two. All of the interviews were transcribed in full using the transcription software CLAN and uploaded onto the qualitative analysis software Nvivo (see 5.2.5 for further details). This provided a simple method for storing, accessing, and coding the data.

To address the research questions and hypotheses outlined in this chapter another series of agent-instructed Lego assembly tasks were conducted. These were very similar to the ones used in Study One. While this section will still discuss the methodology in some detail, the main focus will be highlighting the differences between the two studies.

5.3.1 Agent Design

The interface that provided the instructions to participants was almost identical to that used in Study One (Figure 9). In this study, however, the two interactive buttons were moved further apart. This was to remedy the several occasions in the first study where participants would accidentally click the wrong button. If this were to occur too many times then the results of this study may be jeopardised. These were complemented with symbols indicating the function of each button to further prevent the accidental selection of either.

Figure 9: An example of the start screen in of the Study Two *Aquagon* interfaces.

Each interface again consisted of an HTML file linked to a library of sound files for each instruction, and in each voice. The difference for this study was that instead of .mp3 files .wav were used. This is because .mp3 files are compressed and lose some quality, whereas .wav files are a lossless format and do not have this compression. The only compromise is a larger file size, though this was of no concern. Apart from these changes the interfaces remained the same, and still contained the same information logs tracking the time for individual steps and the number of repeats requested in each task.

The instructions were also borrowed from the first study. The non-vague files were left out and only the vague were used. The instructions for the second synthesised voice were inputted through the same Text2SpeechPro software. Those for the voice actor were provided as one large .wav file. This was edited using the free software Audacity (http://audacity.sourceforge.net/) to create individual sound files for each step of the models. In total there were six different interfaces – one for each voice and model combination. The instructions were stored in the library of sound files linked to their respective interface.

5.3.2 Voice Continuum

In the previous study the one voice being used was Cepstral Lawrence (CL), a synthesised voice developed by the company with which it

shares its name (http://www.cepstral.com). In the research questions it is stated that the aim is to compare synthesised voices to a human voice. To achieve this, another synthesised voice was first required that displayed some differences to the original, preferably with some improvements, though gauging this was quite subjective. Secondly, a human voice was required to represent the highest quality of agent voice used in this scenario. This would allow for the development of a voice continuum, in which notable differences and to a lesser extent quality can be seen across the three different voices (see Figure 10.)

In deciding on the other two voices, it was believed that the voices should sound similar in regards to age and accent in order to remove any compounding factors that may arise. Unfortunately, there was no information as to either of these in regards to Cepstral Lawrence. In an attempt to ascertain these, interview data from the prior study was analysed combined with both researcher and non-researcher opinions on the voice, resulting in the definitions of its age being somewhere between 40-50 years old and its accent as Southern English RP. Another decision made to reduce the number of variables in this study was ensuring that all voices were male. This did not reflect any particular bias or attempt to skew the data, but was simply a measure employed to reduce variables and simplicity of data analysis. Data grouping and counterbalancing, for example, increases with the amount of variables being tested, as does sample size. Moreover, this thesis is not necessarily concerned with voice gender, more with providing some groundwork on attitudes, VL, and voice in HAI that can be built upon in future research. With these in mind, two more voices were able to be found to fulfil the criteria.

Initially, the second synthesised voice was to be one produced by Nuance (http://www.nuance.com/). However, the only similar voice available was that of Daniel, also known as the voice behind Siri and other Mac OS X devices. There was some likelihood that participants had already interacted with this voice before in some form of OS X device, whereas it was doubtful that Cepstral Lawrence had gained the same potential exposure. To remove any prejudice and bias that may have arisen from this pre-task exposure this was deemed an unsuitable option and discarded. The decision was eventually made to opt for the voice Giles made by CereProc (CP) (https://www.cereproc.com). This voice is described as a Southern English RP voice and, like Cepstral Lawrence, operates at a higher quality of audio than its predecessors (22 kHz). There was one alternative to Giles called William, but researchers and non-researchers again decided that the former was more suitable in regards to its similarity with Lawrence. These two voices were the first two steps in the voice continuum.

Similarity between the synthesised voices was desired to an extent, but there was also a need to highlight the differences between them.

Lawrence lacks any personal description from its developers, but the voices as a whole are described as "high quality and natural sound" and having the ability to manipulate phones, lexicon and prosodic features. Giles only had the personal description of "Southern English RP Male" but the voices as a whole had a richer depiction of their abilities, with their voices and text-to-speech software being described as the most advanced in the world. Though this is arguably a marketing ploy, there is also a description of the research and development into the "emotional continuum"[36] and its implementation into the CereProc voices.

To generate another layer of comparison between the two voices, an analysis of assembly instructions spoken by the two yield differences in their prosodic capabilities, which is featured in the preliminary iteration of the voice continuum seen below.

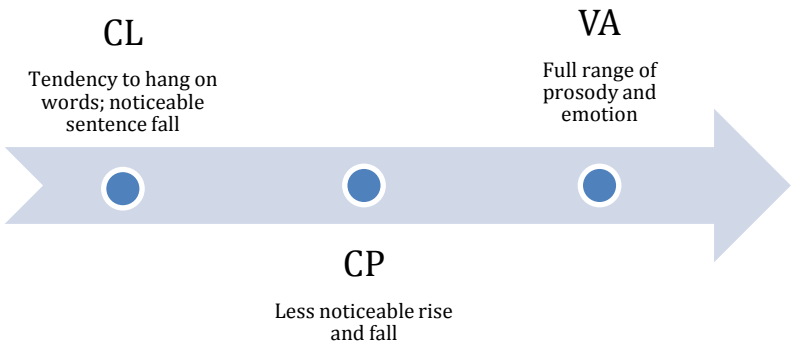

Figure 10: The voice continuum showing examples of prosodic capabilities.

Lawrence had a tendency to produce an unnatural hang on particular words, which was something also noted by participants in the first study, even with attempted correction of its speech. This meant that words such as *basically* would have a notably odd cadence about them. Similarly, there was a lack of sentence fall that is natural in non-question statements by humans. Giles, on the other hand, was able to produce a more natural sounding prosody, although still a far cry from the full prosodic capabilities of an actual human voice.

In choosing the human voice, the decision was made to hire a voice actor (VA). This was done through the company Voice Bunny[37]. Not only are voice actors used in the creation of synthesised voices, but a professional was deemed to have a greater likelihood in producing a consistent sound than an amateur recording. Again, there was little personal information available for the voices, though in specifying a request for English male voices aged between 40-55, three suitable matches were found. After a brief audition, it was again decided that

[36] For more information see https://www.cereproc.com/en/about/randd

[37] For the company website see http://www.voicebunny.com

the voice identified as "Mark" was the most similar to the other two voices[38]. All three voices had somewhat of a Southern RP English accent.

Though this completed the voice continuum shown above, there are some caveats to consider. This is not an accurate scale in regards to the quality between all of the voices. There are subjective measures to also consider in this regard. It is simply a visual representation in the perceived increase human-like features in across three different voices and their ability to produce speech. Similar to the discussion in 2.5.4, this voice continuum here can be seen as a cline, in which voices that are perceived to have common phenomena (e.g. human likeness) are clustered together.

5.3.3 Participants

A total of forty-eight L1 English speakers were recruited for the experiment and reimbursed with a £10 voucher for participating. Twenty-one participants were male (43.8%) and twenty-seven were female (56.2%) and together had a mean age of 24.2 years (SD = 5.56). Participants were reimbursed with a £10 Amazon voucher for their time. They were recruited through email advertisements and a study advertisement website[39].

5.3.4 Procedure

Before the sessions participants were randomly assigned to one of the 12 different group iterations[40]. Because they were required to interact with two voices out of the three there were three main groups comprised of all the pairing combinations – CL & CP, CP & VA, and VA & CL There was also the factors of voice order and model order to consider which in total created the 12 different group iterations to counterbalance the variables. As there were forty-eight participants in total there was an even balance across these iterations with four participants in each.

[38] Voice can be heard at http://voicebunny.com/projects/add_booking/5KKMKQC/856091

[39] https://www.callforparticipants.com/

[40] Each of the twelve iterations was numbered and https://www.random.org/ used to generate a random number between 1-12. This number indicated the iteration the participant was assigned to.

Table 11: The twelve iterations of voice and model order (A = *Aquagon*; N= *Nex*).

Voice Pairing A: CP + CL		**Voice Pairing B: CP + VA**		**Voice Pairing C: CL + VA**	
A1 - CP > CL		*B1 - CP > VA*		*C1 - CL > VA*	
A1 AN (1) IIII	**A1 NA (2)** IIII	**B1 AN (5)** IIII	**B1 NA (6)** IIII	**C1 AN (9)** IIII	**C1 NA (10)** IIII
A2 - CL > CP		*B2 - VA > CP*		*C2 - VA > CL*	
A2 AN (3) IIII	**A2 NA (4)** IIII	**B2 AN (7)** IIII	**B2 NA (8)** IIII	**C2 AN (11)** IIII	**C2 NA (12)** IIII

Each group was presented with two tasks and allows for both a between and within subjects analysis of their interactions. Table 11 shows the distribution of the participants amongst the different iterations. Counterbalancing the voices was particularly important as it has shown that the order in which a participant interacts with a voice can affect their choice of which voice to use again (Lee et al., 2011). While participants do not have a choice over the voice used, it remains an important variable to balance.

Each session had a time limit of fifteen minutes and participants were provided with a physical timer to keep track of this time. This was placed in front of them near the MacBook Pro 10.2 that contained the interface. The session ended when either the fifteen minutes had expired or the participant had completed each of the 48 steps of the model and reached the end of the task. Two camcorders were used to record each task. A Panasonic HDC-SD900 captured the close up shot of the participants' face and was situated towards the opposite end of the table where they were positioned. A Canon Legria HFR306 recorded from the side to capture both the nuances of the model assembly and the interactions with the interface. Although each camera had the capability to record in full high definition, early trials showed the file sizes to be too large to be practical. The .mp4 format, which has a significantly smaller file size, was used as a substitute without having to compromise too much on quality. In total a multimodal corpus of twenty-four hours of video and audio data was created.

At the beginning of each session, participants were briefed on how the tasks would be conducted and what was expected of them. A task information sheet was provided as the final checks for the task area were being completed. Finally, a consent form indicating each participant's willingness to complete the session, be recorded, and have this data used for further publications and research. They were also notified of their ability to withdraw at any point without having to provide a reason. There was on occasion where a participant requested not to be filmed during the tasks. They did, however, agree to audio recordings for the post-task interviews.

Following the signing of the consent forms participants were presented with the first task as decided by their randomised grouping. The two cameras were synchronised and the participants were

instructed to click the Start button available on the interface. This was also synchronised with the timer in front of them being activated which counted down from 15 minutes. Once participants either finished assembling the models, or the timer reached 15 minutes, the task was deemed complete. After each task participants were asked to complete a questionnaire and take part in a semi-structured interview.

5.3.5 Measures

In order to provide a preliminary evaluation of this discussion, a mixed methods approach was used to assess the data collected from the general experiment design. First, as mentioned previously, a questionnaire was given to participants following each assembly task. This contained the questions on the characteristics discussed in the research questions and used a five-point Likert scale (1= Strongly Agree; 5= Strongly Disagree). Some of the questions were negatively loaded and some were positively loaded, and the ordering of the scale flipped for half of the total participants. This meant that some participants had the scale as it is shown above and others had the reverse (i.e. 1 = Strongly Disagree; 5 = Strongly Agree). An additional open question on the user's perception of the age of the agent was also included. These were inputted into the statistical software SPSS following each session.

As well as questionnaires to assess user perception, the semi-guided interviews following each task were also analysed. The video data for each interview was transcribed in CLAN (http://childes.psy.cmu.edu/clan). This allowed for both transcription and viewing of the video and audio data to be conducted simultaneously on the same program. Once the transcriptions were complete they were exported as .txt files. Although forty-eight people took part in the experiment and took part in two tasks each, two of the tasks did not record as intended properly. This resulted in there being only ninety-four videos and transcriptions. The collection of these transcriptions, however, provided a written corpus of spoken interaction containing 60,000 words. The transcription files were inputted into the computer aided qualitative research software Nvivo[41] and analysed with an iterative content approach. Some of these were quantitative and some were qualitative. These files within Nvivo were coded first to indicate which variables each one belongs to, indicating the participant number (without a name to provide anonymity), task number, voice being used and model being assembled. Open coding on the actual interview data is then done to identify key themes throughout the data, such as voice quality from the first study.

The third quantitative measure is also used here. A profile is generated for each file to indicate the following:

[41] This software can be found at http://www.qsrinternational.com/product.

- Attitudes towards the agent's voice
- Whether they noticed any VL being used
- If so their attitude towards its use of VL in general
- Attitudes towards the specific VL items from the framework.

These were then categorised as positive, negative or neutral. This follows a similar approach to other research categorising qualitative responses (Becker-Asano et al., 2014) though focuses on creating a user profile. This ensures a fair spread of 96 items in each of these categories where applicable[42] rather than totalling up the descriptive adjectives being used in total throughout the interview data. This was intended to prevent some skewing of the data. It is thought this may have been a possibility if counting simply the adjectives each participant uses in their interviews, as some may talk more than others for a multitude of reasons and thus unbalance the data[43]. User profiling also allowed the introduction of a new measure of VL recognition as it was evident not all participants registered its use by the vague agent in Study One. It also allowed for some quantifying of this attitudes towards it, as well as attitudes towards the voice. A qualitative approach was still needed to accompany this but the combination of the two created a richer dataset. Quantitative coding focused on coding each transcription according to the overall perception of VL in the interaction and whether the participant had noticed its occurrence, as well as the agent's voice. Qualitative coding attempted to hone in on the reasons why users perceived agents as they did and if there were other patterns of attitudes and identities to be observed.

Analysing the task performance in each task was again conducted using the information logged by the interface. The time in which it took participants to complete the tasks was logged, as were the number of steps that participants had completed. This formed two groups for this metric – those who completed within the allotted time (time to complete) and those did not (steps completed). The number of repeated instructions requested was the same as in the previous study. These were logged on the interface and inputted into SPSS along with the other quantitative measures.

---

[42] By "where applicable" this means that if participants do not notice anything about the vague language use then they cannot by default have opinions on the general and specific use of vague language either.

[43] To expand on this point, it is impossible to fully understand how a user comes to a decision regarding their descriptions, even with rigorous personal profiling. However this method does help to understand how users have identified the agent, and to a less extent why they have done so.

5.4 Quantitative Results

In this section the quantitative results from both the questionnaires and interviews are presented. First, the task performance and survey measures are discussed. These are followed by the quantified data on attitudes towards VL and voices from the interviews.

5.4.1 Task Performance

There was no significance to be found in the differences between the three voices for either the time taken to complete the task or the number of repeats requested by participants. This was also true when analysing the separate groups in terms the two voices they were paired with. There was, however, a significant difference between the two completion groups. A one-way ANOVA revealed in the tasks that were incomplete (N = 24) there were more repeats (M = 10.75, SD = 5.42) than those that were complete (N = 72, M = 6.33, SD = 4.4), F (1, 94) = 16.08, p = < .001.

5.4.2 Survey Measures

A one-way between-subjects ANOVA was conducted to compare the mean values of each attribute used in the questionnaires across the three agent voices. These were all followed with post-hoc Bonferroni corrections. The ANOVA revealed that there was a significant difference in the likeability of the voice, F (2, 93) = 14.77, p = < .001. Post-hoc Bonferroni corrections showed that VA was significantly more likeable than CL (p = < .001) and CP (p = .001). Similar significant differences were observed in how annoying each voice was, F (2, 93) = 8.68, p = < .001, with VA significantly less annoying than CL (p = .001) and CP (p = .002). Significant variation was found in how coherent the voices were rated (displayed as "Incoherent" in Table 12), F (2, 93) = 3.43, p = .036. VA was rated as significantly more coherent than CP (p = .033), though no other differences between voices were observed. Ratings of kindness were significant, F (2, 93) = 3.36, p = .039. Bonferroni corrections, however, revealed no further significant differences between the voices, though the difference between VA and CL was close (p = .058). A significant difference was observed in how much each voice enabled participants to complete the task, F (2, 93) = 4.24, p = .017, with both VA (p = .04) and CL (p = .04) rated as significantly enabling task completion more than CP. Finally, there was a significant difference observed in how humanlike each voice was rated, F (2, 93) = 15.004, p = < .001. VA was rated as significantly more humanlike than CL (p = < .001) and CP (p = < .001), whereas there was no difference observed between CL and CP themselves.

Table 12: ANOVA Results for Study Two.

Voice	Likeable***		Annoying**		Incoherent*		Kind*		Enabled Completion*		Humanlike***	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
CL	3.31	.965	2.72	1.143	3.47	.950	3.06	.914	2.03	.999	3.88	.976
CP	3.66	.971	2.78	1.008	3.25	.842	3.00	.672	2.69	1.223	3.88	1.238
VA	2.34	1.066	3.69	.965	3.84	.954	2.56	.914	2.03	.861	2.50	1.244
TOTAL	3.10	1.138	3.06	1.122	3.52	.940	2.88	.861	2.25	1.076	3.42	1.319

*p values: * = p < .05; ** = p < .001; *** = p < .00001*

Significant differences observed in the characteristics from the questionnaire. The mean averages (M) and standard deviations (SD) are included. Lower mean scores indicate a higher rating for that characteristic.

There were also three significant results observed when comparing those who completed the tasks and those who did not. Agents in the completed tasks were rated as more comprehensible (M = 3.63, SD = 1.03) than agents in the incomplete tasks (M = 3.08, SD = 0.93), F (1, 95) = 5.24, p = < .05. Similarly, complete task agents were seen to allow the completion of the task (M = 2.07, SD = 1.03) more so than incomplete task agents (M = 2.79, SD = 1.06), F (1, 95) = 8.77, p = < .01. Finally, participants were more apprehensive about interacting with agents when they did not complete the task (M = 3.08, SD = 0.93) than when they did (M= 3.71, SD = 1.03), F (1, 95) = 6.98, p = .01.

5.4.3 Vague Language and Voice Perceptions

In assessing the perception of both VL and voice the frequencies of the general attitudes coded in each transcription were totalled. Table 13 shows the frequencies of users noticing VL across the three voices. Note that the total of 94 is the result of missing interview data as a result of a recording error.

Table 13: A comparison of VL being noticed or not across each voice condition.

	Yes	Unsure	No	TOTAL
Cepstral Lawrence (CL)	20	4	9	33
CereProc Giles (CP)	16	3	13	32
Voice Actor (VA)	15	0	14	29
TOTAL	51	7	36	94

In comparing these it was found that in the majority of tasks participants noticed VL, but there remained a significant number where it was not. There was a slight majority in the number of times it was noticed in CL and similarly when it was not noticed in VA. Although there are slight majorities and minorities for either side, this represents a slim confirmation that participants would be less likely to recognise VL in the human recording.

Table 14: Frequency of positive, neutral and negative attitudes towards VL across the three voices.

	Positive	Neutral	Negative	TOTAL
Cepstral Lawrence (CL)	1	4	15	20
CereProc Giles (CP)	1	3	11	15
Voice Actor (VA)	4	8	3	15
TOTAL	6	15	19	50

In assessing attitudes towards VL in each voice there is a clear disparity between the numbers in the negative column (Table 14). There were far more instances of VL being seen as a negative feature of the agent when being used with the synthesised voices. However, despite this favourability there were low numbers in the positive reactions to vague language in VA and indeed in across all three voices. Also, in total only 50 of 96 interactions showed any indication of attitudes towards VL despite the previous table showing 94 mentions of it. Although this initially this seems like a mistake and possible miscoding of data this is likely a result of participants alluding to or mentioning VL but in context of the voice, and not being able to separate the voice from the language or the voice from the agent as a whole.

Table 15: Frequency of positive, neutral and negative attitudes towards the three voices in general.

	Positive	Neutral	Negative	TOTAL
Cepstral Lawrence (CL)	1	12	12	25
CereProc Giles (CP)	1	16	12	29
Voice Actor (VA)	18	8	3	29
TOTAL	20	36	29	83

When looking at the attitudes towards the voices separate from the language, there was a large difference between how the synthesised and human voices were perceived (Table 15). CL and CP synthesised had a greater number of both neutral and negative attitudes given in the interviews, whereas VA had a very large majority of the positive attitudes. The almost even distribution of neutral and negative attitudes towards CL and CP indicates there is somewhat of a middle ground (neutral to negative) where these voices fall. This shows that although there are barely any positive attitudes towards them, they still have non-negative attitudes present. The overwhelming majority of positive reactions too confirms what was suspected in the experiment design, but the leap in numbers from the VL attitudes show that there is still some resistance towards accepting VL in this context even with a voice actor.

5.5 Qualitative Results

This selection highlights the key emerging themes that have emerged from content analysis of the qualitative interview data. These are

discussed in the context of extracts taken from this data. In the extracts the participant numbers are identified (e.g. P11, P40), as are some occasions where the researcher asks a question (RES). To clarify which voice is being discussed in each extract the voice initials (CL, CP, VA) are also included.

One of the aims of Study Two was to understand how participants perceived the VL of an agent using different voices. Voice quality was again a frequent talking point in the interview data, particularly when comparing the synthesised (CL, CP) with the voice actor (VA). The data saw participants discuss the voice both in combination with the VL and without. As seen in the quantitative results the general attitudes towards CL and CP were often described in the neutral-negative side of the spectrum, whereas for VA attitudes were positive-neutral.

5.5.1 General Attitudes Towards Voices

The comments on the quality for all three of the voices did have a fair amount of comments that were seen as neutral. That is they did not appear to show strong opinions for or against the voice:

> *P4: Yeah it was fine. It seemed kind of like an old school computer voice like from the eighties or something. Yeah it was fine. I didn't particularly find it annoying but you know it also wasn't soothing or lovely or anything. (CP)*
>
> *P7: It just sounds like a computer voice so I'm kind of indifferent to the voice. (CL)*

Both P4 and P7, interacting with CL and CP, consider these synthesised voices as fairly typical for something that comes from a computer. P4 even relates it to an older style of voice, perhaps indicating their knowledge of and experiences with higher quality modern alternatives. They both display a sense of familiarity with these voices and this interaction appears to be consistent with their idea of what some types of computer voices sound like. This relation to other voices is described in different ways:

> *P5: It sounded like a machine, robotic voice kind of. So it wasn't that bad. (VA)*
>
> *P31: The actual tone of the voice itself was fine - no different from listening to a sat nav or that sort of thing really, which I suppose we've just kind of become used to so I didn't really notice it. (CL)*

P5 describes the voice actor as sounding somewhat robotic and machinelike, and as such generated a fairly neutral attitude, despite it not being a synthesised voice. P31 discusses CL as being similar to those used in sat nav systems amongst other things, which they

mention is something that "we" i.e. society has become used to. As a result, hearing that sort of voice in this interaction does not bother them too much. P31 has some sort of expectation as to what agent voices sound like, so much so that it has become a norm for them in this space of interaction.

For neutral attitudes towards all three voices, participants appear to either find the quality of the voice sufficient enough for the context of interaction and inoffensive. This may also be combined with a sense of familiarity and relation to other types of voices they have interacted with, or their sense of expectation as to what they should sound like.

Some of the neutral attitudes could be interpreted as being positive, which is a drawback of a subjective classification process. Describing a voice as "fine" could indicate the voice being acceptable and appropriate and therefore positive. However, it is not deemed as a strong positive response here, some of which are discussed further on this section.

There were similarities in the neutral attitudes towards all three voices. For the negative attitudes towards the synthesised voices and the positive attitudes towards the voice actor, there were similarities as to the qualities that were being described, either in a bad or good manner. The clarity of the voice is one such example:

> *P17: As soon as I heard the voice I didn't like it. It was just it was like they were drunk or something because it's so mumbled. (CP)*

> *P41: Yeah it wasn't particularly clear and also it was pretty complicated to understand. The accent wasn't normal. It was very fake and computerised. (CL)*

Both participants show that clarity is a decisive factor in how an agent's identity will be portrayed. P17 goes as far to liken CP to being drunk due its lack of clarity and apparent tendency to mumble instructions. CL was also seen as unclear by P41 who comments on the difficulty in understanding its instructions. Interestingly, although some participants were happy with the computerised voice, P41 sees this as a negative aspect of the agent, especially with its accent. They comment on this being "very fake". This is presumably either in comparison to what an accent should sound like in human terms, or their other experiences with higher quality voices in previous agent interactions. P14 shares a similar view with regards to the robotic qualities of CP:

> *P14: I think it's too robotic personally I think if you're going to have something like this telling you instructions you want it to be quite open, quite informal. (CP)*

For them CP's robotic nature appears to make it unable to create a sociable identity and is instead closed off and formal. This is despite the use of vague instructions, which is similar to the barrier of voice quality that was discussed in Chapter 4. Participants sometimes had different views regarding the negative aspects of the voice actor:

> *P7: I thought it was really annoying. It sounds like one of the recorded voices that asks you about your PPI claim. (VA)*

Instead of commenting on any apparent lack of clarity or quality, P7 instead relates VA to pre-recorded message such as those used in telephone marketing calls. This is still a comparison to other types of agents, only likely those that use human recordings instead of text-to-speech systems and one that has negatively familiar connotations. These negative attitudes towards the voice in general were rare. There were more instances of these attitudes being displayed in combination with comments regarding the language.

Positive attitudes were overwhelmingly seen in favour of the voice actor, though there were positive remarks to be found about CP and CL:

> *P2: It was easy to understand. It was quite clear and I could hear exactly what it meant. (CP)*

> *P4: This one I found much less annoying…there was less of a gap between the human like syntax and stuff and the actual voice. It was still kind of like electronic enough that I didn't feel like it was trying to pretend to not be a machine. (CL)*

P2 simply states that there was enough clarity in CP that it was easily understandable. Admittedly this is not a huge increase in positivity from some of the comments that were coded as neutral, but this attitude moves away from indifference. With P4 their attitudes towards CL are somewhat produced by the comparison between their first task with CP, as well as being influenced by the language. It was a positive contrast from CP, owing to the more successful execution of what P4 perceived to be humanlike syntax, which resulted in a less annoying voice. Again, this is not overwhelmingly positive but there is a marked difference between this and indifference. Overall, these were infrequent between both of the synthesised voices, owing to the negativity surrounding their noticeable machinelike qualities.

With the voice actor the quality was distinct and noticeable:

> *P6: Well I thought that voice was more approachable than the first one because it was kind of less robotic and more human than the first one. (VA)*

P32: The voice seemed nicer, more polite…like BBC News or something. (VA)

P40: It was more natural…like less clunky or stilted. It sounded more like a real person and the way a sentence was read out just sounded more natural. (VA)

Clarity was again an influence on attitudes towards a voice. The voice actor was often described as being clear which contributed towards its ability to sound more natural and humanlike. P40 describes VA as sounding more like a real person, indicating both the differences between VA and the synthesised voices, but also that they don't identify VA as a real person. This may be because they see it as a recording rather than being of a quality closer to human likeness. Although this is not specifically mentioned in the interview, they do later describe their ability to compare VA to an actual person rather than a computerised voice. Nevertheless, it is the quality of pronunciation in particular that creates this positive image. P32's comments show a comparison between other human voices instead of other agents or computer devices, and in turn describe VA as being more polite and nice than CP in their first task. P6 also sees the voice as more humanlike, which made it appear more approachable. These comments are similar to those where CL and CP were viewed negatively because of their lack of human likeness. This would indicate that for some individuals the more similar an agent's voice is to a human the better, at least for its clarity and pronunciation if nothing else.

5.5.2 Combined Effects of Voice and VL

The attitudes towards the general voice of the agent were difficult to isolate from its other features, particularly the language. This is the crux of the decision to investigate the differences between synthesised and human voices. Extracts of interviews used in this section are longer than those used so far and, although they may seem to be tailored towards specific individuals, the details help to emphasise the various observations being made. Participants other than those being quoted often share these observations.

For the synthesised voices similar problems of disparity between voice and language arose:

P4: It sounded like very human syntax but from this very electronic kind of voice, so I think that was the most annoying thing about it…there was a clear gap between how it sounded and what it was saying. (CP)

P29: I mean regardless of like voice. If this was like a voice with a better personality…if it were more human I think it would be

more acceptable but as it is just a robot I'd just it's a bit jarring. (CL)

P44: I think it's just like how it was robotic in it tone it just makes it sound more weird. (CL)

P4's comment summarises the problems that were present in Study One. The disparity between the voice and the language, where there is a disconnection between the synthesised voice and the humanlike language, was also seen in both CL and CP in this study. P29 suggests that improving the quality to make it more humanlike would create less of a disconnection, while P44 comments on the strangeness of having a robotic voice with VL. Again the VL is being assessed along with the perceived robotic nature of the agent and participants sometimes create an out-group identity for it. Because it is "just" a robot it is a strange experience for it to be using language not considered robot like.

Participants were sometimes able to split the voice and the language, and expressed differences in how it contributed to the overall age the agent:

P45: I kind of split it. I split it the sound with what was being said. So in my mind what was being said was kind of a very young approach but the tone of the voice was very adult, so they kind of felt like the script and the voice were two separate voices so to speak. (CL)

Again there is a disparity between the older tone of CL and the younger associations of the vague script for the instructions. In this instance P45 sees this as almost creating two different voices that have two different ages. This was not necessarily a criticism of the agent overall, and P45 does later describe the voice as "pleasant enough". It does, however, further highlight the disconnection that exists for many with the synthesised voices.

Although there were numerous comments displaying negative attitudes towards CL and CP using VL, there were some participants who reacted positively:

P31: It was starting most of its sentences the same way with so or now or whatever it said and that was quite informal. It felt more like sort of natural speech which was a little bit jarring at first because it was natural speech from a very obviously not natural source, but actually once you got used to that it was clear enough… it tried to make it sound too precise you'd worry about getting it exactly the right angle whereas it says pretty much you feel like you, you know, you're open to a bit of, a bit of freedom there. (CL)

For P31 the language took some time to get accustomed with, though saw the use of informal VL as a positive as it allowed for more freedom in assembling the model. This is similar to the appropriate use of VL in Study One. If the vague item was consistent with the action a participant is undergoing then it was sometimes seen as a useful addition, as it allowed for more interpretation and imagine for them. The use of natural speech seemed to become more familiar with P31 throughout the interaction, even if it was an obviously "non-natural source". It would appear that even though there was little expectation of CL using such language, it just required time to produce a sense of familiarity and acceptance.

As expected it was more accepted in the tasks where the voice actor was the instructor.

> *P38: I don't think it helps with the instruction and completing the task but I think it does add to the sort of human feel of the voice and making you feel a bit more comfortable, as though it's a person talking to you and not a computer. (VA)*

The VL created a more humanlike instructor with VA, though P38 did not go as far to say it was a human voice being used. This was despite it not being appropriate for instruction giving or model assembly. Using an appropriate voice with VL can create a more comfortable interaction, if this is something a designer wishes to achieve. This also suggests that in other contexts this humanlike interaction may be received more positively than it is here. This human feel of the voice was also recognised as an effort being made in the agent design:

> *P6: I felt like there was an effort made to have it a little bit more human sounding in the way it spoke as well… it felt like it was trying to not be too cold.(VA)*

Although short of full appraisal in the voice sounding warm, the combination of VL and VA appeared to have less of an imposing identity here. They also described their mind "glazing over" some of the VL elsewhere in the interview, indicating that the use of VL is not out of place and less noticeable than in the synthesised tasks.

In other tasks, however, the VL was glazed over completely:

> *P42: It was clear and didn't sound like it was being rude or forcing you to do something. He knew what he wanted you to do and it made you realise you know rather than kind of umming and ahing…straight to the point which I liked. (VA)*

Interestingly in P42's previous task with CP they describe it as "all umming and ahing" i.e. not being direct and to the point, assumingly

because of the noticeable VL. In their second task with VA it appears to have directness without imposition. This could be the result of the language blending in the voice more than it did in the first task. Although it is seen as being straightforward, the lack of imposition suggests that the VL did not pass by completely unnoticed.

While there was evidence in the qualitative data to suggest VA is better suited to VL than the synthesised voices, there were still reservations in how appropriate it was:

> *P30: It was a bit funny hearing some voice that seems so formal and robotic saying something like looks a little bit like this and something a bit like that, it was like it was a strange combination. (VA)*

Although human, the voice still sounded robotic for some participants. This then still had a similar disparity between that and the language as was seen with CL and CP. P30's description of this being a "strange combination" points to unmatched expectations once again. P19 describes it as "interesting" again because it sounds computerised yet uses human expressions:

> *P19: It's quite strange because you can tell that it's not a person but uses the same kind of expressions, as we do like it said like every now and then and stuff like that. It sounded like somebody had been recorded talking and then it had been made computerised. I don't know, interesting…* It's odd because you expect it to either be a computer or a person and with it being that strange mixture…. *(VA)*

The reaction is almost one of intrigue, although the combination of a supposedly machinelike voice and humanlike language is not perceived as negatively as many of the CL and CP tasks. The voice neither fits the apparent conditions of being a computer or a person; instead it is a "strange mixture" of both. Although this was a somewhat neutral reaction, for others even an agent with a voice actor producing the speech was not sufficient:

> *P23: It was the same as the other one it was using a lot of so and basically. It wasn't formal enough for me…I prefer the formal language because it's not a person. (VA)*

P23 displays their preference for formal language rather than VL because the instructor is "not a person". Even with a professional human recording the interaction is not with a human, and so the agent does not coincide with P23's expectations of language use.

Another participant had similar thoughts of the VA agent not sounded humanlike enough:

P7: I would either prefer that or you know proper I'm a robot instead of kind of a fake halfway thing… Ones that sound actually humanlike or are actually recorded snippets of a person speaking.

RES: You said something that's quite humanlike. What do you think it takes to be humanlike?

P7: I think a lot of it is the smooth flow of words…actually being able to have some kind of cadence to the sentence, so every word isn't said exactly the same or monotone… able to have inflection throughout the sentence and being able to say it at different speeds…being actually to place emphasis on different words in a sentence, dependent on what the point is… which is really difficult to achieve for a machine obviously. (VA)

For them a humanlike voice requires the use of prosodic variety and vocal output, which the voice actor does have. P7 describes their preference for such a pre-recorded human agent or an actual robotic voice, rather than "a fake halfway thing" that they perceive VA to be. They do not recognise the pre-recorded nature of VA, nor its prosodic capabilities, and so attempts at using VL are not well received. This shows the strong association between prosody and identity. For P7 there is a certain association between variety of speech output such as emphasis, and differences in pitch and speed, and being a human. However, prosody is likely not enough, as even with VA they fail to recognise it being a pre-recorded human despite their preference for one. It is perhaps their first interaction with CL that altered their perception of the VA voice, though it is also possible that there is something about the physical makeup of the agent or the interaction space that is preventing their recognition of its capabilities. Elsewhere in the interview they describe its attempts at creating emotion:

P7: If you can't attempt emotion successfully, you probably shouldn't attempt it at all. (VA)

This comment points to VA sounding like a halfway point between a pre-recorded human and a computerised voice, and unsuccessfully attempts at creating emotion because it lacks the full condition of human likeness. In their task with CL they had no such qualms with attempts at emotion, likely because of its obviously synthesised nature.

There were other instances where participants did not believe VA was the most suitable voice, though it was still seen as a preferable option when comparing the two tasks:

RES: So do you think it's still a good idea to have that kind of language in there even with that kind of voice?

P28: I think it can work, because that one worked a lot better. Maybe it just needs to be the right kind of voice for it.

RES: So I'm guessing that one's a bit more suitable?

P28: Yeah. (VA)

P28 mentions the language may just need the right voice. While they show VA as being more suitable, they stop short of saying VA is the appropriate choice here. This may also be an example of the context of interaction not being suitable for the use of VL, even when using a voice actor in an agent.

5.5.3 Identifying Agent and Human Likeness

Comments regarding the humanity of VA highlighted the complex nature of identifying appropriate voice, language and contextual combinations, as well as where the lines between agent and human exist. They were more common for CL and CP, where the question regarding attempts at being human were often raised:

P29: There were lots of basically or effectively and you know little colloquialisms, which I guess were supposed to make it sound more human but just made it sound weird. It made me a little bit uneasy because it was trying to be human but it wasn't human…If a robot is going to be giving me instructions I'd sooner have them like do this, do that as opposed to like well it's basically a bit like you know just put it in like you know whatever. (CP)

P30: Strange when you're hearing it from something that seems so unhuman. (CP)

Here, P29 is familiar with robots using direct language and expects the same in this interaction, particularly with an instructional task. As such they are not appreciative of the VL. They prefer a "robot" being direct in their language use rather than attempting to sound like a human. This is something they believe CP is unable to achieve. Comments on the non-vague agent from Study One reflect the expected direct identity of the language used by synthesised voices, and with CP here this is a similar outcome. Interestingly they describe the use of VL as attempt to sound more human. This is a common theme but it is worth noting that all verbal communication in HAI, HCI and HRI has some basis in human communication – humans can be and are often direct themselves. It may be the undertones of social engagement in the VL that is more entrenched in human interactions than in human-agent interactions. As the CL and CP agents in particular were not associated with VL their use of it sometimes caused confusion:

> *P46: I think it said like a couple of times which isn't what I expected a robot to say so I was then thinking it was saying white or something. So I wasn't expecting a robot to speak like a person so I was trying to figure out what it was saying and realised it was just saying "like"… Because you're not expecting it from a robot, I was there trying to interpret it as something else, but it can add confusion so I would've preferred if it had just been I don't know step one do this, very kind of like a robot. (CP)*

P28's expectation of a "robot" is such that they will not use VL or "speak like a person". This causes them to misinterpret the word *like* for something such as *white*, which was the colour of some of the pieces in the Nex model task were they undergoing. Again, they describe robot like speech as being incremental (in instruction giving) and direct. Having a voice such as CP attempt words such as *like* with unclear pronunciation was not a welcome addition. In other examples the VL was still seen as more human, but not necessarily more natural:

> *P33: I don't know if it was more natural, it was more human. I think the first voice felt like it was attempting to be more natural than it actually was whereas this one seemed to have a lot less of the vague language… I was more in tune with what it actually was as opposed to the first one which seemed to be trying to be something it wasn't i.e. this one almost counter intuitively felt more human by not attempting to be so human. (CL)*

P33 noticed less VL in the CL task than the CP, despite there being the same amount. The use of VL created a more humanlike identity for the agents, though it was not always seen as a natural interaction. Moreover, the use of VL in the CP task was disproportionate, as it appeared to be trying too hard in being human. The CL task, however, seemed more suitable in its use of VL, to the extent that it was perhaps glazed over rather than being a focus of interaction. It had more of a firm identity rather than "trying to be something it wasn't". There may also be the factor of the task order and model being assembled to consider. This was another example of Nex being T1 and Aquagon being T2. Participants would often comment on the relative difficulty of Nex, so perhaps the recognition of VL is greater due to an increased focus on the instructions.

Although P33 was able to have a firm impression of what CL was, for others this was the opposite:

> *P15: I noticed that sometimes seemed a bit unsure of itself almost. Like where it's like you kind of should do this and it's like okay. Well it's a bit disconcerting because the voice is like robotic so you wouldn't think it would have like it would be unsure of itself or not quite sure what to say. (CL)*

The VL became too much of a focus and CL did not portray the sense of expertise and understanding of the task that was expected of it. The robotic voice is associated with knowledge and so indicating imprecision did not create an identity P15 would associate with this type of agent. They did, however, notice the same use and frequency of VL in their second task with CP, but commented on it being more acceptable as it appeared more humanlike and the language "less extreme".

Comparisons between the two agents participants had interacted with were common, and were one of the questions in the second interview. There were other examples of preference between the two synthesised voices:

> *P41: A lot, lot, lot better. I could understand the voice better and he explained it well. The accent was better. (CP)*
>
> *P45: I had a reaction right away to the second one. I don't like this voice, it feels like it's giving me things to do almost as if critical parent and the first one was kind of on my side in comparison but I didn't hear that way the first time round. (CP)*

P41 sees CP as having a better accent and overall being more comprehensible. P41 later described CP as being more humanlike than a computerised voice, in a similar vein to P15. P45, however, had a strong negative reaction to CP early on in their interaction. They described it as being like a "critical parent". CL, on the other hand, appeared to be on their side, though this was only observed in hindsight when comparing the two tasks.

Comparisons often extended to other agents and humans outside of those used in the interactions and included examples from different communities of discourse:

> *P19: It was like the difference between how a doctor talks to you and how the nurse talks to you kind of thing…Yeah it's just the whole language of the last one was more colloquial than this one, although I did like the fact that it specifically said like once you've connected those bits up this is the legs. That was nice. Useful to know, which the other one didn't say but also the other one I found the instructions easier to follow so I didn't necessarily need it. (VA & CL)*

Using a healthcare context as a comparison, P19 describes VA as a nurse and CL as a doctor. VA is seen as more colloquial than CL, despite having the same amount of VL. This provides a sense of a sociable and friendly identity of the VA agent, whereas the CL being a doctor conjures up one of being authoritative and clinical. This may have implications in other contexts where either of these identities are

something a designer wishes to portray in their agent. Further comparisons for CP and VA included well-known voices:

> *P8: Made me think of Stephen Hawking. (CP)*
>
> *P39: It's okay it sounds like Siri the phone so just sounds very friendly, very positive. (VA)*

P8 relates CP to the voice used by Stephen Hawking, which was a regular comment by participants. They also commented on the question of being humanlike in the questionnaire was not something they had considered, due to it being firmly agentlike and not humanlike. P39 relates VA to Siri, which uses voice actors in its development. Because of this association, their familiarity and positive attitude towards Siri is also reflected towards VA. Although both participants related the agents to other familiar voices, there is a great distinction between the qualities of the two. Stephen Hawking's synthesised voice has been in use since the late 1980s, while Siri is a product that originated in the current decade. This highlights one of the extreme gulfs in quality that users of these systems can perceive.

One of the interesting aspects in discussing the differences between the agent voices stems from the question on the agent's age in each questionnaire. This created an insightful discourse regarding agent and human likeness. Participants were asked to given an estimate age for each agent. Observations in the initial dozen or so studies saw that participants often wrote a small description of how the instructor sounded instead of approximating a figure to its age. This prompted an early hypothesis regarding the wording of this question. It was thought that the wording was throwing people off answering the question as intended i.e. a numerical figure, particularly because the use of the word "sounded" may have taken focus away from "how old". Similarly, not using the word "age" may have contributed to this. Also, this was the only question that did not use a multiple choice Likert scale. Instead of an age range that may have been more suitable, participants were given an open-ended comment box question.

Following this observation participants were asked this question in the interviews as well as in the questionnaire. Sometimes this elicited responses regarding the illegitimacy of trying to give an agent an age.

> *P17: I put twenty-five down on the page but I almost felt I wanted to write robot as if it wasn't a real person. (VA)*
>
> *P19: I think this one kind of sounds older… but it might be because it wasn't as friendly and your friends are generally the same age as you…again like not really any age because they're neither of them are proper voices and I don't know whether maybe it's just the language that was. (CL)*

P37: Seems a bit weird. It doesn't feel right for an age to a voice that's robotic. (CP)

Giving an age appears to be a challenge for some of the participants. Although others can liken it to other people such as lecturers, grandparents, friends and celebrities alike, others see an agent (usually verbalised as robot, computer or machine) as something that cannot have an age. This would place having an age something this is part human likeness but not in agent likeness. P19 does acknowledge that it sounds older but because it was "not a proper voice" then it is hard to make an age for it. Interestingly they say both of their interactions were with something that were not proper voices, with includes VA first followed by CL, so even though the VA was such a high quality voice provided a professional human, they still class it as being improper. This suggests that it is not solely the voice that has an effect on user's perception and perhaps the modality of interaction and what the interface consists of. Also, as noted several times in previous sections agents can consist of both synthesised and human voices, so perhaps even a human recording in an agent interface is still firmly seen as non-human, possibly somewhere in between.

In one instance P17 discards the possibility of CP having an age and provides a robotic alternative:

P17: So yeah that's the thing I guess it's almost as if a human has an age as in the years they've lived but a robot has an expiry date. (CP)

An expiry date is seen to be more fitting than an age. In the questionnaire following P17's task with CP, they were close to writing "robot" instead of a numerical figure when prompted about its age. This shows that the identity they have formed for CP here is incredibly non-human and this non-humanness appears to correlate with their dislike for the agent.

Difficulty in assigning an age was also sometimes the result of the prosodic features and vocal output of the agent:

RES: Did you struggle with the fact that you had to put an age to the voice?

P22: A little bit yeah…It's not very humanlike. It's very monotone. (CL)

P38: For the other voice the terminology as I just said, sort of saying like and just and also the more upbeat, cheery nature, I placed it a lot younger. (VA)

P38's interaction with VA was one in which they heard an upbeat voice in combination with terminology they associate someone younger. For P22, the monotone prosody did not provide much insight and took away from creating a humanlike sound. P38 shows that once again the language and the voice can be two separated individually as well as combined together to affect the perceived age and identity of an agent:

> *RES: Why do you think it's difficult to place on age on any of the voices?*
>
> *P43: Because unless it's a human voice with tone you can't judge it. The only reason I mentioned I put it maybe slightly younger than I would is because they used, they said like a bit and it was a little bit more conversational. I think anyway. (VA)*

Again the language is making the voices appear younger. Despite VA's humanlike qualities it is far enough removed from being human for P43 that they cannot assign an age towards it; instead they can only compare it as being younger to CL.

In some cases participants gave specific answers to the question of age, though again the language and tone of voice of the agents were contributing factors:

> *RES: How old did you say this one was?*
>
> *P46: Thirty-five. Because it reminded me of like a TV presenter, that's all I could picture. The other one I put eighteen. Not because of the actual tone of the voice but because of the language that was being used. I just imagined someone sat there with a recorder that was changing their voice into a robot trying to describe how to put this together so that's why I put that. (CP)*

The language was again something that sounded like a younger speaker, and because it was noticed more in P46's first task with VA, they perceived it to be younger. Their interaction with CP was likened to a TV presenter and their familiarity with them affected the age they put down. Interestingly with VA, the language sounded humanlike to the point where P46 likened it to someone reading the instructions and changing their own voice "into a robot". This shows both their strong association with other people, but the manner of the interaction with an interface preventing a full association with human likeness. This may also be the lack of other interaction modalities:

> *P19: I think not having a picture means that you can't really see the age. (VA)*

P19 mentions elsewhere in their interview that identifying the age of a person without seeing them, such as in a telephone conversation, is

difficult as there is only the voice available to make an assessment. As only the voice and a computer interface are being used here it is perhaps understandable why people have difficulty in assigning ages even with the VA agent. Also, this may be part of the reason why participants struggled to see VA as a pre-recorded human and considered it wholly or partly robotic.

The extracts on participants' reactions to the agent's age provide some interesting insights into how they create identity for them. While they do link to some of the other themes seen here this one is particularly intriguing, as some explicit comments on robotic like qualities relate to not being able to have an age, a concept attached to life and indeed to humans. This would suggest in this sense they are still in the realm of robots, machine and perhaps just tools. They are at least in a non-biological category particularly when participants describe voices as robotic. Even when using the human recording, however, some participants still believe that is hard to provide an age for the as it is still just a machine. This raises questions as to whether a human voice alone is enough to create something humanlike if that were to be a designer's aim.

5.5.4 Other Interaction Effects and Continuing Themes

There were examples of the themes discussed in the Study One interview results that also appeared in this data. The frequency of the VL was one of these:

> *P32: Yeah it was okay. The instructions were clear, it was the distracting that it kept saying just and it used a few filling words…it kind of put me off a bit, it made me laugh.*
>
> *RES: So you prefer it without that I'm guessing?*
>
> *P32: Not in every other instruction, like if it was toned down a bit. (CL)*

VL frequency from the old emerging themes showing through here with "just" perceived to be said too much. They say it would be fine if it wasn't used so much, again reinforcing the theme that VL can be fine to an extent but not necessarily when it is seen as overused.

The appropriate use of VL was another theme that sometimes emerged in these results again:

> *P32: When it wasn't particularly clear that felt a little patronising. Just do this, but I can't, what do you mean just? (VA)*

P38: Yeah saying just if you manage to do it, it was fine, but otherwise it felt a bit like it was saying oh this is easy but you're like no it isn't. (VA)

This is similar to some of the data discussed in Chapter 4 where participants would indicate the consistency of the language with the actual undertaking of the step being an appropriate and beneficial use of VL. When it was not consistent, however, it was not well received, which is again the case here. The use of "just" as a minimiser when the instruction cannot be minimised in the eyes of the participant can come across as condescending. Again the individual needs and perceived difficulty of the task are factors that may need consideration.

Another familiar discussion point was of the interaction context. Instruction giving, even with a high quality voice actor as the agent's voice, is not always suitable:

RES: So you'd rather it just got rid of that kind of language and just had?

P33: Yeah, for instructional purposes yes, but that sort of language would be fine for something that's supposed to give you more human and intuitive, something more of a creative process rather than something that's supposed to be instructional. (CP)

Again language would be okay for something non-instructional, such as something creative, but here it is not seen as necessary. Interestingly P33 suggests that a creative process that involves a humanlike intuitive approach to a goal would be more suitable than a logical instructional, step-by-step approach. This again links back to categorising where particular language is suitable and shows some association with a logical incremental procedure with direct language and a creative one with indirect. Sometimes this was an annoyance though other times it was a passing comment on what could be improved:

RES: So you noticed like and basically in the first voice?

P21: I think so. Yeah, yeah I did I definitely noticed the first one I'm trying to think now whether or not.

RES: Is that annoying, good, bad, nothing at all?
P21: Err yeah that in the first one it wasn't annoying, I wasn't annoyed, but it was noticeable. Just as when you're talking to someone and they say like or basically quite a lot you notice it. It's but annoying not it's just it's unnecessary, especially with some instructions that are meant to be concise and for a purpose. There's no need for that kind of thing. (CP)

P43: Probably not appropriate for instructional, I want it to be pure information, but it wasn't off-putting or anything. (VA)

Because instructions are seen by P21 as to be concise, they see VL as unnecessary because it doesn't add anything relating to the instructions – at least those items they recalled. Interestingly, they thought the first CL task had more VL items than second CP when they have the same amount. Perhaps this is something related to the voice but it may also be the novel and unexpected nature of the VL being more noticeable to P21 and perhaps it becoming more apparent in the second task. While P21 believes there is "no need for that kind of thing" in instruction giving, P43 is somewhat more neutral in their attitudes towards the VL. Although they believe it isn't necessary, they nonetheless were not irritated by its use. This may be one of the key differences between the synthesised voices and the voice actor. When it was seen as unnecessary, it appeared more likely to cause a negative reaction in CL and CP whereas with VA the feelings would gravitate more towards the neutral end of the spectrum.

5.6 Discussion

Following on from Study One, this study aimed to evaluate the effect of an agent's voice on the user perception of VL in the same agent-instructed Lego assembly tasks. Three voices were used to create three different verbal instructors that guided participants in assembling Lego models. Cepstral Lawrence (CL) used in Study One was used again here, as was a further synthesised voice CereProc Giles (CP). To compare the effects of synthesised and human voiced agents using VL a voice actor (VA) was hired for the third and final voice. Each participant took part in two tasks with two of the three agent voices. They were given a maximum time limit of fifteen minutes to complete each task. After these they were provided with questionnaires to answer, followed by a semi-structured interview.

In looking at the quantitative measures it was found that the voice actor was perceived as significantly more likeable, coherent and humanlike than the two synthesised voices, as well as less annoying. This coincided with the general beliefs discussed in EH2-1, though it was not as widespread as first posited and was instead limited to four significant results. Similar to Study One, no significant differences in task performance between the three voices, and as such EH2-2 was not observed here.

EH2-3 posited that the acceptance of VL would be greater in VA than in the two synthesised voices. This was observed in the coding of the interview data. Participant interviews were read through and coded on various themes. One of these was whether they had noticed anything in regards to the VL being used. In doing so it was revealed that in 51 out of 94 interactions did participants notice VL. This was less than

expected, although still over half of the total participants. A follow up coding showed that the attitudes towards CL and CP's use of VL was more negative-neutral than VA's, which was positive-neutral. This numbers of positive reactions to VA's use was again less than expected, but showed a marked difference. The final hypothesis, EH2-4, was also observed – participants did display more positive identities towards VA than CL or CP.

The main focus of this study assessed the differences in perception and identity creation between the three voices, though mostly between the professional human voice actor and the synthesised text-to-speech systems. This included the general attitudes towards the voices and their uses of VL. The voice actor had a large majority when it came to the general attitudes of the voice. Its clarity in pronunciation, consistency, and human likeness all contributed towards this positivity. This shared similarities with previous research on human vs. synthesised voices and showed that there is still a strong preference here for the human voice. If improving an agent's likeability and perception of being humanlike is part of a designer's goal, then using a human recording may be the preferable option. If these are not of great concern then a synthesised voice may be preferable, even if only for logistical and financial benefits. The use of language, however, must be carefully considered when using such a voice. Participants sometimes had particular expectations of what words would be used in the agent, and these occasionally were not met. Instead of always referring to the pieces or the actual assembly, the use of VL was not always expected and sometimes caused miscommunication.

The positive attitude towards VA in general was also seen somewhat in its use of VL, though included some indifference. The contrast to CL and CP was noticeable, however, which indicated that voice quality does indeed have a large role in how users perceive vague agents. It appeared more natural coming from the voice actor. There was also less of a disparity between voice and language that was often seen in the synthesised voices, both in this study and the previous. Despite this evident contrast, there was still some resistance to VA's use of VL. Participants were rarely able to notice that it was a pre-recorded human being used, though some comparisons to other types of agents that use the same method were seen. On several occasions they would still refer to the VA agent as a robot or a computer, much as they would both CL and CP. It would appear that voice alone was not always enough to pull the VA agent away from the firm category of agent likeness that the other voices were in. The interaction being with an interface on a laptop is a likely contributor to this, as there is no embodiment or other attempts at human likeness in the agent.

Even in VA tasks participants would often refer to the agents as trying to be human by using the VL, which was seen often in Study One. With the synthesised voices this was more common and often the disparity

between the quality of the voices and the humanlike lexis would create a negative perception. Sometimes CL and CP would be preferred to one another, and were not always perceived negatively, but generally when using VL they appear to have the same drawbacks in their lack of human likeness. With VA's undoubted human likeness, however, there were still drawbacks. Participants sometimes still expected a certain style of language despite it being a pre-recorded human voice. There was likely some familiarity with verbal agents that use human voices, and yet there was not always an acceptance of it speaking to participants as another person would.

The barriers towards its acceptance partly come back to the concepts of agent and human likeness and how they are categorised by individuals. It would appear that when participants perceive the agent as having to conform to their expectations of what an agent is (e.g. their language use, voice), breaking these expectations causes a negative (or at least apathetic) reaction. The apathetic reactions are likely not too much of an issue, though the negative ones can be depending on the goals of the interaction on the part of the agents, their designers, and their users. With synthesised voices in particular, patterns in the data suggest that they are often expected to speak direct and to the point. Introducing VL sometimes caused miscommunication as it was straying from this directness. The highly computerised voices also show little evidence of having the ability to interact socially with users. As evidenced in the interviews there are numerous examples of CL and CP being described as "just a machine" or "just a robot" and as such needs not to try and create a social rapport with their users, as it is not believable. This is a contrast to the non-vague agent in Study One, which was believable, as it appeared to be acting with its constraints as an agent. Participants acknowledged the VL some of the uses in the interviews, including generating rapport and creating a more friendly and likeable persona, yet also noted that CL and CP could not achieve this. Even throughout the interaction the increase in familiarity was not often enough to change these opinions.

Participants had a greater tolerance for VA and likely a wider acceptance of what its constraints consist of, though there were still comments on it being stuck in between human and agent communication. The social effects of VL were noticed in some occasions, namely the friendliness and likeability it produced. Improving the quality of an agent's voice may create a wider space in which it is expected or believed it should inhabit. The perceived identities of all three voices were established in the agent's ability to be appropriate, successful and often operate within particular boundaries.

There were some similar findings in this study that were observed in Study One. Arguably the most important of these to consider is the context of interaction. Even with VA agent tasks and positive reactions

to the VL, the instructive nature of the interaction was often seen as unsuitable for an agent using VL. The increase in familiarity with an agent using VL, in that there were two vague agent tasks, did not do much to alleviate these issues. Successfully incorporating VL in an agent is not solely dependent on improving the quality of the voice. Designing agents with VL, or other communicative strategies in mind, are still likely best received with a recorded human voice. In a less controlled or more leisurely and creative context with a less rigid output and goal in mind, this effect may be even greater.

A more ambitious but potentially crucial goal would be to have an agent that can adapt to its user in regards to language use, especially if this agent were to take on multiple roles other than an instructor. While certain patterns can be analysed in studies such as this it is more difficult to cater for individual preferences. Until such a process is made easier, these data patterns are a useful approach to understanding human-agent interactions and building foundations for future design.

5.6.1 Limitations

There were again some limitations to note in this study. Although there were some significant effects seen in the quantitative survey measures, it would appear that a more thorough approach to designing such a survey is needed. Refining this may require reflection on these and other studies, so that future surveys may include questions on important findings (e.g. to what extent an agent met participants' expectations). The sample size was greater than in Study One and the data gathered considerably large. There is space for a larger size though, particularly when using quantitative measures.

Coding the qualitative data was a limiting factor due its subjective nature. Although care was taken to be thorough and consistent, other researchers may come to different conclusions of the data. Quantifying some of the interview data to highlight attitudes towards voices and VL was a delicate process that may be undertaken differently.

5.7 Summary

This Chapter presented Study Two that compared the user reactions to vague agents using synthesised voices (CL and CP), to an agent using a professional voice actor recording (VA). This tested the quality of the voice, which was one of the biggest obstacles in Study One. The specific method of this study was discussed in light of the general approach in Chapter 3. Improvements of the methodology from the first study were also presented. This included the creation of a 24 hour corpus of video recordings, as well as a 60,000 word textual corpus of interview transcriptions. The quantitative results were then presented, which focused on the agent characteristics as rated by the user, attitudes towards the voices, and attitudes towards the use of VL. This was

followed by a presentation of the qualitative analysis of participants' interviews. The discussion then revealed that the hypothesis of voice being one of the biggest obstacles in VL acceptance was partly correct, though there are issues on the nature of agents encroaching into the language associated with human identities to consider.

6. Implications for Current Theories in Language in Human-Agent Interaction

6.1 Introduction

Following on from the two studies and their findings, this chapter reviews them in light of some of the theories discussed in Chapter 2. The variation of linguistic variables and its effect on perceptions of polite and non-polite communication become evident. So too does the need for a wider linguistic approach when concerning the use of both vague and non-vague language in these ever increasing contexts, the latter of which falls outside the polite spectrum of discourse. Politeness is still relevant, and the ways to approach both this and relational work are discussed. So too are the methods of assessing identity which are seen to be intertwined with the other linguistic theories. Reflections on the use of vague language are also included, as are those on the CASA paradigm and Media Equation. The theories that focused on are politeness, identity, vague language (VL) and the Computers as Social Actors and Media Equation paradigms.

6.2 Politeness and Face

As discussed in Section 2.2, being polite in communication can often involve speakers using a variety of linguistic tools and strategies to not impose themselves on others (Brown and Levinson, 1987). This includes the use of VL (Channell, 1994, Cutting, 2007) as a means of saving face. This is the public self-image projected to others during social interactions and negotiated by all parties as interaction progresses (Goffman, 1967, Goffman, 2012, Goffman, 2002). One of the intended functions when creating verbal agent interfaces that instructed users with vague language was to emulate some of the polite communication used in human interaction. There were four categories of vague language being used that were described in Chapter 3. Adaptors and minimisers were the two that catered most to these attempts at politeness. Both were used with the intention of hedging instructions, though adaptors were also used to reduce assertiveness and minimise imposition on the participants and minimisers to reduce perceived task difficulty and structure the agent's talk.

The line of acceptable communication is drawn and redrawn between different interactions and within the same interaction as it progresses.. Interaction partners negotiate this temporary line and work towards the preservation of the face of both speakers and listeners alike (Goffman, 1967). Goffman notes that it is important to work towards preserving one's own face, which he sees as having a degree of self-respect, and one's recipients, which he describes as being considerate. This mutual acknowledgement lends itself towards tailoring language

in such a way that it avoids face threatening acts, such as with politeness strategies (Brown and Levinson, 1987).

The discussion of politeness in Chapter 2 highlighted the differences between two theories of politeness. Politeness$_1$ consists of emerging notions of politeness in-interaction that are influenced by individual and social factors, and that are negotiated in talk with regards to facework. Politeness$_2$, however, is the wider theory of politeness attributed by academics that sometimes ignores the individual interpretations of politeness in a given interaction and can strive to describe universal theories of the phenomenon. The study of politeness$_1$ can be achieved through conversation and discourse analysis. In the two studies described in the previous chapters, the interactions with the agent are one-way in regards to verbal output. Participants received spoken instructions but could do speak to the agent itself. This meant that it was only the participants who were under any potential face threats as well as mitigations through politeness strategies employed via VL.

Evidence from both studies showed the intended politeness used in the agent's instructions did not always manifest themselves in these ways for the individual participant. Particular agents, such as the one used here, rely on a limited, pre-determined output and lack the quality of judgements that human beings can make in interactions about what illocutionary forces should be used, and when. The pre-determined outputs in these two studies used VL to create attempts at being polite while providing instructions. As explained in the previous chapters this output was always fixed and could not be changed during the interaction to use any different amount of VL. This section will discuss how the findings in Chapters 4 and 5 reflect upon these attempts and what this may mean for politeness theory in human-agent interaction.

6.3 Face and relational work in HCI

In the first study that compared vague and non-vague agents using the synthesised voice Cepstral Lawrence (Chapter 4), the non-vague agent was often preferred over the vague agent its direct style in 63% of the interactions. Conversely, only 8% preferred the vague agent, while remaining 29% were inconclusive regarding any preference. As highlighted in 4.4.3, the 63% who preferred the non-vague agent were those who commented on the vague agent being "convoluted," "insincere," and "fake." Furthermore, in 75% of the vague interactions, participants said they would prefer a human voice, while this was true for 42% in the non-vague interactions. These findings provided evidence that voice can affect a user's desire to interact with an agent when given the choice, making it an important variable in this specific HAI context.

Given that voice was a common aspect in the interviews, as well as in previous research (see 2.5.4), the second study compared synthesised and human voices for the vague agent instructors. Similar themes were discovered for those tasks where synthesised voices were used, however a notable difference was seen when comparing them to the human voice. The tables in 5.4.3 show that 51 out of 94 (54%) of participants noticed the use of VL. These were shared somewhat evenly between the three voices, with CL having 20 of these mentions, CP 16, and VA the remaining 15. Breaking these numbers down reveals the similar theme of voice being an important factor for these vague agents. The comments on VL for the two synthesised voices were overwhelmingly negative with almost identical percentages (CL = 75%; CP = 73%), while the comments for VA were considerably less negative (VA negative = 20%; positive = 27%) but had a noticeable majority that were neutral (53%). The comments in 5.4 reflect the numbers, with VA being referred to as sounding more natural, in that the language correlates to the voice with a greater success than its synthesised counterparts, though it was not immune from criticism.

Agents can and manifest personalities and perform identities onto their users (Lee et al., 2006, Nass and Lee, 2001), as well as have identities created by their users. Paying homage to a user's face is not used in the same way as in human interaction for the reason that, with a lack of sentience, they lack a self-image they need to protect. However, it may contribute to positive perceptions by a human user, providing the voice is perceived to be of good fit, as well as avoid imposing themselves on their users. There remains the question as to whether or not facework is required in human-agent interaction if an agent does not have a face to protect. A lack of sentience and any need for self-preservation in social interactions creates difficulties in identifying appropriate language choice in a given context with a specific user. Agents lack the same abstract reasoning and understanding of how to negotiate the boundaries of social discourse, and so decisions on the attempts at politeness or impoliteness are dependent on their designers, and possibly any algorithms used to formulate the agent's output. There are also again the variables of individuals and context of interaction to consider, and as a result there is likely no definitive answer. Agents instructing users on assembly or directions with a view to facework may not be as necessary as in other contexts such as the handling of sensitive information, such as medical or educational settings.

Participants displayed preferences for voices and language types, however they did not indicate threats to their face. This might partly be due to the context of the interaction (instruction-giving), but may also indicate a relation to a wider agent identity – one that means face threats with agents do not carry the same weight as in human interaction. This may also be because of relations to other indexed categories of instruction giving that are typically direct, so that the

non-vague agent was matching to expectations of both agent identity and instructor identity. It is not to say that agent instructors cannot threaten the face of users. The change to the Tesco checkout (Chapter 2) in voice and language for being too bossy is a real world example of this, and reflects the work of customer-supplier interactions in business encounters (Koester, 2007). It appears that for user preferences towards the agent's behaviour was more of a concern than any potential face threats.

6.3.1 Application of the FTA Equation

Attempts to create universal theories of politeness in HCI face the same difficulties present in human interaction. There exists no inherent politeness, and so a politeness$_2$ approach will not guarantee successful execution of politeness strategies with agents and their users. However, conducting studies on polite agents does provide an insight as to other factors that can benefit or hinder the use of politeness.

Nonetheless, the equation for FTAs can be compared between HCI and HHI. It might be that x here represents a linguistic variable(s) that affects a user's perceptions of its identity or personality, their preference towards the variables that are being used, as well as any potential face threats. This can include whether the preferences towards agent variables are matched or not, which can positively or negatively affect their perception of the agent's identity (W).

If x then includes both face threats, and specific linguistic variables, the equation can be quite similar to the original incarnation discussed in 2.2.1. This equation, as written below, has the following values: W = weightiness of the FTA; D = social distance; P = distance of power; R = ranking of FTA in a particular culture; x = the specific face threat; S = speaker; H = hearer (Brown and Levinson, 1987).

$$W_x = D(S,H) + P(H,S) + R_x$$

The only caveat here would be replacing speaker and hearer with agent (A) and human (H). This implies the equation can be used in contexts where agents and humans are either the speaker or hearer, if not both:

$$W_x = D(A,H) + P(A,H) + R_x$$

In either format, the values of the equations are perhaps better thought of as a representation rather than attributing absolute numerical values to them. When applying these formulae to the context of the results from both studies, however, there are some amendments that need to be made. Firstly, the value of *x* presented in the original formula has little traction with these studies as participant

questionnaires and interviews did not indicate that the agents were performing any FTAs. Consequently, the weightiness of the FTA (W) also has little significance here. Given the results discussed the suitability and appropriate use of vague and non-VL with synthesised and human voices, the value of *x* may be more relevantly seen as a representation of both the vague and non-vague attempts at managing or not managing facework. Similarly, the value of *W* can then be considered the impact of these attempts, which take into account the remaining variables in the equation, which are still relevant in HCI and explained further on in this section.

One of the criticisms of Brown and Levinson's work on politeness was that it focused too much on the polite end of the spectrum (Watts, 2003). Even facework, some scholars argue, is often too concerned with mitigating face threats (Locher and Watts, 2005). Expanding facework to include the whole continuum was one of the key points of *relational work* (Locher, 2004, Locher and Watts, 2005, Locher, 2006, Locher and Watts, 2008). Relational work is defined as follows:

> *"all aspects of the work invested by individuals in the construction, maintenance, reproduction and transformation of interpersonal relationships among those engaged in social practice"*
> *(Locher and Watts, 2008: 96)*

It refers to the interpersonal level of communication rather than an ideational level (Halliday, 1978). Similar to facework, relational work is discursive (Watts 2003; Locher and Watts, 2005; Locher and Watts, 2008) in that the judgements by individuals on what is (im)polite is an on-going development. A user's expectations towards and agent and its linguistic variables develop in a similarly continuous nature, as discussed in 2.4.1. Previous experiences provide the foundations of future expectations (Tannen, 1993 in Locher and Watts, 2005). This can occur at any time within a single interaction as well as between different interactions (Locher and Watts, 2008).

6.3.2 Re-evaluating Social Distance

Combining relational work with the FTA equation covers all agents in both studies and all of their instructions, and allows analysis of both vague and non-vague communication within a single framework. The remaining values of social distance, power, and face threats in specific cultures can then be built on this framework.

The value social distance (D) in the original theory included the frequency of interactions between speaker and hearer and how familiar they are with one another, such as the comparison between friends and strangers. The degrees of similarity and difference between speaker and hearer also contribute to the social distance, and

this can include both stable social attributes and what is being exchanged both literally and linguistically in an interaction (Brown and Levinson, 1987). Given that agents and other machines are an integral part of our daily lives (Jennings et al., 2014), social distance remains a relevant concept in HCI. The frequency of interaction is an important factor that contributes towards the formula for assessing relational work in these contexts, and part of this relates to how personal the agent is for the user.

To clarify, personal here consists of three overlapping elements. Firstly, there is the concept of ownership and its correlation with interaction frequency. Users can interact with agents that exist as part of their personal property, such as intelligent personal assistants in their smartphones, but they can also interact with agents that exist as part of another party's property, for example automated checkout systems in supermarkets. Although a third party will have likely designed the agents for both, they may exist in both personal and non-personal spaces. An agent such as the automated checkout would also be rooted in a particular place, whereas an assistant on a smartphone could occupy any space the user chooses to go.

Secondly, the aspect of choice exists not only in terms of where the interaction takes place but also when it occurs, if at all. With the example of the smartphone assistant, the user instigates the interactions. Siri, for example, is activated by holding down the home button on the user's phone, engaging verbally by saying, "Hey Siri," or by using tactile features with headsets or cars (Use Siri, 2016). Although some of these interactions could occur by accident, they are arguably likely to be instigated by the user's own volition. Such volition may not be an option in other interactions, such as those with telephony agents when a user expects to speak to a person, or when exposed to verbal announcements on public transport. This overlaps with the concept of power, in that either the user or agent may hold the power of the instigation of interaction.

Finally, there is the concept of personalisation or customisation with an agent. The extent to which a user is able to customise an agent may affect their perceived social distance in the interaction. Using the similar examples from before shows how this can also correlate with ownership and how much choice a user has in interaction. The automated checkout discussed here and further in 2.2.2 was changed to reduce the amount of words it spoke to its customers, giving it even more of a limited output. Moreover, in the United Kingdom the same voice and language will be used homogenously across the stores, barring some exceptions in Wales (BBC, 2008). Siri, as well as having greater output can also be customised. A user is able to select a name that Siri will refer them to and correct the pronunciation as necessary, change the gender of the voice, and also its language (Use Siri, 2016). The ability to personalise agents and the choices a user makes in doing

so reflects their *user preference* towards an agent and their features. Even if an agent cannot be personalised, as was the case with those used in the two studies here, the results summarised in 6.1 show that users still have a preference towards the linguistic features of an agent, and these preferences can refer to both individual variables, as well as a combination of them. The tables in 4.4.3 provide examples of this. They also show that user preference can affect a person's desire to interact with an agent again, potentially limiting or increasing their frequency and familiarity.

When relating the personal elements above with the two studies, they were likely to have been minimal in the interactions with the agent interfaces, as these were novel interfaces that participants had not interacted with before. They were also owned by a third party (the researcher) and could not be personalised. However, as expectations are constantly revaluated, the degree of familiarity with the agent will have developed during the course of the task, as well as between the two tasks.

As mentioned earlier in this section, the value of social distance in the assessment of face threats is also discussed as being the level of similarity between speaker and hearer (Brown and Levinson, 1987). Chapter 2 reviewed previous literature that shows that linguistic similarity between a verbal agent (and other devices) and its user is an important aspect in perception and identity creation, with human voices often being preferred to various synthesised alternatives (Mayer et al. 2003; Lee 2010; Georgila et al. 2012). Similarity-attraction theory has shown that in certain contexts users display preferences for computer speech that is similar to their own (Dahlbäck et al., 2001). This mirrors the preferences that people have for interlocutors who are similar in human interaction (Montoya et al., 2008). Chapters 4 and 5 confirmed that this was also true when verbal agents use VL. This further indicates that perceptions of these variables can be specific (just the voice) and in combination with each other (voice and language).

Similarity can be further broken down into the similarities that the agent shares with the user, such as the degrees of human likeness it, and the consistency an agent's behaviour has with a user's expectations. The expectation frames (Tannen, 1993 in Locher and Watts, 2005) that people create based on their previous experiences can influence how people make judgements on relational work. When people are exposed to relational work that may be considered polite, judgements are made as to whether it is considered appropriate or inappropriate for the specific context (Locher and Watts, 2008). These can be negatively or positively marked i.e. they evoke a negative or positive evaluation of the relational work e.g. polite or impolite, and can also be unmarked e.g. provokes no evaluation even if considered appropriate or inappropriate for the context in question. Some

judgements can also be unmarked, in that they may be deemed appropriate or impolite but are unlikely to be commented on by the receiver of the relational work. All the variations of judgements a person makes relates to their on-going perceptions of what is expected in a specific interaction and how the reality of the interaction matches the perceived social norms. These judgements again can be renegotiated as an interaction develops (Locher and Watts, 2005).

Applying this to the data provides a framework for analysing how, and possibly why, participants perceived the vague and non-VL in the synthesised and human voiced agents. The discussion in 4.4 and 4.5, for example, indicated that when participants' expectations of how a synthesised voice agent should speak were not matched when VL was being used, this resulted in negatively marked reactions in the questionnaires and interviews. Comparisons between judgements on interacting with the agents again and the extent to which they prefer a human voice also showed that the non-vague agent had more positively marked judgements regarding interaction preferences (4.4.3). The second study indicated the disparity (see 5.4; 5.5) between expectations of voice and language resulted in more negatively marked judgements for the synthesised voices.

Locher and Watts (2008) discussed that judgements of relational work can also be unmarked, and this is evident in particularly in the tables presented in 5.4.3. These show that 46% of participants were deemed to have not discussed anything of the VL use in the agents. The discussion in Table 13 (5.4.3) first indicates that this percentage of participants did not notice the use of VL. Much of the relational work undertaken in human interaction is unmarked and is unnoticed because it matches our expectations (Locher and Watts, 2005; 2008). This may account for why almost half the participants did not provide comments on the use of VL[44].

Table 14 in the same section (5.4.3) discusses attitudes towards the three agents' use of VL. Three categories were used here – positive, neutral and negative. These partially correlate with the categories of judgements used by Locher and Watts (2008), with the exception of neutral attitudes. Although there is a mention of "neutral impolite" responses being negatively marked, a separate category was created here. As the unmarked responses for VL attitudes had been separated in the previous table, the neutral category represented marked responses that did not lean towards being positive or negative, meaning they can be considered to be *neutrally marked* judgements.

[44] Unmarked responses i.e. not noticing VL from the interview data includes verbal indications from participants that they did not notice anything significant about the agent's language, as well as the lack of any such comments.

These types of responses, for example, include those where the participant indicates language had been a noticeable aspect of the agent, without committing to judgement or describing it using adjectives such as “okay” or “fine”. This helps the understanding of which variables and their combinations contribute what is appropriate and inappropriate for this context. Positive and neutral judgements in this context indicate an appropriate combination of voice and language (although positive is a more favourable reaction), while negative judgements indicate inappropriateness. Categorising these responses relied on the subjective creation of a user profile (5.3.5) and evaluation of the interactions after it had happened. As such, this approach may not always be useful but does provide an option for analysing other similar interactions, with or without modifications to the methodology.

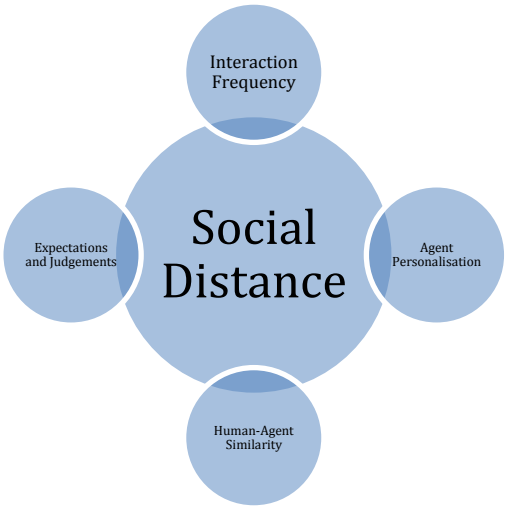

Figure 11: A representation of social distance in HAI.

6.3.3 Agent Power, Culture and Context

Returning to the equation, we can see that similarities between an agent and user, as well as the similarities between expectations and interaction realities contribute to the judgement of relational work. Of the remaining values, the culture of a context (R) is perhaps more relevant here, while power (P) is more challenging to understand when applied to the findings.

Power in human interaction refers to the asymmetric relationship between speaker and hearer (Brown and Levinson, 1987). Examples of this include the social dynamics between a teacher and a student or an employer and employee, for example. Agents mentioned throughout previous chapters often discuss examples including satellite navigation systems, intelligent personal assistants, and automated checkouts. The power dynamic in interactions with these agents is arguably one of the human taking the role of the user, and the agent in an instructor or

advisory capacity. This reflects the roles taken by the participants and agents in the two assembly studies.

Evaluating the data does not make the role of power in participants' perceptions of relational work all that clear in comparison to the other variables in the equation. While the role dynamic of *user-instructor* is at play here, it is not clear with this type of interface who has the power, or if it exists at all. Although the agent has the instructions the user needs to assemble the models, it is the user that controls the turns at talk via the interface. They are able to both progress to the next instruction and repeat the current instructions, whereas the agent simply provides the corresponding output the user has selected.

One extract from the interviews in Chapter 5 gave some insight into how a user may relate power dynamics from human interaction to human-agent interaction. P19 likens the VA agent to a friendlier nurse using informal language, but the CL agent to a doctor that is less friendly but more knowledgeable (5.4.3). This suggests that power is affected by an agent's voice, despite the two agents using the same amount of VL. This was, however, an analogous comparison and not a direct indication of the actual power roles that were undertaken in P19's interactions. The impact of voice can also be seen in other extracts where the combination of a synthesised voice with VL creates a negative judgement. P10 in Study One (p. 76), for example suggests that a person cannot be "chummy" with a machine. This might indicate that for P10 there are limitations to the relationships that can develop in these interactions.

Brown and Levinson in Jaworski and Coupland (2014) describe power in their FTA equation as "the degree to which H can impose his own plans and his own self-evaluation (face) at the expense of S's plans and self-evaluation" (p. 321). The agent is the speaker (S) in this context, and the researcher has decided the plan they have, which, in this case, was to provide the instructions when requested by the users. There is also no self-evaluation on the part of the agent, as it does not have a face to protect. Some users may interact with such an agent and consider it to have a face, but there was no evidence to suggest that this was occurring in this context.

With different agents engaged in different roles, the significance of power may change. One such example would be relational agents, which are "computational artifacts designed to build and maintain long-term social-emotional relationships with users" (Bickmore, 2003: 1). These differ from agent instructors in that they aim to develop a sense of rapport and trust (Bickmore and Gruber, 2010) and manage future expectations (Bickmore and Picard, 2005), as opposed to simply providing information like the agent instructors used here. Their potential has been touted in healthcare, counselling, and educational scenarios (ibid.). In these types of interactions, the application and

assessment of relational work may lean more towards facework than agent instructors do (Bickmore and Cassell, 2001). Relational agents are designed with relationship goals with their users in mind. This shows that the roles agents take would depend on the context in which it is deployed and the decisions an agent designer (Figure 1) during the creation process, which in turn form the affordances it provides to users (see 2.4.3).

Considering the importance of context in the assessment of relational work relates to the final variable (R) in the equation. Brown and Levinson (1987) described R as the ranking that a specific FTA in a given situation, and the degree to which they interfere with both positive and negative face, such as in a specific culture. The situational aspect is important in HCI contexts, as it can be broadened to include any context and agent attributes that are specific to an interaction. For the purposes of this research this focuses primarily on the linguistic variables of language and voice, but this can be expanded to include other variables. This can include those of a linguistic nature, such as those discussed in Chapter 2. They may also include non-linguistic variables, examples of which include the familiarity of a situation or task (such as model assembly), the medium of interaction with the agent (verbal, tactile, mixed) and duration of interaction. This is by no means a definitive list, as the concept of R in this approach is left deliberately open so as to make it applicable as possible when using other research questions and methodologies. These are discussed further in Chapter 7.

An example of what R may look like with the agents in Chapters 4 and 5 can be seen in Table 16. Some of the categories and features within them do overlap, and there are arguments to be made for changing them. Depending on the outlook, the agent type being an instructor, for example, could also be a linguistic variable. Similarly, the duration of the interaction in other non-linguistic variables may be considered a contextual feature. Regardless, it helps define the salient features that can affect perception of relational work in what can potentially be a very large list. This includes those that have been decided by a researcher or designer, such was the case of voice and language here in the research questions, and those that arise from the assessment of interactions.

Table 16: Example of R with the agents used in Study One (S1) and Two (S2). The most salient aspects for these studies are underlined.

Agent Type	Interaction Type	Linguistic Variables	Other non-linguistic Variables
Verbal output (primary) Visual output (secondary) Tactile input Instructor	Laboratory setting Interface on laptop computer One-way verbal dialogue	Voice (male; synthesised (S1, S2) or human (S2) Vague (S1, S2) & non-vague (S1) language Southern English RP (agents) L1 English speakers (users)	Lego model assembly 15 minute max. interactions (S1, S2) / until task completed (S1) Practice task (S1)

6.3.4 Applying the Approach

Taking the concept of relational work and the FTA equation allows for a broader approach in analysing the spectrum of polite and non-polite language used by agents. Altering the descriptions of what the values in the equation represent, and restructuring it slightly, means it can be used as a means of assessing relational work in HCI contexts. Exchanging speaker (S) and hearer (H) for agent (A) and human (H) leaves the equation as follows:

$$W_x = D(A,H) + P(A,H) + R_x$$

Admittedly one of the drawbacks with this equation being so similar is that it may appear to be an attempt at creating a universal means of assessing politeness in HCI. Table 15 summarises the variables that have been discussed in this chapter thus far which combines the values discussed in the original FTA equation, the notion of relational work, and the key variables that have emerged from previous research discussed in Chapter 2 and the results from Chapters 4 and 5. These salient features are of course subject to change in any interaction.

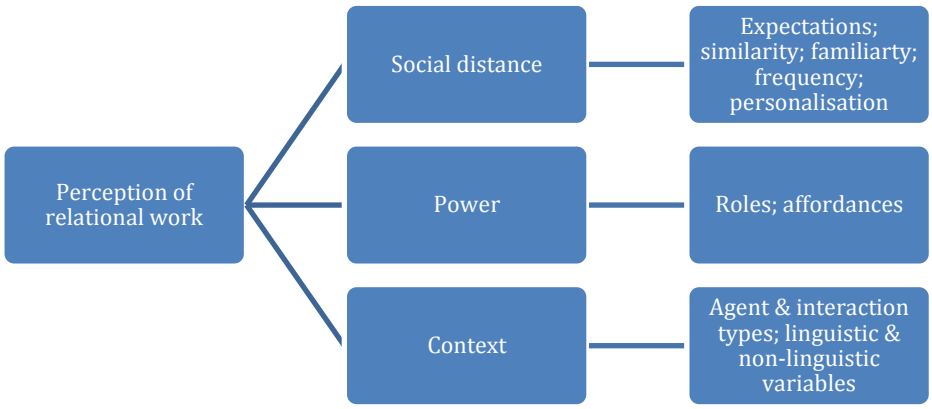

Figure 12: Representing some salient features that can affect a user's perception of an agent's relational work.

This change is important to consider, as given the variation that can be applied to an agent, and the unique experiences of each user, the interaction reality can never be fully guaranteed. As such, attempts to create universal theories of politeness in HCI face the same difficulties present in human interaction. Nonetheless, applying this approach to the findings of Study One and Study Two indicates its usefulness.

In Study One, for example, the linguistic variables of vague and non-VL with a synthesised voice showed a strong preference (63%) for the non-vague agent. Similarly, it showed a higher preference for a human voice with the vague agent interactions (75%) than the non-vague interactions (42%). Combined with the qualitative discussion on the disparity between voice and language with the vague agent (4.4.3), this indicates that expectations of a synthesised voice being non-vague were more common than the same voice including VL. These were impacted too by the other variables in the approach, though the language was the main focal point in this particular investigation. The relational work undertaken by the non-vague agent can be seen to be more appropriate than the work performed by the vague agent.

In Study Two, VL was perceived as more appropriate with the VA agent than either of the synthesised ones, which can be seen by taking Table 14 from 5.3.3 and included both positively and neutrally marked responses as polite or appropriate communication (6.1.3). With this combination, the instances in the interviews where VL was discussed were marked as appropriate 80% of them, while CL and CP very close together at 25% and 27% respectively. This provides further credence to the results from the first study, while also indicating that selecting two different synthesised voices that are close in their linguistic variables (voice age; gender; accent) can create similar user expectations and perceptions of their relational work. Moreover, as Figure 1 (2.4.3) visualises, the interaction reality is affected by the user

expectations as well as the intentions of the agent designer or researcher.

Consequently, when returning to the notion as to whether or not facework is required with agents when they lack a face and any self-evaluation, there are no guaranteed answers. The extent to which it is required requires some feedback from an etic perspective (i.e. user evaluation). It may be able to improve perceptions when of an agent when combined with a human voice, as this may meet more expectations than synthesised alternatives, at least the type used in these studies. It may also be beneficial to include appropriate language use as part of what is judged to be polite communication (Locher and Watts, 2008), as well as what is believable of an agent's behaviour from a user's expectation of their linguistic variables. For example, the synthesised non-vague agent in Study One was not perceived to create any face threats, yet the figures show the synthesised vague agents were rarely considered to be appropriate or believable in their language use.

6.3.5 Future Research of Politeness in HAI

While a universal theory of politeness in HAI may not be attainable, this approach allows researchers to include relational work and the variables that affect how a user's perception. In turn, this can provide guidelines of what can create successful polite communication in these contexts, both for those designing agents and applying specific communicative variables, and those researching and evaluating them in interactions with humans. With this approach being deliberately broad, this can include politeness research specifically. Upon reflection of the results and discussions presented thus far, a differentiation in the definitions of politeness$_1$ and politeness$_2$ may be required in HAI.

Politeness$_2$ in such contexts can be seen to include these guidelines of politeness success, and lack thereof, in interaction. Brown and Levinson's (1987) example of how face threats can be calculated in conversation is an example of guidelines in human-human interaction. The guidelines for HAI include those discussed in this chapter and the previous knowledge attained from research outlined in Chapter 2. Politeness$_1$ in these contexts can include user evaluation of facework and politeness strategies, and how they are part of user preference and user expectation, as well as the longer list of variables that have been covered in the above sections. This also includes the wider spectrum of relational work when appropriate, such as if it becomes a focal point in research or application. Studying patterns of politeness$_1$ evaluations can provide the foundational knowledge for theories of politeness$_2$ in HCI. One such example is the design of studies from the previous two chapters, which were informed by existing research in the area. Hedges and discourse markers, for example were seen to be useful in creating perceptions of likeability in robot helpers, at least when

observers viewed videos of interactions (Torrey , 2009, Torrey et al., 2013). Similarly, politeness strategies were seen to help create the same perceptions in robot instructed drawing tasks (Strait et al., 2014). This contributed to the knowledge of polite communication in HRI, providing a starting point to investigating agents. It also informed the use of hedges and discourse markers as part of these studies. Similarly, the interaction and agent types used were informed by the relative success of both of these previous investigations. Voice was used in both studies, as were instructions, though these were described as both *help* and *advice*. Nevertheless, they were both providing information on completing process (cupcake making; drawing) much like assembling Lego is a procedural task. The results of this study, taken from the participants' evaluations of their interactions (politeness$_1$) have also provided some guidelines (politeness$_2$) for politeness and relational work research in HCI. Patterns that have emerged in the results suggest that human voices are more suited to indirect and polite work than synthesised voices are, which are still appropriate for direct communication.

Future research can further contribute to politeness$_2$ in HCI by focusing on changes to the variables that have been discussed in this chapter. For example, one may focus on altering the linguistic variables of voice and language further by incorporating female voices, as well as male, and alternative approaches to agent. A different approach to employing VL use in this context could also be undertaken, for example by altering the quantity of VL lexis used in an agent's instructions. Similarly, a focus on a wider selection of agent and interaction types could be implemented, with comparisons between laboratory and real world agents, or investigating differences between instruction giving and conversational agents. Some of the more abstract concepts such as social distance and power can be research simultaneously, providing a more detailed contribution to further guidelines and understanding of politeness research in these contexts.

As well as the design and application being changed, the methods of evaluation politeness also contain many variables of their own. A mixed methods approach was used here, which included questionnaires, statistical analysis, and qualitative coding of semi-guided interviews. Similar research has seen the use of a qualitative coding model to analyse adjectives and verbs in open-ended questions (Torrey et al., 2013) and brain-based objective measures using functional near-infrared spectroscopy (fNIRS) (Strait et al., 2014). Other measures, such as the analysis of nonverbal communication and other physiological data could supplement both quantitative and qualitative user evaluations. Building upon the corpus approach used in Study Two would provide another avenue of investigation, particularly as it is multimodal in nature, consisting of both audio-visual and textual data. This could include interactions in the wild,

though the ethical and logistical considerations warrant further detailed discussion.

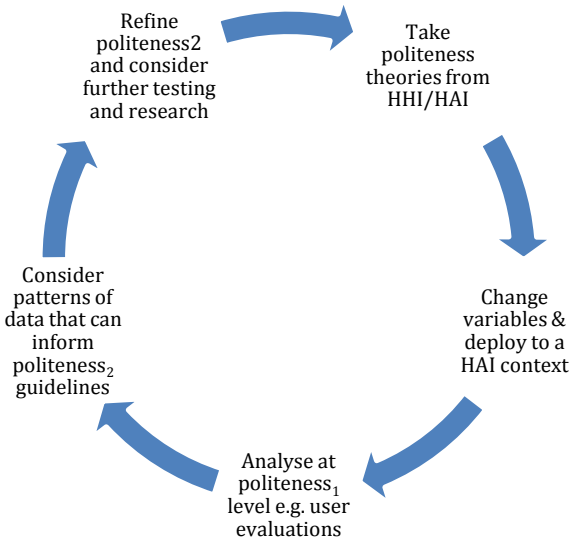

Figure 13: Example representation of politeness research in HAI.

In any evaluation, user preferences and expectations towards agents using politeness will have to consider the individual user. Observing patterns in interaction data, however, can contribute towards an understanding of the salient variables affecting politeness in interactions with agents (Figure 13). As shown from the previous chapters, voice is one of these salient variables that can greatly affect language perception, specifically whether or not it is synthesised or a human recording. Previous research has also highlighted appearance and interaction distance as key "modulators" of people's perceptions towards robots using politeness strategies (Strait et al., 2014). These may also transfer over to interactions with agents and not just robots and merit a change in the agent type for a fair comparison.

There are several politeness$_2$ level guidelines to consider from the results of these studies. If a user does not perceive an agent as capable of facework and using politeness as a result of its voice, then any attempts to do so may be successful as a result. Participants in both studies highlighted this when referring to the insincerity of the attempted facework because the agent is "just a machine" (see 4.4.3 and 5.4 for further examples). Expectations towards human language use changes over time, even throughout a single interaction, as do the judgements of what is acceptable and appropriate communication in relational work (Tannen, 1993; Locher and Watts, 2005). There are the individual user preferences to contemplate, which in agent and computer interactions can be considered part of politeness$_1$, but there is also the wider discussion of how agents are positioned socially in

our lives. In these interactions this can be considered part of politeness$_2$. This includes discussing the very nature of agents and how this relates to language use and facework. If they believe an agent lacks the ability to negotiate social boundaries through language then they may not benefit from the agent's attempts. It may also be that what can be construed as impolite in HCI is not the same as in human interaction.

6.4 Identity

The relational approach that has been discussed so far in this chapter refers to the perceptions of relational work performed in HCI contexts, and how linguistic and non-linguistic variables affect a user's perception of it. This perception of linguistic variables was discussed in Chapter 2 as part of how users view the identities of agents. Hall and Bucholtz (2013) considered facework an important factor in identity construction, and that the two cannot be analysed independently from one another. In discussing the concept of *self,* Watts (2003) argues that self can be labelled as face – as something that develops primarily through social interaction, a discursive notion that draws upon the original Goffman works. Presumably, this could also work for the concept of *other* in an interaction. While agents having a face is a point of argument, them being an "other" and having identities assigned to them is a concept that traces back to the Media Equation and CASA paradigms (Nass et al., 1994, Reeves and Nass, 1996).

Linguistic and non-linguistic variables that affect perception of relational work, discussed in 6.1, can also be applied to assess perception of agent identities. Taking synthesised and human voices as an example, one can see that it affects the perception of attributes when users are asked to evaluate their interactions with agents (4.4.2; 5.3.2). The cycle of politeness research shown in Figure 13 can be applied in a similar fashion to identity. It is still an individualistic matter, relying on unique frames of expectation, but similar types of patterns can emerge from the results. Judgements of an agent's relational work can be part of its identity, and its identity can also affect these judgements.

The discussions of politeness in HCI also overlap upon the issue of identity. In the studies, participants assigned individual identities for the agents, differing often on the variables of language and voice. They also assigned them to wider social identities, usually by indicating it as a "computer" or "machine", and describing what they expect of speakers belonging to these categories. The preferences for combinations of voice and language (see 6.1) were clear to see. VL in the synthesised agents had a majority of negative judgements when marked in Study Two (CL: 73%; CP: 75%). Similarly, in Study One, 33% of participants preferred a human voice when interacting with the vague agent (CL) because of the lack of clarity, and a further 12.5%

indicate this was because of the agent's language. It appears from the findings that expected identities were not being matched in these cases, and instead encroached upon other frames of expected identities that belonged to people. Although there were no pre-task evaluations to propose that participants expected the voice actor agent to attempt politeness or be successful with it, the descriptions in the interview data (e.g. sounding more natural) suggests this may be the case.

Identities can consist of the characteristics of an individual or describe a social group (as well as the expectations of these) to which an individual belongs (Fearon, 1999). By referring to the VA vague language as "humanlike", particularly with the synthesised voices, a number of participants see its functions as belonging to humans and not to agents. This goes some way in explaining the disparity in how appropriate VL use was in the VA agent compared to the synthesised agents.

6.4.1 Individual and Group Identities

Understanding what features may make up the identities of agents and the notion of agent likeness comes back to the creation of identity. For people, the first aspect relates to the society one was born into. This includes social class, geographical location and ethnicity (Hall, 2013). For verbal agents, both the voice and language being used can inform the identities based on these factors. There are also the different agent roles to consider and the purpose for which they were designed. The second aspect of identity creation is made up of the things we choose to bring into our lives and our brought into by others, and ourselves such as family, friends and colleagues (ibid.). Agents are built for singular or multiple purposes, but do not have a choice they for what areas of life they are brought into. An intelligent personal assistant may have multiple roles and be brought into an interaction for direction giving, general queries, or for testing the limitations of its features. They will, however, often be homogenous and mass produced.

Their individual identities and group identities can overlap as a result. Identical agents will share common identities, as will non-identical agents with varying degrees of similarity, which can both be seen as part of a group. They may possess a certain collective individuality too, in that although they are the same, the individuality will be a product of emerging perception by their users.

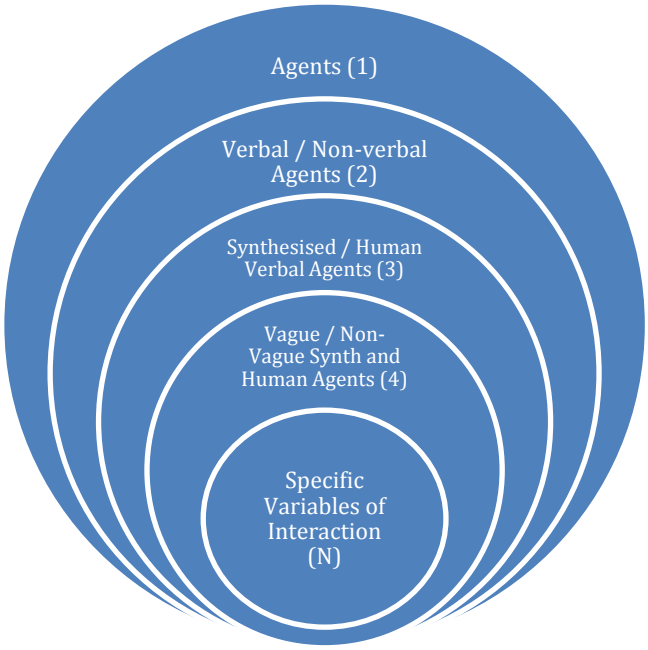

Figure 14: Example of overlapping group identities.

Figure 14 presents a representation of the overlapping between group and individual identities. This is similar to the relationality principle discussed in the second chapter (Bucholtz and Hall, 2005), which posits that identities are not independent and instead are related to one another. The difference here is that this approach shows that the relationality can be somewhat hierarchical, though this does not have to always be the case. In this example, using vague and non-vague agents with different types of voices are a specific part of a wider whole. The negatively marked judgements of VL in synthesised voices (4), for example, can represent a mismatch in expected identities of a wider group (1, 2, and 3). Agents in (4) can also share characteristics with agents in any groups below or above in this type of classification. The group of (N) would contain the very specifics of the interaction and its variables, at least which are salient for the purposes of a particular investigation. They can converge or diverge with the expectations of the categories it is indexed in. For example, if (N) could represent a specific interaction with a vague agent using a human voice in a sat nav, or the same agent but in an intelligent personal assistant. In these scenarios, (N) would also overlap with identities of human voices and perhaps even humans as whole, and not just agents. This may explain why marked judgements of VL in the VA agent were discussed as appropriate 80% of the time, as frames of expected identities were being drawn from groups where VL is either more expected or more appropriate. Those participants who did not make any marked judgements on the VL in the interviews might see VL use as either appropriate or even expected throughout the interaction, though there is also a sense of apathy to consider.

There can be overlapping features of identity that are taken from individual and wider groups. These are not limited to the

representations in Figure 14 either, and further groups upwards in the hierarchy can be observed. Also, in common with identities, relational work, face, and politeness, this is also a process that is redrawn throughout and between interactions. This figure can also be reimagined in a similar style to the two levels of politeness research in HCI. The specific interactions of individuals (N) at a politeness1 type level inform the higher levels on patterns of expectations and what is appropriate what is expected and appropriate on a politeness2 level. Such patterns could be considered collective individual identities that inform collective group identities, and provide designers and researchers with further knowledge on the application of specific variables in interaction and how it affects different levels of identity. Again, these groupings can be specific to an investigation, which need to be clearly and carefully defined (Locher, 2015).

6.4.2 Emerging Identities

Bucholtz and Hall (2005: 588) discuss that emerging identities are perhaps best identified in those cases where "speakers' language use does not conform with the social category to which they are normatively assigned". This is similar to the breaching of norms described in the previous section (Locher and Watts 2008). This could be in several of the levels seen in Figure 14. Negatively marked judgments on VL in synthesised agents may represent violations of expected norms for synthesised voice agents as a group (e.g. 3 in Figure 11). However, if the same user displays these judgments with VL used in human voiced agents, this may be a violation of their wider expectations as to how agents (e.g. 1 in Figure 14) should behave linguistically in contrast with humans. Such a user may believe agents should not be attempting to use VL as a means of facework or politeness strategy, whereas other users may find it more appropriate because it draws on other categories of identity from human interactions.

The quantitative findings discussed earlier in this chapter show the use of VL use in verbal agents often diverges from the norms that people typically associate with an agent's group identity. The vague agents appeared to contradict many of the participants' expectations of how an agent speaks to them, which in turn created emerging identities and changing expectations. It has been discussed many times throughout these past three chapters that a disparity existed between the voice of the agent and the VL. The two using synthesised voices had some equality, as if recognisable features from two similar identity categories were clashing. Evidence for this was seen in many comments of the agents attempting to be humanlike with its language use but not succeeding, as well as the general acceptance of the non-vague agent in the first study conforming to expectations. Common expectations were for the agent instructor to be direct, which has been a mainstay for agent communication in the past. There would appear

to be some indexing of directness as part of the group identity of verbal agents, though when part of this identity is challenged people responded in different ways. Negative responses that discuss the language as being humanlike are intriguing, as all agent speech will have some foundation in human communication – there is simply no alternative. It may be the case that in instruction giving settings, the notion of an agent conducting facework is not always necessary. If it were a sensitive topic an agent were discussing, such as medicine, then the outcome may be different.

For those who responded negatively it appeared to be a similar case as with the other voices in that it was trying to be human and not being successful. Those who received it positively commented on the VA agent sounding more natural. Arguably, in these instances the VA agent's identity is emerging from both agent and human categories – an intermediary between the two. It has the same interaction modality and interface as the other agents, but with the voice actor is provided with the human touch of paralinguistic cues and vocal clarity. The limitations of a machine interface still apply, but the likelihood of language with strong connotations of social negotiation and rapport building being accepted is improved by the human voice.

Despite the relative success of the VA agent, it still encountered some resistance that fuels the discussion of the wider discussion of agents in society. Although the use of a human voice may have been able to draw from indexed categories of human likeness, it remains somewhat of a disembodied voice within a machine. As a result it may not afford the same social capabilities in the eyes of an agent's users. This may be truer of certain agents than others. The one used in the two studies here lacked the dynamic and active ability to adapts its language use as the interaction occurred. In this sense its sense of identity through language use being projected was relatively static and its spoken output finite. During the interaction participants were able to become used to the agents using VL, even if they did not welcome it. If multiple interactions were to occur with an agent that is fixed in its verbal output, emerging identities may change once more and any successful facework accomplished through something such as VL may incur diminishing returns.

While an agent's linguistic output may be limited, their users' previous experiences and perceptions of agents may be less restrained. The emerging identities formed in the perception of the user will have a greater variation than the agent's own output. In the data, participant comments on how agents with specific voices should talk to humans highlighted not only their indexed identity categories, but also their preferences for agent communication. These preferences are part of their own identities when it comes to these interactions. To this end, the emerging identities may be partly seen as the user projecting their own identities onto the agent. For instance, a user who does not find

VL use in synthesised verbal agents appropriate can be said to have this preference as part of their identity. Further research on user preference and its link to identity is required, but this prompts an interesting topic of discussion.

A further point to consider in the creation of these identities is the designer of the agent, which makes the identities somewhat intentional (Bucholtz and Hall, 2005) put from the perspective of a third party rather than the agent itself. This may be a company, researchers or an individual, for example. A third party has a certain impact on the interaction, which may be influenced by the agent's function and intended context of deployment. Identities of agents designed by companies may not be too dissimilar to the identities people project under employment, as they both have commercial interests to consider. An automated checkout and a cashier at a till in the same supermarket are both representatives of that company. Compared to current checkout technology, a human cashier will have a greater potential for adapting to a specific customer's needs through language use, whereas the spoken output of the automated checkout will be relatively limited.

The two studies discussed in this research have shown that verbal agents using VL depend strongly on the voice being used in determining what identities will emerge in interaction. The two synthesised voices that were aligned with identities of directness did not fare as well as the voice actor, which appeared to overlap with identities of successful use of indirectness. Despite this relative success, participants were still cautious of accepting an agent instructor extending its boundaries of expectation. The observable patterns in the data contribute towards understanding these boundaries of identity between humans and agents, though one cannot ignore the individual user and how their perceptions shape the emerging agent identities.

6.5 Vague Language

The discussion on politeness and identity in the agents used in the two studies mainly focus on adaptors and minimisers discussed in the VL model (3.4). This section will also discourse the responses to discourse markers and vague nouns.

In Chapter 3 both adaptors and minimisers have some function of hedging the agent's instructions. Adaptors such as *more or less* and *a little bit* can be used to minimise the imposition of a speaker upon the listener. Similarly, minimisers such as *just* and *basically* accomplish the same while providing some structure to the instructions and attempting to reduce the perceived task difficulty. How these two categories of VL were received mirrors the discussion of identity and politeness, in that it was dependent on the voice. However, when they

were successful, participant feedback indicated that the intended functions sometimes matched the outcome. For example, when the minimiser was used on an instruction a participant found easy: "It *was* just a little twist". They did also achieve some unintended consequences, such as sounding condescending and patronising when a hedged instruction was being used at a stage the participant found difficult, and not as easy as the agent expressed.

Unlike adaptors and minimisers, discourse markers and vague nouns were seldom mentioned in the interview data. Discourse markers included *so* and *now* and were used to structure talk and the process of assembly itself. There were occasional comments, such as "I don't need a so", but these were fairly uncommon. It is perhaps more difficult to assess their impact when they were not heavily discussed, but this may be indicative of their success. This would suggest that discourse markers used in verbal agent instructors are relatively uncontroversial, possibly due to their more functional uses rather than social, at least regards to facework and politeness. Vague nouns were also relatively successful in that there were very little negative responses to them. Given the nature of the assembly task, using vague nouns such as *piece* and *bit* appeared to provide participants with adequate information to construct the Lego models without having to refer to full nouns or noun phrases repeatedly. In terms of inferring shared knowledge and establishing common ground, there was no evidence to suggest that participants felt a common ground was being established. However, one can extrapolate from participants being able to go through the assembly instructions, albeit at different success rates that the vague nouns used were sufficient communication and being more explicit was not necessary. There appeared to be more confusion and discussion over the nouns that were intentionally part of the VL model, such as "connector", but those included in the model were rarely discussed. This suggests that in an instruction-giving context, a verbal agent should be able to use vague nouns with a reasonable chance of success. Success here refers to executing the intended function and not having a number of negative responses towards them. The nature of the assembly models may have contributed towards this success, as many of the pieces were of an obscure and fantastical variety. While this is not unexpected of this type of model, it could mean that participants would not be able to identify pieces themselves without the use of vague nouns. They also clearly referred to constituent parts of the model, and so there was a very finite amount of physical material that they could refer to. If this was the case, then the use of vague nouns is an effective approach for the agent instructors.

For the majority of the interactions, both discourse markers and vague nouns seem to be successful in both studies. Adaptors and minimisers face the problems of their politeness functions discussed in the theory of politeness$_1$. Just as there can be no universal theory for politeness,

there can be no universal guarantee as to how VL will be perceived, particularly when this is one of their functions. There are no guarantees for discourse markers and vague nouns either, though it may be that they are more familiarly categorised with agent identities than the other two features. VL *can* achieve the same functions in HCI as seen in human interaction, but given the context and features of the agent this may be different for the various types of VL involved.

In-group membership is one of the uses of VL. There was a hypothesis that this may bridge the gap between human and agent, in that the agent was using language that can have numerous social functions that might not necessarily apply to agent language, such as face saving and politeness. It seems more likely that any in-group membership was between different group identities (such as verbal agent and human communication) that participants had, rather than the agent and participant being within a group. The success and appropriate use of VL in these contexts are strongly linked to the expectations of these identities and the extent to which they match interaction reality when all variables are presented to the participants.

There is a caveat to consider in the discussion of VL in light of Chapters 4 and 5. While the VL was often perceived negatively with the synthesised voices, and sometimes with the human voice (as discussed earlier in this chapter), there is a possibility that this outcome could be different if the approach to designing the instructions was changed. In designing the instructions (see 3.3), non-vague instructions were first created and then VL added to them. It remains a possibility that using naturally occurring vague instructions from a human source, and then removing the VL to create the non-vague instructions would create different results. However, the clear differences between the synthesised and human voices using VL in Chapter 5 do indicate that voice is a predominant factor in the perception of VL, as well as the linguistic and social boundaries between humans and agents. Given this, some results would arguably remain the same, at least in regards to the differences in voice affecting perception of VL in these agents.

6.6 Computers as Social Actors

The premise of the CASA paradigm and the Media Equation is that people treat computers as they would other people, and that the social rules underpinning communication towards humans are used towards computers too (Nass et al., 1994). Computers can also be perceived to have personalities similar to humans (Nass et al., 1995; Nass and Lee, 2001; Lee et al., 2006) as well as the notions of "self" and "other" (Nass et al., 1994). This was seen in the creation of identities for the agents, even without complex and sophisticated capabilities from the agent. Using just the VL and voice changes resulted in different identities being created and elicited social behaviour from the participants (Lee, 2010). Such small changes are known to evoke a wide range of social

responses (de Graaf et al., 2015). The comparison between the vague and non-vague agent in the first study was further evidence for this. These changes created dramatically different perceptions for some of the participants who noticed them. This is perhaps indicative of language and voice not being small changes at all. The strong ties they both have with perception and the creation of identities. The crossing of the group identity boundaries of agents and humans through changes in language may have contributed to a sense of eeriness as described in the uncanny valley (Mori et al., 2012). This discusses appearance, but it may transcend this and work with language and voice. Agents using language typically associated with human communication, while displaying a clear lack of any other humanity, certainly appears to create a disparity if not eeriness.

While computers are social actors and people do treat them with similar if not the same social rules underpinning human interaction, this does not mean that striving towards human likeness in speech for all verbal agents is the optimal path for development. Lower quality synthesised voices that sound like a machine may be better utilised talking like a machine. This usually means being direct and not using language that appears too humanlike that has social functions alongside transactional ones, such as using politeness strategies. These appear to be better used with a human recording or possibly a high-end synthesised voice. Even with these guidelines, preferences for talking with agents and other machines may be centred on using them to accomplish tasks rather than conversing with (Baron, 2015).

With regards to perceptions of human and synthesised voices, the findings in this study generally agree with previous research suggesting people prefer a human voice to a synthesised alternative (Cowan et al., 2012; Georgila et al., 2012; Mayer et al., 2003; Lee, 2010). Although there is some chance the preference for the voice actor may have been overstated by participants (Mitchell et al., 2011a), the qualitative data regarding the language of agent suggests it is not. Moreover, while this does agree with previous research on voice it also shows that certain methods of delivering politeness, such as through a high volume of VL, is perhaps better accomplished by a human voice.

While the same social rules may apply in HCI, the expectations with agents differ depending on many different variables. The combination of linguistic variables and the context of interaction can alter a user's perception of what language is appropriate, though the process in doing so is a lot like that in human communication. Using theories from linguistics and other interdisciplinary areas remains a suitable approach for understanding our interactions with all manners of agents and other digital entities, as originally proposed over twenty years ago (Nass et al., 1994, Reeves and Nass, 1996).

6.7 Summary

This research has looked at several theories founded in human communication as well as human-computer interaction, and supplemented them with the results of two studies. In analysing the effects of attempted politeness through VL, it emerged that thinking of human-agent interaction in terms of politeness$_1$ and politeness$_2$ is appropriate. Politeness$_1$ is concerned with looking at politeness as it emerges during interaction, given the individual, social and contextual differences that can affect what is polite or impolite in any given interaction. In interactions with verbal agents, this can contain the salient features that affect these emerging notions of politeness. These include a user's expectations of how agents should speak in particular contexts, and what notions of agent identities they bring into an interaction. Poltiteness$_2$ is often associated with the academic discussion of politeness and its wider theories, with a view to creating a universal theory which has since widely been discredited. For HCI this can be seen to include guidelines and knowledge of factors that affect user expectations and preferences, such as the use of voice. It also includes the wider discussion of agent identities and how they positioned socially relative to human identities. The concept of identities is strongly tied to expectations of how agents will behave during an interaction. Despite some success in the voice actor agent, there appears to be some identity boundaries with how they use language. Although they are social actors, agents using language that has strong social functions, including politeness and facework, is not always appropriate for instruction giving. VL categories that have less social and more functional goals were less of an issue in these studies. Discourse markers to structure talk and vague nouns to replace the initial description of assembly pieces were less controversial in their use. When concerning the best use of VL in verbal agent instructors, a human voice is deemed to more natural and humanlike, whereas synthesised voices can create too much disparity and cross the category boundaries of expectation. It may be the case that the more an agent sounds like a machine; the more it is expected to behave like one.

7. Conclusions

7.1 Thesis Overview

This thesis began with the following intended research aims:

> *1) How do users perceive and project identities towards verbal agent instructors that use VL and what contrasts can be seen with human communication?*
> *2) Are there any differences in these identities when comparing vague agents to non-vague agents?*
> *3) Are there any differences in these identities when verbal agents use synthesised and human voices?*
> *4) Does the use of VL in an instruction based task affect a user's ability to conduct a task?*

To investigate these research aims, this thesis first presented a review of relevant literature in Chapter 2. This addressed the background knowledge related to the research aims, including theories of social behaviour towards computers, the concept of identity and its link to language, voice, and vague language (VL). Chapter 3 introduced the VL model and the general approach to applying it in a researchable HAI context. The two specific approaches were discussed in the subsequent two chapters. Chapter 4 compared user experiences with a vague and non-vague agent using a synthesised voice, while Chapter 5 compared vague agents with human and synthesised voices. Chapter 6 then reflected the findings of both chapters on current theories in linguistics and HCI, and proposed approaches and explanations to theories on relational work, facework, politeness, and identity.

7.2 Contributions of this Thesis

7.2.1 Identities in Vague Verbal Agents

The common notion throughout the first three aims of this thesis concerned the concept of identity. The aims were concerned with understanding differences in the identities users created between different agents, as well as the differences between agents and humans. As Figure 14 (6.4) shows, individual agent identities exist as part of overlapping group identities. This in itself is not necessarily different from human identity creation. The relationality principle discussed by Bucholtz and Hall (2005) indicates that identities are interdependent and do not exist in isolation. The fundamental difference is that agents can have different levels of overlap with human identities. As the findings from Chapters 4 and 5 indicate, this is particularly true when agents use VL and can vary dramatically with different agent voices. When VL was negatively marked in the studies, the synthesised vague agents were often criticised for straying from expectations of machine language and into human language. The voice actor, however, was more successful in blending group identities of

humans and agents together, and its use of VL was more likely to be marked as appropriate by the participants. In Study One, the non-vague agent was marked as appropriate as its use of direct language well within the participants' expectations of existing agent identities. The data shows that there are different levels to which agent and human identities can blend, and the use of both voice and language can greatly affect the way in which they are perceived. This could prove to be an important feature when considering other HAI contexts that involve a vulnerable population, such as the elderly, or with relational agents that attempt to create and maintain a social relationship with their users.

A further point on identity discussed in Chapter 2 (Figure 1), was the nature of how identity is formed in agents. Agents often have some form of a designer, which, in this thesis, was also the researcher. Designers have the control of many of the linguistic and non-linguistic variables that the agent consists of. The voice and language of the agent are fundamental parts of this. This is an important point to consider for researchers in HAI contexts, as it is an important distinction from HHI. Other distinctions include agents not necessarily being associated with having an age, or even having any human likeness. There are numerous research gaps to investigate that could further explore the conceptual differences of agent nature.

These findings have important implications for both future research and agent design. Firstly, if opting for a synthesised voice in an agent, the use of language has to be considered carefully. Direct and non-VL were successful in this study, and appeared to correlate with expected agent speech. This may be different in other interaction contexts and with other agent types, but they are some of the important variables to consider. If choosing a professional voice actor, the findings suggest that using VL can be viable. There are no guarantees, however, and agents using "humanlike" language may still encounter negatively marked assessments by their users. The extent to which an individual user is catered is also a worthwhile discussion. As each user brings their own unique collection of experiences and expectations, their interactions with agents may be very different to others.

This thesis also contributes towards the existing body of knowledge comparing human and synthesised voices in agents and other digital interlocutors. Vague agents with a professional human voice are rated more positively than synthesised agents in regards to appropriate language use. This correlates with literature suggesting that a human voice can be more preferable and rate more highly than synthesised alternatives (Mayer et al., 2003; Lee, 2010; Cowan et al., 2012; Georgila et al., 2012; Cowan et al., 2015). The human voice in general was preferred to the synthesised voices (Chapter 5); however the two synthesised voices used in this thesis represent a small sample size of those available.

7.2.2 Building Approaches to Understanding Identity

This thesis has introduced linguistic approaches to both the design process of HAI and the analyses of interactions. A linguistic approach did appear in Torrey et al. (2013), though was not prominent in other similar works on polite advice givers (Strait et al., 2014). Chapter 3 discussed the linguistics focused design of the VL model, while Chapters 4 and 5 featured sections devoted to qualitative analyses of interviews with participants. The VL model provided descriptions of what were thought to be relevant categories of VL items for use in HAI, with references to existing literature from HHI. While they did not always work as expected, often because of voice and language disparity, as well as clashes with agent and human identities, there were instances where it succeeded. In these instances VL appeared to function much like it does in human communication (e.g. minimising instructions in 4.4.4.2). This helped to provide evidence that VL can function in HAI as it does in HHI and that people interacting with vague agents can identify the VL and its uses. The analyses also helped to reinforce the CASA paradigm and Media Equation theories (Nass et al., 1994; Reeves and Nass, 1996).

The analyses have also provided methodological contributions. Qualitative analysis of participants' interviews helped to deliver a greater understanding of their experiences in light of the research aims. Chapter 6 expanded on the approach taken to understanding HAI and introduced the relational approach. This can be used to understand relational work (Locher, 2004; Locher and Watts, 2005; Locher, 2006; Locher and Watts, 2008) undertaken by an agent and its relationship with linguistic and non-linguistic variables. This includes unmarked responses, which featured in the second study (5.4.3). Using the relational approach can encourage the assessment of micro and macro level relationships e.g. if researching power, the roles of agents can be considered and vice versa. This also encourages the use of incremental research building on from previous literature, and investigating the salient variables in any given HAI context. Further research into polite (and non-polite) language in HCI contexts can be wholly experimental or incremental in nature, though the latter approach is favoured here as it is able to draw on both HHI and HCI research while contributing to the cycle of politeness1 > politeness2 cycle (see Figure 13). Such an approach, as discussed in the previous paragraphs, was used in designing Study One and Two.

For the qualitative analyses in Study Two, the transcriptions combined created a 60,000 word corpus. This was supplemented with the video recordings of participants' interactions with the agents, which in Study Two alone was over 24 hours long. The video data was not utilised extensively in this study, as the main source of data came from the

post-task questionnaires and interviews. The purpose of the video data was to aid the transcription of the audio interviews as and when the transcriptions became difficult to fully understand. When segments of the audio became unclear, the videos were observed in order to gain clarity from the lip movements and gestures of the participants, so that their speech could be better understood. This was not a frequent occurrence, however. Further analysis of the video data could be used for future research that focused on the facial, gestural, and verbal responses of participants during the interaction with the agents, though this remained out of scope for this thesis.

The textual corpus was also not analysed with traditional corpus linguistics methods (e.g. collocations or n-grams). However, they provided the necessary information to address the research aims. Its use also stimulates discussion on the pros and cons of a multimodal corpus approach to HAI. This is discussed further on this chapter.

7.3 Limitations and Future Research

7.3.1 Alternative Agent Designs

While there are notable contributions this thesis has made, there are still limitations that can be addressed. These mainly focus on the approaches to design and analyses of the studies. One of the fundamental aspects of both was the voice of the agents, of which the variety was limited. As the focus was on incremental research and keeping agents similar in characteristics particularly in Chapter 5, the scope was left purposefully small. There was no variation in the gender of the voices for example, and little in the way of age and accent. Furthermore, Chapter 5 did not consider the differences between lower and higher end synthesised voices discussed in previous research (Cowan et al., 2012; Georgila et al., 2012). While the two voices were received in similar ways, it is unknown how sophisticated or advanced the participants believe them to be. Although the participants discussed the synthesised voices as being less humanlike, providing them or an additional group with the opportunity to rate them may help build on the literature provided by Cowan et al. (2012) and Georgila et al. (2012). Comparing a wider range of synthesised voices and indeed a wider range of human voices could provide further research insight, as well as potential implications for agent designers.

The language too is another one of the fundamental aspects. While the VL model discussed in Chapter 3 provides detailed descriptions, it was not always received as intended. It was created with saturation in mind i.e. using enough VL to make it noticeable to participants, which may have oversaturated the vague agents somewhat. While there are other factors such as the voice and the nature of agents to consider, exploring alternative methods of applying and research VL in HAI may be useful. This could include the design of vague agents, as well as

researching VL in existing agents. The frequency of VL, which was discussed in Chapter 4 in particular, could be one alternative focus of research. Investigating existing agents may actually strengthen the understanding of a more appropriate VL model and provide further implications for agent designers. As discussed in 6.5, another approach to using VL may be to include a naturally occurring script with VL as the base instructions, before removing the VL to create non-vague instructions. This may create different results to the ones seen in this thesis, though I hypothesise that similar effects regarding voice and boundaries of identity would still appear.

Exploring alternative aspects of agent language is another possibility. Non-vague agents were discussed in Chapter 4, and covered by the relational approach discussed in Chapter 6. Further exploring areas such as direct and even impolite language in agents would create a broader understanding of the linguistics of HAI.

There are other aspects of the agent that were limited in this thesis and could be expanded upon in future research. The agent type for example, was heavily focused on voice as its primary medium of interaction. There was a visual interface, though this was minimalistic and did not provide visual information on the assembling of the Lego models, and tactile input from the participants. In both studies, however, there were calls for inclusion of enhanced visual information. These responses mainly focused on providing participants with some pictorial description of a piece they had to locate, or the process of attaching or assembling that was part of a step. A greater focus on multimodality may affect the way in which an agent's VL is perceived. As some agents combine visual and voice information (e.g. sat navs), this could be a promising avenue of investigation.

One area that this thesis did not explore was that of the adaptability of agents. This has been touched upon in Chapters 4-6, but was not part of the aims or methodology. While there were observable patterns in the responses towards VL and voices in the two studies, there is still the individual to consider. If a user did not respond positively towards VL in either study, there was no means for them to alter the agent. An agent could adapt either implicitly through its own decision making (e.g. algorithms; Wizard of Oz studies), or explicitly via the user (e.g. customisation through interface; verbally during interaction). In either instance, it may help to reduce the social distance between agent and user (6.3.2) and affect their perception of VL and identities they create for the agent. There are further discussions to be had on what can be adapted and who chooses the adaptability and when, amongst other things.

7.3.2 Interactions and Analyses

Having an adaptive agent can change the type of interaction it has with its users and there are other approaches to interaction that have limitations in this thesis. Firstly, as well as agent variation, there was not a fully counterbalanced approach to the participants. While there was a welcome mix, particularly in the second study, there could have been further measures in place to equally weight variables such as gender, age, and background, for example. The age range was also broader than in Study One. There were a large amount of students taking part in both studies, and the age range had a minimum limit of 18 years old. Those under the age of 18 could be investigated, with appropriate consent, as their perceptions of what agent likeness is may differ from older age groups. Older generations have lived through various changes in technology, whereas the younger have grown up with a particular minimum level of technology that they will be used to and expect to interact with (2.3). These variables were not strong focal points of this thesis, but for future research they may be.

While the qualitative data was successful in gathering detailed insights into the user experiences with agents, the quantitative questionnaire measures had less of an impact in both studies. These questionnaires were bespoke for these studies, although they were influenced by those used in previous literature. Other scholars have opted for using attitudinal scales that have been shown to have a significant impact for authors other than themselves. Choosing a statistically sound questionnaire may prove more insightful than tailoring one for each specific set of investigations. It is also worth mentioning that other quantitative measures such as the use of fNIRS have been used in assessing agent factors such as language (Strait et al., 2014). Combining these with a more thorough qualitative approach may lead to more detailed and insightful results.

Regarding analyses, one of the research aims was not observed in as much details as the others. The issue of task performance did not appear to be significant in many of the analyses, and was not focused on during the qualitative analyses of interviews. The issues of identity and relational work, amongst others, emerged as stronger candidates for focus and took precedence over task performance. Reintroducing this in further studies may be useful for agent designer in limited duration environments, and may be an approach worth refining.

The context of the agent interactions was another aspect of the interaction that had an even smaller focus – instruction giving. This was commented upon by participants in Chapters 4 and 5. It was discussed that the vague agents may be perceived more positively in non-instructive contexts, where there is less of a focus of conducting a task, particularly in the limited duration sessions. Interactions with intelligent personal assistants could be one of these alternative

contexts, as the users of them choose when to interact with them. While participants could choose when to receive the information from agents in this thesis, this was under a laboratory setting where completing the task was the purpose. With something like Siri or Google Now, users would expand the contexts in which research could be focused on. Agents taking on other roles than the instructor could also be investigated, which may contribute to a fuller understanding of how agents are positioned socially by their users when taking on different roles. More than one agent or user could also be included in the interaction, which could explore the linguistics of HAI group dynamics.

Another area to investigate is that of agents that can be spoken to as well as use speech themselves. Interactions here were more passive and participants could not talk back to the agent. The use of natural language in agents is becoming more common (Cowan et al., 2015) and so having the human user being a speaking participant too could show a very different interaction than the agents used here. Other agent features that could be altered include the use of embodiment, such as a human avatar with the interface. The exploration of these alternative interaction types could include the use of the abovementioned personal assistants, and take research into HAI and identity creation into the wild and towards real world data.

7.4 Summary

The discussions in this thesis indicate that there are many unanswered questions regarding the nature of agents, and the effects that language has on the creation of their identities by their users. However, the findings suggest that both voice and language are important variables in how identities and attitudes are created. Specifically in the two studies discussed in this thesis, verbal agent instructors with synthesised voices are likely better equipped to produce direct language, rather than vague. Similarly, the same agent types possessing a professional human voice are better at producing VL, though there are no guarantees that it will be marked as appropriate by its users due to differences in agent and human identities. Furthermore, there are the third parties of agent designers and researchers to consider in HAI, as it is not always an interaction exclusive to just agents and users. There are numerous avenues of investigation to explore and numerous ways of collecting data. Future research, however, may have to consider the benefits of a corpus approach and assessing the viability of the idea. As interactions with verbal agents increase, it is important to understand their nuances, and this may require moving data collection towards being relative to the increasing interactions. It is important we comprehend how the linguistic capabilities will affect these interactions, where the boundaries between agent and human likeness exist, and how these will progress as agent technology develops and diversifies.

Guide to Appendices

Appendix A. Full account of the written instructions for each model and agent type.

Appendix B. Supporting documents for Study One: the task information sheet presented to participants, the consent form they then signed and dated, and an example questionnaire that was completed after each task.

Appendix C. Full tables of the interaction preferences discussed in 4.4.3.

Appendix D. Supporting documents for Study Two. These are the same as for Study One in Appendix B, but also include the semi-structured interview guide.

Appendix A: Instructions for both models and agent types

Aquagon Non-vague Instructions

1. Locate the two black feet and place them claws down on the desk.
2. Take the two small black Y shaped pieces with a crossed hole at the bottom.
3. Put these in the rear gap of each foot and attach them to the crossed connector.
4. Push them towards the front of the feet until firmly in place and the round sockets are pointing vertically towards the ceiling.
5. Take two of the yellow pieces with ball joints.
6. Attach the ball joints to the sockets of the black pieces so the yellow socket is pointing vertically in the same way.
7. Pick up the two black pieces with ball joints.
8. Attach them to the yellow sockets so they are again pointing vertically the same way.
9. Keep these black pieces vertical and twist each one 90 degrees to the right. These are the legs.
10. Find the two small dark grey armour pieces.
11. Attach them to the remaining black ball joints on the outside of each leg, so the widest end is closest to the feet.
12. Locate the largest black piece with seven ball joints. This is the body.
13. Attach the narrowest end of the body into the black leg sockets using the two ball joints on the sides.
14. Take a black cylinder, a grey cylinder and a small light grey piece with a curved fin.
15. Place the black cylinder into the bottom round hole at the back of the body piece.
16. Attach the grey piece with the fin to this cylinder so the wider end is closest to the body.
17. Twist this piece so the fin is pointing towards the desk.
18. Place the grey cylinder in the rear hole of the grey finned piece slitted end first.
19. Locate the largest dark grey armour piece.
20. Attach the socket to the middle ball joint on the front of the body so the narrow end points downward.
21. To make the arms take two grey pieces with ball joints, two yellow pieces with ball joints and the two black fists.
22. Connect the yellow joints to the socket of each fist.
23. Connect the end grey joints to each yellow socket.
24. Attach the grey socket of each arm onto the top two black joints of the body.
25. Locate the two small, identical transparent blue pieces and the four curved yellow spikes.
26. Place the big spikes into the outside holes closest to the edge of

each piece.
27. Place the small spikes into the other outside hole.
28. Attach the whole thing to the grey joints on either arm so the end with the bigger spikes is closest to the top of the body.
29. Twist each spike so they point towards the top of the body.
30. Take the remaining yellow piece with a ball joint.
31. Attach the yellow socket to the top black joint of the body so the piece points forwards and is in line with the feet.
32. Take the remaining transparent blue piece. This is the head.
33. Attach the round socket of the head to the joint of the yellow piece so the length of the piece points forward. One crossed hole points forwards and one points vertically towards the ceiling.
34. Locate the large blue and red piece.
35. Attach the crossed red connector to the crossed blue hole on top of the head so that the eyes face forward.
36. Take the tail end and place it inside the back hole of the grey fin at the back of the body.
37. Locate the yellow face piece.
38. Attach it to the remaining hole on the front of the head.
39. Take the remaining grey finned pieces and two blue cylinders.
40. Place the cylinders into the back of each fist using the crossed connector.
41. Attach the finned pieces to the cylinders and twist them so the fins point to the sides.
42. Take the black cylinder and blue swords.
43. Place the cylinder into the remaining hole in the left fin.
44. Take each sword and attach it to the front hole of each fist and point them to the sides.
45. Turn the arms so they point forwards.
46. Connect the two finned pieces together using the black cylinder.
47. Twist one of the fins so it points towards the body.

Aquagon **Vague Instructions**

1. To start with locate the two black feet and place them claws down on the desk.
2. Now take two of the small black pieces that are sort of a Y shape and have a crossed hole at the bottom.
3. Just put these in the rear gap of each foot and attach them to the cross-shaped connector.
4. These should be pushed towards the front of the feet until they are more or less firmly in place. The round socket should point vertically towards the ceiling.
5. So now take two of the yellow pieces with ball joints.
6. Just attach the ball joints to the sockets of the black pieces so the yellow socket is pointing vertically in pretty much the same way.
7. Now pick up the two black pieces with ball joints.

8. Just attach them to the yellow sockets so they are again pointing vertically the same way.
9. So keep these black pieces vertical and just twist each one a little bit 90 degrees or so to the right. These are the legs.
10. Now find the two small dark grey pieces that look something like armour.
11. These should attach to the remaining black ball joints around the outside of each leg. The widest end should be closest to the feet.
12. Now locate the largest black piece that has seven ball joints. This is the body.
13. Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs.
14. Now take a black cylinder, a grey cylinder and a small light grey piece with a curved thing that looks a bit like a fin.
15. Just place the black cylinder into the bottom round hole at the back of the body piece.
16. Now just attach the grey piece with the fin to this cylinder. The end that looks a bit wider should be closest to the body.
17. Just give this piece a little bit of a twist so the fin is more or less pointing towards the desk.
18. The grey cylinder should go in the rear hole of the grey finned piece slitted end first.
19. Now locate the largest dark grey piece that looks like a big piece of armour.
20. Just attach this to the middle ball joint on the front of the body using the round socket. The end that's a bit more narrow than the other should point downwards.
21. Now to make the arms just take two grey pieces with ball joints, two yellow pieces with ball joints and the two small black pieces that look like fists.
22. Just connect the yellow joints to the socket of each fist.
23. The same should be done with the end grey joints in the yellow sockets.
24. The grey socket of each arm should be attached onto the top two black joints of the body.
25. Now locate the two small, identical transparent and the four curved yellow pieces that sort of look like spikes.
26. Just place the big spikes into the holes that are closest to the edge of each piece.
27. Then just place the small spikes into the other outside hole so they are just below the big ones.
28. Now just attach the whole thing to the available grey joints on the arms. The end with the big spikes should be closest to the top of the body.
29. Then just give each spike a little twist so they point towards the body.
30. Now take the remaining yellow piece with a ball joint.

31. Just attach the yellow socket to the top black joint of the body. The yellow piece should point forwards so it's more or less in line with the feet.
32. Now take the remaining blue piece that looks a bit transparent. This is the head.
33. Basically attach the socket of the head to the joint of the yellow piece. The longer part of the head should point forwards. One crossed hole should point forwards and one should be pointing vertically towards the ceiling a bit like the other pieces.
34. So now locate the large blue and red piece.
35. Basically, attach the crossed red connector to the crossed blue hole on top of the head. The eyes should face forward.
36. Now take the tail end and place it inside the back hole of the grey fin at the back of the body.
37. So now locate the yellow face piece.
38. Just attach it to the remaining hole on the front of the head.
39. Now take the remaining grey finned pieces and two blue cylinders.
40. These cylinders should be placed into the back of each fist using the crossed connector.
41. Just attach the finned pieces to the cylinders and then just twist them a bit so the fins point to the sides.
42. Now take the black cylinder and blue sword pieces.
43. Just place the cylinder into the remaining hole in the left fin.
44. Now just take each sword and just attach it to the front hole of each fist. They should both be basically pointing outwards to the sides.
45. The arms then have to be turned so they are more or less pointing forwards.
46. The two finned pieces should be connected together using the black cylinder.
47. Finally just twist one of the fins a bit so it faces the body.

Nex Non-vague Instructions

1. Locate the two orange feet and place them on the desk.
2. Take two of the black pieces with two two ball joints.
3. Attach the ball joints to the sockets of the orange pieces so the black socket is pointing vertically in the same way.
4. Keep these black pieces vertical and twist each one 90 degrees to the right.
5. Take two of the grey pieces with ball joints.
6. Attach the ball joints to the sockets of the black pieces so the grey sockets are pointing vertically in the same way.
7. Pick up the two identical white armour pieces.
8. Attach them to the front joint of each black piece so that the ends with the holes are closest to the grey pieces.
9. Find two of the orange armour pieces that again are the same size.
10. Attach them to the outside of each grey piece so that the ends with the holes are closest to the sockets.
11. Locate the largest black piece with seven ball joints. This is the body.
12. Attach the narrowest end of the body to the black leg sockets using the two ball joints on the sides.
13. Pick up the two remaining grey pieces with ball joints.
14. Attach the sockets onto the top two black joints of the body.
15. Take the two remaining black pieces with ball joints.
16. Attach the sockets to the joints of the grey pieces so they are in line with each one.
17. Locate the large white armour piece.
18. Attach the socket to the middle ball joint on the front of the body so the narrow end points downward.
19. Take the white piece with the ‘H’ in the middle.
20. Attach this to the holes on the top of the white piece.
21. Pick up the small green and the large orange head pieces.
22. Keep the eyes of both pieces facing forwards, and connect the orange crossed connector inside the large piece to the green crossed hole of the smaller piece.
23. Attach the socket inside the green piece to the top black joint of the body so the eyes again face forward.
24. Locate the long, thinnest black piece and a small orange armour piece.
25. Attach the orange piece to the grey socket on the right arm so the end with the holes is closest to the head.
26. Place the black piece inside the orange hole closest to the head.
27. To make the right arm weapon pick up the two crossed red connectors, the black fist and the long orange piece.
28. Connect the red piece to the crossed connector of the orange piece so the other half sticks out.

29. Connect the black fist to the other end of this connector so the socket is pointing away from the orange piece.
30. Connect the other red piece to the crossed hole of the fist.
31. Pick up the large grey piece that is narrower than the other and does not have a black attachment.
32. Connect the crossed hole of the grey piece to the remaining half of the red connector so it lines up with the orange piece.
33. Attach this weapon to the right arm of the body.
34. Pick up the green ball.
35. Place it in between the grey and orange pieces.
36. Locate the remaining orange armour piece and smaller grey armour piece.
37. Attach the orange piece to the grey ball joint on the left arm so the end with the holes is closest to the head.
38. Attach the grey piece to the orange holes so the wider end is closest to the head.
39. To make the left arm weapon pick up the large grey piece and smaller green tube.
40. Connect the widest end of the tube to the smallest end of the black attachment on the grey piece.
41. Attach the socket of the grey piece to the black ball joint on the left arm.Turn the arms so they point forwards.
42. Pick up the long yellow tube.
43. Place one end in the remaining side of the black attachment.
44. Place the other in the bottom left hole at the back of the large white armour piece.
45. Take the long chain and the small grey piece. Place the grey piece in the top hole of the orange armour piece on the right leg.
46. Attach one wide end of the chain to this small grey piece.
47. Rotate this wide end so it is pointing vertically.

Nex **Vague Instructions**

1. To start with locate the two orange feet and place them on the desk.
2. Now take two of the black pieces with two ball joints.
3. Then just attach the joints to the sockets of the oranges pieces. The black joint should just be pointing vertically in the same way.
4. So keep these black pieces vertical and just give each one a little twist 90 degrees or so to the right.
5. So now take two of the grey pieces with ball joints.
6. Just attach them to the sockets of the black pieces. The sockets should be pointing vertically in pretty much the same way.
7. Now pick up the two small identical white pieces that look something like armour.
8. These should attach to the front joint of each black piece. The ends with the holes should be closest to the grey piece.

9. Now find two of the orange pieces that again are the same size.
10. Just attach them to the outside of each grey piece. The ends with the holes should be closest to the sockets.
11. Now locate the largest black piece with seven ball joints. This is the body.
12. Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs.
13. Now pick up the two remaining grey pieces with ball joints.
14. Just attach the sockets onto the top two black joints of the body.
15. Then just take the two remaining black pieces with ball joints.
16. The same should be done with the sockets of the grey pieces so they are more or less in line with each one.
17. Now locate the large white armour piece.
18. This should attach to the middle ball joint on the front of the body and the narrow end should point downwards.
19. Now just take the white piece that has like an 'H' in the middle.
20. Just attach this to the holes on the top of the white piece.
21. Now pick up the small green and the large orange pieces that look like heads.
22. So keep the eyes of both pieces facing forwards and basically just connect the orange crossed connector inside the large piece to the green crossed hole of the smaller piece.
23. Now just attach the socket inside the green piece to the top black joint of the body. The eyes again should more or less be facing forwards.
24. Now locate the small, thinnest black piece and a small orange armour piece.
25. Just attach the orange piece to the grey socket on the right arm. The end with the holes should be closest to the head.
26. The black piece has to then be placed just inside the orange hole.
27. Now to make the right arm weapon just pick up the two crossed red connectors, the black fist and the long orange piece.
28. Basically, connect the red piece to the crossed connector of the orange piece. The other half should stick out at the other end.
29. Just connect the black fist to the other end of this connector so the socket is just pointing away from the orange piece.
30. Then just connect the other red piece to the crossed hole of the fist.
31. Now pick up the large grey piece that is more narrow than the other and does not have, like, a small black attachment.
32. Just connect the crossed hole of the grey piece to the remaining half of the red connector so it more or less lines up with the orange piece.
33. Now pick up the green ball.
34. This should be placed in between the grey and orange pieces.
35. Now locate the remaining orange armour piece and the smaller grey armour piece.

36. Just attach the orange piece to the grey ball joint on the left arm. The end with the holes should be closest to the head.
37. Now just attach the grey piece to the orange holes. The end that looks a bit wider should be closest to the head.
38. Now to make the left arm weapon just pick up the large grey piece and smaller green piece that looks a bit like a tube.
39. The widest end of the tube should connect to the smallest end of the black attachment just on the grey piece.
40. The socket of the grey piece should be attached to the black ball joint just on the left arm.
41. So now the arms should be turned so they are more or less pointing forwards.
42. Now pick up the long yellow piece that looks like a tube. One end should be placed in the remaining hole of the black attachment.
43. The other should be placed in the bottom left hole around the back of the large white piece.
44. Now take the small grey piece and the longer one that looks sort of like a chain.
45. So then place the grey piece in the top hole of the orange armour piece that is just on the right leg.
46. One wide end of the chain thing should be attached to this small grey piece.
47. Now just rotate this wide end a little so it is pretty much pointing vertically.

Appendix B. Study One Supporting Documents

Agent Instructed Assembly Tasks - Study Information Sheet

We are a team of researchers from the Mixed Reality Lab, University of Nottingham working on an EPSRC funded project called ORCHID. An area of research we are particularly interested in is the interaction between people and computers. In the future, the number of computer devices used by humans will increase dramatically, adding complexity to the way we interact with technology.

One approach to help make these interactions as easy as possible is to develop computers with advanced automated functions, with more control over everyday matters in our lives. One of the aims of ORCHID is to explore the different ways in which people react to this automated technology that is able to act on their behalf, and also present them with information of interest/benefit.

As part of our research, we have developed a group of tasks in which you will be expected to construct two Lego models by following spoken instructions received from a computer device – referred to as the *agent*. During each task you will be reliant on the information received by the agent as no visual cues will be provided.

The construction of each model is considered a separate task and each will be followed by a short online questionnaire and debriefing interview. We will be collecting video and audio recordings of you over these two sessions, encompassing the tasks, questionnaires and interviews. The two sessions will be separated by a short break. The expected total length of the study is approximately 90 minutes and you will receive the compensation of a £10 Amazon voucher for your participation.

This data gathered from this study is important, as it will help us develop a better knowledge of the human response to agent instructions and also positively influence future interactive technologies.

All of the data we collect will be held in a secure and safe manner in accordance with the Data Protection Act 1998. Access to this data will be restricted to researchers involved in the project, and will be processed confidentially without ever linking it to your name or identity. It is within your rights to refuse the collection of any of the specific types of data identified above.

The results from the study will be used for publication in academic conferences and journals. You are free to withdraw at any point during, or after, the study and any data collected will be erased from our records. To withdraw, simply inform the researcher during the

study or use the following details to contact us with any queries you might have:

Leigh Clark
Mixed Reality Lab (MRL)
School of Computer Science
University of Nottingham
Jubilee Campus
Nottingham
NG8 1BB

Leigh.Clark@exmail.nottingham.ac.uk

Agent Instructed Assembly Tasks Consent Form – Pre-Task

I confirm that I have agreed to take part in this study, have read the information sheet provided and understand what is involved.
I understand that the study will gather recordings of my participation, and I agree to the use of this data in an anonymised form.
I understand that I can withdraw at any time by informing the researcher conducting this study, and my personal data will be erased from the records.
I confirm that I am over the age of 18.

This is to confirm that I have agreed to take part in this research study:

Signed ……..

Name ……..

Date ……..

Agent Instructed Assembly Tasks Consent Form – Post-Task

I confirm that I have participated in the Agent Instructed Assembly study and I consent to audio and video recordings as well as images that may identify my participation in this research being used for publication purposes such as in conference presentations, papers and academic websites. My name and other personal data will never be linked to any of those publications.

Signed ……..

Name ……..

Date ……..

1. How would you rate the task in terms of the following words?

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Stressful					
Enjoyable					
Mentally Challenging					
Physically Challenging					

2. Is there anything you would change about the task and why?

3. Please rate the following attributes of the agent

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Friendly					
Authoritative					
Trustworthy					
Likeable					
Controlling					
Sociable					
Clear					
Direct					

4. How else would you describe the agent and why?

5. Did the wording of the instructions have any effect on your impression of the agent?

6. What were your feelings towards the agent's voice? Would you have preferred being instructed by a human voice instead?

7. Would you be happy to interact with the agent again?

8. Would you be happy to have the same voice for personal devices e.g. smartphone, sat nav?

Appendix C. Interaction Preferences (Study One; 4.4.3)

Prefer Humanlike? Reasons

	Non-vague	Vague	*Total*
Yes: Little / No Elaboration on Preference	1	1	2
Yes: But Agent Voice is Fine	3	3	6
Yes: Robotic/Lacked Emotion/Voice	4	2	6
Yes: Other	1	1	2
Yes: Easier to Understand	1	8	9
Yes: Because of the Words	0	3	3
No: No Elaboration	2	0	2
No: Though Improved Voice Tone Perhaps Better	0	1	1
Maybe: Depends on Human Voice Style/Type	3	0	3
Maybe: No strong feelings / voice fine	4	3	7
Maybe: Human Could be Better	4	0	4
Reasons Not Clear: Answer lacks confirmation	1	2	3
Totals	*24*	*24*	*48*

Voice for Personal Device? Reasons

	Direct	Vague	*Total*
Yes: Little/No Elaboration	7	6	*13*
Yes: With Specific Tasks	2	0	*2*
Yes: Though would make it less personal	1	0	*1*
Yes: But human voice still better	0	3	*3*
Yes: Yes to voice, but no to the language	0	1	*1*
Yes: Don't care about the voice	1	0	*1*
No: Little/No Elaboration	5	4	*9*
No: But yes for non-personal contexts	1	0	*1*
No: Voice not clear enough	2	0	*2*
No: Prefer another existing voice	1	0	*1*
No: Prefer human voice / more humanlike	0	3	*3*
No: Would be annoying	0	3	*3*
No: Because of the language	0	3	*3*
Maybe: But prefer a human accent voice	1	0	*1*
Maybe: But a bit impersonal	1	0	*1*
Maybe: Depending on context	1	0	*1*
Not Clear: Lack of confirmation	1	1	*2*
Totals	*24*	*24*	*48*

Appendix D. Study Two Supporting Documents

Lego Assembly Tasks - Study Information Sheet

We are a team of researchers from the Mixed Reality Lab and the School of English, University of Nottingham working on an EPSRC funded project called ORCHID. An area of research we are particularly interested in is the interaction between people and computers. One of the aims of ORCHID is to explore the different ways in which people react to this automated technology that is able to act on their behalf, and also present them with information of interest/benefit.

As part of our research, we have developed a group of tasks in which you will be expected to construct two Lego models by following verbal instructions received from a computer, henceforth referred to as the *instructor*. During each task you will be reliant on the information received by the instructor as no visual cues will be provided. These tasks will be followed by a short questionnaire and interview. The total length of the study will not exceed 60 minutes in length and you will receive the compensation of a £10 Amazon voucher for your participation.

All of the data we collect will be held in a secure and safe manner in accordance with the Data Protection Act 1998. Access to this data will be restricted to researchers involved in the project, and will be processed confidentially without ever linking it to your name or identity. It is within your rights to refuse the collection of any of the specific types of data identified above.

The results from the study will be used for publication in academic conferences and journals. You are free to withdraw at any point during, or after, the study and any data collected will be erased from our records. To withdraw, simply inform the researcher during the study or use the following details to contact us with any queries you might have:

Leigh Clark
Mixed Reality Lab (MRL)
School of Computer Science
University of Nottingham
Jubilee Campus
Nottingham
NG81BB 1

Leigh.Clark@nottingham.ac.uk

Lego Assembly Tasks Consent Form

I confirm that I have agreed to take part in this study, have read the information sheet provided and understand what is involved.
I understand that the study will gather recordings of my participation, and I agree to the use of this data in an anonymised form for research and analysis.
I understand that I can withdraw at any time by informing the researcher conducting this study, and my personal data will be erased from the records.
I confirm that I am over the age of 18.

This is to confirm that I have agreed to take part in this research study on the date:

……..

Signed ……..

Email: ……...

In addition to the data analysis, I give permission for data that could identify me
(e.g. photos, video) to be used in publications, conferences, presentations and future research □
Signed ……..

Demographics Questions

1. How old are you / Age range?
2. What is your nationality?
3. What course are you currently studying?
4. What year are you in?
5. Have you assembled Lego before?
    a. If yes, how would you rate your skill level in assembling Lego?
    1 = Very Good, 2 = Good, 3 = Average, 4 = Poor, 5 = Very Poor

Post-task Questionnaire

The paper describes a questionnaire that each participant fills out after each task. The characteristics were deemed as too extensive to include in the full paper and so an example questionnaire is included here. The order of questions 1-18 was randomised and the order of the ratings flipped between participants (Strongly Agree to Strongly Disagree became Strongly Disagree to Strongly Agree). The font has been changed to reflect to match that used throughout the main body of the paper (Century Schoolbook 10pt.).

Lego Assembly 2: Task One
Please rate how strongly you agree or disagree with each of the following statements

***1. The instructions were unimposing**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***2. The instructions enabled me to complete the task**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***3. The instructions were assertive**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

4. I like the voice

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***5. I wasn't able to comprehend the instructions**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***6. The instructions were intelligible**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***7. The voice giving the instructions was humanlike**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***8. The instructions were aware of my needs**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***9. The instructions were precise**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***10. The instructions were unhelpful**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***11. The voice is annoying**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***12. The instructions were rude**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***13. The instructions made me anxious**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***14. If I were assembling Lego again, I wouldn't want the voice around**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***15. The instructions were incoherent**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***16. The instructions are kind**

- ○ Strongly Agree
- ○ Agree
- ○ Neither agree or disagree
- ○ Disagree
- ○ Strongly Disagree

***17. The instructions made me apprehensive of interacting with something similar**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***18. The instructions were not controlling**

- ○ Strongly Disagree
- ○ Disagree
- ○ Neither agree or disagree
- ○ Agree
- ○ Strongly Agree

***19. How old would you say the voice giving the instructions sounds?**

Interview Questions.

With each question, there were opportunities to ask for elaboration on the participants' responses if required. These questions were not always spoken in the same way, and some were omitted depending on how the interview proceeded. As such it was a semi-guided interview. Question 5 was asked following P17 where it appeared the wording of the question in the questionnaire was not providing the answers expected, and so this was followed up in the interviews.

1. How was the task?
    a. Things found easy or difficult
2. Compared to the first task (when doing the second)
3. What were your thoughts on the voice giving the instructions?
    a. Comparisons with first voice (when doing the second)
4. Was there anything about the language that you noticed?
    a. Comparisons with the first agent (when doing the second)
5. Did you have to repeat any instructions?
6. Were you able to put an age to the voice?
7. Were there any observations you made during the task?

References

ADOLPHS, S., ATKINS, S. & HARVEY, K. 2007. Caught between professional requirements and interpersonal needs: Vague language in healthcare contexts. *Vague language explored*, 62-78.

AGRAWALA, M., PHAN, D., HEISER, J., HAYMAKER, J., KLINGNER, J., HANRAHAN, P. & TVERSKY, B. Designing effective step-by-step assembly instructions. ACM Transactions on Graphics (TOG), 2003. ACM, 828-837.

ANDERSEN, G. 1998. The pragmatic marker like from a relevance-theoretic perspective. *PRAGMATICS AND BEYOND NEW SERIES*, 147-170.

BALL, M. & CALLAGHAN, V. Introducing Intelligent Environments, Agents and Autonomy to Users. Intelligent Environments (IE), 2011 7th International Conference on, 2011. IEEE, 382-385.

BARON, N. S. 2015. Shall we talk? Conversing with humans and robots. *The Information Society,* 31, 257-264.

BBC. 2008. *Supermarket tills to speak Welsh* [Online]. Available: http://news.bbc.co.uk/1/hi/wales/7636177.stm [Accessed 22 February 2016].

BECKER-ASANO, C., ARRAS, K. O. & NEBEL, B. Robotic tele-presence with DARYL in the wild. Proceedings of the second international conference on Human-agent interaction, 2014. ACM, 91-95.

BENWELL, B. & STOKOE, E. 2006. *Discourse and identity*, Oxford University Press.

BhamUrbanNewsUK. (2015, August 10). West Midlands: New male voice for Tesco's self -service checkout. [Video File]. Retrieved from https://www.youtube.com/watch?v=tNBH1SRvlos.

BICKMORE, T. W. 2003. *Relational agents: Effecting change through human-computer relationships* (Doctoral dissertation, Massachusetts Institute of Technology).

BICKMORE, T. & CASSELL, J. Relational Agents: A Model and Implementation of Building User Trust. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2001. pp. 396-403. ACM

BICKMORE, T., & GRUBER, A. 2010. Relational agents in clinical psychiatry.*Harvard review of psychiatry*, *18*(2), 119-130.

BICKMORE, T. W., & PICARD, R. W. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(2), 293-327.

BLACK, A.W. 2002. Perfect synthesis for all of the people all of the time. *Speech Synthesis, 2002. Proceedings of 2002 IEEE Workshop on* (pp. 167-170). IEEE.

BRANIGAN, H. P., PICKERING, M. J., PEARSON, J. & MCLEAN, J. F. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics,* 42, 2355-2368.

BRINTON, L. J. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions*, Walter de Gruyter.
BROWN, G. & YULE, G. 1983. *Discourse analysis*, Cambridge University Press.
BROWN, P. & LEVINSON, S. C. 1987. *Politeness: Some universals in language usage*, Cambridge University Press.
BUCHOLTZ, M. & HALL, K. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse studies,* 7**,** 585-614.
BULITKO, V. V. & WILKINS, D. C. Automated instructor assistant for ship damage control. AAAI/IAAI, 1999. 778-785.
CAMERON, D. 2001. *Working with spoken discourse*, Sage.
CARTER, R. 1998. Orders of reality: CANCODE, communication, and culture. *ELT journal,* 52**,** 43-56.
CARTER, R. 2004. *Language and Creativity: The Art of Common Talk*, Routledge.
CASSELL, J. Social practice Becoming Enculturated in Human-Computer Interaction. Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Part III: Applications and Services, 2009 San Diego, CA, USA. 303-313.
CASSELL, J. & TARTARO, A. 2007. Intersubjectivity in Human-Agent Interaction. *Interaction Studies,* 8**,** 391-410.
CHAN, V., RAY, P. & PARAMESWARAN, N. 2008. Mobile e-Health monitoring: an agent-based approach. *IET communications,* 2**,** 223-230.
CHANNELL, J. 1994. *Vague language*, Oxford University Press.
CHENG, W. 2007. The use of vague language across spoken genres in an intercultural Hong Kong corpus. *Vague language explored, 161-181.*
CHENG, W. & O'KEEFFE, A. 2014. 13 Vagueness. *Corpus Pragmatics: A Handbook***,** 360.
CHENG, W. & WARREN, M. 2003. Indirectness, inexplictness and vagueness made clearer. *Pragmatics, 13*(3/4) 381-400.
CLARK, H. H. 1996. *Using language*, Cambridge University Press Cambridge.
CLARK, H. H. & KRYCH, M. A. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language,* 50**,** 62-81.
CLARK, L. M. H., BACHOUR, K., OFEMILE, A., ADOLPHS, S. & RODDEN, T. 2014. Potential of imprecision: exploring vague language in agent instructors. *Proceedings of the second international conference on Human-agent interaction.* Tsukuba, Japan: ACM.
COLLINS, K. 2015. *Tesco's self-checkout robot is becoming less 'shouty'* [Online]. Available: http://www.wired.co.uk/news/archive/2015-07/30/tesco-self-checkout-voice [Accessed 22 February 2016].
COTTERILL, J. 2007. 'I think he was kind of shouting or something': Uses and abuses of vagueness in the British courtroom, *Vague language explored (97-114).*

COULTHARD, M. 2013. *Advances in spoken discourse analysis*, Routledge.
COWAN, B., R., BEALE, R. & BRANIGAN, H. P. Investigating syntactic alignment in spoken natural language human-computer communication. CHI EA '11 CHI '11 Extended Abstracts on Human Factors in Computing, 2011 Vancouver, Canada. 2113-2118.
COWAN, B. R., BRANIGAN, H. P. & BEALE, R. Investigating the impact of interlocutor voice on syntactic alignment in human-computer dialogue. Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers, 2012. British Computer Society, 39-48.
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.
CULPEPER, J. 1996. Towards an anatomy of impoliteness. *Journal of pragmatics,* 25**,** 349-367.
CUTTING, J. 2007. *Vague Language Explored*, Palgrave Macmillan.
CUTTING, J. 2012. Vague language in conference abstracts. *Journal of English for Academic Purposes,* 11**,** 283-293.
DAHLBÄCK, N. & JONSSON, I.-M. Impact of Voice Variation in Speech-based In-Vehicle Systems on Attitude and Driving Behaviour. Human Factors: A System View of Human, Technology and Organisation. Annual Conference of the Europe Chapter of the Human Factors and Ergonomics Society 2009, 2010. 395-408.
DAHLBÄCK, N., SWAMY, S., NASS, C., ARVIDSSON, F. & SKÅGEBY, J. Spoken Interaction with Computers in a Native or Non-native Language - Same or Different? INTERACT 2001, 2001 Tokyo, Japan. 294-301.
DAHLBÄCK, N., WANG, Q., NASS, C. & ALWIN, J. 2007. Similarity is more important than expertise:accent effects in speech interfaces.
DE FINA, A. 2006. Group identity, narrative and self-representations. In: De Fina et al. (eds) *Discourse and Identity*. pp.351-375. [Online]. Studies in Interactional Sociolinguistics. (No. 23). Cambridge: Cambridge University Press. Available from: Cambridge Books Online <http://dx.doi.org/10.1017/CBO9780511584459.018> [Accessed 15 July 2015].
DE FINA, A., SCHIFFRIN, D. & BAMBERG, M. 2006. *Discourse and identity*, Cambridge University Press.
DE GRAAF, M. M., ALLOUCH, S. B. & KLAMER, T. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in human behavior,* 43**,** 1-14.
DOSWELL, J. & HARMEYER, K. Extending the'serious game'boundary: Virtual instructors in mobile mixed reality learning games.

Digital Games Research Association International Conference (DiGRA 2007), 2007. Citeseer.
DRAVE, N. 2001. Vague speaking: a corpus approach to vague language in intercultural conversations. *Language and Computers*, 36(1), pp.25-40.
EELEN, G. 2014. *A critique of politeness theory*, Routledge.
FEARON, J. D. 1999. What is identity (as we now use the word).
FLEISCHMAN, S. & YAGUELLO, M. 2004. Discourse markers across languages. *Discourse across languages and cultures*.
FOGG, B. J. & NASS, C. 1997. Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies,* 46**,** 551-561.
FONG, T., NOURBAKHSH, I. & DAUTENHAHN, K. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems,* 42**,** 143-166.
FORBES-RILEY, K., LITMAN, D. J., SILLIMAN, S. & TETREAULT, J. R. Comparing Synthesized versus Pre-Recorded Tutor Speech in an Intelligent Tutoring Spoken Dialogue System. FLAIRS Conference, 2006. 509-514.
FRASER, B. 1990. Perspectives on politeness. *Journal of Pragmatics,* 14**,** 219-236.
FRASER, B. 2010. Pragmatic competence: The case of hedging. *New approaches to hedging***,** 15-34.
GEORGILA, K., BLACK, A., SAGAE, K. & TRAUM, D. R. Practical Evaluation of Human and Synthesized Speech for Virtual Human Dialogue Systems. LREC, 2012. 3519-3526.
GIBSON, J. J. 1977. The theory of affordances. *Hilldale, USA*.
GLEASON, P. 1983. Identifying identity: A semantic history. *The journal of American history***,** 910-931.
GOFFMAN, E. 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*, Anchor Books.
GOFFMAN, E. 2002. The presentation of self in everyday life. 1959. *Garden City, NY*.
GOFFMAN, E. 2012. The presentation of self in everyday life [1959]. *Contemporary sociological theory***,** 46-61.
GRICHKOVTSOVA, I., MOREL, M. & LACHERET, A. 2012. The role of voice quality and prosodic contour in affective speech perception. *Speech Communication,* 54**,** 414-429.
GROUP, L. P. R. 2011. *Discursive approaches to politeness*, Walter de Gruyter.
HALL, J. K. 2013. *Teaching and researching: Language and culture*, Routledge.
HALL, S. 1996. Who needs identity. *Questions of cultural identity,* 16**,** 1-17.
HALLIDAY, M. A. K. 1978. *Language as social semiotic*, London Arnold.
HALLIDAY, M. A. K. & HASAN, R. 2014. *Cohesion in english*, Routledge.
HÄRING, M., KUCHENBRANDT, D. & ANDRÉ, E. Would you like to play with me?: how robots' group membership and task features influence human-robot interaction. Proceedings of the 2014

ACM/IEEE international conference on Human-robot interaction, 2014. ACM, 9-16.
HAYASHI, M. & YOON, K.-E. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language,* 30**,** 485-540.
HEYLEN, D., NIJHOLT, A., OP DEN AKKER, R. & VISSERS, M. Socially intelligent tutor agents. Intelligent Virtual Agents, 2003. Springer, 341-347.
HOPPER, P. J., & Traugott, E.C. 2003. *Grammaticalization*. Cambridge University Press.
HUGHES, A., TRUDGILL, P. & WATT, D. 2013. *English accents and dialects: an introduction to social and regional varieties of English in the British Isles*, Routledge.
JAWORSKI, A. & COUPLAND, N. 2014. *The discourse reader*, Routledge.
JENNINGS, N. R., MOREAU, L., NICHOLSON, D., RAMCHURN, S., ROBERTS, S., RODDEN, T. & ROGERS, A. 2014. Human-agent collectives. *Commun. ACM,* 57**,** 80-88.
JIANG, J., AWADALLAH, A. H., JONES, R., OZERTEM, U., ZITOUNI, I., KULKARNI, R. G. & KHAN, O. Z. 2015. Automatic Online Evaluation of Intelligent Assistants. *Proceedings of the 24th International Conference on World Wide Web.* Florence, Italy: ACM.
JONSSON, I.-M. & DAHLBÄCK, N. I can't Hear You? Drivers Interacting with Male or Female Voices in Native or Non-Native Language. UAHCI'11 Proceedings of the 6th international conference on Universal access in human-computer interaction: context diversity - Volume Part III, 2011. 298-305.
JOSEPH, J. E. 2010. Identity. *In:* LLAMAS, C. (ed.) *Language and Identities.* Edinburgh: Edinburgh University Press.
JUCKER, A. H., SMITH, S. W. & LÜDGE, T. 2003. Interactive aspects of vagueness in conversation. *Journal of Pragmatics,* 35**,** 1737-1769.
JUCKER, A. H. & ZIV, Y. 1998. *Discourse markers: Descriptions and theory*, John Benjamins Publishing.
KAPTELININ, V. & NARDI, B. Affordances in HCI: toward a mediated action perspective. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012. ACM, 967-976.
KÄTSYRI, J., FÖRGER, K., MÄKÄRÄINEN, M. & TAKALA, T. 2015. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology,* 6.
KISELEVA, J., WILLIAMS, K., JIANG, J., AWADALLAH, A. H., CROOK, A. C., ZITOUNI, I. & ANASTASAKOS, T. 2016. Understanding User Satisfaction with Intelligent Assistants. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval.* Carrboro, North Carolina, USA: ACM.
KOESTER, A. 2007. About twelve thousand or so': Vagueness in North American and UK offices. *Vague language explored***,** 40-61.

KRAUT, R. E., FUSSELL, S. R. & SIEGEL, J. 2003. Visual information as a conversational resource in collaborative physical tasks. *Human-computer interaction,* 18**,** 13-49.
LAKOFF, G. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of philosophical logic,* 2**,** 458-508.
LASERNA, C. M., SEIH, Y.-T. & PENNEBAKER, J. W. 2014. Um... Who Like Says You Know Filler Word Use as a Function of Age, Gender, and Personality. *Journal of Language and Social Psychology***,** 0261927X14526993.
LATINUS, M. & BELIN, P. 2011. Human voice perception. *Current Biology,* 21**,** R143-R145.
LEE, E.-J. 2010. The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior,* 26**,** 665-672.
LEE, E. J., NASS, C. & BRAVE, S. 2000. Can computer-generated speech have gender?:an experimental test of gender stereotype.
LEE, K. M., JUNG, Y. & NASS, C. 2011. Can User Choice Alter Experimental Findings in Human-Computer Interaction? Similarity Attraction Versus Cognitive Dissonance in Social Responses to Synthetic Speech. *International Journal of Human-Computer Interaction,* 27**,** 307-322.
LEE, K. M., PENG, W., JIN, S. A. & YAN, C. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication,* 56**,** 754-772.
LEECH, G. N. 1983. *Principles of pragmatics*, Taylor & Francis.
LLAMAS, C. & WATT, D. 2010. *Language and identities*, Edinburgh University Press.
LOCHER, M. A. 2004. *Power and politeness in action: Disagreements in oral communication*, Walter de Gruyter.
LOCHER, M. A. 2006. Polite behavior within relational work: The discursive approach to politeness. *Multilingua-Journal of Cross-Cultural and Interlanguage Communication,* 25**,** 249-267.
LOCHER, M. A. & WATTS, R. J. 2005. Politeness theory and relational work. *Journal of Politeness Research. Language, Behaviour, Culture,* 1**,** 9-33.
LOCHER, M. A. & WATTS, R. J. 2008. *Relational work and impoliteness: Negotiating norms of linguistic behaviour*, na.
MAES, P. 1994. Agents that reduce work and information overload. *Communications of the ACM,* 37**,** 30-40.
MASCARENHAS, S., PRADA, R., PAIVA, A., & HOFSTEDE, G. J. (2013, August). Social importance dynamics: A model for culturally-adaptive agents. In *Intelligent virtual agents* (pp. 325-338). Springer Berlin Heidelberg.
MATTHEWS, P. H. 2007. *The concise Oxford dictionary of linguistics.* Oxford University Press.
MAYER, R. E., JOHNSON, W. L., SHAW, E. & SANDHU, S. 2006. Constructing computer-based tutors that are socially sensitive:

Politeness in educational software. *International Journal of Human-Computer Studies,* 64**,** 36-42.
MAYER, R. E., SOBKO, K. & MAUTONE, P. D. 2003. Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology,* 95**,** 419.
MCCARTHY, M. 1998 *Spoken Language and Applied Linguistics*. Cambridge University Press
MCCARTHY, M. & CARTER, R. 2006. as visible patterns of interaction. *Explorations in corpus linguistics***,** 7.
MITCHELL, W. J., HO, C.-C., PATEL, H. & MACDORMAN, K. F. 2011a. Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior,* 27**,** 402-412.
MITCHELL, W. J., SZERSZEN SR, K. A., LU, A. S., SCHERMERHORN, P. W., SCHEUTZ, M. & MACDORMAN, K. F. 2011b. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception,* 2**,** 10.
MONTOYA, R. M., HORTON, R. S. & KIRCHNER, J. 2008. Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships,* 25**,** 889-922.
MORAN, S., PANTIDI, N., BACHOUR, K., FISCHER, J. E., FLINTHAM, M., RODDEN, T., EVANS, S. & JOHNSON, S. Team reactions to voiced agent instructions in a pervasive game. Proceedings of the 2013 international conference on Intelligent user interfaces, 2013. ACM, 371-382.
MORI, M., MACDORMAN, K. F. & KAGEKI, N. 2012. The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE,* 19**,** 98-100.
NASS, C., FOGG, B. J. & MOON, Y. 1996. Can computers be teammates? *International Journal of Human-Computer Studies,* 45**,** 669-678.
NASS, C. & LEE, K. M. Does Computer-Generated Speech Manifest Personality? An Experimental Tets of Similarity-Attraction. CHI '00 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2000 The Hague, Netherlands. New York, NY, USA: ACM.
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**,** 171-181.
NASS, C. & MOON, Y. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues,* 56**,** 81-103.
NASS, C., MOON, Y., FOOG, B., J., REEVES, B. & DRYER, D. C. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies,* 43**,** 223-239.
NASS, C., STEUER, J. & TAUBER, E. R. 1994. Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Boston, Massachusetts, USA: ACM.

NICULESCU, A., I. 2011. *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality.* Ph.D., University of Twente, Netherlands.
NORMAN, D. A. 2013. *The design of everyday things: Revised and expanded edition*, Basic books.
OREL, F. D. & KARA, A. 2014. Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market. *Journal of Retailing and Consumer Services,* 21**,** 118-129.
QUAGLIO, P. 2009 *Television dialogue: The sitcom Friends vs. natural conversation* (Vol 36.). John Benjamins Publishing.
PEARSON, J., HU, J., BRANIGAN, H. P., PICKERING, M. J. & NASS, C. I. Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. Proceedings of the SIGCHI conference on Human Factors in computing systems, 2006. ACM, 1177-1180.
PIERCE, C. 1902. Vagueness. *Dictionary of Philosophy and Psychology II*.
PRINCE, E. F., FRADER, J. & BOSK, C. 1982. On hedging in physician-physician discourse. *Linguistics and the Professions,* 8**,** 83-97.
REEVES, B. & NASS, C. 1996. *How people treat computers, television, and new media like real people and places*, CSLI Publications and Cambridge university press.
ROBARTS, S. 2015. *Tesco's self-service checkouts are getting friendlier* [Online]. Available: http://www.gizmag.com/tesco-supermarket-self-service-checkout-voice-phrases/38707 [Accessed 22 February 2016].
ROSÉ, C., WANG, Y.-C., CUI, Y., ARGUELLO, J., STEGMANN, K., WEINBERGER, A. & FISCHER, F. 2008. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning,* 3**,** 237-271.
ROWLAND, T. 2007. 'Well maybe not exactly, but it's around fifty basically?': Vague language in mathematics classrooms. *Vague language explored***,** 79-96.
SCHAAFSTAL, A. M., JOHNSTON, J. H. & OSER, R. L. 2001. Training teams for emergency management. *Computers in Human Behavior,* 17**,** 615-626.
SCHEUTZ, M., CANTRELL, R. & SCHERMERHORN, P. 2011. Toward Humanlike Task-Based Dialogue Processing for Human Robot Interaction. *AI Magazine,* 32**,** 77-84.
SCHULLER, B., BATLINER, A., BURKHARDT, F., DEVILLIERS, L. MÜLLER, C. & NARAYANAN, S. 2013. Paralinguistics in speech and language - state-of-the-art and the challenge. *Computer Speech & Language.* 27(1) pp.4-39.
SMITH, C., DOBNIK, N. C. S., CHARLTON, D., PULMAN, J. B. S., DE LA CAMARA, R. S., TURUNEN, M., BRADLEY, D. B. J., HANSEN, B. G. P., MIVAL, O., WEBB, N. & CAVAZZA, M. 2011. Interaction

Strategies for an Affective Conversational Agent. *Presence: Teleoperators and Virtual Environments,* 20**,** 395-411.
SPENCER-OATEY, H. 2007. Theories of identity and the analysis of face. *Journal of pragmatics,* 39**,** 639-656.
STRAIT, M., CANNING, C. & SCHEUTZ, M. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, 2014. ACM, 479-486.
STRAIT, M., VUJOVIC, L., FLOERKE, V., SCHEUTZ, M. & URRY, H. Too Much Humanness for Human-Robot Interaction: Exposure to Highly Humanlike Robots Elicits Aversive Responding in Observers. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015. ACM, 3593-3602.
SUKTHANKAR, G., SHUMAKER, R. & LEWIS, M. 2012. Intelligent agents as teammates. *Theories of Team Cognition: Cross-Disciplinary Perspectives***,** 313-343.
SUZUKI, N. & KATAGIRI, Y. 2007. Prosodic alignment in human-computer interaction. *Connection Science,* 19**,** 131.
TAMAGAWA, R., WATSON, C. I., KUO, I. H., MACDONALD, B. A. & BROADBENT, E. 2011. The Effects of Synthesized Voice Accents on User Perceptions of Robots. *International Journal of Social Robotics,* 3**,** 253-262.
TANNEN, D. 1993. *Framing in discourse*, Oxford University Press on Demand.
TANNEN, D. 2005. *Conversational style: Analyzing talk among friends*, Oxford University Press.
TERRASCHKE, A. & HOLMES, J. 2007. 'Und tralala': vagueness and general extenders in German and New Zealand English.
TORREY, C. 2009. How robots can help: Communication strategies that improve social outcomes.
TORREY, C., FUSSELL, S. & KIESLER, S. 2013. How a robot should give advice. *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction.* Tokyo, Japan: IEEE Press.
TRAPPES-LOMAX, H. 2007. Vague language as a means of self-protective avoidance: Tension management in conference talks. *Vague Language Explored***,** 117-137.
TURNER, P. 2008. Being-with: A study of familiarity. *Interacting with Computers,* 20**,** 447-454.
Use Siri on your iPhone, iPad, or iPod touch. (2016, February 22) Retrieved from https://support.apple.com/en-gb/HT204389.
VON SCHEVE, C. 2013. Interaction Rituals with Artificial Companions. From Media Equation to Emotional Relationships. *Science, Technology & Innovation Studies,* 10**,** 65-83.
WANG, A. 2005. 'When Precision Meets Vagueness: A Corpus-Assisted Approach to Vagueness in Taiwanese and British Courtrooms', presented *7th Biennial Conference of International Association of Forensic Linguists,* Cardiff.

WANG, N., JOHNSON, W. L. & GRATCH, J. Facial expressions and politeness effect in foreign language training system. Intelligent Tutoring Systems, 2010. Springer, 165-173.
WANG, N., JOHNSON, W. L., MAYER, R. E., RIZZO, P., SHAW, E. & COLLINS, H. 2008. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies,* 66**,** 98-112.
WATT, D. 2010. The identification of the individual through speech. *In:* LLAMAS, C. & WATT, D. (eds.) *Language and Identities.* Edinburgh: Edinburgh University Press.
WATTS, R. J. 2003. *Politeness*, Cambridge University Press.
WOOLDRIDGE, M. & JENNINGS, N. R. 1995. Intelligent agents: Theory and practice. *The Knowledge Engineering Review,* 10**,** 115-152.
ZHANG, Q. 1998. Fuzziness-vagueness-generality-ambiguity. *Journal of pragmatics,* 29**,** 13-31.