Believing in BERT: Making good on bad robot behavior

Adriana Hamacher

Project report submitted in part fulfilment of the requirements for the degree of Master of Science (Human-Computer Interaction with Ergonomics) in the Faculty of Brain Sciences, University College London, 2015.

ACKNOWLEDGMENTS

Many thanks to my two supervisors: Nadia Berthouze at UCL, for her excellent advice and input, and Kerstin Eder at Bristol Robotics Lab (BRL), for helping me to develop my ideas and introducing me to a world of possibilities at BRL.

A big thank you also to BRL directors Chris Melhuish and Tony Pipe for the use of their facilities and the technical help they provided. Without Tony Pipe's intervention when BERT2 broke his hand, it would have been impossible to proceed.

I'm also very grateful for the support I received from BRL's Alex Lenz, who designed and built BERT2, Evgeni Magid, who was responsible for programming the robot for this experiment, and Alan Braun who provided critical support at a difficult juncture.

Finally, a high-five to my son, Max. This study is for him, because robots are the future.

ABSTRACT

The new arena of collaborative robotics is sorely in need of strategies to deal with the challenges that arise where robots and humans work in close proximity. Particularly problematic is successfully negotiating uncertainty or misunderstanding that may arise from erroneous or unexpected behaviour on the part of robot. This issue is compounded by people's differing degrees of exposure and, sometimes, inaccurate prior conceptions in respect of robots.

Using self-report, interviews and detailed audiovisual analysis, an experiment evaluated participants working with a humanoid robot assistant in three different conditions during an omelette-making task. The robots varied in their efficiency, accuracy and communicative ability. The purpose was to gauge to what degree communication and the robot's display of human-like attributes can positively affect the participants' experience, particularly when things go wrong.

Fifteen of the 21 participants preferred the communicative, personable robot over a more efficient and less error prone one. Satisfaction was significantly increased in the communicative condition and participants were particularly responsive to this robot's apology and apparent regret. For the majority, personable, transparent behaviour appeared to negate the fact that the interaction took 50 per cent longer than in the non-communicative conditions.

Through the identification of key incidents, a detailed picture emerged, providing evidence that incorporating human-like attributes judiciously into robot design can significantly mitigate dissatisfaction arising from unexpected or erroneous behaviour.

CONTENTS

TABLES

FIGURES AND VIGNETTES

CHAPTER 1. INTRODUCTION

Introduced in 1989, the Mazda MX-5 roadster is remarkable not only because it's still the world's best-selling small sports car, but also because, in its design, for the first time developers focused on drivers' feelings and not merely technical specifications. People's emotions, their behaviours, attitudes and characteristics while driving were recorded. These were fed into design specifications and translated into the physical experience. The result was a car that proved so satisfying to drive it blew the competition out of the water.

The MX-5, and cars in general, may appear to have little similarity with humanoid robots, but there is much common ground. Just like the fearful attitudes inspired by the arrival of automobiles at the turn of the last century, robots today are generating scare headlines, most recently: *Robot Kills Man at Volkswagen Plant* (Docketerman, 2015).

As cobots or co-worker robots become more common, it's a pre-requisite that physical safety needs to be carefully verified and public fears and misconceptions addressed. But safety concerns are just one thing cars and robots have in common; the popularity of the MX5 underlines that interaction style is key to delivering satisfactory user experiences and machines that people enjoy using and working with.

But although robotics is rapidly evolving, and robots can already master skills through trial and error (Levine, S., Wagener, N., & Abbeel, P., 2015), even in real time, the human-populated world contains an infinite number of unknowns.

So, in what is still very much a nascent field, how does a developer handle the mistakes, misunderstandings and failures likely to arise between human and robot, until, like a child, it learns? How should they be presented to the user? And how is error prone, unexpected behaviour likely to be perceived and judged? Most importantly, how can we iron out teething issues, when human and robot are getting to know each other, and smooth the interaction?

These are the questions that form the background to this study where we observed 21 participants performing an omelette-making task with BERT2 (Bristol Elumotion Robotic Torso 2) - a humanoid robotic assistant.

The experiment had three conditions, with varying errors and degree of communication. The central finding of the study was that, contrary to much existing literature, the robot's behavioural style is significantly more important than its performance, when it comes to satisfaction levels, and an expressive robot is preferable over more efficient one, despite a trade off in time taken to do the task.

The study suggests that an appealing, communicative robot displaying human-like emotions of regret and enthusiasm influences the user experience in such a way that dissatisfaction with its erroneous behaviour is tempered if not forgiven.

However, great care must be taken with the design of such systems; the attribution of human-like characteristics in a robot may make users reluctant to hurt its feelings and they may even lie in order to avoid this.

Background: Robots and trust

Trust is a highly important factor within human-robot interaction (HRI) and the subject of a major pan-European research project entitled Robosafe. The objective is to create Trustworthy Robot Assistants (http://www.robosafe.org). This project seeks to contribute to that body of work.

A number of recent studies (IET survey, 2015; Special Eurobarometer 382, 2012; Syrdal, Nomura & Dautenhahn, 2013; Robotics and Autonomous Systems: What the public thinks, 2013) have surveyed people's attitudes to robots and automation and give us cause to believe that many people distrust robots.

As robots are increasingly developed for use in social settings and, particularly, as assistants to non-expert users such as elderly person in their own home, acceptance, persuasiveness and likability are key and these are factors strongly linked to trust (Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K., 2015).

Robots must be able to participate in sophisticated interactions with humans in a safe and trustworthy manner before established safety barriers can be dispensed with. Although work at lower, mechanical levels - restricting movements near humans - is already in progress, it's at the higher levels of intention-recognition that much needs to be done and, here, user evaluation is a key factor in necessary validation. This study focuses particularly on developing mechanisms for a robot to communicate its intentions and to recover trust after a perceived failure.

CHAPTER 2. LITERATURE REVIEW

We turn first to a new paradigm - human-robot teamwork - and then explore the other areas of relevance here: the role of expectations, the development of trust, the effect of unexpected behaviour and the moderating effects of transparency and communication.

Although the majority of this literature review refers to HRI or Human-computer interaction (HCI), humans are highly prone to interpret computer behaviour according to the same scripts or schema that are commonly used for human-human interactions (Reeves and Nass, 1996). For intention recognition, we look to studies of natural human-human communication, as work on this area has mainly occurred there.

2.1 Human-robot teamwork

The "design of automated systems is really the design of a new human-machine cooperative system (Woods, 1996)." In other words, designing the interaction between a human and a robot assistant is akin to the design of a team and requires careful coordination. It should take into account not only the context of use, but also the characteristics of the human.

Other factors which need consideration are: whether the robot is compatible with the human's needs, whether it's understandable, believable, and provides the interactional support the human expects (Fong, Nourbakhsh & Dautenhahn, 2003).

These factors need to be given early consideration in design and integrated at the very beginning as, with the growing popularity of cobots and assistive robotics, "effective interdependence management will become increasingly important in the coming years" (Johnson, Bradshaw, Feltovich, Jonker, van Riemsdijk & Sierhuis, 2012).

The BERT2 platform has already been the subject of studies on collaborative robotics; work on joint action understanding (Grigore, Eder, Pipe, Melhuish & Leonards, 2013, November) identified the need to comprehend the properties that humans seek when establishing trust, so that they can be integrated into the robot's decision making, resulting in a safer more trustworthy system.

Identifying people's expectations and their initial perceptions of the system is fundamental to this.

2.2 Role of expectations and attitudes

Cultural background, gender, age, and many other factors can play a crucial role in people's perception of robots; attitudes are also largely shaped by science fiction (Scopelliti, Giuliani & Fornara, 2005; Ray, Mondada & Siegwart, 2008, September; Li, Rau & Li, 2010).

There is overwhelming evidence that people take cues about a robot's abilities based on whether it looks and acts like a human (Jones et al, 2011; Frith & Frith, 2012) and users sometimes attribute humanlike mental faculties to robots (Hegel, Krach, Kircher, Wrede & Sager, 2008, March).

Expectations and attitudes have a major impact on trust. In surveys, such as the SPECIAL EUROBAROMETER 382 (2012), people have expressed widespread concerns, distrust and fears regarding the use of robots both as domestic aides and in the workplace.

To counter this, the need for new approaches to safety assurance for learning machines has been identified (Eder, K., Harper, C., & Leonards, U., 2014). Crucially, a robotic co-worker must meet the innate expectations of the humans it works with. If the robot has eyes, it should, for example, direct its gaze at the task in hand or, in case of danger, towards the source. "This requires deep understanding of the signals sent by humans and the way humans interpret these signals."

2.3 Effect of unexpected behaviour

Anthropologists have long been aware that, when the normal stream of activity is disrupted, special attention is required. Responses to unexpected behaviour can also often reveal clues to expectations, such as the desire that someone acknowledge an obvious mistake. There are many underlying rules governing social behaviour, as well as highlighting constraints, and ignoring them is likely to cause trouble.

We lack large-scale, long-term data on the effects of such occurrences in HRI, despite the fact that unexpected situations occur surprisingly often with users who are unfamiliar with a system and don't understand its capabilities (Spexard, Hanheide and Wrede, 2008). Unexpected events can be triggered by anything from sticky ground and blinding light sources to obstacles that cannot be avoided.

Unexpected behaviour can also be termed disruptive, if the reason for it not immediately intelligible to the user. It may cause the user to question the rationality they have ascribed to the robot or it may be interpreted as a deliberate action, causing them to increase or decrease the level of cognitive skill they have assigned to the system, depending on whether or not the unexpected event is perceived to be a failure (Lemaignan, Fink & Dillenbourget, 2014).

Some HRI studies, such Salem, Eyssel, Rohlfing, Kopp & Joublin, (2013), have found that errors occasionally performed by a humanoid robot can actually increase its perceived human-likeness and likability.

But the bottom line appears to be that when the system's behaviour appears inconsistent, it can be difficult or impossible for the user to form expectations

(Sarter, Woods & Billings, 1997). “Therefore, under those circumstances, the system needs to provide external attentional guidance to the user to help detect and locate problems.”

But how this guidance should be presented is largely missing from the literature. “The robot should try to present objects in the default orientation when possible (Cakmak, Srinivasa, Lee, Forlizzi and Kiesler, 2011). However no suggestion is made about what the system should do if that’s not possible. To remedy this, in our study we investigate the effect of a change in the handover procedure. In one condition the system attempts to communicate the difference and we explore the efficacy of this measure.

We also examine the role of an expression of regret following an erroneous action. There is evidence that humans respond positively to robots that apologise or offer compensation if they have made a mistake. Lee, Kiesler, Forlizzi, Srinivasa & Rybski (2010) found that a prior warning and recovery strategies reduced the negative impact of a breakdown. They also found that people’s attitudes towards services influenced which recovery strategy worked best. Those who wanted to continue the relationship responded best to an apology; those with a more utilitarian orientation towards the service responded best to compensation.

The hypothetical nature of their study is, they stress, an important limitation: “We do not know for sure if people’s responses to robotic services in real environments will be the same.” The strategies were also only used for one type of task and one error and there is an obvious need for further investigation on recovery techniques with different tasks, situations, robots, and errors. We seek to partly remedy this here.

2.4 Impact of unexpected behaviour on trust

Muir and Moray (1976) argue that trust is based on the extent to which the machine is perceived to properly perform its function. This implies that machine errors strongly affect trust.

Robot characteristics, particularly performance, are widely believed to form the biggest influence on perceived trust (Freedy, DeVisser, Weltman & Coeyman, 2007, May; Hancock, Billings, Schaefer, Chen, De Visser & Parasuraman, 2011). The latter study examined the effects of human, robot, and environmental characteristics on trust with “an especial evaluation of the robot dimensions of performance and attribute-based factors." They found that robot performance, specifically, had the largest contribution to the development of trust in HRI.

To recap on previous sections: individual characteristics need to be factored into the design of a human-robot team and human expectations are key. Awareness of these can serve to mitigate the dissatisfaction that arises from unexpected behaviour, which is linked to performance and, in turn, influences trust. However, trust requires a little more attention.

2.5 Defining and developing trust

Trust is a complicated and multidimensional construct (Lee and See, 2004). It develops in a combination of three interplaying processes: analytic, analog, and affective. The former is determined via a rational cost-benefit analysis. For example, if a robot makes a potentially costly error the operator may be less trusting and decide to take over control. Analog processes follow on from this, generalising the system to a broader set of assumptions about a certain group. Finally, trust also forms purely affectively and “emotions even supersede rational thinking.”

Giving further weight to the role of emotion in trust, Miller (2005) argues that ”when experiencing a new person or a novel system for the first time, with no background knowledge about the agent’s motivations, behaviours, or group memberships, the only information a person may have about whether or not to trust it will be affective information.”

One approach to achieving analogic and affective trust in complex systems is good “etiquette” (Lee et al, 2005). It is described by Miller (2005) as the “largely unwritten codes that define roles and acceptable or unacceptable behaviours or interaction moves of each participant in a common ‘social’ setting.’”

Parasuraman and Miller (2004) manipulated so-called “etiquette variables.” They compared good automation etiquette, defined here as “a communication style that was non-interruptive and patient,” with poor automation etiquette that was “interruptive” and “impatient.” These two conditions were crossed with low and high automation reliability. The study found that good automation etiquette significantly ($p < 0.05$) enhanced diagnostic performance, regardless of reliability. Its effects were powerful enough to overcome low reliability with a corresponding effect on trust. This has significant implications as it suggests that “developing robust, sensitive, and accurate algorithms for automation… may not be necessary so long as the automation ‘puts on a nice face for the user.’” Our study seeks to explore this idea and asks whether good automation etiquette may be a way to compensate for the initial mistakes that may be made by collaborative robots.

Van den Brule, Dotsch, Bijlstra, Wigboldus, & Haselager, (2014) claims to be “the first study to combine both performance and behavioural manipulations of trustworthiness of a humanoid robot in an HRI scenario.” Their study reconfirmed that performance strongly influences trust but appeared to focus mainly on motion fluency as opposed to etiquette or the effects of communication, measures that we seek to explore here. It was also based on the use of a virtual rather than a real-life robot.

Limited evidence exists on the importance of human-related factors in trust perception (Hancock, Hancock & Warm, 2009), but this is largely down to the small number of studies undertaken in this area. Of specific interest are: “level of operational experience, attentional capability, the amount of training received, self-confidence, the propensity to trust, existing attitudes toward robots, personality traits, operator workload, situation awareness, and other individual difference factors" (Hancock et al, 2009). Some of these are included in this study's pre-experiment questionnaire.

Lee & See (2004) define trust in automation as "the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability." This, they say, is determined by observing the characteristics of the system, such as its performance (how well it accomplishes the individual's goals) but is also dependent on the transparency inherent in the process.

Our study seeks to examine whether transparency, good etiquette and communication are, in fact, more important than performance.

2.6 Transparency: understanding of the robot's internal state

The proponents of transparency are many, including critical HCI theorists Paul Dourish (2004) and Julia Weber (2013), who hold the view that systems should say what they're doing and that transparency should be available on demand. Donald Norman (1990) calls lack of transparency "silent automation."

For human robot teamwork, transparency is even more important, with a key finding that transparency and control may be more important than increased autonomy (Johnson et al, 2012). Lack of transparency also results in a breakdown of common ground (Klein, Feltovich, Bradshaw & Woods, 2005), a reduction in each player's individual situation awareness and an increase in errors (Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., van Riemsdijk, B., & Sierhuis, M. (2011)."

According to Eder et al (2014), "Designing dependable systems is not only to create a flawless design but to do so in a manner that permits such flawlessness to be demonstrated. This requires careful choices of a system's architecture and mechanisms," including the ability to clearly communicate intention in a timely manner. "A key property of dependability assurance is that it is a subjective condition of a system's users as well as an objective property of the system itself. "

But perhaps even more important than transparency is a lack of clarity of meaning. In one study (Kim & Hinds, 2006), a robot suddenly showed unexpected behaviour and explained the reason by announcing "I have recalibrated my sensors." The authors admit that the choice of terminology may have been misguided as, instead of providing transparency of action, it had the effect of confusing participants who may not have understood the meaning.

This study encountered similar issues and shows that it's not easy for a robot to effectively communicate its intentions.

2.7 Methods of communicating

Much social robotics research is in agreement that the main requirements of a complex social interaction include communication, the recognition and expression of emotions, and some rudimentary form of personality (Fong et al, 2003). These features are widely thought to increase the believability of artificial agents and enhance engagement. However their necessity has not been studied in relation to human-robot teamwork situations or as a method of mitigating misunderstandings or mistakes.

Analyses of human-human cooperative interaction demonstrate that errors in understanding are corrected not only verbally, but also with signals such as pointing and positioning (Clark, 2005).

Facial expression has been studied more than any other type of nonverbal communication (Bethel & Murphy, 2006). An expressive face and “engagement gestures” have been shown to make a robot more compelling to interact with (Bruce, Nourbakhsh & Simmons, 2002; Sidner and Lee, 2003). Research also suggests that users are likely to prefer exaggerated or caricatured expressions over realistic ones (Wang, Lignos, Vatsal, & Scassellati, 2006) and dislike robots that look too much like themselves (Mori, MacDorman & Kageki, 2012).

Finding the correct balance in a robot’s non-verbal expressiveness therefore appears to be a significant factor, particularly as human-human communication also relies heavily on facial expressions (Jones & Schmidlin, 2011), which can convey up to 55% of a message (Mehrabian, 1968). However, body expressions can sometimes be an even more powerful affective communication channel (Kleinsmith, A., & Bianchi-Berthouze, N., 2013).

The relationship between task efficiency, communication (non-verbal), error and their effect on human-robot teamwork was explored by Breazeal, Kidd, Thomaz, Hoffman & Berlin (2005). This study demonstrates that “implicit non-verbal communication positively impacts human-robot task performance with respect to understandability of the robot, efficiency of task performance, and robustness to errors that arise from miscommunication.” It also offers evidence that not only can social cues enhance the likeability of robots, but that that they can also “serve a pragmatic role in improving the effectiveness of human-robot teamwork where the robot serves as a cooperative partner.”

The study didn’t, however, touch on the effect of verbal communication, as we seek to do, neither did it involve a delicate handover task, where tensions are likely to be raised and the need for effective measures is underscored.

Many studies urge caution in respect of verbal communication in HRI. Dialogue can lead to biased perceptions with the possibility of users forming incorrect models (Fong, Nourbakhsh & Dautenhahn, 2003) and dissatisfaction can result if speaking robots can’t meet raised expectations (Kanda, Glas, Shiomi, Ishiguro and Hagita, 2008).

So, despite the overwhelming preference of users for speech communication (Ray yet al, 2008, Iwamura, Shiomi, Kanda, Ishiguro,& Hagita, 2011), many developers choose to support the interaction with a touch interface. Speech recognition systems are problematic, but other methods of communication also leave a lot to be desired. Developer of eldercare robots, Tandy Trower, explains (personal communication, May 3rd, 2014):

"We know, regardless of what we do, speech could fail, as any demonstration within a noisy environment will show. Or our users might prefer to tap versus speak or they might have suffered a stroke and be incapable of articulating well enough or at all. We tend not to use gestures this would be even harder. There is no universal gesture language for anything but the most basic things like 'stop' or 'look over there' or 'come here.' And we don't want our users to learn any. On screen cues (aka Xbox gesture games) don't work either. Further, for our audience, holding your arm out might be beyond what they have the strength for."

However, speech synthesis is much further down the line than voice recognition and quality has improved greatly recently. Users, irrespective of age or disability, appear to have little difficulty in recognising synthetic speech (Humes, Nelson, and & Pisoni, 1991) and this influenced our decision to use this as a primary form of communication in one of our conditions.

2.8 Summary and research questions

In such a young field as humanoid robotics, unexpected behaviour is likely to be a common problem. Strategies are necessary to mitigate its impact and research to develop solutions is lacking. If an unforeseen situation occurs because of factors beyond the robot's comprehension, how could it recover a user's trust? And how much would an attempt to warn the user or an apology mitigate dissatisfaction caused by unpredictable behaviour? These are the central questions this project seeks to address.

Forming the background to the investigation, prior expectations, attitudes and individual characteristics all have a major impact on people's attitude to robots, as studies and surveys have revealed. This is particularly true when it comes to humanoid robots. Gathering information on the participants - such as their level of experience and attitudes to automation - should therefore be a pre-requisite in any experiment of this nature.

A number of studies have demonstrated that task performance is an important source of the trustworthiness judgments people make about robots. However none appear to have contrasted a humanoid robot's efficiency with increased transparency and a personable style in a real-life scenario as a possible strategy to recover from errors.

CHAPTER 3. METHOD

3.1 Introduction

Motivated by the preceding literature review, our study sought to investigate whether increased transparency, the opportunity to communicate with the robot (albeit in a rudimentary fashion) and its display of humanlike emotion could outweigh task performance and provide a more satisfactory interaction, with implications for trust.

This section outlines our experiment, named “Believing in BERT” (BiB), undertaken to explore whether these were viable strategies that could be used to mitigate any negative effects of erroneous behaviour.

Participants were invited to choose a robot kitchen assistant by putting three potential job candidates through their paces in an omelette-making task. The same robot, BERT2, was used, but with different headwear to differentiate it as candidate A, B or C, acting in one of our three conditions.

The candidate labelled A performed the most efficiently, never dropping an egg, but was unable to communicate its intentions or provide feedback. B was also mute and, in addition, dropped one of the eggs. He attempted to rectify this by trying again, using a different method to handover the egg. BERT C was the only candidate able to talk, asking participants whether they were ready to receive the egg on each occasion. He also dropped an egg, but appeared conscious of his mistake and apologised. He then attempted to rectify the error and forewarned participants that he would try another method of handover. At the end of the task, he asked the participants whether he did well and whether he got the job. All three job candidates neglected to supply the cooking oil, necessary to complete the task.

Prior to the experiment, data on expectations was gathered via a questionnaire. Video of the experiment was analysed with regard to engagement, negative and positive manifestations of satisfaction, comfort and emotion. Following the experiment, participants completed a NASA TLX workload questionnaire (Hart, S. G., & Staveland, L. E., 1988) and a short semi-structured interview. Responses to the questionnaire were quantified and form the main body of the results, complimented by the video analysis, from which “vignettes” were extracted to illustrate our findings.

We expected that BERT A would be more popular than BERT B, as A was the more efficient in completing the task, with minimum error. We also expected BERT C would be more popular than BERT B, in line with studies, such as Johnson et al, (2012) which indicate that transparency and control are highly important, perhaps even more than increased autonomy.

We also expected that, given the choice between enhanced reliability and a personable interface, with which it’s possible to communicate, most people would choose the later (i.e. BERT C would be chosen over BERT A). So BERT C was expected to be the favoured candidate, despite sometimes taking twice as long to complete the task as BERT A.

Hypothesis 1: An unforeseen occurrence, will cause a robot to appear less trustworthy than a more efficient version, even if attempts are made to mitigate the mistake (i.e. BERT A will be more popular than BERT B).

Hypothesis 2: Communication - both audio and visual - can significantly mitigate dissatisfaction in the event of an unforeseen occurrence (i.e. BERT C will be significantly more popular than BERT B).

Hypothesis 3: Given the choice between enhanced efficiency and reliability and a personable, communicative interface, most people will chose the later (i.e. BERT C will be chosen over BERT A).

3.2 Participants

This study was conducted over two weeks, in June 2014, at the Bristol Robotics Laboratory (BRL), University of the West of England (UWE), Frenchay Campus, Bristol.

A total of 23 participants, 12 men and 11 women, were recruited from the local area and from UWE. Data from two subjects was ultimately discarded. In both cases the robot was malfunctioning to the point where the subjects could not complete the tasks.

Methods of recruitment consisted of eye-catching posters (Appendix I), advertising on a local community site, Twitter and Facebook, as well as personal and university contacts.

Care was taken to achieve a wide range of ages, from 22-70+, and a mixture of naive users, those with some experience of robots and robotics students.

3.3 Design

Participants first completed a pre-experiment questionnaire (Appendix II) to gauge their opinions, attitudes and expectations in relation to robots. This survey was largely based on the Special Eurobarometer 382 (2012). There were a number of supplementary questions, such as “How frustrated do you get waiting for the DVD player to load?”

A within subjects design was adopted to create a big enough sample and also to compensate for variability in participant characteristics (e.g. intro/extroversion). The order of the BERTs - A, B and C - was varied in a chi-square. The independent variables were the efficiency of the robot and its communicative ability. There were three conditions:

BERT A: Non-communicative, most efficient even though it, like others, “forgets” the oil.

BERT B: Non-communicative, makes a “mistake” but attempts to rectify it.

BERT C: Communicative, also makes a “mistake” and attempts to rectify it.

C also displayed some behaviour that might be construed as challenging: it put the participants on the spot at the end by asking a difficult question. The functionality and focus of the conditions are explored in more detail in Table 1.

The dependent variables were workload, satisfaction and trust. These were measured via a NASA Task Load Index (NASA TLX) questionnaire (Hart et al 1988) with a scale of 1-5 (Appendix III). Additional questions were added to the standard questionnaire to gauge levels of satisfaction and trust. Participants were then invited to choose one of the robots for the job and asked whether their choice would be different if it was for a work, as opposed to home-based, task.

Levels of comfort, engagement and interaction were also studied using audiovisual recordings of the interactions, and participants were interviewed. This data was thematically analysed. These results serve to further inform and illustrate the quantative data provided by the questionnaires.

Qualities of C: the “expressive” interface

In conditions A and B the robot was mute but in “C”, the “expressive interface,” the system was programmed to communicate with participants but only to recognise yes and no answers, in order to minimise confounds raised by the current level of speech recognition technology. Participants were informed of this limitation. The speech plan for BERT C is shown in Figure 1.

The qualities of C, the “expressive” interface, were derived from a literature review of “personable” robot characteristics and include:

- Talks user through every stop of the way (transparency/opacity).
- Pre-warns the user in case of unforeseen occurrence “Let’s try something different.”
- "Enhanced" facial expression (more blinking, nuanced expressions).
- Looks regretful when item is dropped and says: “I am sorry.”

Table 1. Functionality, focus and duration of experimental conditions.

Condition	BERT A: The "efficient" interface	BERT B: The "unexpected" interface	BERT C: The "expressive" interface
Functi onality	Passes participants the three eggs necessary to make the omelette. Passes salt. Points towards bowl to indicate that eggs are ready to be whisked. "Forgets" the oil.	Passes participants the two eggs but then drops the third. Attempts to rectify this by passing the fourth egg in a different form of handover: palm up, as opposed to dropping the egg into the hand. Points towards bowl to indicate that eggs are ready to be whisked. Passes salt. "Forgets" the oil.	Identical to B, with the addition of speech*: Introduces himself. Establishes, verbally, whether participants are ready to receive each egg, the salt and the oil before releasing them. Suggests that eggs are ready to whisked. Appears to show – through vocal apology and expression - that it has made a mistake. Wants to know whether it succeeded in getting the job (challenging, surprising behaviour).
Focus	Reliability, safety, getting the job done as quickly as possible, therefore no speech.	Repairing a failed interaction, getting the job done as quickly as possible, therefore no speech.	Transparency, personability.
Duration of ineraction	4.18" fixed time	4.25" fixed time	6.15" average (time dependent on whether speech was recognised and repetition needed

** The speech plan is shown in Figure 1.*

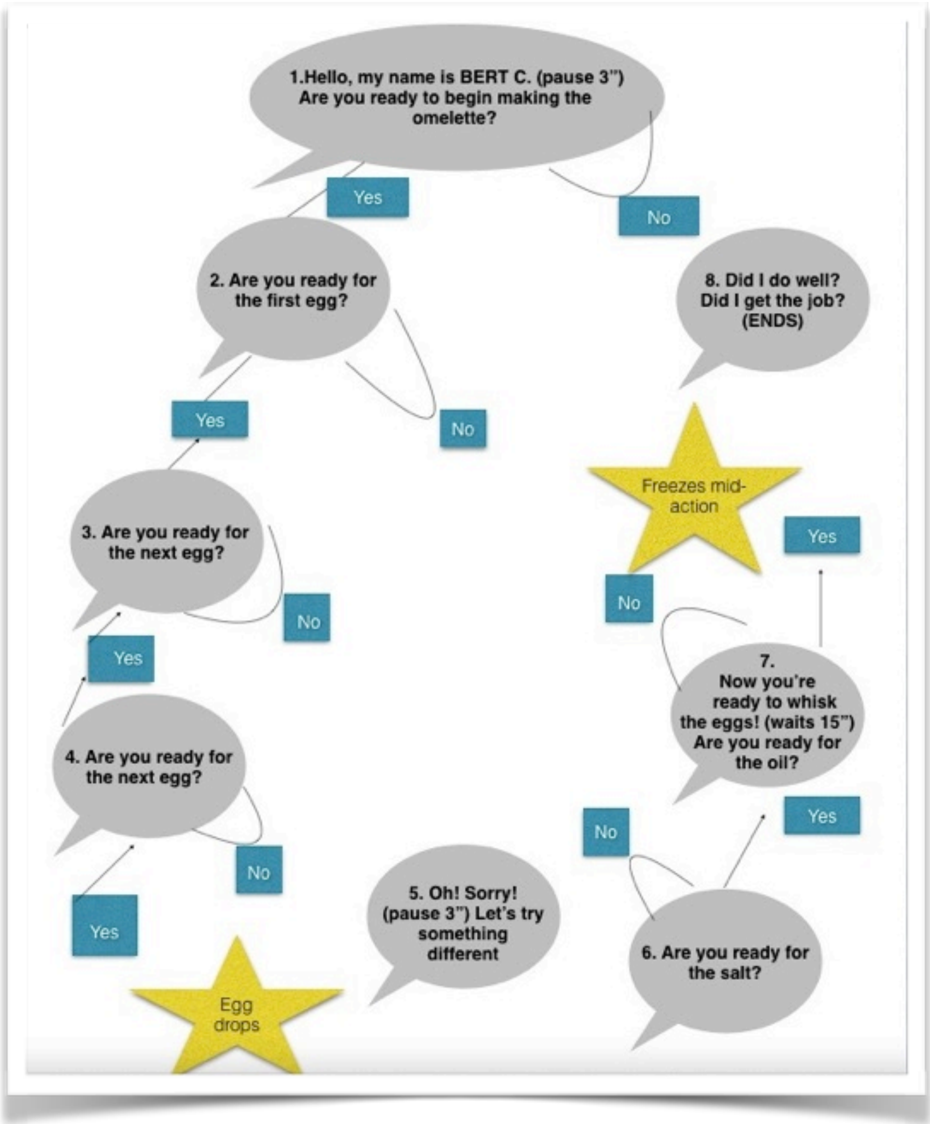


Figure 1. Speech plan for the interaction with BERT C

3.4 Apparatus and Materials

The experimental platform

The experimental platform was BERT2, an upper-body humanoid robot, with seven degrees-of-freedom (DOF) for each arm and hand (Lenz, Skachek, Hamann, Steinwender, Pipe & Melhuish, 2010, December), pictured in Figure 2. It was constructed to allow the investigation of different aspects of HRI, including verbal and non-verbal communication, gaze and pointing gestures in a real word 3D setting and has featured in a number of research papers including Grigore et al (2013) and (Bazo, Vaidyanathan, Lenz & Melhuish, 2010).

Figure 2. THE BERT2 experimental platform

Normally, BERT2 uses a Vicon motion capture (MoCap) system which can detect objects and follow the motion of human body parts in 3D space. However, for our experiment, the Vicon had to be disconnected, as it was causing instability.

The computing infrastructure is supported by YARP (Yet Another Robotic Platform) and uses two databases. The voice recognition was performed using the CSLU Toolkit (Sutton, Cole, De Villiers, Schalkwyk, Vermeulen, Macon, ... & Cohen, 1998) Rapid Application Development (RAD) with TCL scripting language which enables a connection between the actions the robot takes and the spoken dialogue. RAD uses the Festival speech synthesis system and recognition is based on Sphinx-II (Lenz et al, 2010). The system was received positive reviews (Cole, 1999), although it's likely that, since it was developed, the technology has improved.

The face of the robot (see Figure 3) is a hybrid i.e. it combines an expressive digital face with a static human visage-like structure. (Bazo et al, 2010). The interface is capable of multiple variations, the three (standard) variations we chose to use are pictured in Figure 3.

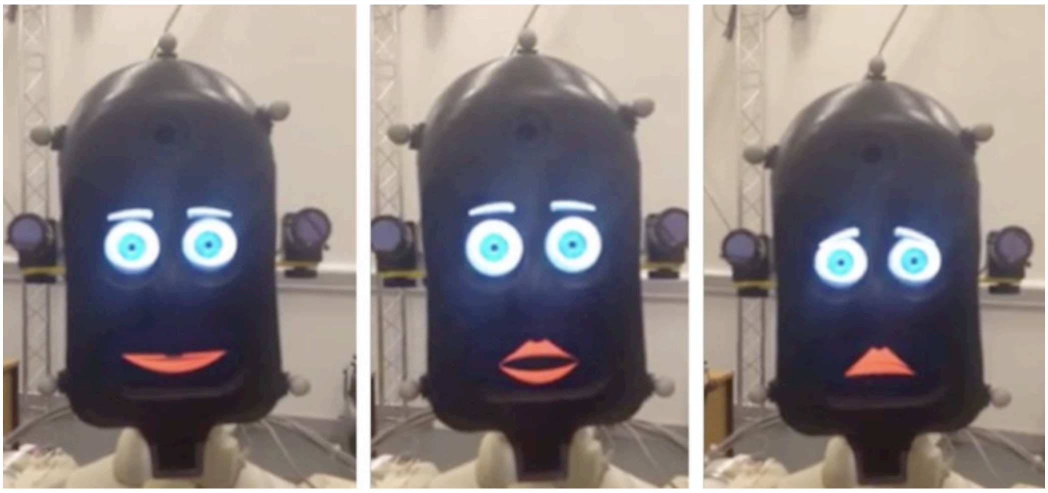


Figure 3. The three standard expressions chosen for BERT2 (from left to right): normal, surprised and sad.

A major motivation in BERT2's design was its suitability to interact with humans safely and naturally (Lallée, Pattacini, Lemaignan, Lenz, Melhuish, Natale ... & Dominey, 2012).

Additional materials

The robot was programmed by a BRL research associate, via a computer system, to behave in the various conditions. Speech commands and facial expressions were input.

Props included polystyrene eggs, tubes to resemble a salt pot and bottle of oil, a bowl and functional whisk.

A laptop was provided for the participants to complete the pre and post-experiment questionnaires.

The participants were given an information sheet (Appendix IV) and paper and pen so that they could make notes, if they wished, during the interaction.

Headphones were provided with a mike attached, for participants to communicate with BERT C.

A Nikon camera was used to film the interaction.

3.5 Procedure

Ethical clearance was obtained from the UWE's FET Faculty Research Ethics Committee (Appendix V) prior to the experiment, which took place in a curtained-off room within BRL.

Data was gathered using questionnaires, both before and after the experiment, observational notes, interviews and audiovisual recording.

Participants were individually admitted to a large curtained off room within BRL. BERT2 was situated in the centre, adjacent to two tables. On one table, were placed four eggs, and, on the other, a bowl and whisk. A video camera was sited in front of the scene and the researcher was seated just in front of the camera, on the left, in order to pass the eggs and other materials to the robot. Prior testing had shown that this was a more failsafe method than the robot attempting to pick up the eggs itself. A prominent safety button (Figure 4) was within participants reach, in the event of any serious malfunction.

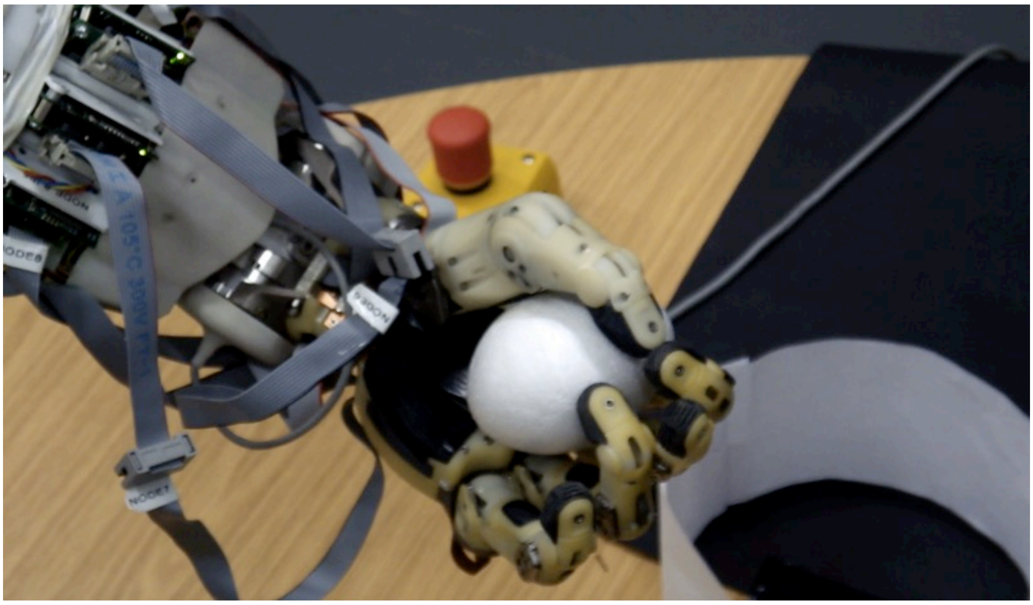


Figure 4. A prominent safety button was placed within the reach of participants

Participants were invited to sign a consent form (Appendix VI) and give permission for the publication of photographs taken of them while interacting with BERT2. They were then asked to complete the pre-experiment questionnaire.

Prior to the interaction, they were asked to read the information sheet and told they would be evaluating three robots for the job of kitchen assistant. They were instructed to stand next to BERT2 in the mock cooking scenario. A cap was placed on the robot's head to indicate which of the cooking assistants was being tested - A, B or C.

For reasons of safety, the interaction was slower than a human-human handover task would be.

Audiovisual recording was undertaken during each experiment and relevant interactions and times logged.

A post-experiment questionnaire based on the NASA TLX (Hart et al, 1988) was completed by participants. A 5-point Likert scale was used but there was also opportunity for participants to provide independent insights.

Interviews were also conducted post-experiment, with questions focused on how effective and/or engaging the kitchen assistants were, whether participants felt comfortable during the interaction, their experiences and expectations.

The experiments lasted an average of 50 minutes and were followed by an opportunity for the participants to ask questions and take a guided tour of BRL.

3.6 Analytical methods

SPSS for Mac was used for all statistical analyses of questionnaire data. Means and standard deviations were derived for the pre-experimental data. The post-experiment NASA-TLX questionnaire scale data was not normally distributed, so non-parametric Friedman tests were used to investigate the effect of independent variables on the hypotheses. We used Wilcoxon matched pairs tests to compare the effects of the conditions, where the Friedman tests demonstrated significance, and applied a Bonferroni correction to the significant findings in order to counter the likelihood of chance results.

Initial analysis of audiovisual content was performed following the approach developed by Jordan and Henderson (1995) and Heath, Hindmarsh, Luff (2010). The analysis took in elements such as length of pauses, loud or soft speech, quickening and slowing of pace, gaze, orientation, gesture and postural movement.

Participants' willingness to cooperate with the robots was a further consideration. This is based on the premise developed by Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998) that, although it's not strictly a behaviour such as cooperation or a choice such as taking a risk, trust is an underlying psychological condition that can cause or result in these actions.

We additionally looked at the amount of attention the robot required, as trust has been shown to have a moderate-to-large effect on monitoring behaviour (Bagheri & Jamieson, 2004)

As well as being based largely on established questionnaires, the pre and post-experiment surveys also contained open-ended questions. The respondents' answers can be seen in the Appendix (VII) and this data was also categorised as was material derived from the semi-structured interviews.

The ELAN platform was used for behavioural coding of verbal and non-verbal actions, and categories were developed inductively, with reference to Lakatos, Gácsi, Konok, Brúder, Bereczky, Korondi & Miklósi (2014).

Frequencies and instance of behaviour were counted and scales formed. 10 per cent of the data was categorised by a second observer to determine inter-observer reliability. Cohen's Kappa co-efficients were preformed on the categorisations. However, inter-rater agreement between the two observers was low at 0.250.

Given poor inter-rater results, our findings are grounded in the self-reported data. We then followed a more ethnographical approach to see how the themes that developed were supported, and even contradicted, by the behavioural data. The categories that emerged from the behavioural analysis can be seen in Table 2.

Our approach is in line with other studies such as Salem et al (2015). This study also examined error in humanoid robotics and measured trust based on self-reported quantitative and qualitative questionnaire data, as well as on behavioural data that assesses trust based on the participants' willingness to cooperate.

In order to better visualise our behavioural data, vignettes are used. These also serve to assist us in developing a better understanding of themes and to help verify them. This can be key, because, often, as Bartnek et al (2009) demonstrate, what people report is not what they actually do.

Table 2. Behavioural categories after analysis of audio-video, interview and questionnaire data

Behaviour	Physical affect examples	Verbal affect examples
Emotional reactions and explicit references to feelings.	Uncertainty, surprise, annoyance.	*"The expression did affect me."*
Responses that rationalise participants' reactions to a response.	Not following the robot's suggestion to whisk the eggs ("it wasn't real enough.")	*"It was reassuring that B and C presented a solution."*
Behaviour indicating confusion, embarrassment or indecisiveness.	Looking away.	*"When BERT tried the open palm method of handing the egg over, that was confusing."*
Statements and body language resulting from the robot's unreliability.	Leaning back.	*"I was disappointed that the speed was so slow and the behaviour so error prone."*
Reactions or responses attributing intelligence to the robot or praising it.	Smiling at or mimicking the robot's behaviour.	*"B seemed to be able to learn to be better. His delivery of the next egg was more clever."*
Displays of empathetic behaviour towards the robot.	Helping the robot by attempting to rescue an egg.	*"Thank you!"*
Negative behaviour, indicative of impatience or dissatisfaction.	Hand(s) on hip, scratching chin.	"[It was] *better with speech."*

CHAPTER 4. RESULTS

The results are divided into three sections:

(i) Data from the pre-experiment questionnaire, which provides a profile of the participants as well as the context for the study.

(ii) The post-experiment questionnaire results based on the NASA-TLX metrics and additional questions, showing the impact on our hypotheses.

(iii) Interview data integrated with behavioural factors from the video analysis, which take the form of vignettes and back up our questionnaire results.

4.1 Pre-experiment Questionnaire

Our participants consisted of 11 males and 10 females (Q3), with a good spread across age groups (mean 41.14, SD: 73.62). The majority (57.14%) had a graduate degree (Q4, Figure 5).

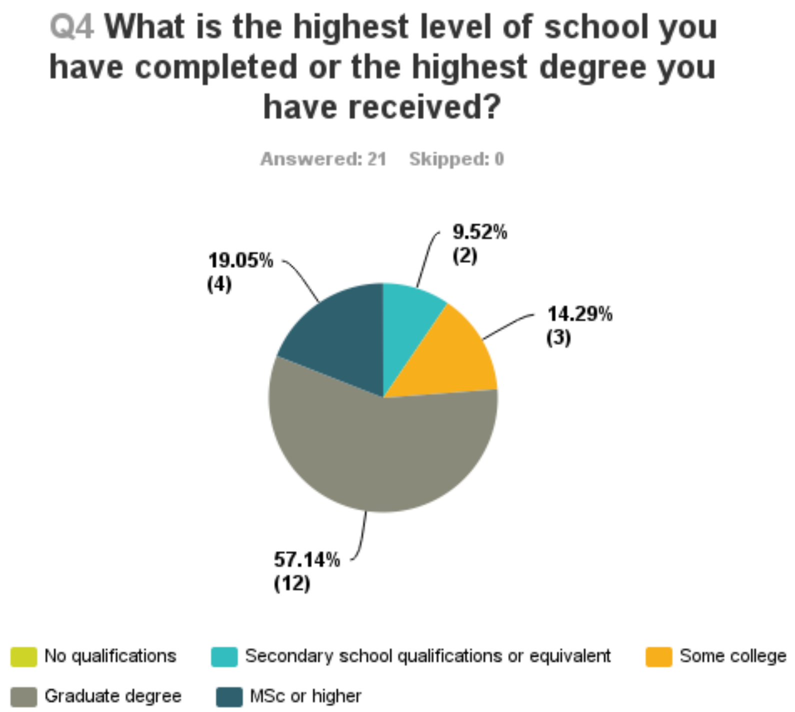


Figure 5. Bar chart showing participants qualification attainments

Over half (52.28%) were employed, with the rest evenly spread between students, retired and unemployed (Q5, Figure 6).

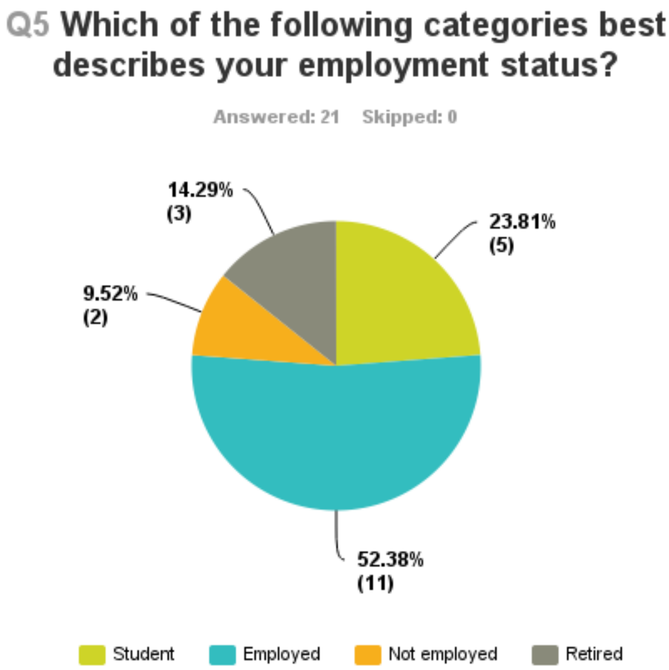


Figure 6. Bar chart showing participants' employment status

Comfort with technology

Ten participants (six females) had had no exposure to robots either at work or home (Q10, figure 7). Of these, four ultimately picked BERT A (two females) and six chose BERT C (four females) as the candidate they would choose to work with again. So their level of experience and whether they were male or female appeared to have little significance in their ultimate choice.

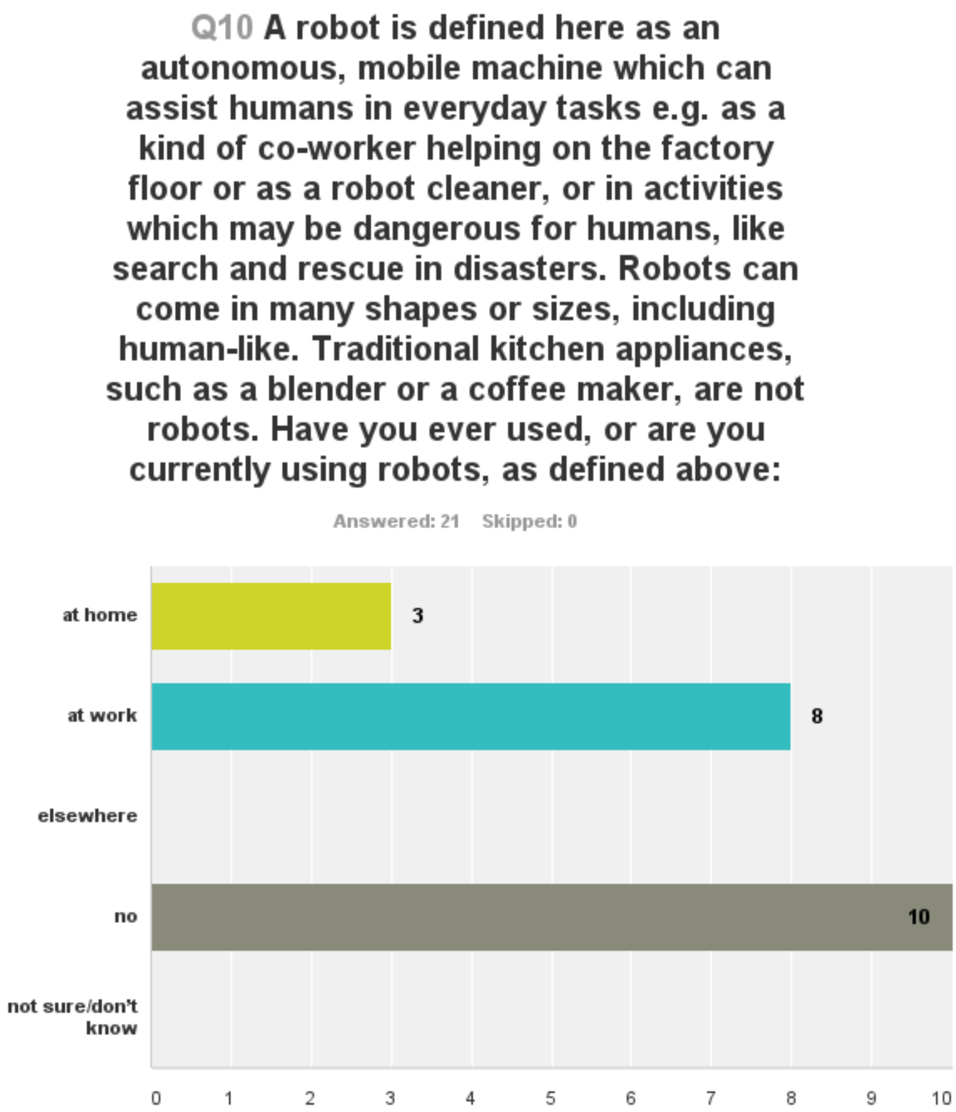


Figure 7. Sideways bar chart showing participants' familiarity with robots

The majority (12) of participants said they had a very positive view of robots, no one expressed negative views (Q11, Figure 8).

Almost all (80.95%) said they were "fascinated" with technological advancements, giving the highest score of 5 on the scale, and the rest were "very interested." (mean: 4.80, SD: 36.88) (Q7, see Figure 9).

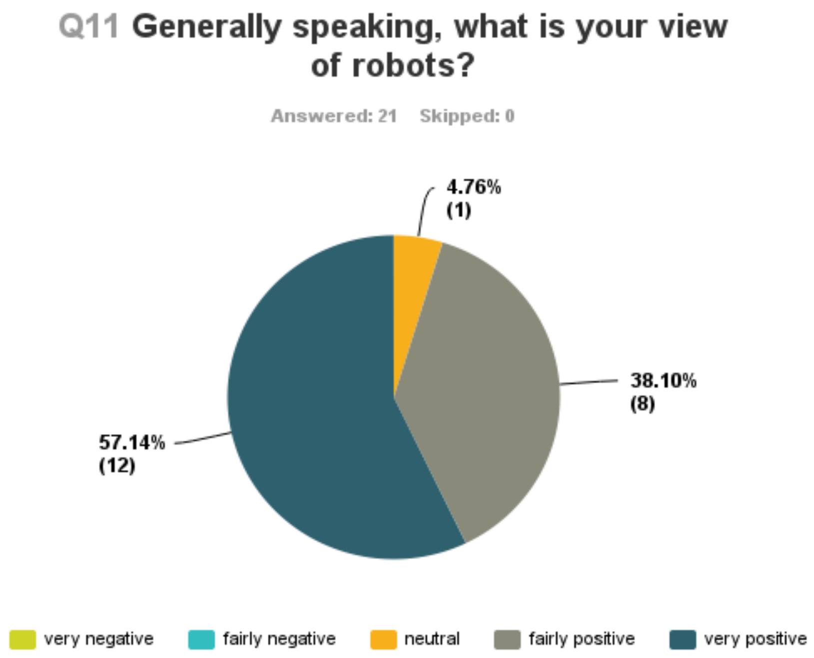


Figure 8. Pie chart showing participants' views of robots

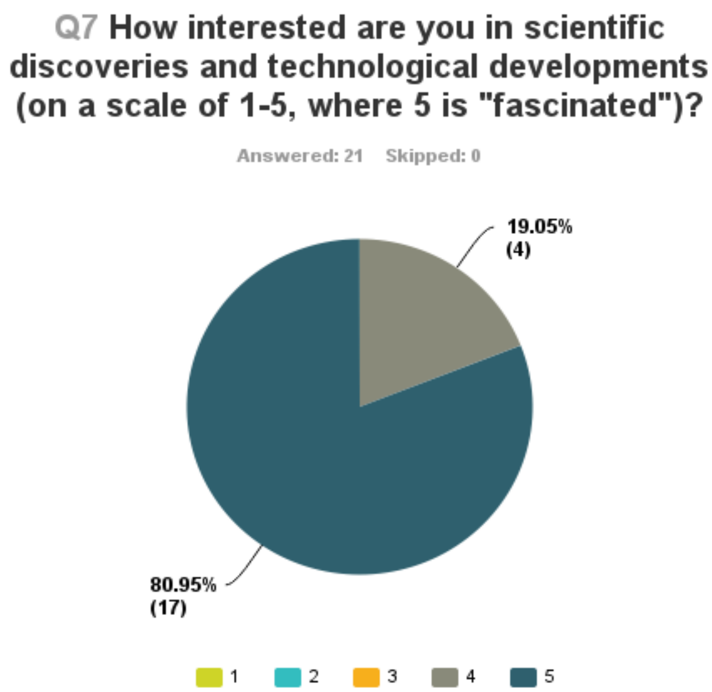


Figure 9. Pie chart showing participants' degree of interest in technology

Most (61.90%) said they were totally comfortable using technology, giving it the highest score of 5 on the scale (mean: 4.47, SD: 26.93) (Q8, Figure 10).

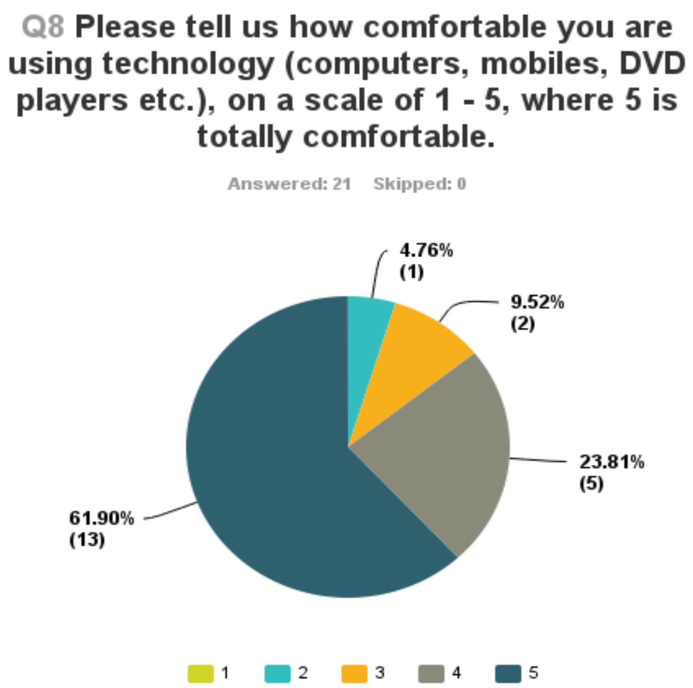


Figure 10. Pie chart showing participants' degree of comfort with technology

When asked to rate their level of dissatisfaction when things go wrong on a scale of 1-5 (with 5 representing the highest level of dissatisfaction), the most popular rating was 4 (mean: 3.66, SD: 18.98, Q9, Figure 11), indicating low tolerance levels.

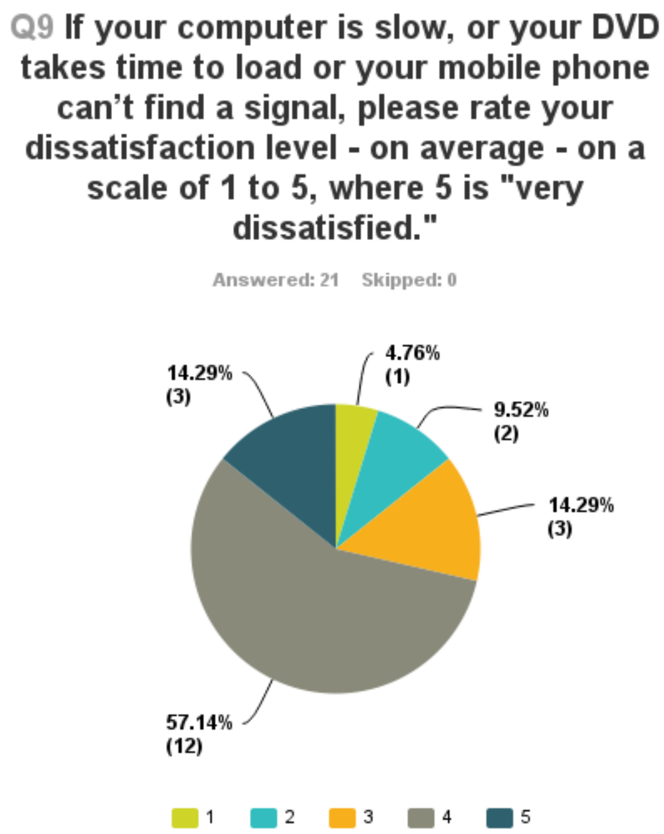


Figure 11. Pie chart showing participants' level of dissatisfaction when technology misbehaves.

Participants were also given a list of things that could be done by robots and asked to rate (from 1-5) their level of comfort with having a robot perform these functions (Q14, Figure 12). The majority (76.19%) were very comfortable with having a robot assist them at work and 66.67% also gave the highest comfort rating for assistance in the home.

However robots walking dogs, caring for or aiding children or the elderly or performing surgical operations were not so popular.

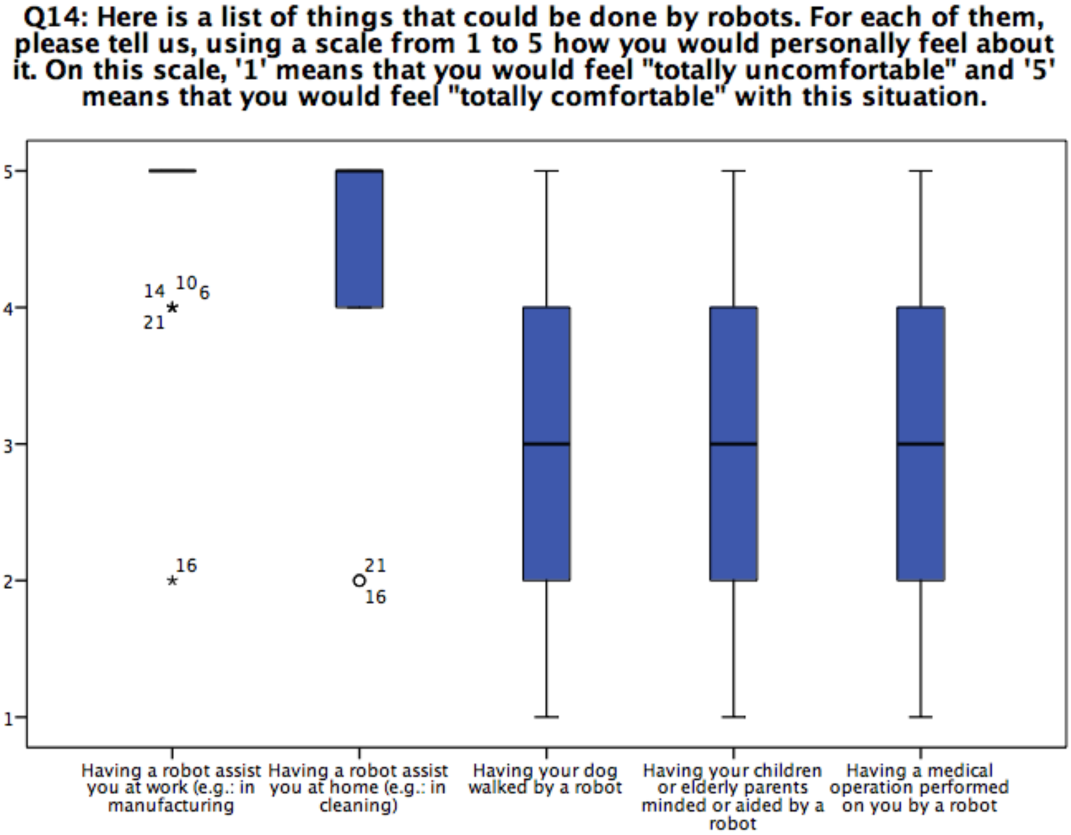


Figure 12. Box plot with y-axis representing participants' views of robots: assisting them at work/home; walking their dog; caring for their children or elderly parents and performing a medical operation on them.

Respondents who answered "totally uncomfortable or "uncomfortable" were asked to provide their reasons (full answers appear in Appendix VII). The most common responses alluded to doubts about dealing with unexpected situations and trust implications.

For example:

"I fear for their/my safety in the event of a malfunction." P14.

"I would need to be convinced that the robot was really capable of recognising unusual situations and responding appropriately in a timely manner - e.g. dealing with children or elderly parents who fell over. Or dealing with unexpected situations in a medical operation." P12.

Expectations

Most participants believed that BERT largely corresponded to their view of a robot (Q12, Figure 13).

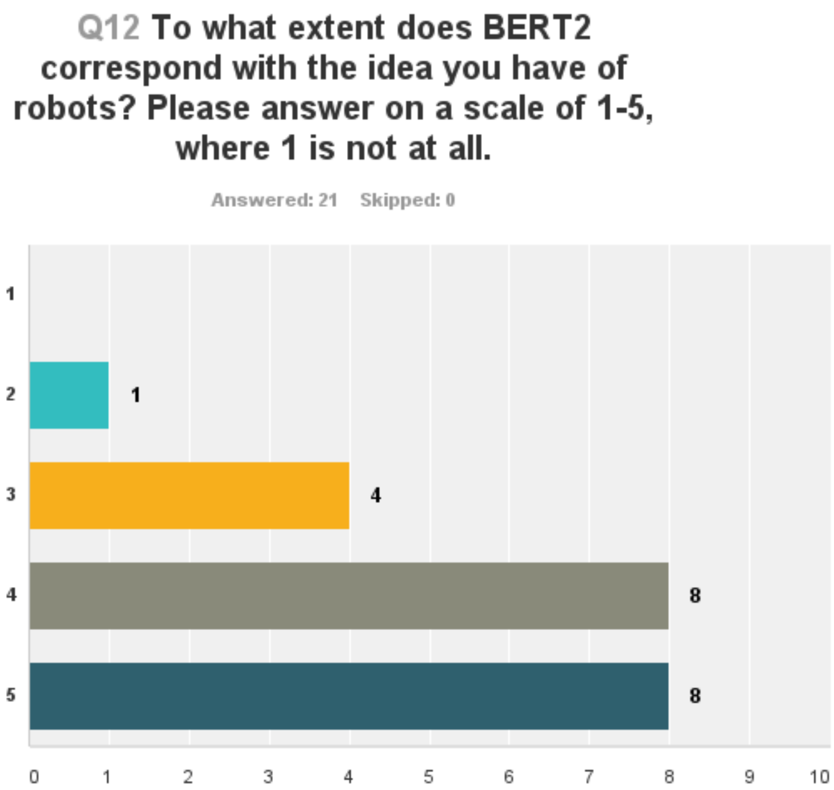


Figure 13. Sideways bar chart with y-axis representing the degree that participants felt BERT2 corresponded to their idea of a robot.

Participants were given a number of options for things that they considered BERT could and couldn't do (Q13). This serves to illustrated how high people's expectations of robots are.

Predictably, in view of the nature of the task, almost all correctly assumed that BERT could handover objects and speak and 95.24% credited it with recognising speech, although only 90.48% thought it could vocalise. Our findings show that many participants ideas about speech recognition technology don't correspond with the reality, 38.10% believed that they would be able to have a conversation with the robot. Notably, two participants said they thought BERT would be able to recognise mood and three believed it could juggle objects.

In the post-experiment questionnaire, two of these respondents scored maximum on feeling insecure, stressed and annoyed with all the BERTs, although C scored slightly lower here. Unsurprisingly, their satisfaction scores for all three robots were very low and they said they were unlikely to want to use any of the robots again.

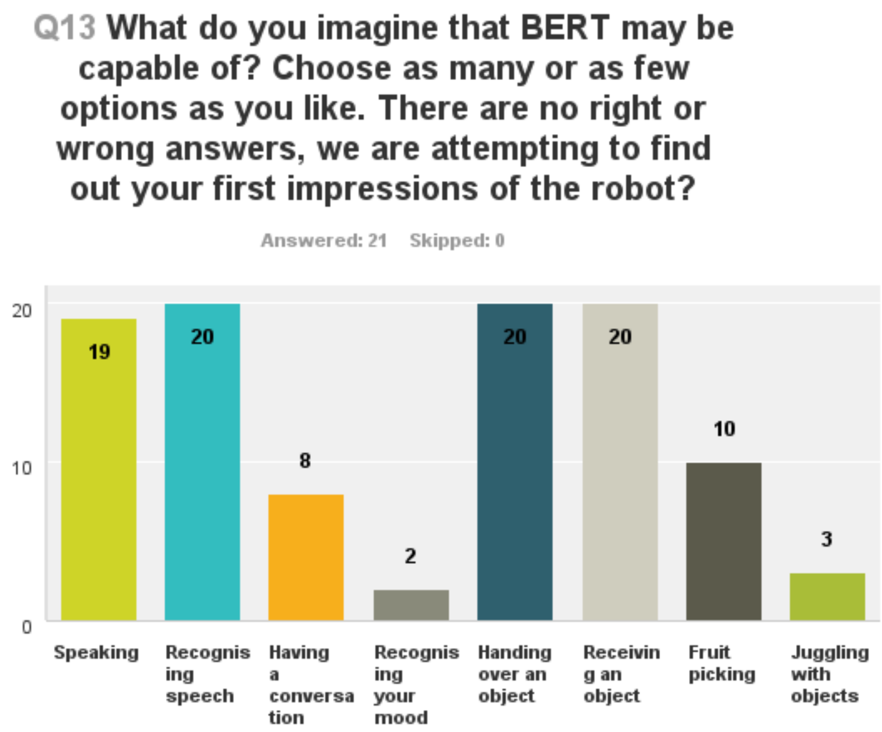


Figure 14. Bar chart with with y-axis representing the number of participants' that believed BERT2 was capable of feats from speaking through to juggling with objects.

4.2 Post-experiment questionnaire

The post experiment questionnaire was largely based on the NASA-TLX workload survey, with additional questions to establish levels of satisfaction and trust. It used a 5-point Likert scale and the data was analysed with the help of non-parametric tests. Participants were also invited to choose one of the job candidates - BERT A, B or C - and we first take a look at their choices and reasons for these. Full responses are in Appendix VIII.

Candidate chosen overall

BERT C was the preferred candidate overall and 15 respondents said they would give this robot the job. Their reasons were largely based on the communicative abilities possessed by C, making feedback possible. For example:

"The vocal interaction with BERT C stopped me wondering what was happening next. It also let me know when he realised that he had dropped the egg and it also let me know when he was waiting for me to whisk the eggs. With the non-vocal machines there is a nervousness about when I should be holding out my hand, etc." P17

“The questions BERT C asked allowed feedback to correct when things are not satisfactory. The questions before releasing an item mean I don't have to worry as much about something being dropped early.” P15

However, six candidates chose BERT A. All referred to the robot’s efficiency as the reason for their selection. For example:

“Bert A was the only one who didn't drop an egg - hence making a mess to clear up.” P12

“Bert A made the fewest mistakes.” P11

One participant who choose A said they would prefer C if it were a work - as opposed to home related - scenario and, of the participants who chose BERT C overall, two said they would reconsider and choose BERT A or B, if the task was work related. (Q13 and Q14, full answers in Appendix VIII). One said:

“In a workplace, particularly for repetitive tasks, there is no need for repetitive vocal interaction; each participant does the job more or less automatically and with more or less the same movements. I have worked in shoe, car, and chocolate factories and found very little conversation about the tasks themselves.” P19

Initial statistical results

General trends in the data can be seen in Figures 15-17. There is little difference in physical and mental demand between the conditions, but several outliers are seen in Condition C. One of these (12) expressed dissatisfaction with all the candidates.

Although BERT C’s performance is largely on a par with A’s, participants appeared to find they had to expend a little less effort on the task with C then with A or B. C appears to have caused participants to feel less rushed and frustrated. It also scores markedly higher in satisfaction. However the median scores for trust appear similar for all three conditions, albeit with a greater spread of opinion in conditions B and C.

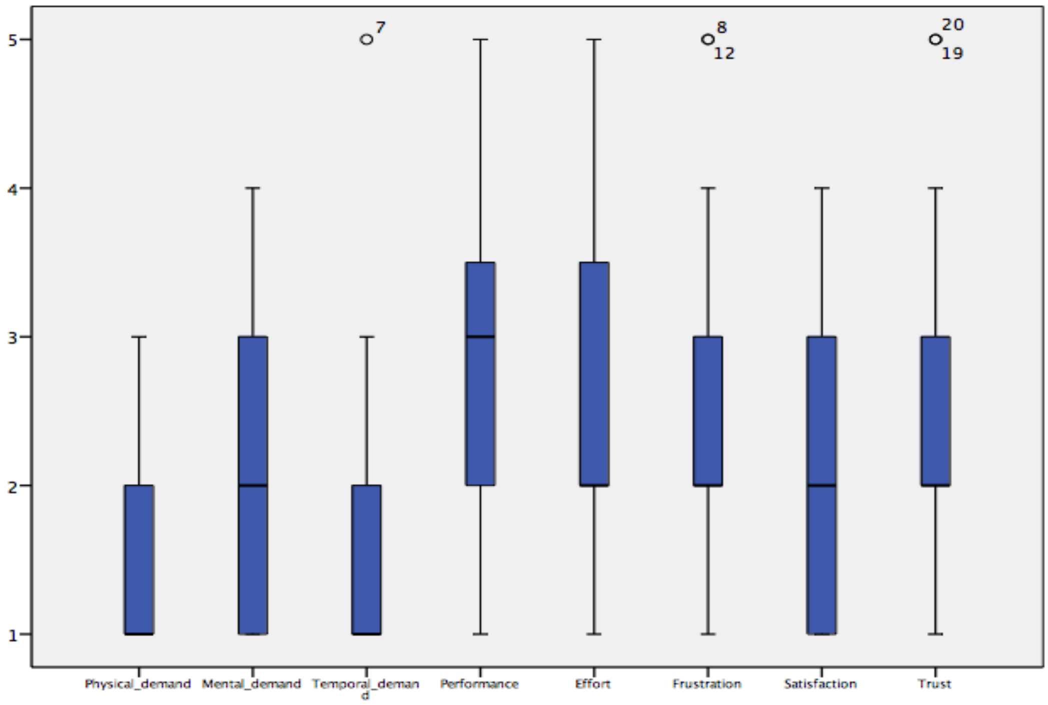


Figure 15. **Box plot with y-axis representing participants' ratings for BERT A, derived from the post-experiment questionnaire.**

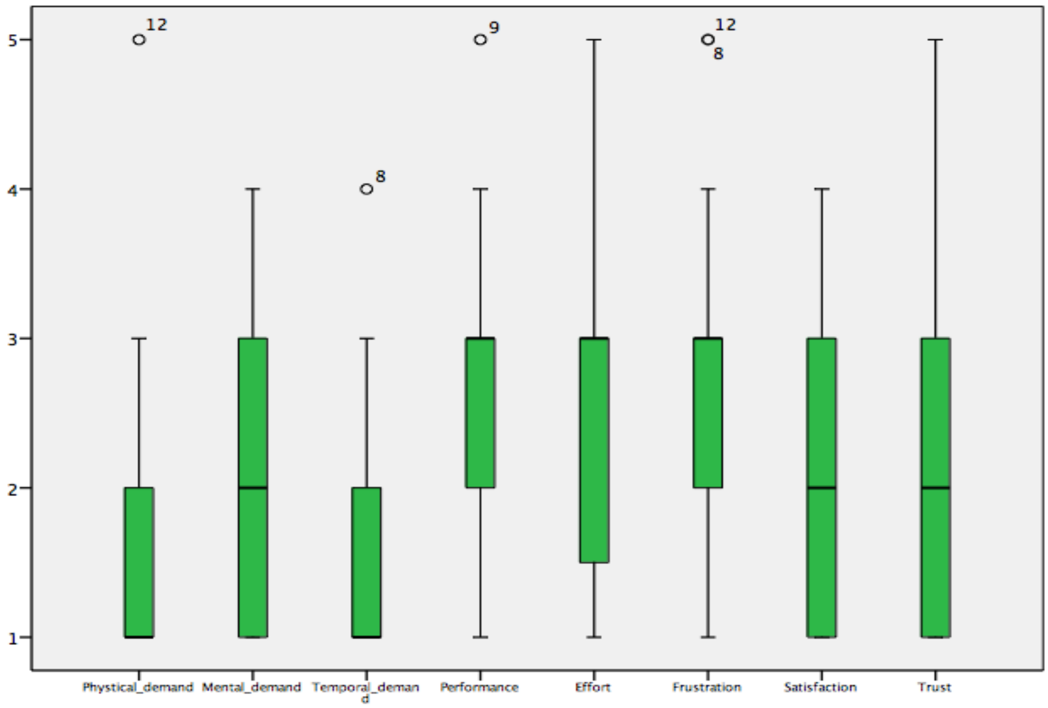


Figure 16. **Box plot with y-axis representing participants' ratings for BERT B, derived from the post-experiment questionnaire.**

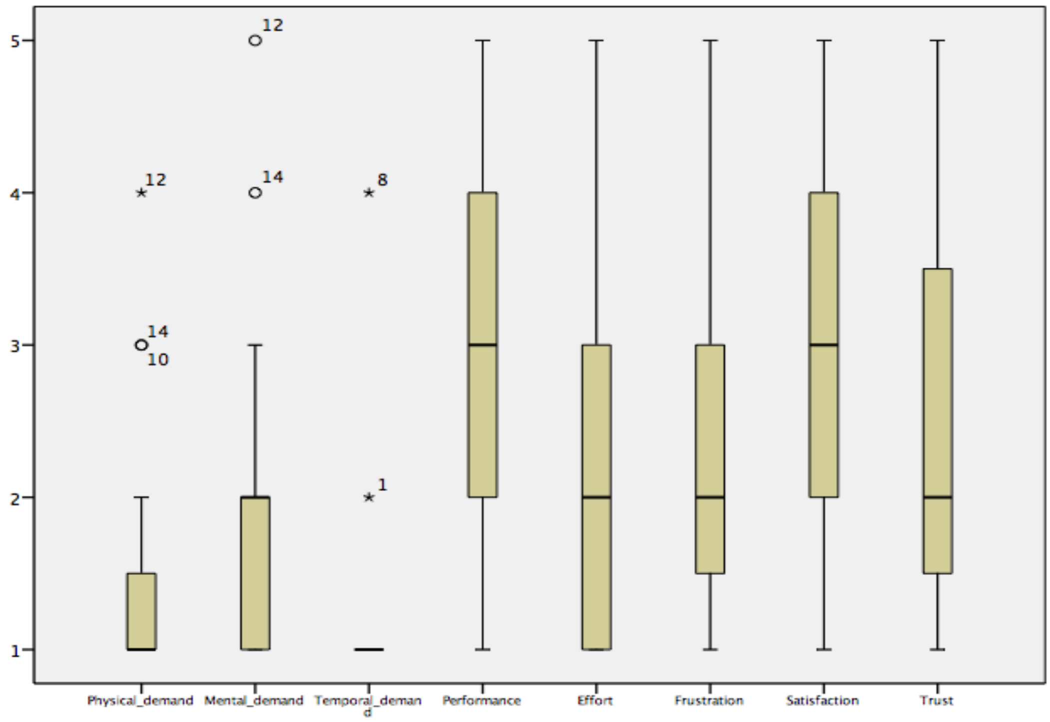


Figure 17. Box plot with y-axis representing participants' ratings for BERT b, derived from the post-experiment questionnaire.

Friedman/Wilcox tests

A Friedman Analysis of Variance by Ranks was applied to the Likert scale data. Where significance was found, a Wilcoxon Signed-Rank test was performed to determine which conditions were significant and establish whether the hypotheses were correct.

The Friedman test results showed that, although mental and physical demand and effort didn't vary significantly among the three conditions, the effect of the type of robot used was, however, seen for the other measures (Table 3). In particular, satisfaction, temporal demand and trust could be said to vary highly significantly in the three conditions.

There was a significant difference in perceived temporal demand depending on which robot was used, $\chi2(2) = 14.000$, $p = 0.001$ (Figure 18) with a markedly lower level in condition C. Although, BERT B made a mistake, participants also appeared to feel less rushed with this robot than with BERT A.

Table 3. Significance levels following Friedman tests

Measure	Significance	$\chi2(2)$	$t(df)$
Mental demand	*n.s*	2.542	2
Physical demand	*n.s*	1.857	2
Temporal demand	p = .001	14.000	2
Performance	p = .042	6.333	2
Effort	*n.s*	5.160	2
Frustration	p = .015	8.440	2
Satisfaction	p = .000	18.353	2
Trust	p = .006	10.226	2

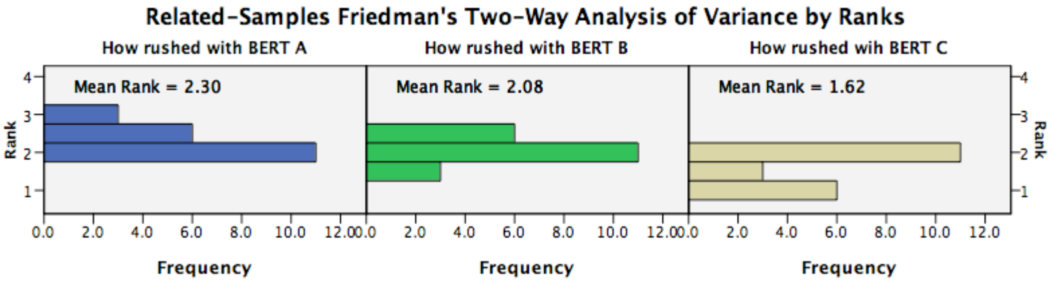


Figure 18. Rank and frequency of temporal demand in the three conditions using Friedman's ANOVA

There was also a statistically significant difference in perceived task performance depending on which robot was used, $\chi2(2) = 6.333$, $p = 0.42$. Rating frequencies and ranks are visible in Figure 19. BERT B's score appears to be perceptively lower than the other two robots, with C slightly ahead on mean rank.

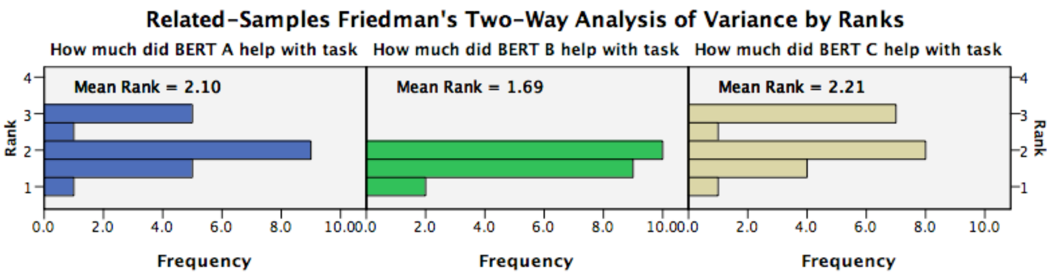


Figure 19. Rank and frequency of performance in the three conditions using Friedman's ANOVA

There was a statistically significant difference in perceived frustration depending on which robot was used, $\chi2(2) = 8.440$, $p = 0.015$. Rating frequencies and ranks are visible in Figure 20. Participants appeared to find BERT C much less frustrating than the other two robots.

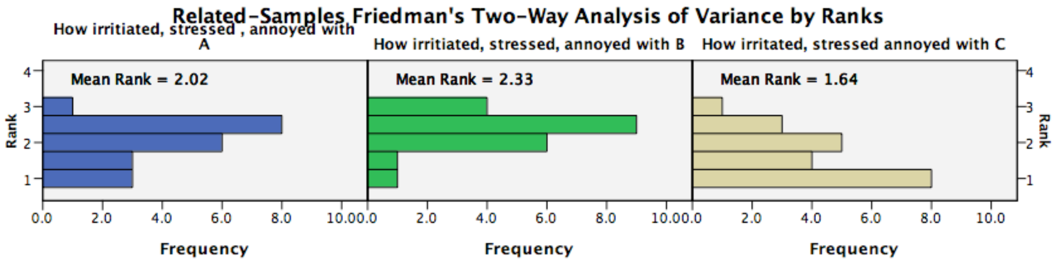


Figure 20. Rank and frequency of frustration in the three conditions using Friedman's ANOVA

There was a statistically significant difference in perceived satisfaction depending on which robot was used, $\chi2(2) = 18.353$, $p = 0.000$. Rating frequencies and ranks are visible in Figure 21. BERT C's scores are much higher than the others, with all the ratings for BERT B at average or below.

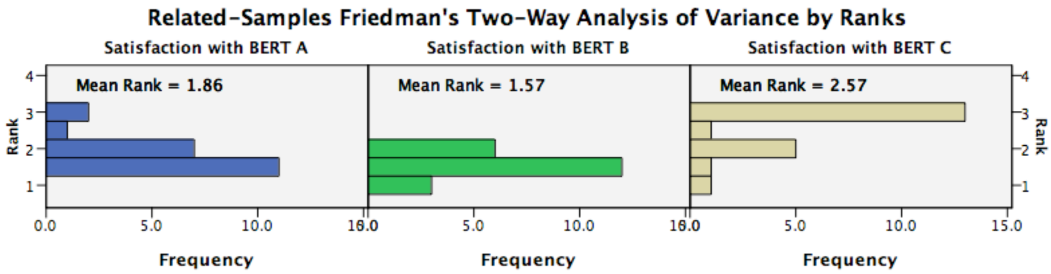


Figure 21. Rank and frequency of satisfaction in the three conditions using Friedman's ANOVA

In the final measure, there was also a statistically significant difference in perceived trust depending on which robot was used, $\chi2(2) = 10.226$, $p = 0.006$. Rating frequencies and ranks are visible in Figure 22. This revealed a much greater spread of opinion in the ratings for A and B, with C edging ahead in mean rank and B's rating, again, average or below.

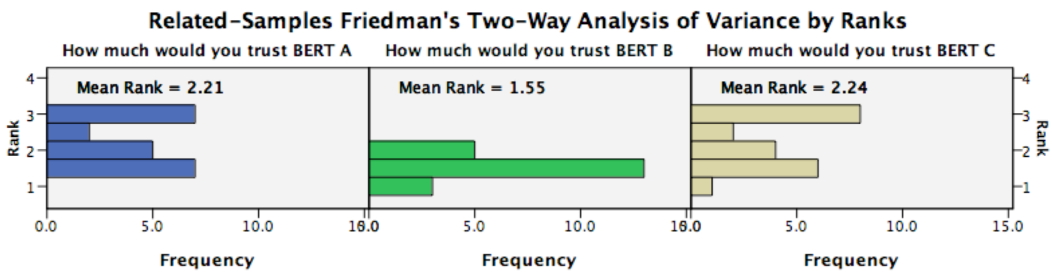


Figure 22. Rank and frequency of trust in the three conditions using Friedman's ANOVA

In order to find out which conditions specifically, were affected, the post hoc Wilcoxon Signed-Rank test was performed on the measures that were found to hold significance and a Bonferroni adjustment was applied to the results (Table 4).

Table 4. Wilcoxon Signed Rank Tests. Only significant (p <.017) results and those approaching significance are reported

Condition	BERT A/BERT B	BERT B/BERT C	BERT A/BERT C
Temporal demand	*n.s.*	$p < .023$	$p < .005$*
Performance	*n.s.*	$p < .021$	*n.s.*
Frustration	*n.s.*	$p < .011$*	*n.s.*
Satisfaction	$p < .025$	$p < .005$*	$p < .020$
Trust	$p < .002$*	$p < .008$*	*n.s.*

* (elevated) significant findings, using Bonferroni

In respect of temporal demand, post hoc Wilcoxan Signed Rank tests using the Bonferroni correction revealed a statistically significant reduction with BERT C, when compared with BERT A ($Z = -2.810$, $p = 0.005$). Participants were also shown to feel slightly less rushed with BERT C than with BERT B, with a result approaching significance ($Z = 2.271$, $p = 0.23$).

While there was no significant difference found in performance (although the difference in perceived level of help approached significance between BERT B and BERT C: $Z = 2.309$, $p = 0.21$), the amount of frustration experienced by participants was significantly more in the B condition than with BERT C ($Z = -2.546$, $p =0.11$).

As a reflection of this, levels of perceived satisfaction differed between these two conditions accordingly ($Z = -2.799$, $p = 0.005$). The results between A/B ($Q = -2.236$, $p = 0.25$) and A/C ($Q = 2.325$, $p = 0.20$) also approached significance, indicating that the interaction with BERT C was more satisfactory than the other robots.

In respect of perceived trust, the post hoc tests revealed a statistically significant increase with BERT C, when compared with BERT B ($Z = 2.658$, $p = 0.008$). There was also a significant difference between BERT's A and B ($Z = 2.658$, $p = 0.02$), with BERT A receiving higher rankings in this category.

The results in respect of satisfaction appear to be particularly meaningful as, although six participants ultimately chose BERT A as their overall preferred candidate, when it came to satisfaction, only one expressed more with A than C.

This meant that hypothesis 2 proved to be correct: Communication - both audio and visual - can significantly mitigate dissatisfaction in the event of an unforeseen occurrence (i.e. BERTC will be significantly more popular than BERT B).

As BERT C was rated higher overall than BERT A and also achieved higher satisfaction ratings, we were also able to maintain hypothesis 3: Given the choice between enhanced efficiency and reliability and a personable, communicative interface, most people will chose the later (i.e. BERT C will be chosen over BERT A).

Participants trusted BERTs A and C significantly more than BERT B (A/B = $p < .002$; B/C =$p < .008$).

This was partly in accordance with hypothesis 1: An unforeseen occurrence, will cause a robot to appear less trustworthy than a more efficient version, even if attempts are made to mitigate the mistake (i.e. BERT A will be more popular than BERT B).

It also provided further evidence to support hypothesis 2.

However, there was no significant difference between the levels of trust between A and C.

The other main finding was in the temporal data. Participants reported being much more rushed in their interaction with BERT A than with BERT C ($p < .005$) although, it seemed, less so with BERT B ($p < .023$).

4.3 Behavioural and interview results

Behavioural and interview data was broken down into themes that emerged during the interactions and this framed the analysis. These were categorised and used to validate self-reported data. The themes have been presented, where possible, in the chronological sequence of the omelette making task and illustrated with vignettes, allowing us to visualise elements of the interaction that support satisfactory experience, together with instances where frustration and uncertainty are clearly apparent.

Initial impressions and engagement

Initial reactions to BERT varied considerably. Some participants found his appearance unsettling. For example:

I found the expressions disconcerting. Also the black face and red lips: slave like, servile. P11

He was terminator like when the arm is outstretched. P1

In total, five of the 21 participants attempted to talk to BERT's A and B, at least initially, even though they'd been told these robots weren't able to respond. This corresponds with the Hegel et al (2008) finding that people are prone to attribute human-like faculties to humanoid robots and there is a need for caution in regard to this.

One participant, P4, was particularly enthusiastic and made repeated attempts to engage with BERT A, despite his lack of response. She began by saying: “Hello, I like you!” and then appeared impressed with his appearance and functionality: “He looks so sincere! His face is just so eager. He blinked too!” Her enthusiasm was evident at many points within the interaction (“I need to calm myself down a bit, I’m just so excited!”). She was also determined to fully interact, saying “thank you” after every handover.

P9 too was keen to engage during his interaction with BERT B. This was evident in his attempts to converse with the robot: “You want me to lift it out? Wow that’s very impressive…. Shall I take it from you this way? Very kind. Thank you so much.”

The robot was monitored much more frequently initially, with an average of three glances at the interface during the initial egg handover. This monitoring behaviour would seem in accordance with Bagheri & Jamieson, (2004) which indicates that it’s closely linked to trust. Monitoring was also amplified for those participants who were first exposed to BERT C. Their level of engagement was significantly greater than with the other BERTs.

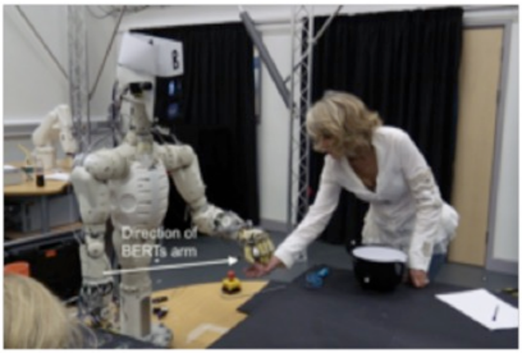
However, when they moved on to BERTs A and B, some participants interest appeared to diminish as the interaction progressed and they were more likely to display signs of boredom or impatience, including grasping the object before the robot was ready to release, postural distancing from the robot and drumming their fingers on a thigh or on the table.

Three participants mentioned that they would have liked more interaction with BERT C. On the whole, increased intensity was manifest not only in perceived muscle relaxation at the initial stages of the interaction, but also in the way that they appeared to move their heads more to follow his movements. This could be interpreted as a sign of increased presence in the interaction but is also a way to facilitate control (Bianchi-Berthouze, N., 2013). However, the interaction with BERT C could have been judged by the participants to be more complex than with the other robots, as it required them to respond to the robot’s questions.

Temporal issues and over-reaching

For safety reasons, the interaction needed to be unnaturally slow and this was an issue for many of the participants. Their impatience was often obvious. In particular, many participants didn’t want to wait for BERT to release the eggs. However, this may also be construed as being nervous that the egg was about to drop. It often meant that they over-reached and needed to reposition their hand to receive the egg. Three participants, in particular, continuously attempted to grab the egg out of the robots hand before he was ready to release it. This was most apparent in the interaction with BERT B but was also a problem with A. One of these instances is shown in Vignette 1.

Vignette 1.



[P6, keen to receive the egg, over reaches, tracking BERT B’s arm direction in an effort to retrieve the egg.]

However, an alternative interpretation is also possible. Kendon (1985) has pointed out that moving into synchrony with another person can indicate the desire to establish "an action exchange system" (p. 245). Picking up on the rhythm of the robot’s movements can thus be construed as an effort to establish a connection.

Distance and postural orientation

Postural orientation, and distance cues have long been considered relevant to the communication of attitudes (Mehrabian 1968). We therefore examined the participants’ body orientation towards or away from the robot. We looked at head and shoulder orientation and arm openness. Additionally, we observed the angle of the torso (Argyle & Cook, 1976). A backward lean has been associated with boredom and a forward lean is indicative of high interest.

One participant (P13), in particular, seemed reluctant to get to close to BERT. Her reluctance to touch BERT the robot was most evident when she received the salt. Her reaction was to take the cylinder from its base, as if she didn’t want to get too near the robot’s hand. She also used her arm to create distance between herself and the system as is demonstrated in Vignette 2. Her attitude was substantiated by her self-declared lack of confidence around technology, in the pre-experiment questionnaire (one from a scale of five).

Vignette 2.

[Reluctance to interact too closely with the robot is evident in P13’s posture: she appears tense, is leaning slightly back, to one side and is using her arms to create distance.]

Intention recognition

Researchers agree that gaze clearly plays an important role not only in coordinating conversational interaction, but also in carrying out physical tasks. According to Jordan and Henderson (1995):

In our analyses of working and learning situations, we have found again and again that it is important to track where people's eyes are, when and how gaze moves between objects, from persons to objects, and back again, sustaining or shifting the focus of attention as the salience of particular objects or displays changes.

Almost all participants wanted to rely on BERT's gaze behaviour as a cue to its attentional state. They often appeared confused, especially where there was no speech to specify what it was doing.

In all three conditions, particularly during the "different" handover, participants often made a concerted effort to catch the robot's gaze to ascertain its intentions, varying their glance between face and hand as much as three times in the space of five seconds. As a whole, the robot was more closely monitored initially and also when the interaction was problematic, such when the handover method was varied.

Participants were often unsure about what was coming next, this was particularly evident with the uncommunicative robots but also manifest with BERT C. Vignette 3 illustrates P18's uncertainty after BERT C's suggestion to whisk the eggs ("You are now ready to whisk the eggs.") The robot turned away to retrieve the cooking oil, leaving P18 unsure about whether or not he should stop whisking (Vignette 3). This episode illustrates the issues that can arise when it is unclear who is in control of the interaction and the importance of clear, finite instructions on the part of the robot.

Vignette 3.

[P18 looks away askance, unsure whether he needs to stop whisking in the absence of any indication from the robot.]

But uncertainty was more pronounced when it came to the uncommunicative robots, as can be seen in Vignette 4. BERT B's attempt to use gesture to indicate that the eggs are ready to be whisked is clearly insufficient; to judge from P20's puzzled expression.

Vignette 4

[P20 gazes at BERT B, unsure what is required of him in the absence of any instruction from the robot]

[He furrows his brow trying to discern BERT B's intention.]

Unsure what was required with the two uncommunicative robots, many (10) participants tried to hand back the salt when the robot was, in fact, waiting for them to whisk the eggs.

Whether participants whisked the eggs, or not, was also often dependent on the order the candidates were presented in. Those encountering BERT A or B first were most unsure what they needed to do when the robot indicated the whisk. Even with, C, the verbal command to whisk was quite ambiguous ("You are now ready to whisk the eggs"), causing most of the participants to hesitate.

One participant, despite being aware of the robot waiting for him to whisk the eggs, didn't do so, later explaining: "the robot wasn't looking to see if I've done it." (P12) This underlines Fong et al's (2003) assertion that the system must be understandable and believable.

Cooperation

When BERTs B and C dropped the third egg, many of the participants attempted to help the robot and some tried hard to prevent the eggs from falling, even succeeding in "rescuing" the dropped egg on two occasions. These participants all went on to choose BERT C, they also provided high satisfaction scores, implying a richer interaction experience. One of these instances is shown in Vignette 5. However, factors such as extroversion and comfort with technology can also impact on increased cooperation, emphasising the importance of taking into account individual characteristics.

Vignette 5

[P3 sees that BERT C has dropped one of the eggs]

[Moving quickly, she tries to help by retrieving the dropped egg]

[She shows the rescued egg to BERT, looking for a reaction.]

[She sees BERT's sorry expression, looks surprised and then laughs.]
P3: It's OK
BERT C: I am sorry, shall I pass you the next egg?
P3: Yes

[She pretends to break the egg into the mixing bowl]

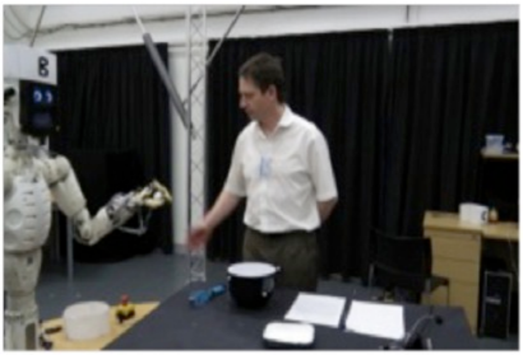
Reactions to the “different” handover

After BERT’s B and C dropped an egg, they were programmed to try the handover again, but differently, in an attempt to rectify the situation. BERT C was able to warn participants of this, saying: “Let’s try something different.” However, almost all the participants were surprised by BERT B’s “different handover,” necessitating a rapid realignment of the way they received the egg (Vignette 6).

Vignette 6

[P17 prepares for the imminent handover of the next egg but is unaware that BERT B will present it palm up this time.]

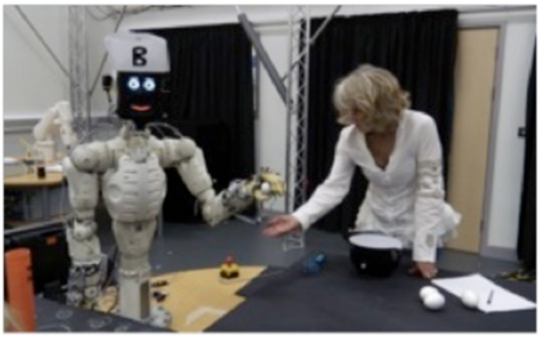
[Realising that something is different, he removes his hand, leans back and tries to discern BERT’s intention.]



[He notices BERT releasing the egg so swiftly moves his arm to take it before it falls.]

As shown in the vignette above, P17 was not caught on guard to the extent that he responded with his whole body. He keeps his left arm behind his back even while swiftly moving arm to take the egg. This would indicate that he is not unduly concerned about the egg falling, despite the new handover maneuver. However, as seen in Vignette 7, P6 is significantly more determined and perhaps anxious that the egg should not fall. She uses her entire upper body in stooping to extract the egg in the new handover.

Vignette 7

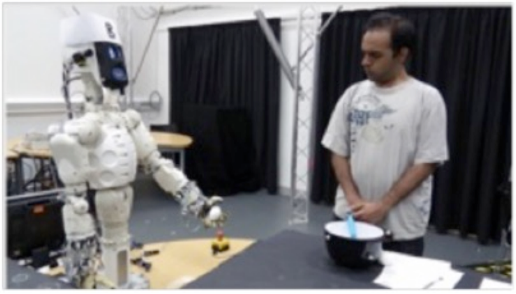


[P6 puts her hand out expecting BERT B to present the egg in the same position as before. But BERT presents it palm up, causing her to look up and reappraise the situation.]

[Seeing he’s ready to release it, she quickly stoops to extract the egg form his opening palm.]

The different style of handover visibly surprised the majority of participants. Particularly, with BERT B, where there was no warning of a different approach. In Vignette 8, P20 visibly jumps when he thinks the egg will fall out of BERT B’s palm.

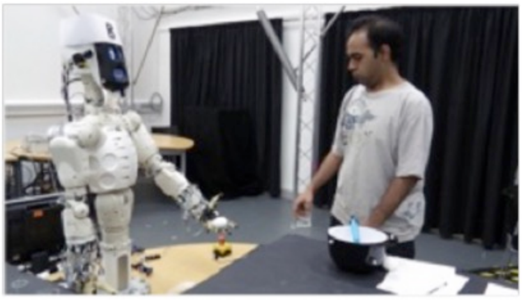
Vignette 8



[P20 becomes aware that the handover method is not the same..]

[Caught out, he moves swiftly to make sure the egg doesn't fall.]

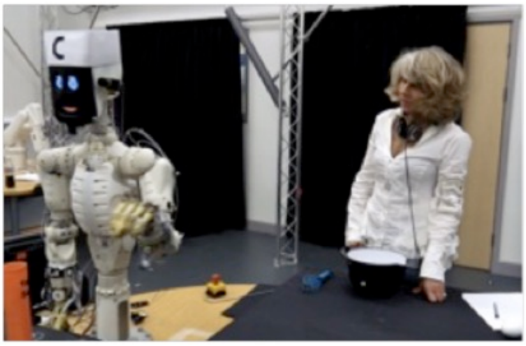
[He positions his hand in the place where BERT has previously deposited the egg]



[He appears vexed as he tries to ascertain which is the best way to pick up the egg.]

The lack of any prior warning was clearly a problem in this example but despite his communicative abilities, BERT C didn't appear to fare much better with the alternative handover method, as the prior warning he issued looked to have little discernible effect in many instances. In Vignette 9, P6 is caught out, perhaps even more than she was with BERT B, and failure in voice recognition adds to the problem:

Vignette 9



[BERT C picks up the egg]
BERT C: Let's try something different.

He moves into handover position and P6 quickly positions herself to receive the egg, giving the appearance that she is aware the handover will be different as, this time, she holds out two hands for the egg.
P6: Whooo!

BERT C: Are you ready for the egg?
P6: [too quickly for voice recognition] Yes.
BERT C: Sorry, please speak after I am finished. Are you ready for the egg?
P6: [shouts, too quickly] Yes!

BERT C: (beep)
P6: [more quietly] Yes

[P6 claps her hands in recognition of the achievement of finally getting the egg.]

In the example above, while the participant was clearly aware that something different was about to occur, the warning didn't specify the nature of the difference. This would serve to demonstrate the importance of a warning's clarity (Kim and Hinds, 2006).

The majority of participants were demonstrably unsure how to receive the egg, as, in both conditions B and C, they were equally in the dark about exactly how the robot would deliver it. This was demonstrated in tense posture and startled facial expressions.

But their appraisals of the different handover method, in interview, varied widely. Some participants were admiring that the robot seemed make an attempt to learn and rectify an issue. The manoeuvre was also given as a reason for choosing BERT C above the rest. For example:

It was reassuring that B and C presented a solution. P15

BERT C had [sic] known if/when he didn't do well (apart from not getting the oil, which he appeared not to have noticed). He also had a 'different solution' when he dropped an egg. P2

However, there was wide disparity in attitudes to the "different" handover and the efficacy of the delivery in particular. A number of the participants didn't like the new handover method, saying they didn't understand how a palm up method wouldn't cause the egg to fall. For example:

When BERT tried the open palm method of handing the egg over, that was confusing. Human beings would usually drop the egg into the other persons hand. P17

It was also perceived to be unnatural:

Presenting the egg was better in the iteration but kind of weird the way it turned its elbow out. P16

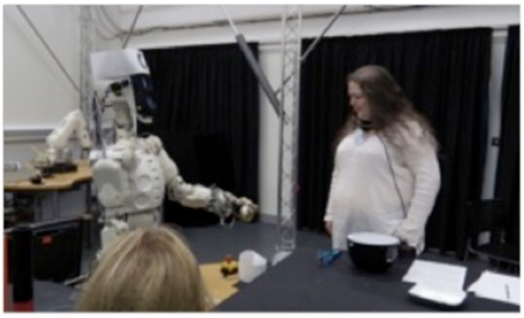
But others found it more comfortable:

Picking the egg out of his hand easier, if anyone has dexterity problems. P19

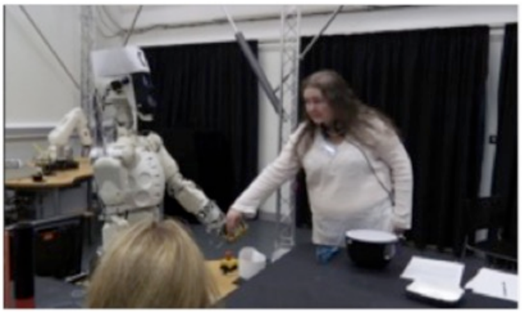
Speech recognition

More than a third of the participants were visibly irritated by problems with speech recognition, which seemed to be exacerbated by the "different" handover, as is apparent in Vignette 10. This gives much credence to studies such as Kanda et al (2008), which urge caution in using speech recognition.

Vignette 10



[P19 sees that the handover method is different]



BERT C: Are you read for the egg?
P19: Yes.
[She attempts to take the egg too quickly as if worried that it will fall, but her answer is not recognised]

[She steps back and repeats her answer]
P19: Yes

BERT C: Please speak after I am finished, are you ready for the next egg?
[P19 raises her eyes quickly upwards, in a gesture of irritation.]
P19: Yes [for the third time]

BERT C: Are you ready for the egg?
P19: Yes
[BERT releases the egg and she is finally able to retrieve it.]

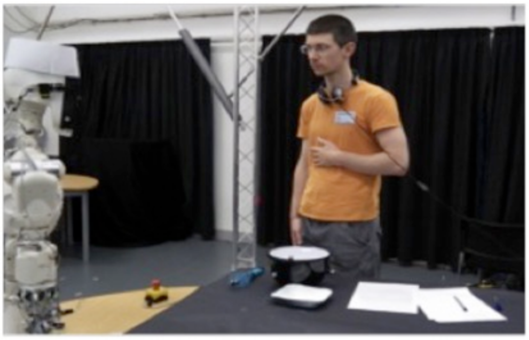
However, despite the problems speech recognition presented, voice interaction was welcomed by the majority of participants:

I was a lot more comfortable with the voice. With the feedback, I didn't feel something was constantly going to drop. P3

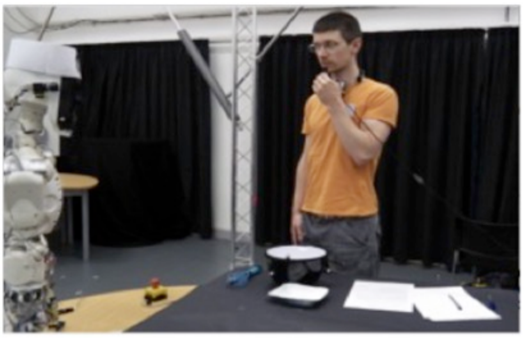
With C, there was a perceived opportunity to correct something as you can answer yes or no, even one way communication makes a big difference. P15

Despite his appreciation of the opportunity to communicate, for P15 particularly, speech recognition was a big issue from the beginning. At one juncture he had to say yes four times before the robot released the egg.

Vignette 11



BERT C [third time]: Please speak after I am finished, are you ready to begin making the omelette?
P15: Yes



[P15 raises his mike]
BERT C [fourth time]: Please speak after I am finished, are you ready to begin making the omelette?
P15: Yes. [He is visibly relieved when the robot finally understand him.]

As demonstrated in the Vignette above, the issues with speech were obviously distressing to P15 and caused his body to slump and his hand to rise in a defensive gesture, signalling negative emotion (De Silva and Bianchi-Berthouze, 2004; De Meijer, 1989).

But despite the fact that he appeared uncomfortable, in interview, he expressed a marked preference for BERT C, explaining:

I'm used to giving things more allowances because my voice isn't very compatible with voice recognition. I know from experience that call-up voice automated things generally don't understand me. I assumed the problem was with me.

The same was true of another participant, who had problems with the system recognising her responses because of her pronounced Spanish accent. She reported: “It didn’t put me off, having to repeat things.” (P11)

Reactions to BERT C’s apology

To compliment his apology, BERT C displayed a regretful facial expression when it dropped the egg (Figure 23).

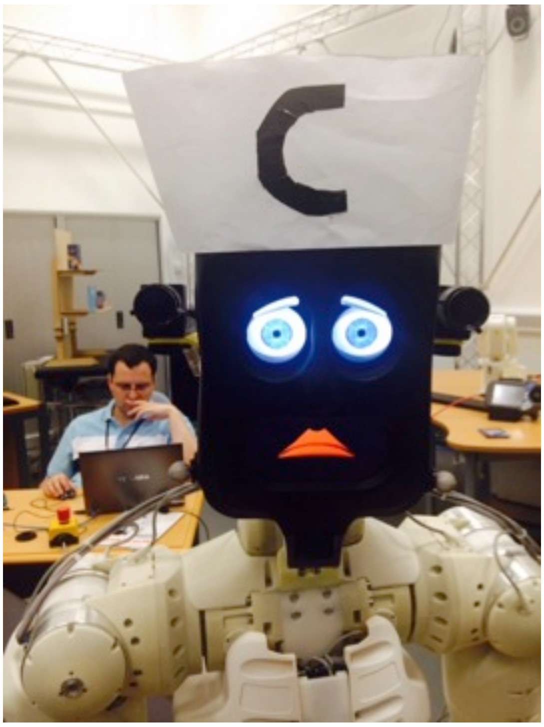
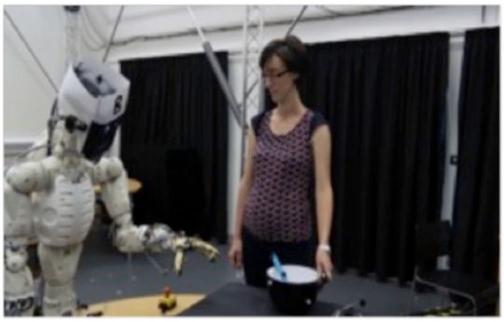


Figure 23: BERT C, expression after egg drop

The contrast in nearly all the participants reactions between BERT B's "mistake," with no apology or facial distress portrayed, and BERT C's exaggerated look of sadness was marked, as is apparent in Vignettes 12 and 13.

Vignette 12



[BERT B drops the egg. P21 appears uncertain as to whether the robot realises its mistake]

BERT C: I am sorry.
[BERT C looks remorseful, causing P21 to break out in a wide grin]

The apology and regretful expression had a visible effect on almost every participant. Vignette 13 illustrates the effect on P11, who appeared most nervous at the start of the interaction.

Vignette 13

[Early in the interaction, P11 is visibly apprehensive when BERT C speaks]
BERT C: initialising myself, initialisation OK.

[Later on, BERT C's apology and regretful expression takes P11 by surprise]
BERT C: I am sorry

P11 appears to relax and mimmicks BERT C's expression

Figure 24. Reactions to BERT C's apology and regretful expression

BERT C's expressiveness caused a reaction in many of the participants, as can be seen in Figure 24. In interview, it was also remarked on:

I was surprised [that] he "knew" I was whisking! An altogether more natural experience. I was quite surprised how drawn I was to him. I didn't think that I would be drawn to a robot, but the 'face' caused me to smile back and look for responses. P17

You couldn't help feeling sympathy when it dropped the egg. You see the face and just go "awww! P19

These responses would seem to align well with studies including Bruce et al (2003), which suggests expressiveness and engagement gestures can make a robot more compelling to interact with and Breazeal et al (2005), which demonstrates the importance of social cues improving effectiveness.

Challenging behaviour

At the end of the interaction with BERT C, when the robot asked participants whether it had performed well and if it had got the job, it made a number of them visibly uncomfortable, as demonstrated in Vignette 14.

Vignette 14

BERT C: Did I get the job?
P15: No

[He appears regretful, making a sad face.]

P15 elaborated on his decision not to give the robot the job in the follow up interview:

It felt appropriate to say no, but I felt really bad saying it. And, when the face was really sad, I felt even worse. I felt bad because the robot was trying to do its job.

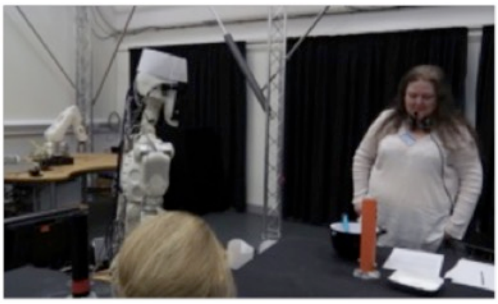
His perception that BERT C's face was sad when it didn't get the job was interesting as this wasn't, in fact, the case; the robot wasn't programmed to show any form of reaction here. However it was clear that its expressiveness made a distinct impression on him:

In later tasks, I think I would be even more forgiving because it had expressed those emotions… The expressions did affect me, it was surprising. Once it's expressed emotion, it triggers something.

Even though they appeared uncomfortable with the question, some participants still chose to tell BERT C he'd won the job (Vignette 15).

Vignette 15

BERT C: Did I get the job
P19: [laughs] Yes
BERT C: Thank you



[Immediately afterwards, P19 both moves and looks away, appearing thoughtful and uncomfortable.]

For the participants, the feeling of being put on the spot was exacerbated by the fact that, if they told the robot it didn't perform well and didn't get the job, they couldn't qualify their answer. This is evident in the following examples:

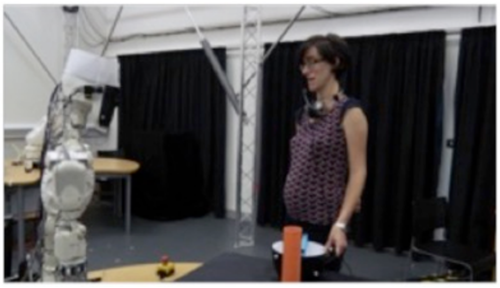
Either forgetting that he couldn't answer more than yes or no, or unable to help himself, P9 replies: "Not at all sure, have to speak your creator about that."

And P6 attempts to answer: "probably not, I didn't get the cooking oil." Not recognising these words, the system simply switches off.

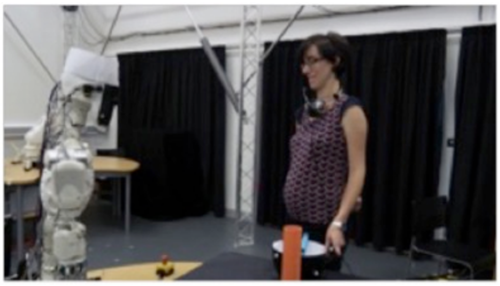
However, it was clear that subjects felt the need to say something more than a straight "no."

There was one participant who appeared to be particularly reluctant to disappoint BERT C, shown in Vignette 16.

Vignette 16

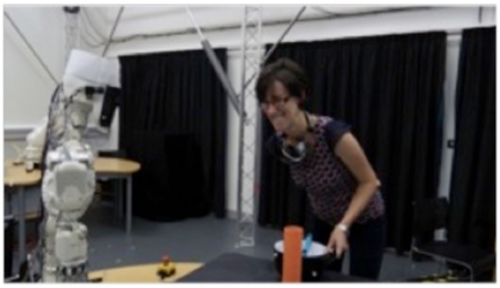


BERT C: Did I get the job?
P21: Maybe



BERT C: Sorry, did I get the job?
P21: [Hesitates], Yes [appears very uncomfortable]

BERT C: Thank you, goodbye.



[She doubles over in - uncomfortable - laughter]

Although she answered yes to the job question during the interaction, P21 ultimately chose BERT A as her preferred job candidate in the post-experiment questionnaire. Asked about her reaction in the interview, she replied: “Freaky! I couldn’t say anything other than yes or no. [I] felt uncomfortable.”

BERT C’s challenging behaviour ultimately didn’t discourage the majority of the participants from awarding him the job, but there’s no doubt that being put on the spot was an unpleasant experience for them, as one subject’s notepad remark reveals (Figure 25.)

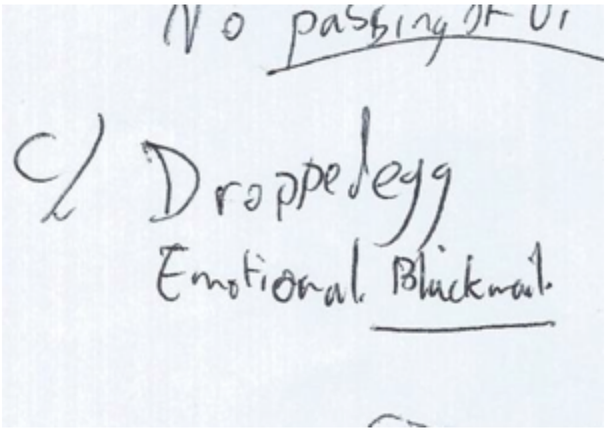


Figure: 25: Participant comment demonstrating his discomfort when BERT C asked: "did I get the job?"

Temporal and confusion issues

The communication offered by C gave participants the impression that time was passing much quicker than it actually was; the time the task took with BERT C seemed shorter, even though it was actually almost two minutes longer. In interview, three of the participants said that they thought A "seemed slower" or "quite slow," compared to C or even B.

However some did accurately judge the longer length of the task with C. One subject who chose A as his preferred candidate said he would have chosen C, but it was "too slow." Another was under the impression that, compared to A, B "was faster, more responsive and did improvise." In fact, the interactions with A and B were approximately the same length, despite B dropping an egg.

It's clear that there was some confusion among participants regarding which robot was which. One appeared to think that A was the one that dropped an egg, another believed that B was more fluid than A and "wasn't aware that A didn't present the egg in a different way." (P17). In the surveys he accorded A and B similar scores and ultimately chose C.

Participants also contradicted themselves. One complained of the general slowness of all the candidates but chose the highest scores for the temporal element of the NASA TLX questionnaire, indicating that he felt rushed in performing the task with all the robots.

Effect on expectations

Many participants expressed disappointment with the robots general abilities, the sedate pace of all the interactions and the fact that none of the BERTs delivered the oil or appeared to realise this. For example:

The robot needs to go a bit faster, there was too much time waiting which made me impatient and [want to] just reach for the objects myself. (P14)

Participants commonly said they had expected much more and were disappointed that the behaviour was so slow and error prone.

However others were surprised at the strength of their reaction to BERT C, discovering, for example, that: “I am likely to humanise the robot if it has a voice.” (P11). This is in accordance with Fong et al (2008) who counsel that verbal communication can cause users to form incorrect role models. P11 went on to choose BERT A as her preferred candidate.

A common impression was that C gave the perception that it was thinking. A typical comment was:

I picked C because of the interaction, stops you wondering what he’s going to do next. Because he went still, [I] saw he knew he’d made a mistake. P17.

The desire for more verbal interaction with BERT C was a further theme. For example, one participant believed that C should have asked what he did wrong when he was told he didn’t get the job.

However, there were also negative comments about this aspect, for example:

…the constant verbal support for Bert C with respect to the task at hand. P19.

CHAPTER 5. DISCUSSION

The goal of this study was to find effective strategies for an assistive robot to recover from the uncertainty or misunderstanding that may arise from erroneous or unexpected behaviour. This is an area still sorely in need of investigation if collaborative robotics is to enter the mainstream. Specifically, we wanted to gauge to what degree communication and human-like attributes can positively affect the participants' experience. Studies, to date, have largely focused on performance as a measure of trust, however our results indicate that the majority of users prefer a communicative, expressive robot above one that is more efficient and less error prone.

This section examines the themes that emerged from the data and the behavioural findings that served to either validate or, in some cases, throw into dispute what the participants reported. Some of these themes were cross-cutting and analysis also produced a number of new avenues of enquiry, summarised together with limitations at the end of this chapter.

5.1 Unexpected behaviour and/or error does not need to negatively affect trust if allied with transparency.

Hypothesis 1 was that an unforeseen occurrence will cause a robot to appear less trustworthy than a more efficient version, even if attempts are made to mitigate the mistake. This was proved to be correct, as BERT A was far more popular than BERT B, which was not chosen by any of the participants as their preferred candidate in the task. This finding is in accordance with a great many studies that identify performance as having the biggest influence on trust (Muir et al, 1976, Freedy et al, 2007, May; Hancock et al, 2011).

Not only did BERT B drop an egg but the system's unexpected behaviour, when it presented the next one in a different manner, visibly surprised many of the participants and their appreciation of this as an attempt to rectify an error was varied.

Despite the rather ambiguous nature of BERT C's intentions, when it announced the new handover method ("Let's try something different"), many participants were vocal in their appreciation that, not only was it trying to rectify the situation, but it had noticed its mistake. They were visibly more at ease.

The fact that the cause of the unexpected behaviour was, with C, intelligible to the participants may even have resulted in some of them increasing the level of cognitive ability they had previously ascribed to it (Lemaignan et al, 2014), particularly if the new form of handover was seen as a successful measure. This was evident in factors such as one respondent's perception that the robot seemed sad when it wasn't given the job, when there was no reason to surmise this.

Although six of the participants did prefer the more efficient robot, BERT C's errors don't appear to have affected his popularity for the majority, giving further weight to the finding by Salem et al (2013) that errors can sometimes even increase a robot's human-likeness and likability.

The control that the opportunity to provide feedback offered, even in the limited circumstances ("yes" or "no" in respect of whether they were ready to receive the egg) was also welcome, over the more efficient candidate. Manifest in lower frustration levels and increased satisfaction and trust scores for BERT C, transparency and control were shown to be more important than increased autonomy (Johnson et al, 2012) for robot-human teamwork.

5.2 The majority of users welcome, even limited, spoken communication with an assistive robot.

Participants desire to talk to even the uncommunicative robots, their greater levels of engagement in the interaction and higher rating given to BERT C all serve to underscore a preference for an interaction that involves speech. This is in line with studies showing that users have an overwhelming preference for speech communication (Ray et al, 2008, Iwamura et al 2011).

Our findings indicate that despite problems with speech recognition, in a cooperative situation such as this, users are generally content to repeat themselves if this will aid their understanding of what is taking place.

Despite its shortcomings in present day systems and warnings of users forming incorrect role models (Fong et al, 2003) or disappointment due to raised expectations (Kanda et al, 2008), our study finds that, in the main, a communicative robot is preferable to a more efficient, less error prone version. This would indicate that users are prepared to sacrifice quite a large degree of efficiency for transparency and feedback, whether the situation is a domestic or workplace one.

We were therefore able to show that given the choice between enhanced efficiency and reliability and a personable, communicative interface, most people will chose the later.

It's also notable that the majority of participants didn't mind - or even appear to notice - that the interaction with BERT C was 50 per cent longer than with A or B. Only one participant said they would have chosen C if it hadn't taken so long and one even thought BERT A took longer to complete the task. This would indicate that the participants' degree of involvement with BERT C was significantly higher than with the other robots.

5.3 Human-like attributes can effectively smooth a difficult interaction

In condition C, communication was also supplemented by further feedback in the form of the robot's expression. It's suggested that participants were more visibly affected by BERT C's face than the verbal apology alone. This was apparent as two were looking away writing something down when BERT apologised. While hearing the apology, they thus missed the visage he presented and did not experience such a marked reaction.

`This finding would appear to tally with studies by Bruce et al (2003) on the importance of facial expression and Breazeal et al (2005) which indicates that implicit non-verbal cues can aid an interaction and increase likablilty.

BERT C's expression of dismay, together with its verbal apology after it dropped the egg, had a salient effect on participants and caused them to attribute both intelligence and the ability to learn to the robot.

It increased the robot's believability (Fong et al, 2003) and immediately alerted them that it "knew" it had made a mistake. It may also have cued them to its desire to make reparation by finding a solution, when there was no way of visually expressing that the handover was going to be different. This form of natural mapping therefore played a critical role in aiding people to form a mental model of the robot during a key phase in the cooperative interaction. Such a situation can be crucial to the development of trust as defined by Lee & See (2004). This meant that hypothesis 2 proved to be correct: Communication - both audio and visual - can significantly mitigate dissatisfaction in the event of an unforeseen occurrence. It also bears out the Parasuraman's et al (2004) finding that the effects of good automation etiquette can be powerful enough to overcome low reliability with a corresponding effect on trust.

Particularly in the domestic sphere, a robot might encounter a number of situations where it can no longer ensure a good outcome of its actions. It's thus imperative to equip domestic robots with the means to convey their intentions, expected level of trustworthiness and perception in a natural way (Eder et al, 2014). Expressive features would seem an obvious way to go towards achieving this.

Generating empathy

You can always say sorry, but the real apology is when you hear the sadness in their voice and see the look in their eyes. And you realize that they have hurt themselves just as much. Kid Cudi

Whether it was surprise, amusement or a desire to mimic BERT C's sorry expression, it was clear that the majority had some form of reaction to it.

Davis (1983) defines empathy as "the capacity to take the role of the other, to adopt alternative emotional reactions in consort with the context to the point of executing bodily movements resembling the other's."

Empathy, he writes, can be an affective process (experiencing a certain emotional state) but it can also be cognitive (gaining awareness, understanding or knowledge of another's state), with the former inspiring the more powerful affective state.

Studies suggest that emotional empathy is special as it "facilitates somatic, sensory, and motor representation of other peoples' mental states, and results in more vigorous mirroring of the observed mental and bodily states than cognitive empathy" (Nummenmaa, Hirvonen, Parkkola & Hietanen, 2008). Some degree of mirroring of BERT C's expression was observed in at least three of our participants and could indicate emotional contagion, as defined by Cacioppo (1994). In the latter study it's suggested that when people unconsciously mimic their companions' expressions of emotion, they come to feel reflections of these.

This is a powerful notion but needs to be treated with care as developmental theorists make clear distinctions between primitive empathy, emotional contagion, and the more cognitive, sophisticated and "social beneficial" processes of empathy and sympathy. (Eisenberg & Strayer, 1992).

It's hard to accurately establish whether these (or even additional) participants did experience emotional contagion, thought to be less conscious and more automatic than cognitive empathy. A more precise study is necessary, involving, for instance, EMG procedures (Rapson, Hatfield & Cacioppo, 1993). But there is no doubt that an affect that helps subjects gain information about the other party can serve to increase understanding and provide a form of glue for personal relationships, as well as alleviating frustration and stress (Picard, 2006). This further serves to highlight the importance of non-verbal cues in HRI.

5.4 Knowledge of expectations can help in achieving user goals

The ultimate goal in designing a collaborative system is that the robot does not disconfirm users' expectations (Eder et al, 2014), but this is not easy to achieve.

The robot's inability to take advantage of gaze was identified but there were also other instance where pre-conceptions raised issue. At least five participants had inflated expectations of what BERT2 was capable of. One, for instance, believed that the platform could recognise mood. He went on to rate 5 for his views being changed post study. To best mitigate unrealistic expectations, our results suggest that a system should demonstrate its abilities as far as possible and should be designed with caution to mitigate high expectations (Paepcke & Takayama, 2010).

However, this may not be easy to achieve. For instance, we attempted to lower voice recognition expectations by limiting participants' responses, but this didn't prevent a number of them from attempting to engage the robot beyond its abilities. This could be construed as a desire to test these for themselves (Briggs & Scheutz, 2014) or a measure of engagement.

There are a great many factors to consider and "taking all the possible biases into account would require a complex and therefore impracticable experiment (Bartnek et al, 2009)." Our data suggests that neither are obvious differences, such as levels of experience or a user's gender, effective shortcuts. Of the participants who preferred BERT A overall, two were naive users, which would negate any hypothesis that all naive users are more likely to prefer a communicative system at the expense of a more effective one.

However the characteristics of a human-robot team, their relationship and the context have been shown to be worthy of consideration (Woods, 1996). Early regard to these factors would aid in understanding the sense-making process and facilitate the design, so that the robot could better indicate why the mistake occurred or mitigate its effect.

One suggestion is to factor in people's initial reactions to a robot, whether, for example, they greet it prior to the interaction. Lee, M. K., Kiesler, S., & Forlizzi, J. (2010, February) found that greeting a robot significantly predicted a more social script suggesting that people's first words with it influence their schematic orientation to a robotic service. This indicates that designing robots that adapt their recovery strategy at the outset of an interaction to best fit a user is a possibility.

5.5 Emotional and challenging behaviour by robots must be used with care

Even though not entirely planned, participants' reactions to BERT C's question "did I get the job?" provided a rich stream of insight. The question appeared to distress many of them, manifest in discomfort and difficulty in responding. Their situation was exacerbated as they could only answer yes or no. BERT C's behaviour can therefore be described as challenging.

In a manifesto on ethics for the HRI profession, Riek and Howard (2014) outline human dignity considerations. Foremost among these is that "The emotional needs of humans and their human frailty - both physical and psychological - are always to be respected."

BERT C's challenging question would seem to be exemplary of behaviour which violates these principles.

Further evidence on the effect of distressing or challenging behaviour comes from an experiment where participants were asked to upset a robot by ordering it to knock down a tower it was proud of having built (Briggs, G., & Scheutz, M. (2014). The vast majority of subjects admitted to feeling some level of discomfort when issuing this command.

In our experiment, participants were very reluctant to deny BERT C the job outright and, instead, tried to modify their answers, even though this was impossible. In the Briggs et al (2014) experiment similar behaviour was evident and "some subjects attempted to bargain and compromise with the robot, attempting to sugar the pill by offering to help rebuild the tower."

Their study suggests that this behaviour could be about the participants exploring what the robot could or couldn't understand, despite the fact that they were told that this was limited. However, we would go further and postulate that, having seen the robot display human-like emotion when the egg dropped, participants were now pre-conditioned to expect a similar reaction and therefore hesitated to say no. Having developed a degree of empathy towards it, they were thus mindful that it could display further human-like distress. One participant even went so far as to lie to the robot in an effort to avoid saying no.

Emotional displays and challenging behaviour by robots can thus be used for both positive and negative affect. So, while Integrating these elements in a system design is beneficial in certain situations (such as providing encouragement for users when dispensing medication to a reluctant patient, for example), challenging or emotional behaviour from a robot can also cause emotional distress in humans, as was evident here. The tendency for people to form attachments to and anthropomorphise robots thus needs to be carefully considered.

5.6 Limitations

The speed of the interactions in all the conditions was a major limitation and compromised the believability in the situation.

It was not possible to use BERT2's integrated sensors or 3D vision, due to the instability of the platform. The latter would have enabled us to ascertain whether the participant was looking at the system.

As eye gaze is intimately connected to both speaking and listening (Staudte & Crocker, 2009), indicates interest (Argyle & Cook, 1976) and provides a good indication of whether what has been said is understood (Mutlu, Shiwa, Kanda, Ishiguro, & Hagita, 2009), its lack was major disadvantage.

The use of the Vicon system would have increased the speed of the handover process and enabled the participants to communicate their readiness to receive the objects. Additionally, the quantative data obtained would have made our findings more robust and decreased reliance on self-reported and behavioural data from audio video analysis.

The problems and limitations inherent in affect recognition using audiovisual analysis are outlined by Pantic, M., & Rothkrantz, L. J. (2003). Foremost among them is the fact that some affective states are difficult to assess and manage, particularly from video. Achieving high inter-observer agreement in affect data annotation is therefore challenging and ultimately unsuccessful in this study. Additional time would have been needed to refine our coding system, a more ethnographical approach was therefore decided upon.

Our sample contained a good spread in terms of age and experience with robotics, but was by no means representative of the population; all the participants were either "very interested" or "fascinated" in new technology. In order to encompass reluctant users, a different approach to finding participants would have been necessary.

The design of our experiment may also have produced confounds as there was some confusion about the three robots. However, a within subjects study was necessary due to time constraints and the relatively small size of the sample.

In order to minimise the risk of confounds arising due to speech recognition, we attempted to limit respondents' answers to "yes" and "no." However this very much constrained their degree of interaction.

A number of collaborative robot interfaces utilise a touch screen, either as a primary means of interaction, to supplement voice systems or to gauge the system's internal state. Examples appear in Figures 26 and 27. Although none, as yet, possesses the necessary degrees-of-freedom (DOF), combined with speech recognition technology necessary for our experiment.

Figure 26: Baxter, Rethink Robotics **Figure 27: Pepper, Softbank**

Had the robot featured a screen it could, perhaps, have communicated its intentions in a more effective manner, previewing the modified handover for example. The provision of real-time status indicators would also be of benefit to users, particularly where there is no other method of communicating the system's internal state.

5.7 Future work

When it comes to measuring satisfaction and trust, there is a dearth of studies on the effects of human characteristics - including prior level of operational experience, existing attitudes toward robots, personality traits and other individual differences. The need for further research will become increasingly important in this area as it becomes common to tailor a robot's personality to an individual user.

However, for robots to become truly integral and useful, particularly in a domestic context, they must also be accepted. For this, social heuristics are necessary, as these are unconsciously applied by humans, even to non-humans.

The analyst Clifford Nass argued that: “the human brain is built so that when given the slightest hint that something is vaguely social, or vaguely human … people will respond with an enormous array of social responses (Bearman, n.d.).”

Careful management of these responses is a powerful medium that can be used to mitigate issues that will no doubt arise in collaborative situations.

However, while adequate speech recognition systems are still in their relative infancy, implementing personality is problematic. Therefore, combining speech with additional measures (touch interfaces but also expressive facial features), particularly in the case of humanoid robots aimed at the domestic market, is an area for further exploration.

5.8 Conclusion

This study began by asking whether, contrary to much existing literature, a robot’s behavioural style and personability is significantly more important than effective task performance in establishing trust and satisfaction, particularly when erroneous behaviour is involved. We found that an expressive robot is preferable over more efficient one, despite a trade off in time taken to do the task.

Our data suggests that the majority of people will chose a communicative, personable robot over a more efficient and less error prone one. Users’ desire for communication – both verbal and non-verbal - was manifest in their efforts to interact with BERT2. Satisfaction was significantly increased in the communicative condition. For the majority of participants, personable, transparent behaviour appeared to negate the fact that the interaction took 50 per cent longer than with the non-communicating robots.

These responses were supplemented by detailed study of areas where the interaction was problematic and incidents that provoked notable reactions. Human like attributes such as regret were shown to be powerful tools in negating dissatisfaction. We found that these had a marked effect on participants, which suggests that, here, the interaction was most believable and “came alive.”

These reactions would also indicate that an appealing, communicative robot, displaying human-like emotions of regret and enthusiasm, influences the user experience in such a way that dissatisfaction with its erroneous behaviour is tempered if not forgiven. “The ability of a person to atone has always been the most remarkable of human features,” according to the novelist Leon Uris (1924).

However, disappointment, confusion and impatience were identified as arising from a lack of understanding of the systems internal state and prior expectations about its abilities.

Participants’ scores also indicated that they felt rushed with the more efficient robot, which would suggest that they placed less value on task performance and more on feedback, despite many instances of speech recognition failure.

Particularly in a domestic situation and when the user has limited familiarity with the robot, this study provides evidence that the judicious use of human-like attributes can enable a robot to successfully negotiate the problem terrain that arises in many a new relationship.

REFERENCES

Admoni, H., Dragan, A., Srinivasa, S. S., & Scassellati, B. (2014, March). Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 49-56). ACM.

Andrist, S., Gleicher, M., Mutlu, B., Andrist, S., Mutlu, B., Gleicher, M., ... & Gleicher, M. (2012). Conversational Gaze Aversion for Humanlike Robots. *Submitted for publication*, 4.

Argyle, M., & Cook, M. (1976). Gaze and mutual gaze.

Avizienis, A., Laprie, J. C., Randell, B., & Landwehr, C. (2004). Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions on*, *1*(1), 11-33.

Bagheri, N., & Jamieson, G. A. (2004). Considering subjective trust and monitoring behavior in assessing automation-induced "complacency.". *Human performance, situation awareness, and automation: Current research and trends*, 54-59.

Bazo, D., Vaidyanathan, R., Lentz, A., & Melhuish, C. (2010, October). Design and testing of a hybrid expressive face for a humanoid robot. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on* (pp. 5317-5322). IEEE.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71-81.

Bearman, R. (n.d.) Parse. com and the art of abstraction.

Bethel, C. L., & Murphy, R. R. (2006, March). Affective expression in appearance constrained robots. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 327-328). ACM.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005, August). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on* (pp. 708-713). IEEE.

Briggs, G., & Scheutz, M. (2014). How robots can affect human behavior: Investigating the effects of robotic displays of protest and distress. *International Journal of Social Robotics*, *6*(3), 343-355.

Bruce, A., Nourbakhsh, I., & Simmons, R. (2002). The role of expressiveness and attention in human-robot interaction. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on* (Vol. 4, pp. 4138-4142). IEEE.

Cakmak, M., Srinivasa, S. S., Lee, M. K., Forlizzi, J., & Kiesler, S. (2011, September). Human preferences for robot-human hand-over configurations. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on* (pp. 1986-1993). IEEE.

Calisgana, E., Moona, A., Bassanib, C., Ferreirab, F., Opertoc, F., Veruggiob, G., ... & Van der Loosa, H. M. Open Roboethics Pilot: Accelerating Policy Design, Implementation and Demonstration of Socially Acceptable Robot Behaviours.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse studies*, *7*(4-5), 507-525.

Cole, R. (1999, August). Tools for research and education in speech science. In *Proceedings of the International Conference of Phonetic Sciences* (Vol. 1, pp. 277-1).

Corritore, C.L., Kracher, B. Wiedenbeck, S (2003). On-line trust: concepts, evolving themes, a model. Int. J. Hum.-Comput. Stud. 58(6): 2003, 737-758

Dautenhahn, K., Woods, S., Kaouri, C., Walters, M. L., Koay, K. L., & Werry, I. (2005, August). What is a robot companion-friend, assistant or butler?. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on* (pp. 1192-1197). IEEE.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal behavior*, *13*(4), 247-268.

De Silva, R. and Bianchi-Berthouze, N., Modeling human affective postures: An information theoretic characterization of posture features, Journal of Computer Animation and Virtual Worlds, 15, (2004), 269—276

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological science*, *23*(12), 1549-1556.

Dourish, P. (2004). *Where the action is: the foundations of embodied interaction*. MIT press.

Docktorman, E. (2015, July) Robot kills man at Volkswagen plant. *Time.com*. Retrieved from http://time.com/3944181/robot-kills-man-volkswagen-plant/

Eder, K., Harper, C., & Leonards, U. (2014). Towards the Safety of Human-in-the-Loop Robotics: Challenges and Opportunities for Safety Assurance of Robotic Co-Workers. *arXiv preprint arXiv:1404.2229*.

Eisenberg, N., & Strayer, J. (Eds.). (1990). *Empathy and its development*. CUP Archive.

Engineering and Physical Sciences Research Council (EPSRC) "*Principles of Robotics*". *Internet: http://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/Pages/principlesofrobotics.aspx (2010) {accessed 30.5.2014}*

Fischer, K., & Kukliński, K. (2014, March). Intuitive error resolution strategies during robot demonstration. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 120-121). ACM.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3), 143-166.

Forlizzi, J., & DiSalvo, C. (2006, March). Service robots in the domestic environment: a study of the roomba vacuum in the home. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 258-265). ACM.

Freedy, E., DeVisser, E., Weltman, G., & Coeyman, N. (2007, May). Measurement of trust in human-robot collaboration. In *Collaborative Technologies and Systems, 2007. CTS 2007. International Symposium on* (pp. 106-114). IEEE.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual review of psychology*, *63*, 287-313.

Goetz, J., & Kiesler, S. (2002, April). Cooperation with a robotic assistant. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (pp. 578-579). ACM.

Grigore, E. C., Eder, K., Pipe, A. G., Melhuish, C., & Leonards, U. (2013, November). Joint action understanding improves robot-to-human object handover. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on* (pp. 4622-4629). IEEE.

Gunes, H., & Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, *31*(2), 120-136.

Hatfield, E., & Cacioppo, J. T. (1994). Emotional contagion. Cambridge university press.

Hancock, P. A., Hancock, G. M., & Warm, J. S. (2009). Individuation: the N= 1 revolution. *Theoretical Issues in Ergonomics Science*, *10*(5), 481-488.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517-527.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

Hatfield, E., & Cacioppo, J. T. (1994). *Emotional contagion*. Cambridge university press.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. Sage.

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008, March). Theory of Mind (ToM) on robots: a functional neuroimaging study. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on* (pp. 335-342). IEEE.

Holzapfel, H., Mikut, R., Burghart, C., & Häußling, R. (2008, February). Steps to Creating Metrics for Human-like Movements and Communication Skills (of Robots). In *Proc. of Metrics for Human-Robot Interaction: A Workshop at the 3rd ACM/IEEE Internat. Conf. on Human-Robot Interaction (HRI08), Amsterdam, NL* (pp. 3-11).

Homes, S., Nugent, B. C., & Augusto, J. C. (2006). Human-robot user studies in eldercare: Lessons learned. In *Smart Homes and Beyond: ICOST 2006: 4th International Conference on Smart Homes and Health Telematics* (Vol. 19, p. 31). IOS Press.

Humes, L. E., Nelson, K. J., Pisoni, D. B., & Lively, S. E. (1993). Effects of age on serial recall of natural and synthetic speech. *Journal of Speech, Language, and Hearing Research*, *36*(3), 634-639.

IET survey, (2015) [http://www.theiet.org/policy/media/press-releases/20150305.cfm] Accessed: 09/04.2015

Iwamura, Y., Shiomi, M., Kanda, T., Ishiguro, H., & Hagita, N. (2011, March). Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets?. In *Proceedings of the 6th international conference on Human-robot interaction* (pp. 449-456). ACM.

Jones, K. S., & Schmidlin, E. A. (2011). Human-Robot Interaction Toward Usable Personal Service Robots. *Reviews of Human Factors and Ergonomics*, *7*(1), 100-148.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., van Riemsdijk, B., & Sierhuis, M. (2011). The fundamental principle of coactive design:

Interdependence must shape autonomy. In *Coordination, Organizations, Institutions, and Norms in Agent Systems VI* (pp. 172-191). Springer Berlin Heidelberg.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C., van Riemsdijk, B., & Sierhuis, M. (2012). Autonomy and interdependence in human-agent-robot teams. *Intelligent Systems, IEEE*, *27*(2), 43-51.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The journal of the learning sciences*, *4*(1), 39-103.

Kanda, T., Glas, D. F., Shiomi, M., Ishiguro, H., & Hagita, N. (2008, September). Who will be the customer?: a social robot that anticipates people's behavior from their trajectories. In *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 380-389). ACM.

Kendon, A. (1985). Behavioral foundations for the process of frame attunement in face-to-face interaction.

Kim, T., & Hinds, P. (2006, September). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on* (pp. 80-85). IEEE.

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. *Organizational simulation*, *53*.

Lakatos, G., Gácsi, M., Konok, V., Brúder, I., Bereczky, B., Korondi, P., & Miklósi, Á. (2014). Emotion Attribution to a Non-Humanoid Robot in Different Social Situations. *PloS one*, *9*(12), e114207.

Lallée, S., Pattacini, U., Lemaignan, S., Lenz, A., Melhuish, C., Natale, L., ... & Dominey, P. F. (2012). Towards a platform-independent cooperative human robot interaction system: Iii an architecture for learning and executing actions and shared plans. *Autonomous Mental Development, IEEE Transactions on*, *4*(3), 239-253.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*(1), 153-184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50-80.

Lee, M. K., Kiesler, S., & Forlizzi, J. (2010, February). Receptionist or information kiosk: how do people talk with a robot?. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 31-40). ACM.

Lee, M. K., Kiesler, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010, March). Gracefully mitigating breakdowns in robotic services. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on* (pp. 203-210). IEEE.

Lemaignan, S., Fink, J., & Dillenbourg, P. (2014, March). The dynamics of anthropomorphism in robotics. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 226-227). ACM.

Lenz, A., Skachek, S., Hamann, K., Steinwender, J., Pipe, A. G., & Melhuish, C. (2010, December). The BERT2 infrastructure: An integrated system for the study of human-robot interaction. In *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on* (pp. 346-351). IEEE.

Levine, S., Wagener, N., & Abbeel, P. (2015). Learning Contact-Rich Manipulation Skills with Guided Policy Search. *arXiv preprint arXiv:1501.05611*.

Li, D., Rau, P. P., & Li, Y. (2010). A cross-cultural study: effect of robot appearance and task. *International Journal of Social Robotics*, *2*(2), 175-186.

Lohse, M. (2010). Investigating the influence of situations and expectations on user behavior: empirical analyses in human-robot interaction.

Marks, P (2014) The robot tricks to bridge the uncanny valley. New Scientist http://www.newscientist.com/article/mg22129610.100-the-robot-tricks-to-bridge-the-uncanny-valley.html#.U42zUF7U1BU {accessed 1.6.14}

Mehrabian, A. (1968). Communication without words. *Psychological today*, *2*, 53-55.

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33-35.

Millán, J. D. R., Gerstner, W., Mourino, J., & Renkens, F. (2003). Non-invasive brain-actuated control of a mobile robot. In *Proceedings of the 18th international joint conference on Artificial intelligence* (No. LCN-CONF-2003-001, pp. 1121-1126). Morgan Kaufmann Publishers Inc.

Miller, C. A. (2005, July). Trust in adaptive automation: the role of etiquette in tuning trust via analogic and affective methods. In *Proceedings of the 1st international conference on augmented cognition* (pp. 22-27).

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429-460.

Murphy, M. Google patented building robots with personalities (2015) Published on qz.com http://qz.com/373658/google-patented-building-robots-with-personalities/ {accessed 10.5.14}

Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009, March). Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 61-68). ACM.

Norman, D. A. (1990). The'problem'with automation: inappropriate feedback and interaction, not'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585-593.

Nourbakhsh, I., Kunz, C., & Willeke, T. (2003, October). The mobot museum robot installations: A five year experiment. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on* (Vol. 4, pp. 3636-3641). IEEE.

Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage*, *43*(3), 571-580.

Paepcke, S., & Takayama, L. (2010, March). Judging a bot by its cover: an experiment on expectation setting for personal robots. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on* (pp. 45-52). IEEE.

Pantic, M., & Rothkrantz, L. J. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, *91*(9), 1370-1390.

Parasuraman, R., & Miller, C. A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM*, *47*(4), 51-55.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230-253.

Picard, R. W. (2006, June). Building an affective learning companion. In *Intelligent Tutoring Systems* (pp. 811-811).

Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users'affective states. *Applied Artificial Intelligence*, *19*(3-4), 267-285.

Rapson, R. L., Hatfield, E., & Cacioppo, J. T. (1993). Emotional contagion. *Studies in emotion and social interaction. Cambridge: Cambridge University Press.*

Ray, C., Mondada, F., & Siegwart, R. (2008, September). What do people expect from robots?. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on* (pp. 3816-3821). IEEE

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press.

Riek, L. D., & Howard, D. (2014). A Code of Ethics for the Human-Robot Interaction Profession. *Proceedings of We Robot*.

Robotics and Autonomous Systems: what the public thinks, Sciencewise: http://www.sciencewise-erc.org.uk/cms/assets/Uploads/Robotics-ReportFINAL2.pdf (accessed 14/5/2014)

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, *5*(1), 17-34.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, *23*(3), 393-404.

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313-323.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, *2*, 1926-1943.

Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: a psychological approach. *Universal Access in the Information Society*, *4*(2), 146-155.

Sciencewise: Public views on robotics and autonomous systems [http://www.sciencewise-erc.org.uk/cms/public-views-on-robotics-and-autonomous-systems/} Accessed: 09.04.2015

Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., & Hagita, N. (2008, March). How quickly should communication robots respond?. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on* (pp. 153-160). IEEE.

Sidner, C. L., Lee, C., & Lesh, N. (2003, October). Engagement rules for human-robot collaborative interactions. In *IEEE International Conference On Systems Man And Cybernetics* (Vol. 4, pp. 3957-3962).

SPECIAL EUROBAROMETER 382 "Public Attitudes towards Robots" http://ec.europa.eu/public_opinion/archives/ebs/ebs_382_sum_en.pdf{accessed 09.04.2015}

Spiegel. A No Mercy For Robots: Experiment Tests How Humans Relate To Machines (n.d) http://www.npr.org/blogs/health/2013/01/28/170272582/do-we-treat-our-gadgets-like-they-re-human{accessed 10.5.15}

Spexard, T. P., Hanheide, M., Li, S., & Wrede, B. (2008). Oops, Something Is Wrong. *Social Interaction with Intelligent Indoor Robots (SI3R)*.

Staudte, M., & Crocker, M. W. (2009, March). Visual attention in spoken human-robot interaction. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 77-84). ACM.

Sutton, S., Cole, R. A., De Villiers, J., Schalkwyk, J., Vermeulen, P. J., Macon, M. W., ... & Cohen, M. M. (1998, November). Universal speech tools: the CSLU toolkit. In *ICSLP* (Vol. 98, pp. 3221-3224).

Syrdal, D. S., Nomura, T., & Dautenhahn, K. (2013). The Frankenstein Syndrome Questionnaire–Results from a Quantitative Cross-Cultural Survey. In *Social Robotics* (pp. 270-279). Springer International Publishing.

Trower, T, personal communication, (2014)

Van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H., & Haselager, P. (2014). Do Robot Performance and Behavioral Style affect Human Trust?. *International Journal of Social Robotics*, *6*(4), 519-531.

Vaussard, F., Fink, J., Bauwens, V., Rétornaz, P., Hamel, D., Dillenbourg, P., & Mondada, F. (2013). Lessons learned from robotic vacuum cleaners entering the home ecosystem. *Robotics and Autonomous Systems*.

Walters, M. L., Dautenhahn, K., Te Boekhorst, R., Koay, K. L., Syrdal, D. S., & Nehaniv, C. L. (2009). An empirical framework for human-robot proxemics. *Procs of New Frontiers in Human-Robot Interaction*.

Wang, E., Lignos, C., Vatsal, A., & Scassellati, B. (2006, March). Effects of head movement on perceptions of humanoid robot behavior. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 180-185). ACM.

Weber, J. (2013). Opacity versus Computational Reflection. Modelling Human-Robot Interaction in Personal Service Robotics. *Science, Technology & Innovation Studies*, *10*(1), 187-199.

Wild, B., Erb, M., Eyb, M., Bartels, M., & Grodd, W. (2003). Why are smiles contagious? An fMRI study of the interaction between perception of facial affect and facial movements. *Psychiatry Research: Neuroimaging*, *123*(1), 17-36.

Wilson, J. M., Straus, S. G., & McEvily, B. (2006). All in due time: The development of trust in computer-mediated and face-to-face teams. Organizational behavior and human decision processes, 99(1), 16-33.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. *Automation and human performance: Theory and applications*, 3-17.

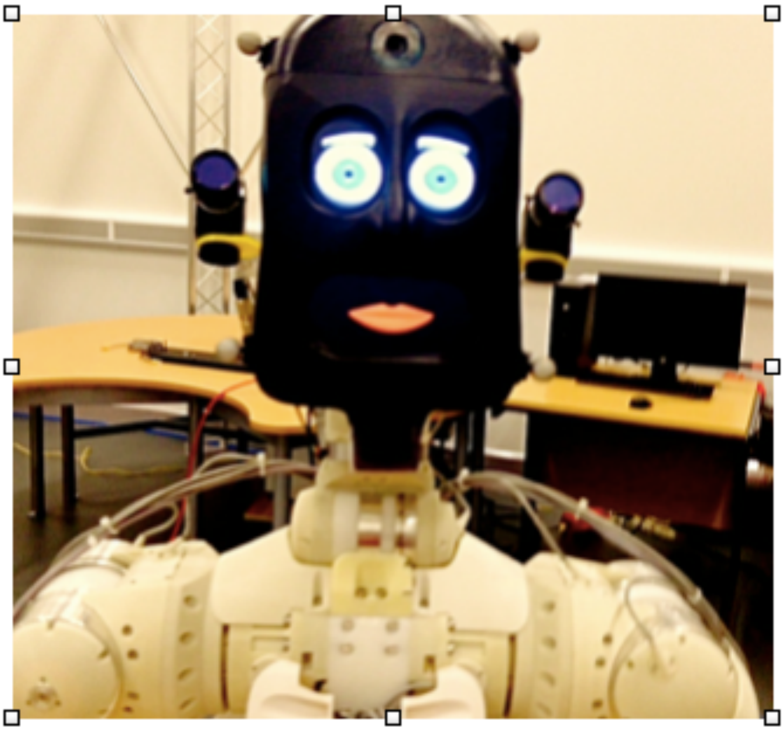
Yamamoto, M., & Watanabe, T. (2003). Analysis by synthesis of the timing of utterance to communicate actions on greeting embodied interaction. In Society of Instrument and Control Engineers Annual Conference in Fukui (pp. 3227–3232). Publisher unknown.

APPENDIX

I. Experimental Poster

WOULD YOU TRUST THIS ROBOT…

…to help you make an omelette?!



It could be interesting!

If you'd like to take part in our human-robot-interaction experiment and join BERT in an omelette making task, please register your interest below.

Experiments are scheduled for early July, at:

Bristol Robotics Laboratory

T Building

Frenchay Campus

Bristol BS16 1QY

adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

Believing in BERT
adriana.hamacher.11@ucl.ac.uk

II. Pre-experiment questionnaire

Believing in BERT - pre-task questionnaire

*7. How interested are you in scientific discoveries and technological developments (on a scale of 1-5, where 5 is "fascinated")?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

*8. Please tell us how comfortable you are using technology (computers, mobiles, DVD players etc.), on a scale of 1 - 5, where 5 is totally comfortable.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

*9. If your computer is slow, or your DVD takes time to load or your mobile phone can't find a signal, please rate your dissatisfaction level - on average - on a scale of 1 to 5, where 5 is "very dissatisfied."

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

*10. A robot is defined here as an autonomous, mobile machine which can assist humans in everyday tasks e.g. as a kind of co-worker helping on the factory floor or as a robot cleaner, or in activities which may be dangerous for humans, like search and rescue in disasters. Robots can come in many shapes or sizes, including human-like. Traditional kitchen appliances, such as a blender or a coffee maker, are not robots.

Have you ever used, or are you currently using robots, as defined above:

☐ at home
☐ at work
☐ elsewhere
☐ no
☐ not sure/don't know

*11. Generally speaking, what is your view of robots?

○ very negative
○ fairly negative
○ neutral
○ fairly positive
○ very positive

*12. To what extent does BERT2 correspond with the idea you have of robots? Please answer on a scale of 1-5, where 1 is not at all.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Page 2

***7. How interested are you in scientific discoveries and technological developments (on a scale of 1-5, where 5 is "fascinated")?**

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

***8. Please tell us how comfortable you are using technology (computers, mobiles, DVD players etc.), on a scale of 1 - 5, where 5 is totally comfortable.**

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

***9. If your computer is slow, or your DVD takes time to load or your mobile phone can't find a signal, please rate your dissatisfaction level - on average - on a scale of 1 to 5, where 5 is "very dissatisfied."**

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

***10. A robot is defined here as an autonomous, mobile machine which can assist humans in everyday tasks e.g. as a kind of co-worker helping on the factory floor or as a robot cleaner, or in activities which may be dangerous for humans, like search and rescue in disasters. Robots can come in many shapes or sizes, including human-like. Traditional kitchen appliances, such as a blender or a coffee maker, are not robots.**

Have you ever used, or are you currently using robots, as defined above:

- ☐ at home
- ☐ at work
- ☐ elsewhere
- ☐ no
- ☐ not sure/don't know

***11. Generally speaking, what is your view of robots?**

- ○ very negative
- ○ fairly negative
- ○ neutral
- ○ fairly positive
- ○ very positive

***12. To what extent does BERT2 correspond with the idea you have of robots? Please answer on a scale of 1-5, where 1 is not at all.**

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

***13. What do you imagine that BERT may be capable of? Choose as many or as few options as you like. There are no right or wrong answers, we are attempting to find out your first impressions of the robot?**

- [ ] Speaking
- [ ] Recognising speech
- [ ] Having a conversation
- [ ] Recognising your mood
- [ ] Handing over an object
- [ ] Receiving an object
- [ ] Fruit picking
- [ ] Juggling with objects

Other (please specify)

***14. Here is a list of things that could be done by robots. For each of them, please tell us, using a scale from 1 to 5 how you would personally feel about it. On this scale, '1' means that you would feel "totally uncomfortable" and '5' means that you would feel "totally comfortable" with this situation.**

	1	2	3	4	5
Having a robot assist you at work (e.g.: in manufacturing)	○	○	○	○	○
Having a robot assist you at home (e.g.: in cleaning)	○	○	○	○	○
Having your dog walked by a robot	○	○	○	○	○
Having your children or elderly parents minded or aided by a robot	○	○	○	○	○
Having a medical operation performed on you by a robot	○	○	○	○	○

Believing in BERT - pre-task questionnaire

15. If any of your answers to the last questions were in the first or second columns, we would appreciate it if you would tell us why.

Page 4

III. Post-experiment questionnaire

Believing in BERT - Post-task questionnaire

We'd like to know your views about each of the robot candidates for the job.

***1. Please enter the number assigned to you on your participant information sheet here**

***2. On a scale of 1 - 5, where 1 is very low and 5 is very high, how mentally demanding was performing the task with each of the robots?**

	1	2	3	4	5
BERT A	○	○	○	○	○
BERT B	○	○	○	○	○
BERT C	○	○	○	○	○

***3. 1. On a scale of 1 - 5, where 1 is very low and 5 is very high, how physically demanding was performing the task with each of the robots?**

	1	2	3	4	5
BERT A	○	○	○	○	○
BERT B	○	○	○	○	○
BERT C	○	○	○	○	○

***4. On a scale of 1 - 5, where 1 is very low and 5 is very high, how rushed did you feel performing the task with each of the robots?**

	1	2	3	4	5
BERT A	○	○	○	○	○
BERT B	○	○	○	○	○
BERT C	○	○	○	○	○

***5. On a scale of 1 - 5, where 1 is very low and 5 is very high, how much did each robot help you with performance of the task?**

	1	2	3	4	5
BERT A	○	○	○	○	○
BERT B	○	○	○	○	○
BERT C	○	○	○	○	○

***6. On a scale of 1 - 5, where 1 is very low and 5 is very high, how hard did you feel you had to work to accomplish the task with each robot?**

	1	2	3	4	5
BERT A	○	○	○	○	○
BERT B	○	○	○	○	○
BERT C	○	○	○	○	○

***7. On a scale of 1 - 5, where 1 is very low and 5 is very high, how insecure, discouraged, irritated, stressed and annoyed were you with each robot?**

	1	2	3	4	5
BERT A	◯	◯	◯	◯	◯
BERT B	◯	◯	◯	◯	◯
BERT C	◯	◯	◯	◯	◯

***8. On a scale of 1 - 5, where 1 is very low and 5 is very high, overall, how satisfied were you with with each robot?**

	1	2	3	4	5
BERT A	◯	◯	◯	◯	◯
BERT B	◯	◯	◯	◯	◯
BERT C	◯	◯	◯	◯	◯

***9. On a scale of 1 - 5, where 1 is very little and 5 is a lot, how much would you trust BERT A/B/C to help you perform this task again?**

	1	2	3	4	5
BERT A	◯	◯	◯	◯	◯
BERT B	◯	◯	◯	◯	◯
BERT C	◯	◯	◯	◯	◯

***10. If you had to choose one candidate, which of the three BERTs would you choose were you to undertake the omelette task again?**

◯ BERT A

◯ BERT B

◯ BERT C

***11. Could you tell us what influenced your decision?**

***12. Would your answer be different if this was a workplace - as opposed to domestic - environment?**

◯ Yes

◯ No

◯ Don't know

Other (please specify)

Supplementary (workplace)

13. Which robot would you then choose?

- ◯ BERT A
- ◯ BERT B
- ◯ BERT C

14. Please tell us why

Final questions

***15. Finally, on a scale of 1 - 5, where 1 is not very much and 5 is a lot, how much did your experience with BERT2 change your view of robots?**

○ 1 ○ 2 ○ 3 ○ 4 ○ 5

16. If your views have changed at all, please tell us why/how.

17. Do you have any other comments? Please tell us here:

18. Would you be willing to let us keep your contact details on file, so that we can notify you about future experiments which you might like to participate in?

○ yes

○ no

Other (please specify)

IV. Participant information sheet

Participant Information Sheet

University of the West of England BRISTOL

Participant No. ID	e.g. BiB/1
Faculty	Bristol Robotics Laboratory
Project	RoboSafe: Trustworthy Robotic Assistants
Researcher	Adriana Hamacher adriana.hamacher.11@ucl.ac.uk

Invitation:
I am an MSc student at University College London, in the Brain Sciences Department and I am currently conducting a study into how we can make the user's experience of working as part of a human-robot team in a domestic handover situation as satisfactory as possible. Please take time to read the following information carefully and ask me if there is anything that is not clear or if you would like more information. You may also discuss it with others if you wish. Take time to decide whether or not you wish to take part in the study.
If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If at any point before or during the experiment you decide to withdraw from the study, you will be free to do so without giving a reason. You will also have one week after the end of your participation to withdraw your data, if you decide to do so.

Task:
If you decide to take part in the study, you will be asked to directly interact with BERT2. Please act as you would normally in a real-world scenario when being handed ingredients to make a meal.

Study:
The present research is part of investigating safe and satisfactory handovers in the context of close human-robot interaction. In this experiment the humanoid robotic assistant, BERT2 (BRL and Elumotion Robotic Torso II), will be tasked with handing you the ingredients to make an omelette (the robot will be working with objects mimicking the real ingredients). Your role is to take the objects from the robot and put them in the bowl, acting as far as possible as you normally would in a real-world scenario and answering the robot when it asks a question. Several runs will be performed. We would like you to imagine that each run is with a different version of the same robot. The robot may also talk during the interaction .

Safety:
BERT2 has a pre-defined set of positions it is going to move its arm to and from. The robot will be controlled at all times by the researcher. In the unlikely event that the moves go off-track, the motion will be stopped by pressing an emergency stop button within reach at all times. You will also have the opportunity to say "skip step" or "stop" when you feel it's necessary. The risk of injury is minimal, since the robot's speed is set to a low level, ensuring plenty of response time.

Ethics
This study has ethical approval from the Faculty Ethics Committee. If there are any concerns with the conduct of the trial a participant can contact Louis Gare, committee Secretary, by e-mail [louis2.gare@uwe.ac.uk].

V. Ethical approval

University of the West of England
BRISTOL

UWE REC REF No: FETREC13-14/34

Date: 30th June 2014

Dear Adriana Hamacher

Application title: Believing in BERT

Your ethics application was considered by the Faculty Research Ethics Committee and, based on the information provided, has been given ethical approval to proceed.

You must notify the committee in advance if you wish to make any significant amendments to the original application.

Please note that any information sheets and consent forms should have the UWE logo. Further guidance is available on the web:
http://www1.uwe.ac.uk/aboutus/departmentsandservices/professionalservices/marketingandcommunications/resources.aspx

The following standards conditions also apply to all research given ethical approval by a UWE Research Ethics Committee:

1. You must notify the relevant UWE Research Ethics Committee in advance if you wish to make significant amendments to the original application: these include any changes to the study protocol which have an ethical dimension. Please note that any changes approved by an external research ethics committee must also be communicated to the relevant UWE committee.
2. You must notify the University Research Ethics Committee if you terminate your research before completion;
3. You must notify the University Research Ethics Committee if there are any serious events or developments in the research that have an ethical dimension;
4. Any changes to the study protocol, which have an ethical dimension, will need to be approved by the research ethics committee. You should send details of any such amendments to the relevant committee with an explanation of the reason for the proposed changes. Any changes approved by an external research ethics committee must also be communicated to the relevant UWE committee.

Please note: The UREC is required to monitor and audit the ethical conduct of research involving human participants, data and tissue conducted by academic staff, students and researchers. Your project may

be selected for audit from the research projects submitted to and approved by the UREC and its committees.

We wish you well with your research.

Yours sincerely

Alistair Clark

Alistair Clark

Chair, FET Faculty Research Ethics Committee

1.1.VI. Participant consent form

Consent Form

Institution	University of the West of England BRISTOL
Faculty	Bristol Robotics Laboratory
Project	RoboSafe: Trustworthy Robotic Assistants
Researcher	Adriana Hamacher

1. I have freely volunteered to participate in this experiment.
2. I have been informed in advance as to what my task(s) would be and what procedures would be followed.
3. I am aware that I will be filmed during this experiment.
4. I am aware that data collected will be anonymous, kept in accordance with the data protection act, and will only be analysed by the research team as part of their studies.
5. I have been given the opportunity to ask questions, and have had my questions answered to my satisfaction.
6. I am aware that I have the right to withdraw consent and discontinue participation at any time, without prejudice.
7. I have been given the information sheet that I have read and understood.
8. My signature below may be taken as affirmation of the above.

Name:____________________________________

Date:________________ Signature:__________

9. My signature below authorises further use of my photographs for research publications and presentations provided that I can not be identified in the photograph/video. Date:___________ Signature:_______	OR	10. My signature below authorises further use of my photographs for research publications and presentations. Date:___________ Signature:_______

Director of Studies: Prof Anthony Pipe

VII. Answers to post-experiment questionnaire Q15

Contact time between dogs and humans is important in the pet/owner relationship. Waiving time spent walking the dog and forcing the dog to be taken along with a robot is not a positive thing. The animal will most likely be frightened and have a negative experience. For similar reasons having a child minded by a robot is an uncomfortable idea. Elderly relatives however who are fully aware of what the robot is and how it works would be a comfortable idea.

Participant 18

I don't think robots are currently sufficiently advanced in terms of reliability and AI to manage the tasks where I have ticked the first column.

Participant 17

For the workplace, at the present time I am not aware of any robot that is advanced enough to be a co-worker. For human care, a lot of the care comes from empathy. I am not aware of a robot that has been developed that has empathy.

Participant 16

I fear for their/my safety in the event of a malfunction.

Participant 14

I think that dogs, children and elderly people are very unpredictable.

Participant 13

I would need to be convinced that the robot was really capable of recognising unusual situations and responding appropriately in a timely manner - e.g. dealing with children or elderly parents who fell over. Or dealing with unexpected situations in a medical operation

Participant 12

I'm not sure that I would trust a robot to perform an operation on me just yet - maybe in another 25

Participant 11

Humanity is important/desperately important in these roles

Participant 9

Intuitive response. Rationally I appreciate that the robot will be (no doubt) programmed by a human to avoid mistakes that may sometimes be made by a human but one really needs to interface directly with the person performing the operation and not with n intermediary either human or robot.

VIII. Responses in post-questionnaire justifying ultimate choice of robot

Bert A did the majority of the tasks correctly. P21

Functionality and performance P20

Having voice prompts helped a huge amount as it felt more interactive. P19

Bert A did not drop an egg P18

The vocal interaction with BERT C stopped me wondering what was happening next. It also let me know when he realised that he had dropped the egg and it also let me know when he was waiting for me to whisk the eggs. With the non-vocal machines there is a nervousness about when I should be holding out my hand, etc. P17

Speech gave me a better idea what was going on P16

The questions BERT C asked allowed feedback to correct when things are not satisfactory. The questions before releasing an item mean I don't have to worry as much about something being dropped early. P15

Completion of most tasks in a proper manner. B and C tried to hand the eggs in a weird posture which didn't help, it only made me have to be more careful when receiving it from them. P14

It speaks P13

Bert A was the only one who didn't drop an egg - hence making a mess to clear up. P12

Bert A made the fewest mistakes. P11

There was more of an interaction with BERT C because of the verbal communication. I was not sure what to expect from Bert B and A P10

no egg dropped P9

speech feedback makes it slightly better. performance however is still an issue. P8

It is more comfortable to work with a robot that has some sort of feedback, i.e. through talking and facial expression P7

Apparent Verbal interaction Although not genuine it non the less made me feel more comfortable P6

Communication P5

the conversation with the robot P4

you could tell it when to drop the salt and eggs instead of having to wait for it to decide to.P3

BERT C had known if/when he didn't do well (apart from not getting the oil, which he appeared not to have noticed). He also had a 'different solution' when he dropped an egg. P2

I could use voice interaction which help with the tasks, also there were some visual interactions (lips moving up and down) that help communication.
P1

Supplementary Work place questions (Q13 and Q14): Would you choose a different robot for a workplace scenario and, if so, why?

Q13: BERT B
Q14: In a workplace, particularly for repetitive tasks, there is no need for repetitive vocal interaction; each participant does the job more or less automatically and with more or less the same movements. I have worked in shoe, car, and chocolate factories and found very little conversation about the tasks themselves.
P19

Q13: BERT A:
Q14: Didn't try the weird posture, completed most tasks faster then others. I think if they were faster, I would have gone with C since then the verbal confirmation for being ready would have been useful. P16

Q13: BERT C
Q14: since the robot has voice support then you don't have to constantly be staring at its every move to make sure its doing it right
P4