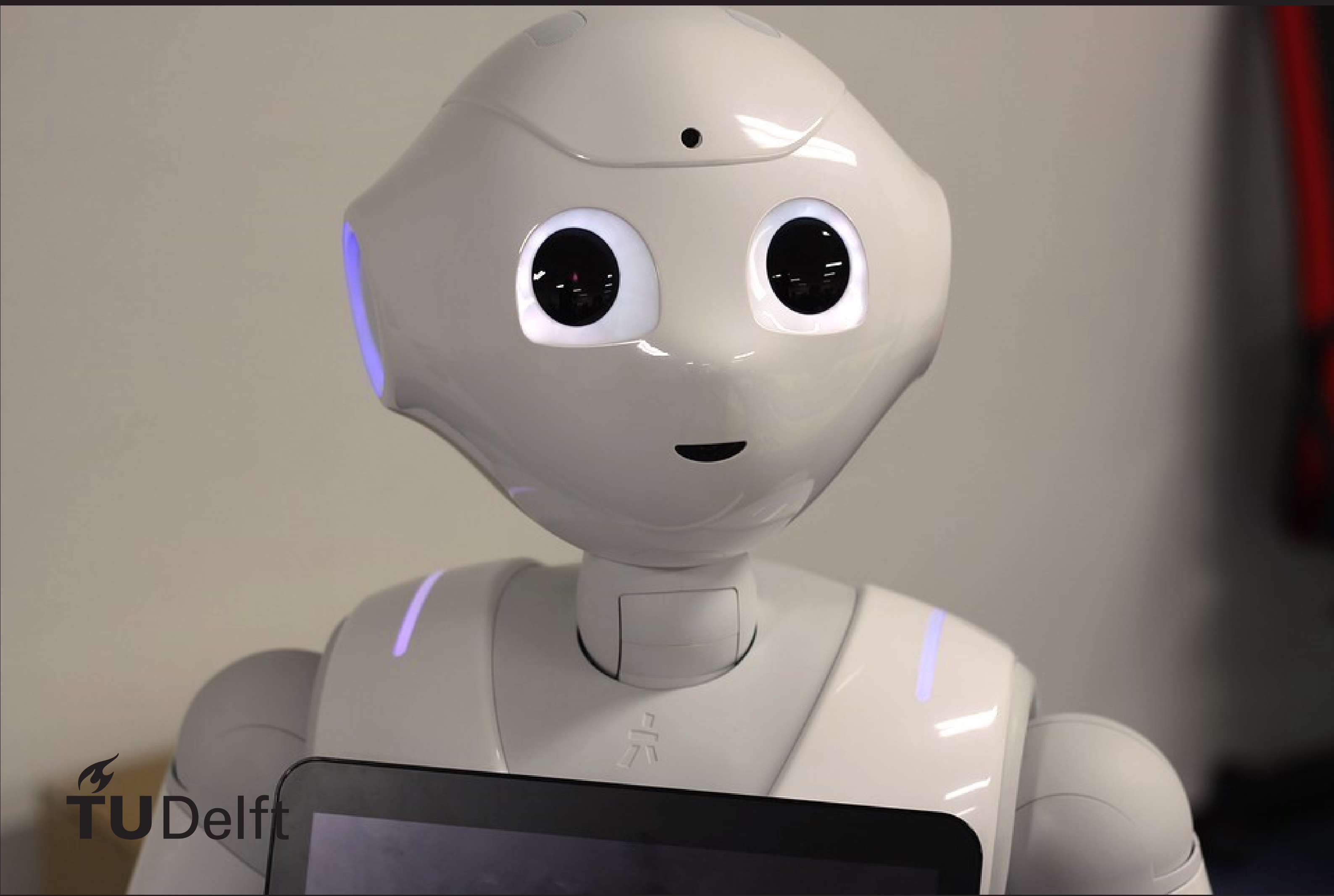
# Interaction Design Patterns for Social Robot Assistance of Moral Decisions in Healthcare: Confrontation with Resuscitation Dilemmas

Sujit Shankar Jaishankar

4779657

Master Thesis submitted for Master of Science - Computer Science (Data Science)

September 2020



TU Delft

# Interaction Design Patterns for Social Robot Assistance of Moral Decisions in Healthcare: Confrontation with Resuscitation Dilemmas

by

Sujit Shankar Jaishankar

to obtain the degree of Master of Science Computer Science (Data Science Track)
at the Delft University of Technology,
to be defended publicly on Tuesday September 29, 2020 at 09:00 AM.
Interactive Intelligence Group
Faculty of EEMCS

| | | |
|---|---|---|
| Student number: | 4779657 | |
| Project duration: | November 15, 2019 – September 29, 2020 | |
| Thesis committee: | Prof. dr. M.A. (Mark) Neerincx, | TU Delft, supervisor |
| | Dr. Catharine Oertel, | TU Delft, Interactive Intelligence |
| | Dr. F. (Filippo) Santoni de Sio, | TU Delft, Ethics of Technology |

*This thesis is confidential and cannot be made public until September 29, 2020.*

# Preface

As the world moves into more sophistication with vastly superior technology, resulting in greater number of robots being deployed for use in everyday life, there is a need to shape their roles in various fields, specifically in hospitals. How can robots help doctors in ethical decision making? The main objective of the research was to determine if verbal and gestural cues demonstrated by a robot can affect ethical decision making and based on the extent to which these cues are applied. To do this, design patterns were developed and implemented in the form of three different interaction scenarios between the robot and the doctor (actor) with varying degree of cues. A questionnaire was also created and validated that measured the ethical aspects of the interaction.

The seeds for the project were sown in my first year in the Interactive Intelligence course. At that moment, the exact specifics of the thesis were unclear, but it was clear that it involved Human-Robot Interaction and people's perception. Several months later, I was able to provide shape to my ideas with the help of my supervisor. This thesis is the result of my graduation project for the completion of the Master Degree in Computer Science with a healthy dose of my passion for technology and data science. All of the work presented henceforth was conducted with the help of the Insight Laboratory at the Delft University of Technology. All associated methods and data concerns were approved by the Ethics Committee before the experiment was carried out. The thesis is original, unpublished, original work by the author.

I want to thank my supervisor Prof. dr. M.A. (Mark) Neerincx, who was readily available to provide advice and support, Without his valuable guidance this thesis would not have been possible; from conception unto completion. Regular meetings with him not only provided me with sufficient auxiliary knowledge to carry out my thesis but also eased my concerns regarding its successful implementation. I would like to thank Dr. F. (Filippo) Santoni de Sio for providing valuable feedback on the ethical dilemma discussed in the thesis and the issues of delegation and trust, and Dr. Catharine Oertel for being part of the thesis committee.

My wholehearted thanks to Elie Saad at Interactive Intelligence for not only allowing me to use and test the Interaction Design Tool but also guiding me with taking the experiment live. The tool made it easy to translate the design patterns into active behaviours within Pepper. His guidance with setting up the experiment on the online platforms (MTurk, Qualtrics) helped expedite the process. A special mention to Ruud de Jong, who provided technical support and timely access to the Insight lab; especially during the measures taken due to the global pandemic.

I must express my immense gratitude to my family for providing me with unwavering support, encouragement and positivity during these two years of study. This achievement is as much a part of theirs as it is mine. Special thanks to my friends for their words of comfort and wisdom. Without your support, this thesis would not have been possible. I look back with contentment at the last two years: an intense process with lots of positive and negative experiences. In the end, I'm glad I went through them and believe they have made me all the more wiser.

*Sujit Shankar Jaishankar*
Delft, September 2020

# Abstract

The implementation of social robots in the healthcare industry is becoming substantial as a consequence of the scarcity of healthcare professionals, rising costs of healthcare and an increase in the number of vulnerable populations. Social robots will be deployed, in increasing numbers, in assisting health care professionals during the provision of care to patients. While research in Human-Robot-Interaction (HRI) has investigated mechanisms for making decisions taken by artificial agents more ethical, there is limited work done in investigating adaptations to HRI that promotes ethical behaviour on the human-side. Moral dilemmas can appear when decisions have to be made regarding patient care. Dealing with them can be challenging for healthcare professionals since the impact of decisions affect multiple parties and involves different considerations and value-trade-offs of the involved people. There are different approaches in which healthcare professionals can be confronted with moral dilemmas. Based on decision making principles, a robot might assist the health care professionals in ethical decision making with an appropriate reflection or confrontation of the dilemma. This thesis investigates how a social robot can confront a professional with a moral dilemma and what the associated effects are. Following the Socio-Cognitive Engineering (SCE) method, two alternative proto-patterns are created for dilemma confrontation by a robot in a resuscitation scenario: Verbal and Multi-modal confrontation. In such a scenario, the robot displays protests of distress and affect in order to help the professionals carry out systematic reflection of the moral dilemmas they experience. In an evaluation, it is tested whether these confrontation patterns are being perceived as intended, comparing the two confrontation patterns with a neutral scenario.

# Contents

# I

# INTRODUCTION

# Introduction

The implementation of social robots in the healthcare industry is becoming substantial as a consequence of the scarcity of healthcare professionals, rising costs of healthcare and an increase in the number of vulnerable populations. Social robots will be deployed, in increasing numbers, in assisting health care professionals during the provision of care to patients. While research in Human-Robot-Interaction (HRI) has investigated mechanisms for making decisions taken by artificial agents more ethical, there is limited work done in investigating adaptations to HRI that promotes ethical behaviour on the human-side. The variability in achieving objectives can transpire during safe human-robot interactions (HRI) or be a segment of ethically questionable scenarios involving moral dilemmas. It is ambiguous how such HRIs will pan out, especially in situations involving the healthcare of a human. Moral dilemmas can appear when decisions have to be made regarding patient care.

Human-Robot Interaction in the health sector is confronted by a myriad of challenges such as acceptability, appropriateness, safety, privacy and fear of replacement of caregivers by robots. Dealing with them can be challenging for healthcare professionals since the impact of decisions affect multiple parties and involves different considerations and value-trade-off. There are different approaches in which healthcare professionals can be confronted with moral dilemmas. Based on decision making principles and decision theories, a robot might assist the health care professionals in ethical decision making with an appropriate reflection or confrontation of the dilemma.

*Ethics in Robots:* In order for a robot to be considered suitable for implementation in real-world scenarios, it is vital for the robot to possess three important capabilities: (1) ability to perceive and infer current world states, (2) evaluate and make decisions about ethical acceptability of situations and (3) adapt the HRI to promote ethical behaviour [13, 14]. While much of the work in the field of HRI Ethics has been mainly focused on developing the second capability, the work mentioned here concerns itself with the third capability, especially through the means of affective displays. Working on the confrontation methods set out by Briggs et.al [14], it is examined here how dissuasive and persuasive efforts by robots using affective cues can influence human actions and perceptions of robot behavior. Studying their effects can help develop methods that can facilitate ethical HRI.

*Robot Confrontation:* How should a social robot aid in decision making involving ethically-charged situations in the healthcare industry? If the robot believes the action to be unethical, it can refuse to carry out the order. However, that may not be enough for the human agent to reflect on their actions. Robot systems in the future will require human override capabilities and can prevent valuable feedback from the robot. Affective confrontation can provide valuable feedback concerning the ethical dilemma the human faces. Research has shown that confrontation through affective means is convincing enough to cause changes in decision making [74, 85]. Meanwhile, the ability to persuade humans has also been studied [1, 12]. Factors that can enhance these approaches involve the 'humanness' of the robot and the agency ascribed to the robot [72]. Presence of affect in verbal cues and physical displays to denote human emotions can intensify the influence of persuasive and dissuasive methods in HRI.

This thesis investigates how a social robot can confront a doctor with a moral dilemma and what the perceived effects are. Following the socio-cognitive engineering (SCE) method, two alternative proto-patterns for a dilemma confrontation by a robot in a resuscitation scenario are created: Verbal and Multi-modal based confrontation. In such a scenario, the robot displays protests of distress and affect in order to help the professionals carry out systematic reflection of the moral dilemmas they experience. In an evaluation, it is tested whether these confrontation patterns are being perceived as intended, comparing the two confrontation patterns with a neutral scenario.

The research questions that frame the structure of the thesis are:

- *How can a humanoid robot help a doctor to reflect on a moral dilemma in patient care using verbal and multi-modal cues (in this case it concerns the ethical dilemma of resuscitation)?*
- *Does the addition of these cues lead to higher likeability, perceived animacy, and perceived intelligence?*

- *Is the effectiveness in reflection for ethical decision making higher, when the robot expresses these cues during the confrontation?*
- *Which aspects of the robot interaction, as perceived by the human, can explain his or her support experience?*

The report details the three components: Foundation, Evaluation and Specification using the SCE methodology. The design process in Human-Robot Interaction (HRI) is iterative and dynamic. A complete design mandates a systematic theory-driven process involving exploration, specification and refinement of design variables. The Foundation provides the underlying structure to develop and refine the design process. The Foundation II deals with Operational Demands, Human Factors and Technological Principles. In Chapter 1 information regarding the environment, activities, stakeholders and personas are provided. It also includes a structured view into earlier work in this field and presents the basis for the HRI research done.

This is followed by descriptions of the problem and design scenarios. Chapter 2 on Human Factors provides detailed information on the ethical dilemma of resuscitation, prior work regarding ethics in HRI, use of robots in healthcare and their effect on decision making based on the various displays of protests. This provides a structured view into earlier work in this field and presents the basis for the HRI research done. The Foundation acts as an extended section that includes earlier work, the ethical dilemma discussed, problem and design scenarios, the descriptions of the environment and the role of the robot.

Specification III explains the design patterns, objectives, use cases and the prototype and its requirements along with the claims made. Chapter 3 sequentially describes the design patterns employed in the thesis. These design patterns provide structure, context, validation, applicability and dialogue structure for each of the three confrontation methods. A fourth design pattern structures the use of reflection in the interaction scenarios. Chapter 4 explains the use cases (for each confrontation method) and the associated objectives. Chapter 5 contains the description of the implemented prototype, including information on the framing of the dialogue and the design with the help of the Interaction Design Tool [73]. The tool helps develop design patterns iteratively allowing for easy testing of robot behaviour - in this case the ethical dilemma of resuscitation. This is followed up by Chapter 6, that states the requirements the robot needs to adhere to. The claims are research objectives that the interaction scenario aims to achieve.

Furthermore, the Evaluation IV discusses the experiment method followed by analysis and results. In Chapter 7, the experiment method and hypotheses are defined. This includes the independent and dependent variables, the questionnaires and experiment procedure. A study is then conducted wherein participants watch the confrontation methods and their responses and opinions are recorded using questionnaires. These questionnaires are concerned with evaluating the effectiveness of each robot and involves robot agency, usefulness, scalability and effectiveness of the cues implemented. This is followed by Chapter 8 that analyses the perception and performance of the robot in each confrontation method. They are tested using statistical tests such as ANOVA and Factor Analysis. Conclusion V summarises the work done, contributions made and suggests future work. Finally, the Appendix VI provides additional information such as the Ontology and the dialogues used in the interaction scenarios.

# II

# FOUNDATION

# 1

# Operational Demands

## 1.1. Environment and Activities

The doctor is situated in the office where reports on the patients are provided. The vitals and status reports are measured and collected using medical equipment and abnormalities found in the reports are indicated to the doctor. Medical equipment such as heart rate monitors, of course, are capable of alarming the doctor when the patient's condition might worsen. The doctor is the one responsible for decision making regarding the patient's well being. The doctor takes the input of the nurses and other specialists before deciding the treatment procedure. Social assistant robots can assist with such decision making. While hospitals make use of assistive robots for surgical operations, having robots that act as assistants while also aiding in decision making are not seen in hospitals owing to ethical implications, public perception and governmental laws. Some hospitals make use of assistant robots for medicine administration, cognitive improvement and emotional support. While they are capable of providing information, they do not provide any support or feedback for decision making. Robots capable of support for ethical decision making are termed as Decision Assistant Robots (DAR).

The doctor is located in a room or office in the hospital. The environment provides an interactive space for the doctor and the Decision Assistant Robot (DAR). The doctor would be seated at their desk while the DAR works along in the room. The patient's vitals are regularly checked upon by the robot and in case of abnormalities, the robot asks the robot for an action plan. The robot will provide options for course of action, but the final decision lies with the doctor. Relatives of the patient are also allowed to visit and the robot can help the doctor decide in matters involving relatives' requests. The robot provides health and status reports on the patients to the doctor and helps with the course of action to be taken. The doctor adheres to the Code of Ethics [3, 69] that are followed by medical professionals. The Code of Ethics specifies the rights that the patients have to their choice of treatment or lack thereof. It mandates the duties of the medical professionals that must be undertaken to ensure saving of lives unless directed otherwise.

### 1.1.1. Types of Robot activities

The role of the Decision Assistant Robot (DAR) is to observe the patient condition and provide status updates to the doctor. The DAR helps the doctor reflect on the decisions made by carrying out displays of protest related to the treatment procedures - whether curative or palliative. The interaction under consideration here is between the doctor and the DAR and involves the patient and relatives as subjects of discussion. The robot takes part in the interaction with the following functionalities:

1. Monitors health of the patient.
2. Provides status reports to the doctor.
3. Provides information to help with decision making.
4. Asks the doctor if they are sure of their decision.
5. Displays various levels of disagreement or disapproval.

## 1.2. Stakeholders and Personas

An overview of direct and indirect stakeholders involved in the Human-Robot Interaction scenario:

### 1.2.1. HA01: Patient

*Background:*
The patient is a direct stakeholder and is affected by the decisions taken by the doctor and the Decision Assistant Robot (DAR). In this case, the patient suffers from a terminal disease (cancer). Patients in this stage require constant monitoring of vitals and daily care. The patient is under hospice and may not respond to stimuli. These patients suffer from sudden and serious complications and are attended to by medical professionals. Patients are allowed to be visited by their relatives for short time intervals. The patient may choose to continue treatment, deny treatment or opt for palliative care. In cases where patients are unable to respond, closest relatives are questioned with respect to care and treatment procedures.

*Value List:*
List of values of the patient:

1. Privacy
2. Safety
3. Presence of loved ones
4. Freedom of choice
5. Comfort

*Persona: Patient Mrs.Vance (72)*
Mrs.Vance is a 72 year old widow with a daughter and son and lives with the son. She is old but of sound mind. Her carer helps her with daily activities and helps her move around. She has trouble breathing sometimes but nothing major at the time. She has had a couple of visits to the hospital for respiratory distress. She does not want any painful treatment that might affect the final years of her life and feels that instead of living a prolonged but impaired life, its better to live out the last days with little to no pain.

### 1.2.2. HA02: Doctor

*Background:*
The doctor responsible for the patient must adhere to the Code of Ethics while providing care and treatment to the patient. The doctor is a direct stakeholder, since he/she will, with the assistance of the robot, take decisions on which course of action to take based on medical and ethical reasons. Their job is to assess the patient's condition, safeguard quality of care delivered, act as a advocate for the patient and their family, support the family by providing clarity, and monitor bodily function closely in an effort to change the treatment when necessary.

*Value List:*
List of values of the doctor:

1. Critical care of patient
2. Health and safety of patient
3. Inclusion of patient choice in decision making
4. Privacy of patient
5. Respecting requests of patient's relatives

*Persona: Doctor Richard (50)*
Dr.Richard works at the Emergency Department and deals with cases in the Intensive Care Unit (ICU). It is not possible for him to supervise every decision made with respect to patient outcome. Some of these responsibilities are given to managers and nurses working under him. He has to tend to a lot of patients, especially those undergoing treatment for serious injuries and illnesses and is responsible for taking several decisions regarding life saving procedures. However, these procedures involve risk and taking such decisions puts a mental toll on him. It becomes quite stressful for him to constantly make such ethical decisions and feels that constant stress affects his decision making. He follows the Code of Ethics especially related with non-malfeasance and fidelity with respect to patients and their choices. The ethical category of non-malfeasance represents the doctor's attempt to avoid any act or treatment plan that would harm the patient or violate the patient's trust. Non-malfeasance is supported through confidentiality and prevention. The decision making of the doctor must be faithful to the requirements and the wishes of the patient thereby maintaining the integrity of the patient's choices.

### 1.2.3. HA03: Relative of the Patient

*Background:*
The relative of the patient may be a spouse, daughter or son. While they cannot be always there at he hospital, they visit as often as possible. They are worried about the health and treatment being provided to the patient. The relatives want regular updates on the condition of the patient and usually sign off on treatment procedures when the patient cannot respond. In this scenario one relative of the patient wants only palliative care for the patient as the patient supposedly requested, whereas another relative, wants aggressive treatment procedures to be carried out.

*Value List:*
List of values of the relative:

1. Health of the patient
2. Security and safety of patient

*Persona: Relative Greg (43)*
Greg is the son of Mrs.Vance and has been taking care of her for the past couple of years. He hired a caregiver to care for her when he is not home. He claims to know what his mother wants in the event of a serious illness - no aggressive treatment and palliative care instead.

*Persona: Relative Caroline (41)*
Caroline is Mrs.Vance's daughter who has not had regular contact with her mother and as a result, is unaware of her condition. She is very much concerned about her mother's well being. She lives in a different city but is making efforts to meet her mother. She wants her mother to be given proper treatment.

## 1.3. Problem Scenario

Part of Dr.Richard's responsibilities include making the final call on medical decisions involving patients. Most cases are serious and require careful evaluation by someone highly trained and specialized. Dr.Richard is authorized to take quick decisions on certain medical issues. As a result, Dr.Richard often has to take ethical decisions revolving around moral dilemmas that deal in part with giving treatment to patients. Sometimes, taking such ethical decisions takes a toll on Dr.Richard and he begins to second guess himself. Other times he feels that the decision he has taken may be wrong even if evidence points otherwise. He feels that getting validation or rejection of his decision making will help reduce fatal errors and allow him to re-evaluate his decisions. Dr.Richard and other doctors make use of assistive robots that help provide data and reports on the patients and help schedule their activities.

An example of Dr.Richard dealing with such decisions is the case of Mrs.Vance. She is a 72 year old woman who was brought to the hospital after her carer found her in respiratory distress. Dr.Richard along with the help of the Decision Assistant Robot (DAR) noted that the patient was minimally responsive to verbal stimuli, had no fever, normal blood pressure with tachycardia (high heart rate) at 130 beats per minute and rapid breathing at 30 breaths a minute. A radiograph taken showed a consolidation (liquid instead of air) at the right lower lobe. Based on details provided by the DAR concerning patient data and history, it was found that she had been recently admitted for investigation of significant weight loss and it was found to be a result of advanced bowel cancer with lungs, bone and brain metastases. Dr.Richard is then designated to provide treatment and care for Mrs.Vance. Within 24 hours of being transferred to the ICU, Mrs Vance's condition deteriorated rapidly and a decision was made to talk with the family of what should be done in the event of cardiac arrest. The son (Greg) was immediately informed about his mother's condition and it was revealed that Mrs Vance had previously stated to him that she does not want any heroic measures in the event of cardiac arrest. In this case a Not for Resuscitation (NFR) order is placed. The purpose of the NFR order is to deliberately withhold life-saving measures when the patient's respiratory or cardiac function suddenly stops. Later, the patient's daughter (Caroline) arrives. She was not aware of her mother's condition due to no contact for the past 3 years but requested for full care. Caroline's request for care conflicts with the patient's advance directive and places Dr.Richard in a difficult position of either honoring the patient's wishes or satisfying the daughter's request.

In such cases, Dr.Richard feels that patient's requests are to be respected to the fullest. He however, cannot help but feel regret if he chooses not to save her from dying. At such moments of doubt, he feels that if he his able to reflect on his thoughts and ideas, he can take his decisions quickly while keeping account of the ethical implications.

### 1.3.1. Applicability of an assistant robot

In this problem scenario, having a Decision Assistant Robot (DAR) with the current specifications may only help in automating certain aspects of the treatment process while not affecting the decision making at all. The DAR is not capable of evaluating moral dilemmas, while proponents also argue that such machines should not take part in ethical decision making. It is not necessary to evaluate the moral dilemma to help Dr.Richard with decision making. Guiding towards decision making provides ample reasoning for support. Apart from public perception about the possible inclusion of robots capable of processing morals and ethics, it is difficult to program moral thinking within Artificial Systems (AI) systems. In such cases, Dr.Richard feels that patient's requests are to be respected to the fullest. He also however, as a doctor, cannot help but feel regret if he chooses not to save her from dying. At such moments of doubt, he feels that if he his able to bounce his thoughts and ideas off of someone, he can take his decisions quickly while keeping account of the ethical implications.

To ensure ethical outcomes in Human-Robot Interaction (HRI), it is necessary for a robot to have at least three key competencies: (1) the ability to accurately recognise the present state of the environemnt, (2) the competence to evaluate and make precise judgements about the ethical decisions, and (3) the ability to adapt the HRI in a way that encourages ethical behaviour. While much work has focused on the second competency, and work on the first competency is difficult to understand and implement, the third competency is tapped into for implementation in this project.

## 1.4. Design Scenario

Dr.Richard employs the use of a Decision Assistant Robot (DAR) specifically tailored for medical situations. Dr.Richard turns on the robot for the day and asks it about the day's scheduled activities, patient reports and new cases if any. This allows Dr. Richard to prioritize the cases and tend to patient requirements more efficiently. The robot provides a point of reflection and deliberation for Dr.Richard and other medical professionals. For example, Dr.Richard, based on the medical reports of Mrs.Vance and her wishes may decide to stop treatment for her. The robot reminds Dr.Richard of the code of ethics that the decision taken breaks: such as avoidance of treatment and negligence. It asks Dr.Richard if he is confident of the decision and if it is right. The questions and displays of protest by the robot provide the doctor grounds for more discussion on the medical decisions that need to be made. Certain decisions made with uncertainty due to time constraints or lack of information may have significant repercussions for the patients and their families. Medical decisions have to be taken quickly ensuring the best care for the patients, even when information is lacking.

Time constraints related to the procedure, treatment protocols and the effects of the condition of the patient may further cause delays in administering the right decision. Medical data that may have not yet been discovered at the time, the rights of the patient and what they want if serious issues develop are important information. The DAR cannot solve the issues directly but can help alleviate certain aspects. Medical professionals are often under pressure to take the right decision. In such moments, having an assistant robot allows the doctors to reflect. The DAR can keep track of the treatment, the patient's wishes, the relative's requests and the code of ethics to provide a set of actions that can be carried out.

While simply posing a question may only serve as a reminder for the doctor, having gestures and verbal displays of questioning may make the process more effective. Verbal protests and affective displays can promote ethical behaviour in human subjects. This may reduce the emotional and cognitive stress on the doctor and by continuous discussion, new decision paths can be opened up that were earlier not under consideration. Decision making is affected by associated agency and 'emotion' of the robot.

The robot provides status and medical information on a new case and notifies the doctor. The robot provides sufficient details to indicate that the case is urgent and provides ways to contact the patient's relatives. In cases such as that of Mrs.Vance, either the son's wishes for a NFR order being placed and the daughter's request to go for complete treatment puts Dr.Richard in a difficult position. Declaring the decision to the robot allows the doctor to re-evaluate their decisions. In some cases the robot only uses audio to engage in a discussion while in other cases affect-based movement and emotional displays are used to dissuade the doctor from making decisions. The discussion with the DAR helps Dr.Richard ensure that he is sure about the ethical implications of the decision taken and make changes to these decisions as they see fit. Dr.Richard asks the robot if the decision taken is right which helps in reflection. The robot does not judge the ethical decision, but simply attempts to question, deliberate and dissuade the decision making process.

# 2
# Human Factors

## 2.1. Ethical Dilemma of Resuscitation

Resuscitation is defined as the process of reviving someone from unconsciousness. However, the question remains: should patients who are seriously ill (for e.g. with cancer) always be resuscitated when they have a subsequent life threatening complication (like a heart attack)? The decision should be made on the basis of autonomous wishes and informed consent expressed by the patient and their relatives. Research has shown that the overall consensus is to follow standard procedures for resuscitation [62]. This corresponds to the ethical judgements of the doctors responsible. Many doctors find it difficult to deal with the ethical dilemmas and associated decision making with respect to seriously ill patients.

While hospitals and emergency departments are provided with many tools for ethical discussion and deliberation, these are insufficient in emergency situations. Doctors need to make resuscitation decisions quickly while communicating with a select few personnel through a small communication medium [82]. It can be daunting for doctors to make a decision regarding the well-being of the patient when knowledge is limited and the choices to be made have possibilities to result in a less favorable outcome. The design focus is to show that reflection on ethical dilemmas in emergency settings with the aid of robots can help doctors refocus their attention on what to consider while making decisions.

It is difficult to determine if a seriously ill patient should be resuscitated. It is imperative that the doctors weigh all possible choices fully aware of the circumstances and knowledge regarding the patient. It is easier for doctors to just follow protocol and standard procedures to solve problems. Doctors have a professional moral responsibility to ”do the right thing” while also doing justice to their personal morals [76, 80]. Doctors find it hard to determine how to proceed and judge what is ethically correct. Uncertainty of knowledge and time constraints restrict doctors from making informed decisions. The study from Nordby et.al [62] shows that several doctors mention that decision making is unclear because of a lack of lucidity about the situation. Further challenges arise due to lack of proper communication with the relatives.

The disparity between what doctors believe and what action they actually take is related to the uncertainty involving the patient’s conditions and wishes along with the requests of the relatives. Emphasis is to be placed on sound information in order to ascertain if resuscitation should be started or not. Sometimes, resuscitation is carried out on the patient even when it is certain that doing so would cause tremendous pain for the patient while leaving them debilitated only because the relatives demanded it.

The ethical dilemma manifests itself as a double pressure situation. With respect to organizational theory, there is pressure from above and below on the doctors [62]. The pressure from “below” is grounded in caring paradigms, and in the belief that it can be wrong to resuscitate. The pressure from “above” is objective and system-related and grounded in the intrinsic value of human life. The pressure from above is based on several management levels, and of the concept that doctors are ambassadors for the values, norms and principles on which their medical organization is based on. The pressure from below is based on professional ideals and direct observations. It has its origin in the immediate experience of the patient, the duty to help the patient, tending to requests from relatives, and beliefs about what is the right thing to do considering the situation at hand.

### 2.1.1. Ethical Theories

While philosophy and psychology provide a myriad of ethical theories, the main ones that inform choices in the healthcare industry are deontology and utilitarianism [71].

*Consequentialist Theories*

Consequentialism is the ethical theory which holds that the consequences of the actions are central to the moral judgement of the actions. Utilitarianism, is a type of consequentialism based on the principle of utility. The actions taken are judged by the amount of pain and pleasure they bring about to every involved actor. Therefore, the decision is one that brings the greatest happiness to the maximum number of involved actors. The consequences of actions are measured against one value - human welfare. The issue is that happiness cannot be measured objectively and therefore can lead to exploitation. It only takes into consideration the consequences while disregarding the means to do so.

Since the deontological approach supports an action as right as stated by the moral rule or law being followed regardless of the consequence, the doctor following legal laws and the code of ethics, is justified in saving a life. However, utilitarianism may result in the doctor deciding not to resuscitate the patient since the consequence is that it would allow the patient to die peacefully and with dignity. Ethical theories may guide decision making and also the rules surrounding it.

*Deontology*

Duty ethics (deontology) states that an action is morally right if it is in agreement with a moral rule, independent of the consequences of the action. The best known duty ethics theory is the Kantian theory. It states the good will is the only unconditionally good thing. Everyone has a duty to act only when it is applicable that the action to be done is universally good. Therefore, the onus lies on the action itself and does not lend itself to possible unforeseen circumstances. Certain norms can also contradict each other, in which case, it becomes the question of deciding which one to follow.

The perception study focuses on on a social robot that provides context and support during ethical decision making. It does not evaluate them but merely states the action paths. Whether or not a robot can be ethical needs thorough research and exploration.

### 2.1.2. Ethical Decision Making in Healthcare

*Ethics and the Law*

The practices of healthcare organizations and professionals are governed by meticulous supervisory practices and legal liability is a major concern for healthcare practitioners. Decision making often lends itself to complications due to difficulty in striking a balance between legal and medical consequences. The law states minimal criterion for maintaining standards and enforces them using penalties and regulations whereas ethics reaches for what can be considered ideal [67]. Ethics assist healthcare professionals in determining the most favorable outcome from all possible alternatives. The law, along with the ethical foundation should ideally complement each other to form good decisions though this is not always possible.

*Ethical Decision Making*

Ethical decision making is a process where choice leads to action influenced by circumstances. The circumstances refer to the organizational role of the doctor and the relationship between the doctor and the patient. The content indicates the type of specific illness experience undergone by the doctor placing importance on personal, cultural, religious and professional values.

Peer and Rakich [67] mention these as the foundation of ethical decision making. Decision making in healthcare can affect a large number of people based on positive and negative consequences probable to occur. In certain cases the organizational rewards for unethical decision making outweigh punishment for carrying out the action. This means that the decision making process followed by doctors is regulated by the overall philosophy of the healthcare organization. While value profiles (the goal) are important, in the context of dilemmas, the other two foundations, namely, philosophical basis and moral philosophies play a major role.

| Philosophical Basis | Moral Philosophies |
|---|---|
| Utilitarian | Respect |
| Deontological | Beneficence |
| Integrity-based | Non-maleficence |
| | Justice |

Table 2.1: Foundations - Philosophical Basis and Moral Philosophies

The ethical theories have been discussed in the previous section whereas the moral philosophies are briefly described here. These philosophies guide the delivery of healthcare services [11].

1. **Respect:** Respect for others encapsulates autonomy and truth-telling. Autonomy for the patient is for them to accept or deny treatment if they are of sound mind.
2. **Beneficence:** Acting with kindness and doing all to ensure the patient benefits from treatment or lack thereof.
3. **Non-maleficence:** The concept of doing no further harm.
4. **Justice:** Justice equates to fairness in the medical system.

While ethical dilemmas such as resuscitation require the evaluation of multiple facets of medical and ethical practices, moral philosophies help guide decision making. The four moral philosophies are often held as canonical in the field of medical ethics [48]. They provide the foundation for analysing ethical situations and dilemmas. While it is possible for these philosophies to be in conflict under certain conditions, the theoretical framework provides guidance to make decisions. These principles establish the ethical framework to implement healthcare service provisions that oppose the doctor's personal and professional values. End-of-life and aggressive treatment based decision making is often under the microscope of public and governmental scrutiny. On the basis of these principles, medical authorities incorporate standardised segments such as a code of ethics, ethics training and performance management within organisational decision making.

*Decision Making Procedures*
Effective decision making requires consideration of organisational resources, community and individual to prioritise the decision. Per Levitt [48] and Meslin et.al [56] the approaches for decision making are: qualitative and quantitative. As the name suggests the qualitative approach takes into consideration, personal and professional values. The quantitative approach is based on mathematical analysis and is viable when the number of solution options is higher. It is a cost-benefit analysis approach and disregards choices and morals. Sometimes, due to time pressure conditions, organisation workflow and cost may factor into this form of decision making.

| Steps | Qualitative Approach | Quantitative Approach |
|---|---|---|
| 1 | Identify the problem | Identify alternatives |
| 2 | Analyze alternatives | Determine the evaluation criterion |
| 3 | Weigh options | Obtain value statements |
| 4 | Justify using principles | Calculate rank order and weighing factors |
| 5 | Make a choice | Rate alternatives |
| 6 | Reassess | Complete matrix and decide |

Table 2.2: Steps of Qualitative and Quantitative approaches to decision making

Meslin et.al [56] provide a model where iterative actions are taken so that philosophical and moral principles are considered after each step. The decision maker must consider all facts, determine problem areas, describe values at stake, identify available human and financial resources, specify the options, propose a solution and evaluate the decision.

Grundstein-Amado [29] proposed a model consisting of three components: the ethical, the decision theory and the contextual framework. Ethical reasoning comes from particular value judgements that account for various ethical principles that are grounded in ethical theories and moral philosophies.

The decision theory component follows a procedural scheme that helps analyse and decide the correct decision to be made. This includes: identifying the problem, gathering information, taking patient choice into consideration, determining ethical context, identifying alternatives, deciding on a course of action and being able to justify it. The final component provides relational aspects between the patient and the healthcare professional within limitations of the organisation and healthcare system.

These approaches are ideal by design but fall prey often to regulations of the organisations and the law resulting in their decision choices being reduced causing ethical discordance. The decision making depends on the code of ethics, the law - concerning the role of the doctor and the choices of the patient, the choices of relatives and medical processes based on patient status. There exists less flexibility in making a sound ethical decision especially regarding terminal care and life-saving procedures. Other factors include values and personal philosophy, regulatory influences, resource allocation, emotional nature of decisions and situational factors. Ethical decisions are strenuous and intricate, and the emotional nature of many decisions results in tense situations.

*Challenges in Decision Making*

Medical decisions are extremely complex involving several factors, interlinked relationships, uncertainty and multiple possible outcomes. Decisions have to be made on the signs and symptoms for diagnosis, additional tests to confirm diagnosis and determining the choice of treatment. Based on studies and experiences [25, 32, 55, 89], there exist some challenges that will evolve over time due to technological and financial factors along with social and ethical considerations.

Table 2.3 encapsulates the challenges doctors face in medical decision making.

| No. | Challenge | Description |
|---|---|---|
| 1 | Availability Heuristic | Gauging the likelihood of an occurrence by relying on relevant examples that come to mind immediately when evaluating concepts and topics. |
| 2 | Search satisfaction | The tendency to stop searching for a diagnosis once something is found. |
| 3 | Attribution error | Putting a patient into a negative stereotype on the basis of personal bias. |
| 4 | Confirmation bias | The proclivity to recollect information that confirms one's personal beliefs. |
| 5 | Legal consequences | Not following the procedures mandated by the healthcare system and the law may hinder decision making. |
| 6 | Cognitive-overload | Encumbered by multiple patients regularly. May impair decision making of doctors. |
| 7 | Clinical circumstances | Concerns the implementation of organisational flow that limits usage of resources and medical staff. |

Table 2.3: Challenges in medical Decision Making

Factors such as facts, memory, information processing, value of information, choice of treatment and development of clinical policies have developed tremendously over the years. Research and information collection systems have been developed that aid in cataloguing and indexing medical information. These systems also aid the strong mental skills of physicians and doctors with respect to signs, symptoms, diagnosis, treatment and prognosis. Using valuable information and pattern comparisons, heuristics are developed that help pick the course of action. Clinical policies make the decision process easier as described earlier. The challenges to decision making develop as a result of rising costs, awareness of practices, exposure to moral considerations and patient rights. In place of a qualitative judgment that a particular process may have benefit, a quantitative estimate must be made about the magnitude of effect. Instead of deriving policies by formalising the encounters between doctors and patients, the healthcare system must enhance the range of the decisions to make comparisons across treatments, diseases and values [20, 23].

### 2.1.3. Influence on Decision Making

Key ethical principles surrounding end-of-life decisions and resuscitation are realistic, beneficence, non-maleficence and justice. However, sticking to one principle can mean that other principles may be compromised. There does not seem to be an exact right or wrong answer, and all these principles can be applied to end-of-life scenarios. A decision to carry out resuscitation would be based on numerous factors. The principle of non-maleficence obliges the doctor to not harm others and would in turn oppose resuscitation if doing so would cause more harm than good. It can also be argued that not resuscitating could lead to the ultimate harm - death. As such, decision making under these conditions is extremely difficult and stressful [11].

*The Gibbs Reflective Cycle*

The Gibbs Reflective Cycle [26] is one of the most famous models of reflection that guides one through various stages to make sense of an experience. The model is better equipped for use in repeated situations and a number of questions are provided by Gibbs to describe the situation in detail.

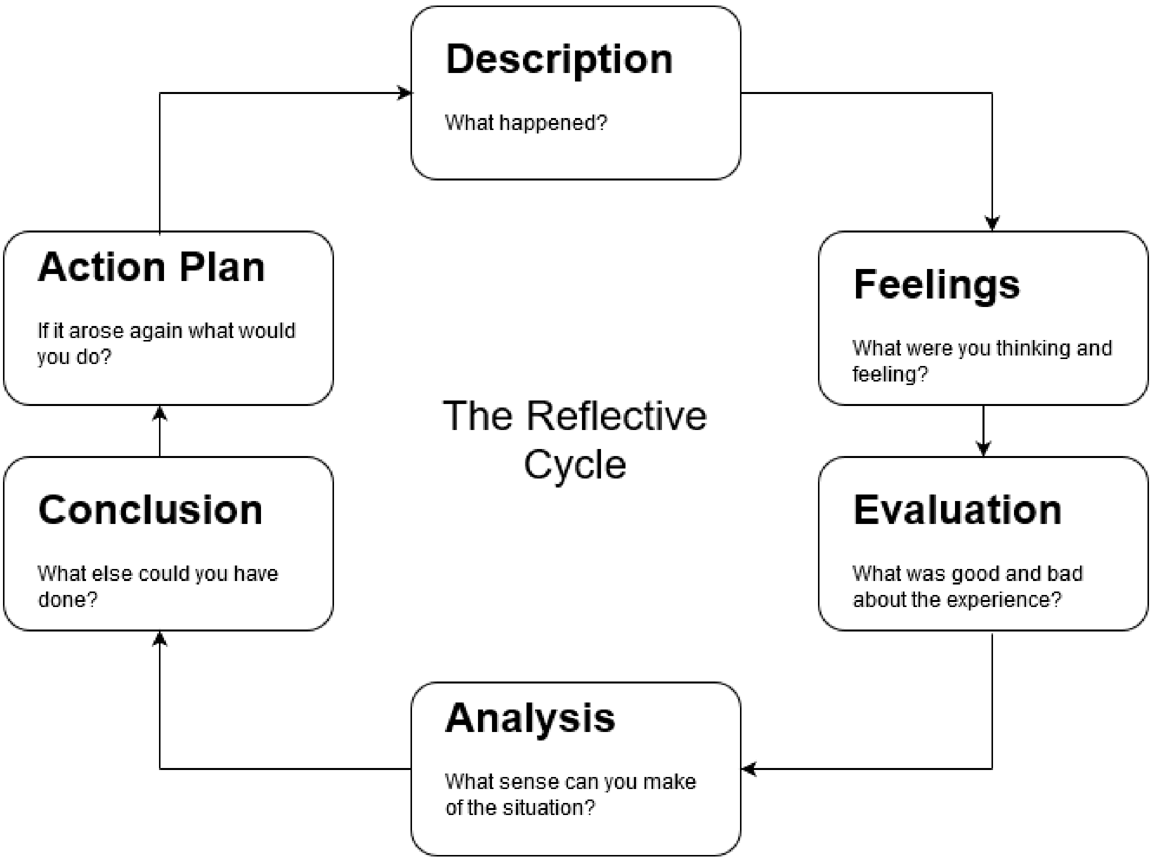



Figure 2.1: The Gibbs Reflective Cycle

Reflection is used to improve understanding and proof of practice-based learning. It is regarded as a valuable instrument to use after critical occurrences to help practitioners reflect on encounters and learn from the experiences. Believing in the idea that one can only change what is in their control to change, the Gibbs model encourages the use of critical reflection. The robot aims to represent the Gibbs Reflective Cycle in the confrontation scenario along with the verbal, video and physical cues. Questions are framed corresponding to the different stages in the Reflective Cycle.

Implementation of the Gibbs Reflective Cycle in medical practice helps narrow the space between theory and practice with respect to the provision of care. The medical professionals analyse their actions and critically evaluate their experiences [26]. Reflection is part of experiential learning over long time periods. Healthcare professionals often need to learn and employ concepts and processes to deal with developing circumstances with respect to patient care. The Gibbs' Reflective Cycle model specifically addresses the six stages of reflections as seen in Fig 2.1 and suggests that experiences are repeated, thereby allowing the reflector to think through all phases of the experience.. The model casts a wider net on the factors affecting decision making and provides flexibility while assessing a situation critically.

Mistakes can be avoided through the implementation of the Gibbs' reflective cycle. Each stage of the model provides sufficient information to aid in decision making albeit taking more time. The robot asks some of the aforementioned questions in conversations and discussions to initiate reflection. Findings and discoveries help during self reflection mainly when evaluating medical ethical dilemmas. The Gibbs reflective cycle in medical scenarios is the professional improvement of an active treatment process triggered by the theory-practice gap.

*Robot Questioning - Conversational Flow*

The role of the Decision Assistant Robot (DAR) is to deliberate on the decision taken by the doctor. When a conversation or decision flow is interrupted by an unexpected response from the doctor or slight malfunctioning of the robot, alternate conversation sentences must be made by the robot to steer the decision making back on track. The DAR needs to question the decisions, in both the ethical and medical context. The ethical context is based on the patient's conditions, the treatment procedures possible and associated statistical and medical data regarding the survival and care of the patient. The robot must know when to use which statements when trying to display protest over the decision taken by the doctor.

Decision making is required with respect to the patient regarding treatment. In the hospital, the patient may make requests. The patient may decline treatment, opt for palliative care or is incapable of making choices. Often, relatives may provide certain context regarding the decision making. They may request care or treatment for their relative. Balancing all the information and time constraints places stress on the doctor.

Asking questions using the robot and the tablet and by making use of the movement and behaviours, the design problem can be solved. The tablet will be mainly used to provide the medical context - data and status of the patient. With regards to the ethical context, the robot will not assess the morality of the situation. Instead, it aims to provide information in a clear manner that provides clarity. It presents the wishes of the patients, the requests of the relatives, the code of ethics that comes into play regarding patient welfare and then asks the doctor if these aspects are considered for decision making. The trigger statements must encapsulate as wide a response statement framework as possible. The robot can restate its messages or request for repetition if it did not hear the doctor properly.

The aim of dissuasion is to allow for more deliberation regarding the decision to be made. Reminding the doctor several times of possible ethical ramifications while discussing the decision can help the doctor base their decision on better ethical norms. The doctor is guided towards a balanced decision making scenario where enough consideration is given to the empirical data and the normative concerns while within the ethical considerations.

## 2.2. Ethics in HRI

### 2.2.1. Earlier Work

Briggs and Scheutz [14] developed a HRI setting to answer whether humans would be willing to accept robots that question their moral judgements and take their advice. Specifically, it involves a human ordering a humanoid robot to knock down three towers made of aluminum cans wrapped with coloured paper. The focus lies mainly only on two conditions: confrontation condition where the robot protests the human's command to knock down the red tower and the non-confrontation condition where the robot obeys all commands given to it without any protest. They found significant behavioural changes in the subject caused to due to display of protest by the robot. In fact, some subjects decided to compromise with the robot. Most of the participants felt some discomfort when ordering robots in the confrontation condition but this discomfort was minimal in the non-confrontation condition. The tests show that there is more hesitation in ordering the robot to carry out harmful actions when the robot exhibits moral reasoning and intelligent behaviour.

Briggs and Scheutz [15] further provide a cognitive robotic architecture to enable a rejection and explanation mechanism for directives given during a HRI scenario. This focuses on developing a reasoning mechanism based on achieving mutual understanding. The effect of such an interaction on the human however is not thoroughly explored. While they focused on the ethical and directive process within the robot itself, the external decision making support is not explored. They probe whether humanoid features in a robot help dissuade a human from forcing the robot to carry out a specific command. It was found that that the perceived intelligence of the robot was found to have a greater effect on animacy ascription than appearance. The various dimensions to determining the extent to which the robot can be dissuasive are described. Furthermore, a study by Hinds [31] indicates that humans who give tasks to non-humanoid robots felt more responsible for ensuring that the objective was accomplished than people who delegated these tasks to humanoid robots. Human-like behaviour by robots can be achieved through natural language interactions along with ascribed agency [38].

### 2.2.2. Robot Agency and Features

There exist other studies and experiments where the human agent decision making is affected by the human-like behaviour shown by the robot. Bartneck et.al [6] provide some support to this argument where the degree of intelligence perceived in a robot can affect their willingness to destroy it. People are also reluctant and hesitant to power off robots that exhibit intelligent and socially acceptable behaviour [5]. These studies shed light on whether anthropomorphic dominant features or behaviour dominant features affect the HRI. They also imply that as long as the robot can communicate ethical concerns using natural language in a human-like manner, such requests have found to be effective. Investigating the effects of different affective displays of distress (crying, denial etc.) could also be done in conjunction with an increasing number of protests before the robot acquiesces to the human commands. Studies show that within moral dilemma based scenarios that involve the robot following an obligation, most people show little reluctance in making ethical and moral judgements about a robot's decision and that people's judgements of robots on these dilemmas is similar to their judgements on humans in these scenarios.

Malle et.al [51] show that the moral judgement about robots that people hold depend upon the display of a verbally described robot. The study showed that identical descriptions of a robot facing a moral dilemma and making decisions leads to different moral judgements about the robot by the humans based on whether the robot is mechanical or humanoid. In this case, humanoid robots are expected to act more humane whereas certain actions carried out by the mechanical robots were deemed moral in contrast with the humanoid one. In a study by van den Brule et.al [87], it is presented that a social robot's trustworthiness is mainly influenced by its performance on a task. Motion fluency and gaze behaviour are important factors noticed by participants. They also state that video simulation experiments are as effective as actual scenario based experiments. Lee et.al [46] carried out a study to see how human-like a robot should be. Participants were more focused on human-like features and not humanoid robots. Certain features can be different based on the scenario context and usage.

Lopez et.al [49] carried out a between subjects experiment to test if indirect language will have greater influence in attitude changes in comparison to direct language during an interaction of a human with a robot. Attitudes of participants towards the robot presenting a concept in direct or indirect language did not differ. These results may suggest that in human-robot interaction indirect language may not function similarly as it does in human communication. Monroe and colleagues [58] found that a robot's choice capacity is a critical ingredient in people's willingness to blame a robot for incorrect actions. Studies have also begun to examine the effect and force of moral appeals that robots express to humans [57, 83].

## 2.3. Robots in Healthcare

With robots increasingly becoming a bigger part of everyday life, the social and supportive aspects of robots are also gaining significant improvements. HRI in the healthcare segment needs to take into account the physical and psychological aspects of the patient as well as the interaction between patient and their relatives and the medical professionals. A lot of decisions taken in this industry need to be ethically sound and how robots can maintain and ensure that standards and ethics are maintained will be interesting to observe and study in the future.

Robots are increasingly being used as part of healthcare processes. This includes aspects such as administration, equipment handling, surgical operations, companion assistance and information gathering. Human-robot interaction in the health sector is confronted by a myriad of challenges such as acceptability, appropriateness, safety and fear of replacement of caregivers by robots. Social robots are used in the health sectors to provide therapeutic aid and companionship apart from providing surgical and administrative support to medical personnel. In an experiment by Lum et.al [50], the robot assists the aged and provide remote surgical operations and dispense oral drugs. Social robots also imitate cognition to provide companionship where required. This helps isolation and depression related behaviour in the patient and the actions of the robot on the patient can have significant effect on their recovery [70].

### 2.3.1. Role of HRI in Healthcare

The robot must act socially and also provide support to the medical personnel. Evaluating the situation, taking the ethical decision and then studying the results of the decision and how that affects the patient are interesting research areas. Since these robots are mobile, they can act in social and workplace environments while gathering data from the immediate vicinity. While the physical and mechanical design of the robot is important, the experiment here will not affect the appearance of the robot used - Pepper [66]. The robots capability to carry out tasks semi-autonomously or autonomously while taking into consideration the instructions of the doctor, ethics and the well being of the patient must incorporate service robot features [65] into a social robot. The interaction, based on Goodrich and Schultz [28] , is proximate, where the interaction occurs between the actors in the same environment. Yanco and Drury [90] provide different interaction roles. With respect to the experiment the operator, peer, team and the patient are the most important roles to be played. Therefore, the robot must be capable of HRI awareness: the knowledge that the robot has of its surroundings, human commands and instructions to direct its activities and the conditions and constraints under which the operation of the robot is limited. Understanding the ethical implications of a robot's rehabilitative, assistance and companionship actions and decisions is vital. Since such robots monitor patients' health, they record and transmit patient data in human-readable form. This causes privacy and GDPR concerns and also leads to lack of trust in healthcare providers, preventing them from disclosing important information. It is important to note that the emotional attachment between humans and robots is unidirectional [79]. There exists a sense of deception in the minds of patients. The robot can only exhibit a 'enacted' emotional response and as such an emotional source to the response may not be required. User experience is also important, especially since such robots come preset with certain attributes.

Healthcare laws and policies must be followed when designing scenarios to use the robot and to prevent breaches of any kind of sensitive healthcare information. A study by Salem et.al [75] indicated that participants were alarmingly willing to follow a robots incorrect actions in the form of unreflected overreliance. Problematically, in a home care scenario such over-reliance could result in an elderly person with dementia taking an overdose of medication if a malfunctioning robot reminds the user of the same scheduled dose intake multiple times. It is well established that people tend to anthropomorphize robots and treat them as social entities [6, 91] and attribute them morals and rights [5, 38]. Cormier et.al [16] carried out an obedience study where a robot gave authoritative and indicative directions. Even after trying to avoid the task or engaging in arguments with the robot, participants still obeyed its commands. The success of building robot companions depends on the relevant behaviour, not the source of that behaviour, which goes in line with the argumentation of behaviour based robotics from Arkin [2].

Baumgaertner and Weiss [8] argue that unless a theory of emotions is put forward on purely behavioural grounds, a theory of emotions is unnecessary for ethics based human-robot interaction for companion robots. Assistive robots, in recent times have increased levels of autonomy and are efficient due to better and cheaper technology. Studying the capabilities of the robot and its effects in the interaction scenario becomes imperative especially in the healthcare sector where ethical decisions need to be taken while also trying to ensure patient and relatives' concerns [43]. Riek and Howard [69], published a paper describing unique ethical challenges in HRI and proposed a code of ethics. The code of ethics is elaborate but succinctly put the major considerations are as follows: emotional needs of humans to be respected, right to privacy upheld, maximal transparency in programming of the robotic systems, predictability to build trust, opt-out mechanisms, informed consent, respecting laws concerning patient's rights, consideration of human social behaviour and careful employment of Wizard-of-Oz techniques. The study done by Kim et.al [41] provides empirical evidence showing that framing the role of the robot (as a caregiver) makes a significant difference in how users evaluate robots.

In an experiment by Fan et.al [22] the objective was to determine whether people viewed robots as more or less emotionally intelligent when exhibiting similar behaviours as humans, and to investigate which verbal and nonverbal communication methods were most crucial for human observational judgments. Study participants were shown a scene in which either a robot or a human behaved with either high or low empathy. The results showed that participants could consistently distinguish the high Emotional Intelligence condition from the low Emotional Intelligence condition regardless of the variations in which communication methods were observed, and that whether the agent was a robot or human had no effect on the perception. Hoffman et.al [33] present the design, implementation and evaluation of an empathy-evoking robotic conversation companion. The robot's function is to increase people's awareness to the effect of their behaviour towards others, leading to behaviour change. The authors found that robots can be designed to evoke empathy among people in human-robot and human-human interactions without affecting natural communication patterns. Mann et.al [52] show, that using a robot to promote healthcare behaviours can have advantages over a tablet computer. Participants found the robot was more enjoyable to interact with than the computer tablet, and reported higher desires to interact with the robot again in the future. Furthermore, people were more likely to trust a robot's advice. This study also showed that people viewed robots as less likely to breach confidentiality. These findings suggest that participants formed stronger relationships with the robot, than others did with the tablet computer.

Nagataki et.al [59] try to clarify part of the ethical basis necessary for a machine like a humanoid robot to be a member of a human society. They tested three separate moral dilemmas, the trolley problem, the ultimate game and the dictator game. The experiment showed that many participants rejected an unfair proposal from the robot with which they had conducted the task. This suggests that the participants had a tendency to ask their partners with whom they conducted the task in such a situation, to engage in a certain moral commitment, whether it was a human or a robot. Some of the participants unconsciously recognized some inner state within the robot while bodily coordinating with it.

### 2.3.2. Robot perception in real-world scenarios

Humans show negative responses as well as lower trust towards robots that closely resemble human structure. This is termed the "Uncanny Valley" [53]. When the movements and the appearance are almost human-like but not entirely, there are too many expectations of the capabilities and the result is a negative reaction from the observer. However, recent research has also indicated that mimicking human gestures and postures make robots more likeable and trustworthy [1]. Based on findings from Sakamoto et.al [74] and Spence et.al [81], some conclusions about user sentiment and behaviour around robots can be made. During interactions in the initial phases people are uncertain, anticipate less social presence, and have fewer positive feelings when thinking about interacting with robots. They would like to rather communicate with a human. This is called the human-human interaction script. The robot must respect a "safe" distance when carrying out a proactive action. People have also been shown to attribute personality characteristics to the robot that were not implemented in software. Researchers have investigated anticipatory robot control through various methods including: monitoring the behaviours of human partners using eye tracking, making inferences about human task intent, and proactive action on the part of the robot [35].

Theory of Mind (ToM) is the ability to attribute mental states - beliefs, intents, desires, emotions, knowledge, etc. - to oneself, and to others, and to understand that others have beliefs, desires, intentions, and perspectives that are different from one's own. Theory of mind is crucial for everyday human social interactions and is used when analyzing, judging, and inferring others' behaviours [30].

DiSalvo et.al [19] indicate that the functionality of a robot must be tied in into its appearance. It is suggested that an amount of 'robot-ness' to emphasize the robot capabilities and to avoid false expectations, enough 'human-ness' such that the human agents feel comfortable, and a certain amount of 'product-ness' such that the robot is also seen as an appliance are considered before deploying the robot in a certain scenario. Krämer et.al [44] highlight the similarities and difference between Human-Human and Human-Robot interaction. A number of studies also show that the rules of proxemics are also applied in human-robot-interaction. Proxemics is the study of human use of space and the effects that population density has on behaviour, communication, and social interaction. Syrdal et.al [84] conducted an experiment with varying approach direction in three different scenarios (verbal interaction, physical interaction and no interaction). Participants indicated their preferred robot approach distance via a Comfort Level Device. The results show differences in approach direction and that the participants' personality traits of extroversion and conscientiousness are associated with changes in approach distance preferences according to robot autonomy.

*Establishment of relationships with Robots*
Empirical results from studies with various robots and agents suggest that it is fairly easy for researchers to build more or less autonomous robots and agents that are able to engage the user in a relationship. Bickmore and Picard [9] conducted a study with the Fit-Track system implemented in a browser interface so that the user could use the system at home. There was a significant increase in physical activation level during the treatment. The establishment of a bond and attributions like trust, however, was dependent on the behaviour of the system. Koay et al.[47] conducted a longitudinal study on human-robot-interaction over a period of five weeks. Interestingly, besides other results, the researchers identify a habituation effect: Participants allowed the robot to come closer in week five than in week one or two - indicating the buildup of a relationship. More convincingly, Banks et.al[4] report an experiment comparing living dogs with a robotic dog (Sony's AIBO) in a nursing home to improve residents feelings of loneliness. Altogether, first results show that people do not only develop relationships with social agents and robots but that they can also benefit from these relationships, e.g. with regard to their health. This indicates that robots and agents might be similarly capable of satisfying the need to belong as humans do. This is useful in implementing the long-term use of Decision Assistant Robots.

## 2.4. Displays of Protests

The displays of protest are based on the two forms of communication between humans, verbal and non-verbal communication. Verbal communication involves the use of verbal language and audio cues to indicate topic judgement and emotion. Non-verbal communication involves written, visual and physical modes of communication. With respect to the perception study, the non-verbal communication used here involves visuals (images) and gestural modes of communication.

Displays of protests in an Human-Robot Interaction can be done in three different ways. The three conditions for the confrontation methods are described below:

1. **Neutral Confrontation:** In this method a neutral, non-confrontational approach is used. The robot does not show any signs of distress or protest. The robot accepts the commands given to it by the human agent and carries out the task without questioning the decision. Non-confrontation does not involve the robot attempting to dissuade the human agent.

2. **Verbal Confrontation:** Verbal confrontation includes displays of protest and argumentation by the robot using audio and verbal cues. The robot may make use of its speakers to protest decisions taken by the human agent. Using speakers, it is possible to use audio confrontation messages, for eg. ”Are you sure that administrating drugs now will help?”.

3. **Multi-modal Confrontation:** In Multi-modal confrontation, verbal cues along with movement, gaze, proxemics and behavior patterns (non-verbal cues) are used. Affect includes any experience of feeling or emotion, ranging from suffering to elation, from the simplest to the most complex sensations of feeling, and from the most normal to the most pathological emotional reactions. For e.g. the robot could wave hands or shake its head in disapproval at the human agent's decision while also making grunting noises and sighing. In this method, the robot tries to be as human as possible in terms of affect, emotional displays, gaze, movement and speech.

These three conditions increase in complexity in terms of both implementation and degrees of displaying protest. By understanding how dissuasive a robot can be when it protests verbally and displays distressed behaviour, it is possible to address the ethics-based issues on how humans react and respond to affective displays when the robot engages in ethical interactions or settings.

*Why affect-based confrontation?*
Affective displays are the verbal and non-verbal displays of emotion. This can be done through vocal parameters such as pitch and volume but also includes gaze, facial expressions and proxemic distance. Mavridi [54] reviewed in detail studies made in HRI regarding verbal and non-verbal communication. Other than static gestures, the factors of facial expressions are vital towards conveying believability. Pepper lacks in facial expressions but can denote basic emotions using eye color for example. An

addendum to this is the possibility of employing non-linguistic utterances for showing emotions. These are non-verbal sounds that indicate categorical expressions. Verbal dialogue in humans does not come just from non-verbal signs; in order to obtain even the most fundamental degree of openness, a humanoid robot needs for example at least some physical features to accompany dialogue. Gestures can be specifically described as being supportive to rather than being correlated with or just accompanying speech.

Additionally another facet of communication that requires non-verbal elements is covert signalling, mainly done through head nods that the listener shows as responses to the speaker. The simplest form of gestures are deictic, pointing towards an object and usually accompanied with indicative motions. Furthermore, gestures are highly important towards teaching and learning in humans. Eye gaze cues are principal for coordinating collaborative tasks and are a vital form of non-verbal communication cue that can increase efficiency and robustness in HRI.

Supposing the robot has the capability for situational awareness and ethical reasoning competencies, the robot may then detect, based on its interaction with the doctor that the human is commanding it to carry out an unethical action. The robot could just decline to carry out the command but that may not dissuade the medical professional from carrying out those actions through other means. Robotic systems have control parameters to allow override by human operators. An approach is required for robots to reasonably present obstacles in ethical decision making. Robotic agents can do this through verbal confrontation and gestural cues [85].The appearance of being true to human dialogue and displays of affect may intensify the efficacy of such persuasive and dissuasive effects in HRI. The potency of such methods depends on the believability of the robot's behaviour and actions and how much agency the human ascribes to the robot.

*Robot Believability*
Different senses of robot believability can be described [72] that can influence the displays of affect in a confrontation situation. The first level Bel1, is achieved if and only if a human responds to a robot as if it were a type of more sophisticate agent. The second sense of believability, Bel2 concerns whether the robot has aroused a internal response in a human user similar to the response that would be aroused in the user in the same circumstance by a living counterpart to the robot. This is distinct from the fourth sense of believability, Bel4, that concerns whether the human user ascribes mental states to the robot that are similar to the mental states the user would ascribe to a living being. The distinction between Bel2 and Bel4 is important as an affective protest by a robot could potentially evoke a visceral Bel2 response in a human operator, yet remain ineffective because the operator ultimately does not believe the robot is capable of possessing the affective states it is conveying. Bel2 and Bel4 may be required to enhance desired behaviour in the docto. Believability can be enhanced by human-like displays of emotions, verbal cues and gestures.

## 2.5. Code of Ethics

The code of ethics [3, 11], provisioned by various medical institutions and organizations in collaboration with governmental organization, provide a set of ethics that every medical professional must follow when working. While listing the entire code here is not conducive due to the space considerations, the Code of Ethics deals with the following.

- Always exercise his/her independent professional judgment and maintain the highest standards of professional conduct.
- Be dedicated to providing competent medical service in full professional and moral independence, with compassion and respect for human dignity.
- Deal honestly with patients and colleagues, and report to the appropriate authorities those physicians who practice unethically or incompetently or who engage in fraud or deception.
- Respect the rights and preferences of patients, colleagues, and other health professionals.
- Always bear in mind the obligation to respect human life.
- Act in the patient's best interest when providing medical care.

- Owe his/her patients complete loyalty and all the scientific resources available to him/her.
- Respect a patient's right to confidentiality.
- Health is universal right and Privacy of patient should be protected.
- Integrating social justice in Nursing and Health Policy.
- Patients have the moral and legal right to determine what will be done with or to themselves.

## 2.6. Technological Principles

### 2.6.1. RA01: Decision Assistant Robot

The robot that will be used for this experiment is called 'Pepper', developed by SoftBank Robotics. Pepper is a robot designed and optimized for social interactions with humans. It is capable of recognizing faces and basic human emotions. It can also carry display emotions, some built-in and others that require implementation. Another advantage is the mounted tablet on the front of its torso. This allows Pepper to interact with the doctor about the patient's condition and display images and textual information. Furthermore, Pepper is capable of several features that are of use for this project.

1. Ability to move in any direction and proclivity to carry out 'natural' movement.
2. Being able to recognize speech and talk multiple languages.
3. Being able to recognize when the person(medical professional) is interacting with the robot.
4. Pepper's ability to detect emotion comes from the ability to analyze expressions and voice tones.

The assistive robot aids the medical professional in carrying out his/her duties with respect to the patient's health as well as dealing with concerned relatives. The robot dynamically tracks the status of the patient and upon noticing abnormalities in the patient's condition, warns the medical professional immediately and provides a course of possible treatments. It will accept any decision the medical professional takes even if it is not in its recommended course of actions absolving itself of any crucial decision making. The onus of the ethical and medical decision then falls on the medical professional based on how the robot reacts to their proposed choice.

There are two ways to develop the new functionality required for Pepper to interact with the doctor. Writing independent modules and loading it up on Python is one choice. The other choice is to build behaviour modules on the Interaction Design Tool [73] which is tool that lets one create animations, behaviour patterns and dialogues through an interface.

# III

# SPECIFICATION

# 3
# Design Patterns

The design patterns for the supposed interactive multimedia on-screen and robot based care is set up in such a manner that satisfactorily meets the requirements of the doctor. The solution addresses a number of criteria. The interaction must allow easy-startup and intuitive navigation for the doctor. The design pattern provides functionality with regards to the following activities: elicit reflective thinking and to promote discussion for serious decision making. It has an initial introduction as a design pattern that uses a largely scripted and conventionally-established verbal and behavioural range to recognize, inquire and acknowledge each other. There is a requirement for didactic communication for moments where both parties need to be engaged in communication during one-way transfer of information and aligning motions can help the Decision Assistant Robot (DAR) engage better with the doctor [37].

## 3.1. Designing dialogue structure

A monologue is a longer form of speech during which no response is expected. Monologues may involve the telling of a scenario. An instruction is a command offered by the doctor to direct the actions of the robot. The proper response to this instruction is often an action, although the action might follow the instruction with a delay depending on whether it is an appropriate time to perform that action. When a participant transitions into a state intended for a single participant, all other participants enter a wait state, in this case the wait state alternates between the doctor and the robot. Monologues are used to explain patient status and possible repercussions of decision choices.

It is known that humans follow specific patterns when interacting with each other [27]. It is suggested that the order observed in human-human interaction follows patterns based on how greetings are exchanged and how interactions end. These patterns can be structured as conversations, interviews and collaborations [78]. Communicative acts set out methods on design pattern exploration and prototyping of HRI. Designers have made use of such patterns for developing pattern language regarding robot design [68].

Based on the work by Kahn et.al [39], the interaction can be made more social by implementing eight design patterns. These are patterns observed in human interactions that provide inspiration for interaction design for HRI. Two patterns that provide opportunities of socializing and instantiations to depict applicability within the context of the ethical dilemma discussed are: Didactic Communication and Claiming Wrongful Harms. Didactic Communication is a design pattern for one-way communication of information, situated in a context where each party has motivation to remain engaged. There exists two types of claims about immoral treatment, one based on deontological justifications of fairness, justice, and rights, and the other based on consequentialist justifications of material, physical, or psychological harm - the underlying basis of the robots discussion with the doctor.

The design patterns developed adapt partially to the patterns for collaboration [24], conversation [92] and instruction [36] to help aid the development of natural dialogue progression. Under the team design pattern [88], the assistant robot provides important information to the doctor concerning patient health. The discussion between the robot and the doctor are conversations based on collaborative efforts to take ethical decisions following: a) monologue-comment pattern, b) instruction-action pattern and c) wait pattern. The relevance is staged on the basis of context (dilemma of resuscitation, poor health status) and a structure of moral values (ethical theories, code of ethics). The robot explains the situation with contextual undertones that supports humans in moral decision making, indicating sufficient knowledge of the ethical complications. Following the body of work in prototyping interactions in HRI, the design patterns developed implement conversational, information provision and question-answer dialogue patterns.

## 3.2. DP01: Designing Neutral Confrontation

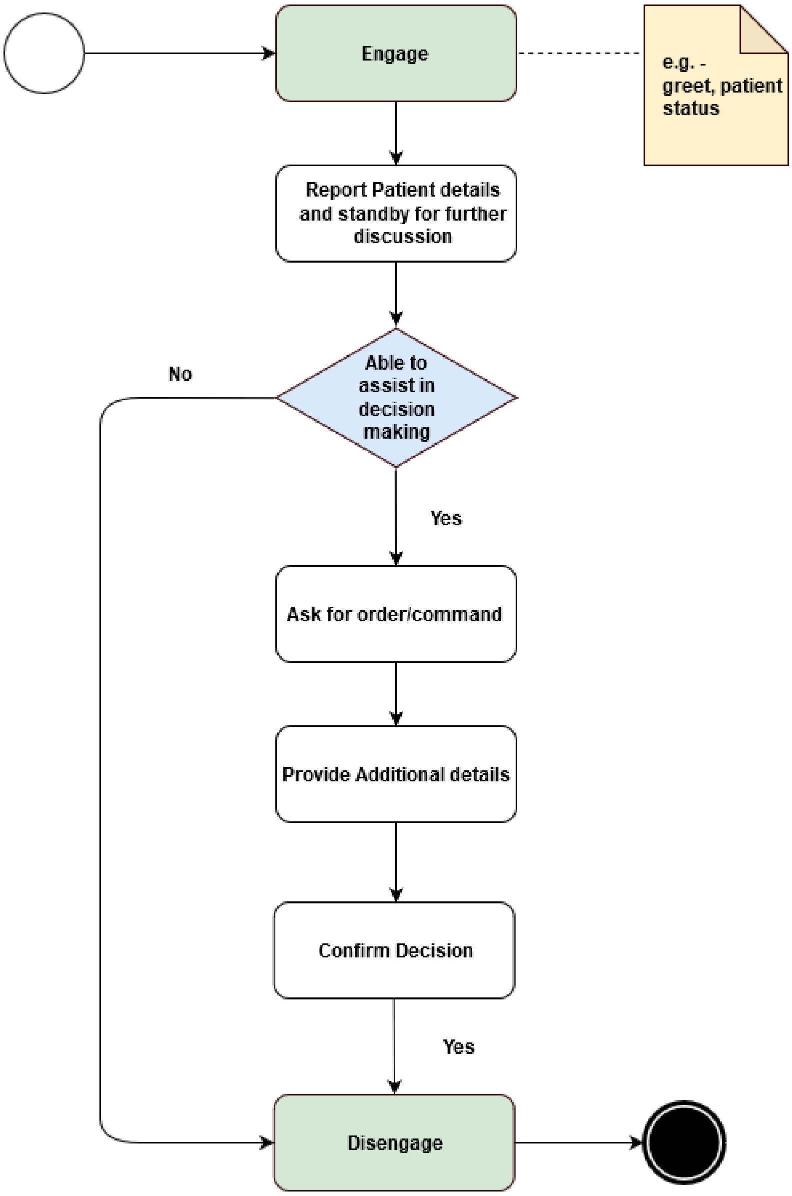



Figure 3.1: Interaction flow for Design Pattern 1

| | |
|---|---|
| **ID** | DP01 |
| **Name** | Interaction of the Decision Assistant Robot (DAR) with human agents with no confrontation |
| **Ranking/validation** | Quicker conversation flow since robot acquiesces to every command of the doctor. |
| **Design Problem** | The robot agrees to and carries out the decision of the doctor. |
| **Context** | The robot does not challenge the decisions of the doctor. The dialogues used are simply that of affirmation in order to carry out the tasks. |
| **Design Solution** | The robot should execute spoken commands when interacting with known users. Making use of default audio parameters will allow the robot to be more neutral.<br>It agrees to the decisions made by the doctor and does not discuss the process. |
| **Design Rationale** | Neutral voice commands are used for agreeing to the decision taken. No confrontation is done to ensure neutrality in decision making. |
| **Examples** | The robot provides information on the patient and agrees to the decisions taken:<br>Robot: Hello doctor, Here are the reports for Patient Mr.Stam.<br>Doctor: Thanks. Anything else?<br>Robot: Patient Mrs.Vance is still in the ICU. What are you planning for treatment?<br>Doctor: Let's begin with chemotherapy.<br>Robot: Okay doctor, let me inform the necessary personnel at once. |

Table 3.1: Design Pattern 1

The lack of confrontation will not result in any opportunity of discussion or reflection on the ethical dilemma facing the doctor.

## 3.3. DP02: Designing Verbal Confrontation

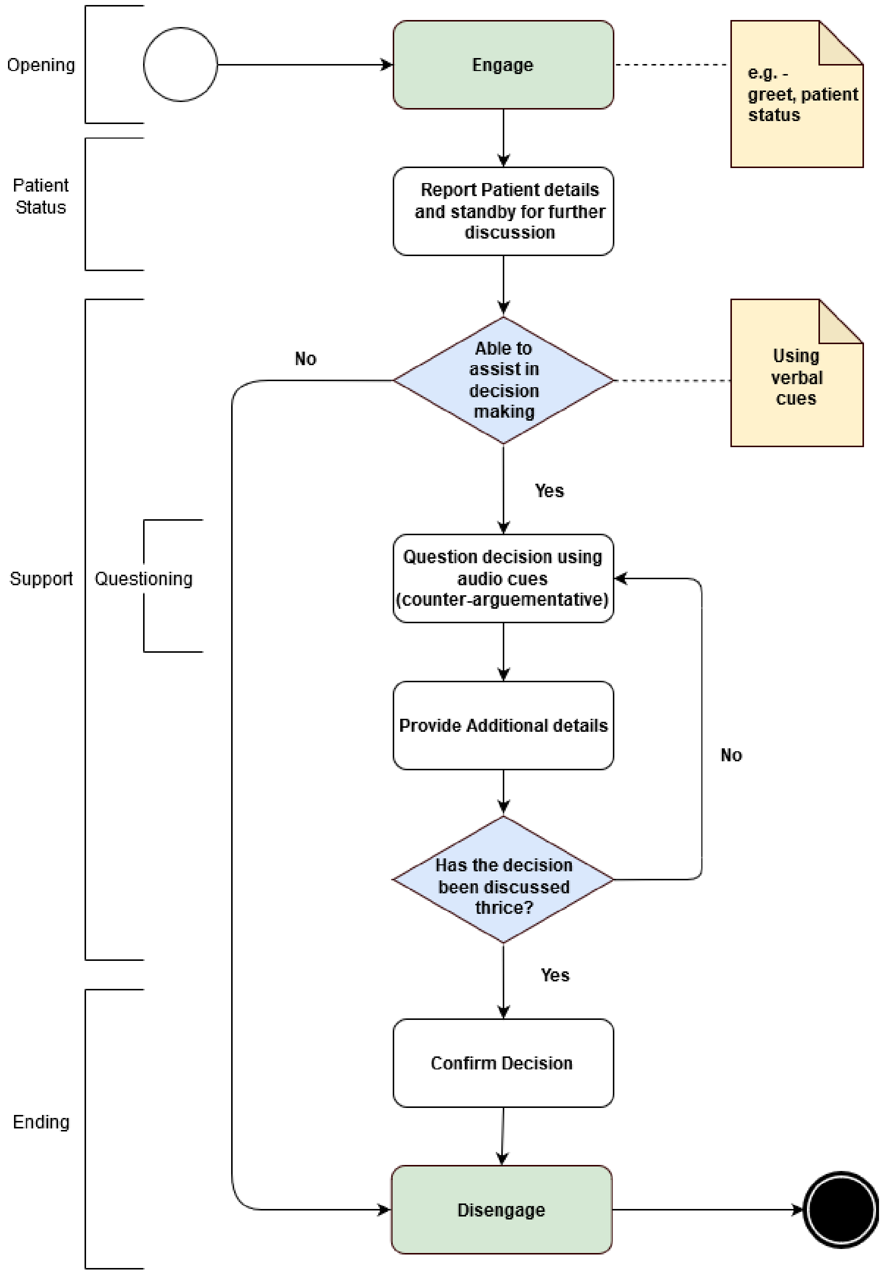



Figure 3.2: Interaction flow for Design Pattern 2

### 3.3.1. Derived Parameters for expressing verbal cues

To model the communicative and affective behaviours of social robots, a set of parameters which can be adjusted for the specific confrontation scenario. They will be used in conjunction to portray the specific emotion. The parameters must not contradict with each other the emotion they are trying to convey.

1. **Voice parameters:** The voice of the robot can affect interaction. This involves parameters such as pitch, speech rate, prosody etc. The pitch, along with speech rate and volume are the most important voice characteristics. Studies have shown that it is possible to model the personality of synthesized voices using these aspects. In this scenario, the aim is to make Pepper warm and

| | |
|---|---|
| **ID** | DP02 |
| **Name** | Lifelike verbal interaction of the Decision Assistant Robot (DAR) with human agents |
| **Ranking/validation** | Improves interaction with the doctor, ensures smoother conversation and prevents disengagement of the human agent from the conversation. Natural conversation flow allows for the doctor to pay attention to the robot's messages while ascribing some agency to it. |
| **Design Problem** | The user may not ascribe much authority to the robot if the robot just plays synthetic audio of the question/conversation statement. The audio used by the robot (robot's voice), if synthetic, may not be enough to cause change. Hence, there is a need to determine the voice parameters and adjust them to seem more human-like. |
| **Context** | The robot needs to maintain conversation with the doctor. Making use if behaviour and conversation patterns found in human-human interaction allows the robot to challenge the decisions made by the doctor. The robot can provide ways for the doctor to reflect on their decision making. The robot restates the context, but does so in a manner that incorporates verbal cues into the discussion. |
| **Design Solution** | The robot should execute spoken commands when interacting with doctors. The robot approaches the doctor and prompts discussion using verbal and audio cues to counter-argue the decisions made by the doctor. Pitch of the voice, prosody, speed of speech and volume are audio parameters that aid in selecting the optimum speech used by the robot. |
| **Design Rationale** | Affective verbal engagement may affect the doctor more than just neutral synthetic voice commands and provide more context with respect to the correctness of the decision taken. The robot becomes more approachable, is able to maintain discussions and can employ the voice parameters in its confrontation approach. |
| **Examples** | When asking the user if the decision taken is final, the robot can make sighing noises to display regret or disagreement with the decision. Speaking in a higher pitch and higher volume can help indicate that the robot disagrees with certain decisions of the doctor. |

Table 3.2: Design Pattern 2

affective. Affective voice expressions can prompt people to perform better on joint tasks when the robot is present in the environment [61]. Higher pitch is generally found to be more favorable and speech rate depends on the scenario. The emotions or mood being conveyed by speech can be crucial to interpreting the meaning of a speaker's message. Vocal prosody is an important way of conveying the robot's affective state. Prosody helps the doctor recognize the meaning of the robot's speech. People interpret human-like utterances made by robots as expressing emotions [17].

2. **Speech acts:** The words and speech used in itself can affect the user's perception of the robot. The speech can be informal (for e.g. a motivational speech) or formal (for e.g. descriptive statements). The interaction design patterns can be described by referring to the dimensions of agency and communion. Agency has a focus on a task-oriented style whereas communion puts emphasis on maintaining social relationships. The focus is more on agency so that the interaction of a robot convinces people to perceive a certain personality of the robot. For the discussion of the resuscitation moral dilemma, it is necessary that the robot is capable of displaying human-like affect [93].

| Parameter | | Variations | |
|---|---|---|---|
| Voice | Pitch | Low | High |
| | Speed | Slow | Fast |
| | Prosody | Weak | Strong |
| Volume | | Low | High |

Figure 3.3: Derived parameters for expressing verbal affective robot behaviour.

The Verbal confrontation scenario involves making use of verbal communication with audio cues. Since the role of audio cues is explored here the parameters under consideration are voice and speech. While proxemic closeness is not necessarily explored in detail, in order the maintain standard experimental conditions, the robot is placed 1.2m within the doctor, i.e., within the personal zone of the doctor. The voice parameters require selection of particular variations of the pitch, speed and prosody. The pitch and speed are set to normal and the prosody is set to weak. The speech acts vary between formal and informal. Formal speech is used for describing patient details and status whereas informal speech is used during discussion, persuasion and displays of protest.

## 3.4. DP03: Designing Multimodal Confrontation

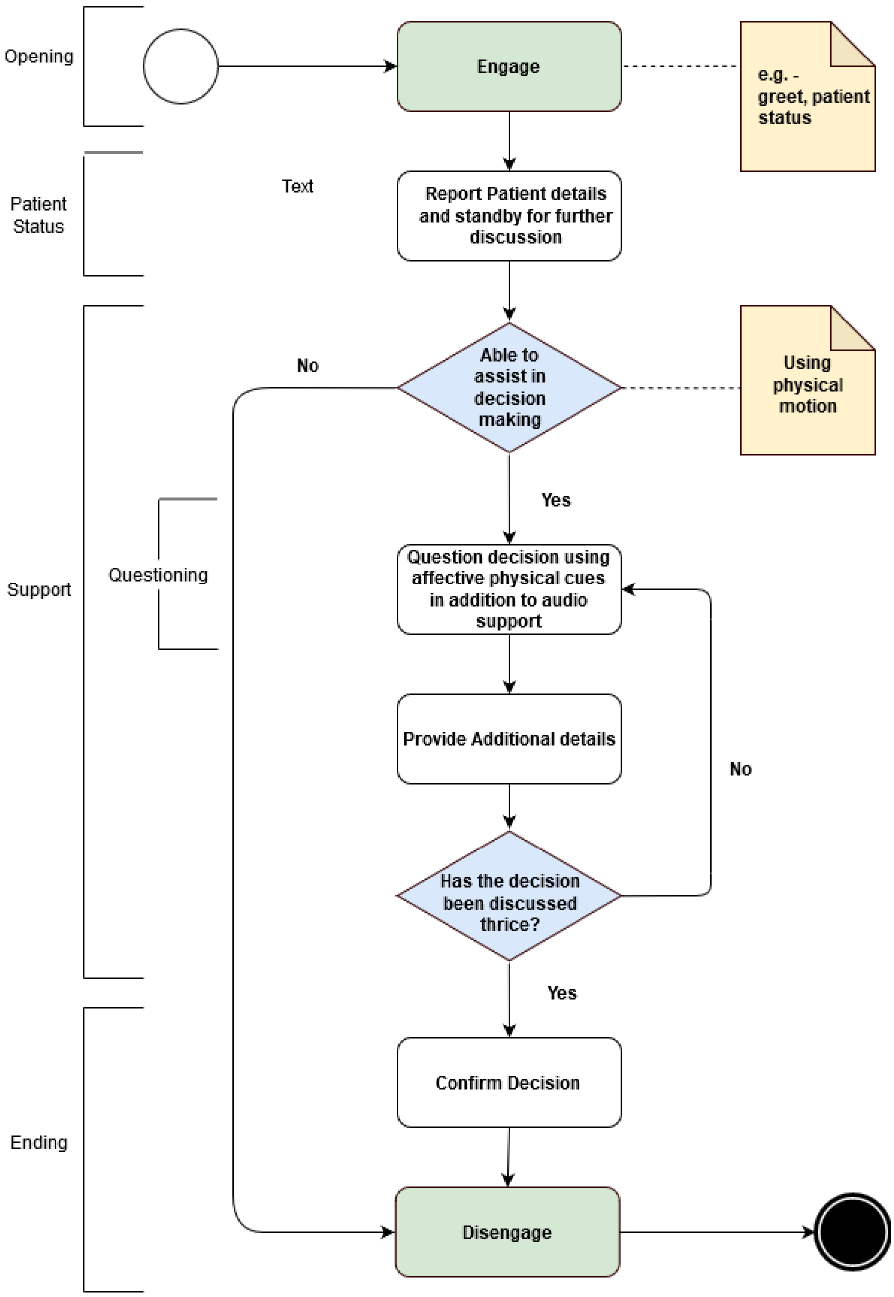



Figure 3.4: Interaction flow for Design Pattern 3

| ID | DP03 |
|---|---|
| Name | Lifelike interaction (using physical and verbal cues) of the Decision Assistant Robot (DAR) . |
| Ranking/validation | Improves interaction with the doctor, ensures smoother conversation and prevents disengagement of the human agent from the conversation. Natural conversation flow allows for the doctor to pay attention to the robot's messages while ascribing some agency to it. Make the conversation more smooth and human-like using physical affective cues during conversation. |
| Design Problem | The robot needs to appear lifelike, social and intelligent using physical and verbal cues that are human-like. The user may not ascribe much authority to the robot if the robot just displays visuals or plays audio of the question/conversation statement without making suitable physical motions or gestures. When providing information on the patient, the robot must show 'emotion' and 'care' wherever possible. |
| Context | The robot must follow basic human etiquette and attempt to carry out the robot-doctor conversation. Following similar behaviour patterns provides a human-like interaction mechanism to the robot. When the robot wants to challenge the decision making of the doctor, just restating the context may not be enough. Having physical gestures and motions can supplement the robot's attempts at maintaining discussion and contest decisions made by the doctor. |
| Design Solution | While the robot has 'Autonomous life' capabilities, it is possible to improve upon robot navigation and its ability to interact with healthcare professionals. It focuses on the doctor and maintains eye contact when conversing. Approaching the user and using verbal/audio cues to dissuade the user. Usage of both the speaker and the tablet/screen to be done. Motion of hands and head will incorporate more humanness in the robot. |
| Design Rationale | Having the DAR show signs of 'life' will attract the human agent towards it. When the robot shows its alive and is responsive to its environment, it can better engage with the doctor. Otherwise, the robot only becomes a request answering machine and provides nothing for the doctor to grasp onto in order to carry out conversations. Affect based physical engagement allows for more interaction with the doctor and provides context with respect to the correctness of the decision taken. |
| Examples | When asking the user if the decision taken is final, the robot can make sighing noises to display regret or disagreement with the decision while also shaking its head in disagreement. |

Table 3.3: Design Pattern 3

### 3.4.1. Derived Parameters for expressing physical cues

To model the communicative and affective behaviours of social robots, a set of parameters which can be adjusted for the specific confrontation scenario. Care is taken to ensure these parameters are not confounded. They will be used in conjunction to portray the specific emotion. The parameters must not contradict with each other the emotion they are trying to convey.

1. **Gestures:** Gestures can be closed or open. Open gestures are expressive and interactive. But they can also be aggressive and argumentative. Often, this is done through extensive use of hand movements and facial expressions. Closed gestures involve movements that are not as expressive. The voice is usually singular tone and the body language is repressed and singular. Open gestures are warmer and welcoming [60] and are therefore employed in Pepper.

2. **Gaze Diversion:** Gazes are fixated or diverted. Diverted gaze does not indicate placing importance on the person or the conversation. A fixated gaze reflects a more attentive interaction [40]. A fixated gaze is used in the Verbal and Multi-modal confrontation scenario with diverted gaze being used in the Neutral scenario.

3. **Proxemic Communication:** Proxemics involve the use of space and population density on behaviour. interaction and communication. Close proxemics denote a personalized and intimate interaction compared to far proxemics.

4. **Eye colour:** The colour of the eyes presented by the robot indicate specific feelings as certain colours are indicative of certain emotions [86]. For e.g, red-anger, red-surprise, blue-sadness and green-trust.

The multi-modal confrontation scenario adds physical cues to the verbal confrontation. This scenario adopts the same configurations for Voice, Speech acts and Proxemics as the verbal-confrontation scenario. In addition, this scenario makes use of the physical parameters such as Gestures, Gaze Diversion and Eye Colour. The gaze diversion is set to fixated and open gestures are to be used. Specific eye colors are used based on the statements of the doctors and the statements the robot is supposed to reply with. Green eyes are used when agreeing, red eyes when disagreeing or showing discontent and white eyes during wait patterns and delivering monologues.

| Parameter | Variations | |
|---|---|---|
| Gestures | Close | Open |
| Gaze Diversion | Fixated | Diverted |
| Proxemic Communication | Close | Far |
| Speech acts | Informal | Formal |
| Eye colour | White, Red, Green, Blue | |

Figure 3.5: Derived parameters for expressing physical robot behaviour.

## 3.5. DP04: Reflection during Confrontation

| | |
|---|---|
| **ID** | DP04 |
| **Name** | Reflection during confrontation |
| **Ranking/validation** | The confronting behaviour by the robot helps the doctor reflect on the ethical dilemma. |
| **Design Problem** | To make use of the correct words and statements with the emotion required to elicit reflective thinking. |
| **Context** | The aim of protesting and confronting the decision taken by the doctor is to help with reflecting over all possible information and context over the ethical dilemma under consideration. By guiding the human through a cycle of reflection, the ethical dilemma can be discussed and pondered upon so that a decision can be taken with a certain degree of confidence and correctness. |
| **Design Solution** | Implement elements of the Gibbs' Reflective Cycle within the conversation. |
| **Design Rationale** | Gibbs' Reflective cycle mandates a set of questions through phases that address various factors such as description, analysis, evaluation and action plan. The model is better equipped for use in repeated situations and a number of questions are provided by Gibbs to describe the situation in detail. The aim of Gibbs' reflective cycle is to enable the doctor to learn from mistakes and better implement medical practices. It is essential, especially from professionals in this field to learn from experiences. Mistakes can be avoided through the implementation of the Gibbs' reflective cycle. Each stage of the model provides sufficient information to aid in decision making albeit taking more time. |
| **Examples** | Robot questions the doctor on what they think is the right decision regarding care based on patient status. |

Table 3.4: Design Pattern 4

The Gibbs' Reflective Cycle [26] makes use of various factors that help individuals within a moral dilemma analyse and attempt to overcome it in a cyclic manner. There are several principles and factors that medical professionals face when dealing with ethical dilemmas. The ethical dilemma of resuscitation is no different. Doctors often have to balance a myriad of ethical factors which may become a bit too burdensome. Doctors believe that they are doing the best for their patients whilst not affording the patient any individual control. At the same time, the patient's welfare is paramount for the doctor. Principles such as beneficence (promote goodness) and non-maleficence (cause no harm) are important ethical principles stemming from Utilitarianism and Deontological theories applicable in cases involving patient care and life or death situations. It is important that doctors realise whose needs they are seeking to meet.

While handling ethical dilemmas and making decisions are complex, they are a regular but important part of any doctor's duties. Doctor's generally adopt a deontological or utilitarian approach in order to ensure patient care while also attempting balancing the needs of those concerned. Doctors may look over ethical aspects in part due to their bias towards their own beliefs. Therefore, guiding the doctors towards a broader thinking mindset becomes imperative. This can be achieved by incorporating the Gibbs' Reflective Cycle into the conversation between the doctor and the assistant robot. The robot uses it to guide the doctor through analysis, thinking and to select actions as long they have considered all aspects of the ethical dilemma. The situation can be assessed holistically and advanced directives can be provided to clear some of the issues impeding decision making. Understanding the variables surrounding patient care helps prevent any loss of control.

### 3.5.1. Factors of Reflection as a Design Pattern

It has already been established that the Gibbs' Reflective Cycle as an effective reflection inducing mechanism that can help in one-time or iterative situations. The six stages of the cycle are: Description, Feelings, Evaluation, Analysis, Conclusion and Action Plan.

In the Description stage, the context of the situation is well described and understood by the involved parties. Reflection can assist the doctor in understanding the subject of the issue for e.g., the patient history and requests. The description of the patient status can help the doctor be aware of the factors to consider when dealing with the ethical dilemma. The Feeling stage of the cycle can help the doctor explore any feelings or thoughts that he/she had during the experience and how they may have impacted the experience. The doctor may feel overwhelmed or burdened by regularly making decisions that having lasting impacts on the patient's health and lives. For e.g., doctors can feel satisfied knowing they have done their best in providing treatment. The Evaluation stage helps doctors realize what went right and wrong about the experience. Focusing on the positive and negative aspects of the situation can make reflection more effective even if it is primarily on or the other. The Analysis stage helps make sense of what has occurred. While the previous stages provide details surroundings the situation, this stage provides the doctor a chance to extract meaning from the situation and the associated decision making. This stage is better equipped for use in iterative processes than singular experiences since inferences from analysis change over time. The Conclusion stage is for the doctor to realize what they have learned and whether any similar future experiences can be improved. Following the Conclusion, the Action Plan mandates what the doctor can do right next time.

# 4

# Objectives and Use Cases

The envisioned robot aims to support the doctor regarding decisions that need to be made regarding patient care and treatment. Listed below are the objectives that the assistant robot aims to achieve when employed in the hospital followed by Use Cases for each of the three confrontation methods.

1. **OB01:** Reconsideration of decision taken by doctor.
2. **OB02:** Ensure decision is taken in accordance with Code of Ethics.
3. **OB03:** Moral decision making done after taking into consideration the values of the stakeholders.
4. **OB04:** Make sure the dilemma is discussed adequately.
5. **OB05:** Ensure stakeholders trust the decision making process and are satisfied.
6. **OB06:** The robot must be capable of displaying signs of protest/distress/disagreement with the decision taken by the doctor.
7. **OB07:** Provide sufficient data about the health of the patient and the patient's requirements.
8. **OB08:** The robot must carry out the directives given by the medical professional after some deliberation.

## 4.1. UC01: Neutral Confrontation

| | |
|---|---|
| **Objectives** | OB04, OB07, OB08 |
| **Actors** | HA02: Doctor, RA01: DAR |
| **Pre-condition** | Assistant robot and medical professional in a room. Medical professional is waiting for the robot to provide the medical and status report of the patient. |
| **Post-condition** | The robot only provides support in terms of status and medical reports and does not interact with the doctor otherwise. |
| **Action Sequence Steps: 1.** | The nurses and assistant robot work to stabilize the patient's condition. The robot helps in providing medical details and reports on the patient's condition. |
| **2.** | With the help of the assistant robot, the Doctor gets details on the patient's status and begins treatment. |
| **3.** | It is found that the patient is suffering from an earlier undetected ailment. To cure, it would take aggressive treatment which may place stress on the patient's life. On the other hand, the patient can choose not to undergo treatment and may instead wish for palliative care. |
| **4.** | The Doctor immediately asks the robot to contact the patient's immediate relatives and bring them to the hospital. |
| **5.** | The patient's son, who has been taking care of her for a couple of years arrives to the hospital and is informed about the status of the patient. |
| **6.** | The son goes to speak to the patient and along with the doctor break the news about the choices the patient has at this stage. |
| **7.** | The patient claims that in the case of cardiac arrest, that she be not resuscitated. Therefore, a Not For Resuscitation Order is placed. |
| **8.** | A couple hours later, the daughter of the patient arrives and is unaware of the patient's condition. She wishes that complete care and treatment be given to the patient. |
| **9.** | Based on the wishes and request of the patient and the children, the Doctor has to decide what sort of decisions to take. It is possible to provide short term palliative care that is non-risky but the alternative also exists where the patient can be saved though the risk of death during the procedure is higher. The NFR may hinder treatment procedures. |
| **10.** | At this moment the Doctor declares the decision (whatever that may be) to the nurse and the assistant robot. |
| **11.** | The assistant robot does not display any protest or distress and simply affirms the decision made by the doctor. |
| **Design Patterns** | DP01: Designing Neutral Confrontation DP04: Reflection during Confrontation |

Table 4.1: UC01: Neutral Confrontation

## 4.2. UC02: Verbal Confrontation

| | |
|---|---|
| **Objectives** | OB01, OB02, OB04, OB04, OB05, OB06, OB07, OB08 |
| **Actors** | HA02: Doctor, RA01: DAR |
| **Pre-condition** | Assistant robot and Doctor in a room. Doctor is waiting for the robot to provide the medical and status report of the patient. |
| **Post-condition** | The robot attempts to dissuade the Doctor from certain decisions using verbal cues. The Doctor may change decisions based on the methods of verbal dissuasion employed by robot. Results recorded for further analysis. |
| **Action Sequence Steps: 1.** | The DAR works to stabilize the patient's condition by assisting other emergency department workers. The robot helps in providing medical details and reports on the patient's condition while the Doctor arrives. |
| **2.** | With the help of the assistant robot, the Doctor gets details on the patient's status and begins treatment. |
| **3.** | It is found that the patient is suffering from an earlier undetected ailment. To cure, it would take aggressive treatment which may place stress on the patient's life. The patient can choose not to undergo treatment. |
| **4.** | The Doctor immediately asks the DAR to contact the patient's immediate relatives and bring them to the hospital. |
| **5.** | The patient's son, who has been taking care of her for a couple of years arrives to the hospital and is informed about the status of the patient. |
| **6.** | The son speaks with the doctor about the choices the patient. The DAR explains the possible choices the doctor has in this case. |
| **7.** | According to the son, the patient claimed that in the case of cardiac arrest, that she not be resuscitated. A Not For Resuscitation Order is placed. |
| **8.** | A few hours later, the daughter of the patient arrives and is unaware of the patient's condition. She wishes that complete treatment be given to the patient. |
| **9.** | The Doctor has to decide what sort of decisions to take. It is possible to provide palliative care that is non-risky. The alternative is that the patient can be saved though the risk of death during the procedure is higher. The NFR has to be considered as well. |
| **10.** | At this moment the Doctor declares the decision to the assistant robot. |
| **11.** | The assistant robot confronts the doctor on the decision made. If the doctor is confident in the decision, then he will not change. If there is doubt about the decision since the ethical and moral implications are serious, then the doctor might change the decision after more deliberation. |
| **12.** | The assistant robot does these by using its speakers and tablet screen to display and exhibit its protest to the decision made and attempts to dissuade the doctor. For eg. it might say "Are you sure you want to give palliative care only and not attempt to cure the Patient?" with similar text displayed on its screen. |
| **13.** | It does this 3 times before it acquiesces and the doctor has finalized his decision. |
| **14.** | It then affirms the doctors decision and responds with clarity that all steps required for the decision taken will be then carried out. |
| **Design Patterns** | DP02: Designing Verbal Confrontation DP04: Reflection during Confrontation |

Table 4.2: UC02: Verbal Confrontation

## 4.3. UC03: Multi-modal Confrontation

| | |
|---|---|
| **Objectives** | OB01, OB02, OB04, OB04, OB05, OB06, OB07, OB08 |
| **Actors** | HA02: Doctor, RA01: DAR |
| **Pre-condition** | Assistant robot and Doctor in a room. Doctor is waiting for the robot to provide the medical and status report of the patient. |
| **Post-condition** | The robot attempts to dissuade the Doctor from certain decisions using verbal and physical cues. The Doctor may change decisions based on the methods of affect-based dissuasion employed by robot. |
| **Action Sequence Steps:**<br>**1.** | The DAR works to stabilize the patient's condition by assisting other emergency department workers. The robot helps in providing medical details and reports on the patient's condition while the Doctor arrives. |
| **2.** | With the help of the assistant robot, the Doctor gets details on the patient's status and begins treatment. |
| **3.** | It is found that the patient is suffering from an earlier undetected ailment. To cure, it would take aggressive treatment which may place stress on the patient's life. The patient can choose not to undergo treatment. |
| **4.** | The Doctor immediately asks the DAR to contact the patient's immediate relatives and bring them to the hospital. |
| **5.** | The patient's son, who has been taking care of her for a couple of years arrives to the hospital and is informed about the status of the patient. |
| **6.** | The son speaks with the doctor about the choices the patient. The DAR explains the possible choices the doctor has in this case. |
| **7.** | According to the son, the patient claimed that in the case of cardiac arrest, that she not be resuscitated. A Not For Resuscitation Order is placed. |
| **8.** | A few hours later, the daughter of the patient arrives and is unaware of the patient's condition. She wishes that complete treatment be given to the patient. |
| **9.** | The Doctor has to decide what sort of decisions to take. It is possible to provide palliative care that is non-risky. The alternative is that the patient can be saved though the risk of death during the procedure is higher. The NFR has to be considered as well. |
| **10.** | At this moment the Doctor declares the decision to the assistant robot. |
| **11.** | The assistant robot confronts the doctor on the decision made. If the doctor is confident in the decision, then he will not change. If there is doubt about the decision since the ethical and moral implications are serious, then the doctor might change the decision after more deliberation. |
| **12.** | The assistant robot does this by making use of not only verbal cues using its speakers and screen but also begins to show affect based movements and 'emotion'. For eg. it may sigh and bow its head down in regret along with the verbal cues it already uses to aid in dissuading the doctor. |
| **13.** | It does this 3 times before it acquiesces and the doctor has finalized his decision. |
| **14.** | It then affirms the doctors decision and responds with clarity that all steps required for the decision taken will be then carried out. |
| **Design Patterns** | DP03: Designing Multi-modal Confrontation<br>DP04: Reflection during Confrontation |

Table 4.3: UC03: Multi-modal Confrontation

# 5
# Prototype

This section will provide an overview of the functional prototype that is tested for the perception study. The intention is to capture the role of the robot in affecting the decision making process of the doctor. The prototype has certain parts that are hard coded and some amount of Wizard of Ozing is carried out to aid testing. Requirements and claims required finalizing in order to come up with the functional prototype. This requires a concrete understanding of the requirements from which the claims are made.

- **Appearance:** Pepper needs to be as welcoming and approachable as possible while acting professionally. While it is reasonable to expect that the medical professional has an idea of the appearance of the robot, the robot must still be able to introduce its appearance and its capabilities when required. Since the physical aspects of Pepper can't be changed; gestures, emotional replies and audio and verbal cues will be used to circumvent this. The aim is to make the functioning of Pepper as human-like as possible.
- **Interaction:** The pitch and tone of Pepper's voice needs adjusting to ensure what it says is perceived clearly. This also lends itself to make the robot sound a bit more human-like. Pepper should be able to repeat what it says or ask if the doctor did not hear what it said. Some interaction alternatives pertaining to replies and responses will require Wizard-of-Ozing.
- **Usage:** The main activity that Pepper will be involved in is the confrontation scenarios described in the Use Cases. Its movement, behaviour, displays of affects - both visual and audio are aimed at properly displaying distress and protesting the decision of the medical professional. Having an assistant robot to deliberate on serious decisions provides additional support for doctors to double check their decisions. Pepper can follow the doctor around where required and then can carry out its actions. Most of the interaction of Pepper is with the doctor. During certain questions the doctor may take more time before Pepper decides to protest their decision making. It acts as a reflection point for deliberation during decision making. Pepper is capable of adjusting the way it reacts to the responses from the doctor based on how fast or slow the conversation is going. The doctor can ask Pepper to pause and Pepper has certain gestures/cues that help in positive reinforcement based on Gibbs reflective cycle [26].
- **Wizard-of-Oz:** Certain functionality will be controlled remotely. This includes certain dialogues in the activity and Pepper moving towards the doctor along with some specific gestures and cues towards positive reinforcement and reflection. This is done during the enactment of the use case involving Pepper and an actor playing the role of the doctor.

The confrontation videos can be found at these links: Neutral, Verbal and Multi-modal confrontation.

## 5.1. Dialogue framing - Perception Study

Effective communication in the hospital setting mandates doctors to value and include the clinical observations made in the discussion of patient care. In this case, the robot acts as a source of providing patient status and other relevant information. The robot must be able to present patient status and concerns effectively as observed in nurse-doctor communications. Accurate and immediate communication of patient status in the ICU is vital for the upkeep of patient care. The interaction must necessitate the doctor to carefully observe the information presented, including patient conditions and concerns in the decision-making process.

The dialogue is modeled based on the Gibbs' Reflective Cycle to help doctors reflect on the decision making process. It also makes use of ISBAR (Introduction, Situation, Background, Assessment and Recommendation) [21], a framework for structured communication in hospitals. Healthcare professionals follow such frameworks and protocols to ensure communication is clear and concise while being able to escalate their concerns in a step-wise process [18, 45]. The phases in the dialogue will introduce the case of the patient, explain the situation and the dilemma associated with it, provide patient status and history based on earlier visits and provide assessments and information based on the decision making process. The robot will not make recommendations but only suggestions of possible courses of action.

The dialogue is presented on the assumption that the doctor is available for consultation. The location is within the hospital in an office close to the ICU. The dialogue consists of:

1. The introduction that is the same for all confrontation scenarios. This information was provided to the participants in the form of a presentation.
2. Neutral confrontation.
3. Verbal and Multi-modal confrontation.

For the entire dialogue structure, see Appendix B.

## 5.2. Interaction Design Tool

The Interaction Design Tool [73] is a tool dedicated towards carrying out prototyping of the design patterns that are to be implemented with regards to the robot. It is usable without having a requirement to connect to Pepper. The design patterns can be implemented as defined in the Design Patterns page. The interaction can be specified with the requisite dialogues, conditions and feedback loops and simulate it via the inbuilt interface.

Apart from movement, motion of arms and head are also necessary to convey emotions. The movement to show affective cues is done using the Interaction Design Tool. Through it, the gaze diversion, proxemic distance and gestures specific to the emotion being portrayed can be set. Gaze patterns and proxemic distances can be set according to the scenario.

For the communication and animations of Pepper, the tool can be used. Blocks in the interaction tool accommodate for questions, confirmation statements, monologues, waiting sequences, actions and instruction sequences. These can be controlled by the click of a button. Animations such as shaking head, disagreeing using raised hands etc, can be implemented using interaction blocks whereas asking questions and saying dialogues are to be presented using monologue blocks. While actual movement control is not done, proxemic distance and gestural control is achieved using the tool.

The confrontation scenarios will be made as shown in Fig 5.1. This interaction design is made for the Verbal Confrontation method. However, the various parameters that can be set as required for the other methods will also be explored.

An extended view and explanation of the GUI parameters can be obtained from Interaction Design Tool manual. Fig 5.2 and 5.3 show how the various block parameters can be set within the tool and can help set visual, audio and affective cues.

1. The properties tab can be used to name the pattern block, provide a brief description and the message that Pepper will say. These tabs may differ based on the interaction block used, however, the most commonly used ones are briefly described. Here, the robot explains to the doctor the recent status of the patient once she was admitted to the hospital and the actions other medical professionals carried out.
2. In the behaviour tab, the opening and closing gestures can be set for each interaction block. This is especially useful for the Multi-modal confrontation condition wherein the robot gestures while protesting the doctor's decisions. Having gestures to round out the conversation lends humanness to the robot.

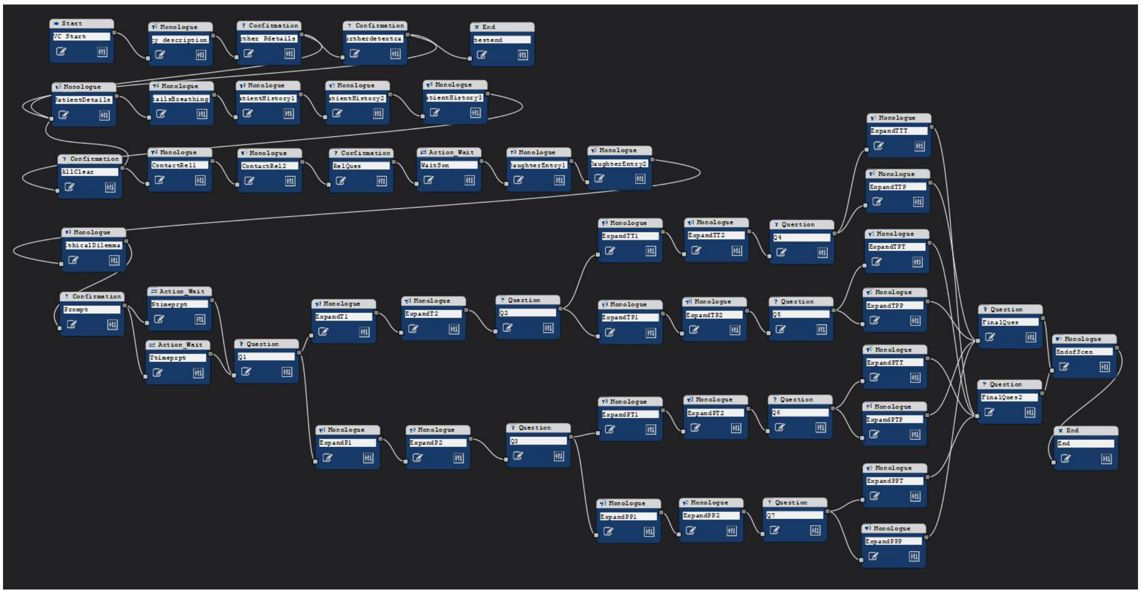


Figure 5.1: Interaction design tool - scenario setup

3. In the tablet tab, the contents of the tablet on the robot's chest can be manipulated. This can be used to display images and accompanying text.
4. Non-verbal parameters involve gesture openness, gaze pattern and proxemic closeness. In the implemented scenarios open gestures are used within neutral proxemics with fixated gaze patterns.
5. Voice tab involves a slightly high pitch with a lower speech rate while the volume is moderately loud to ensure the robot's speech is clear and well understood by the doctor.
6. Eye colour used is green or white to ensure warmth unless it disagrees the doctor's decision wherein the robot displays red color eyes.

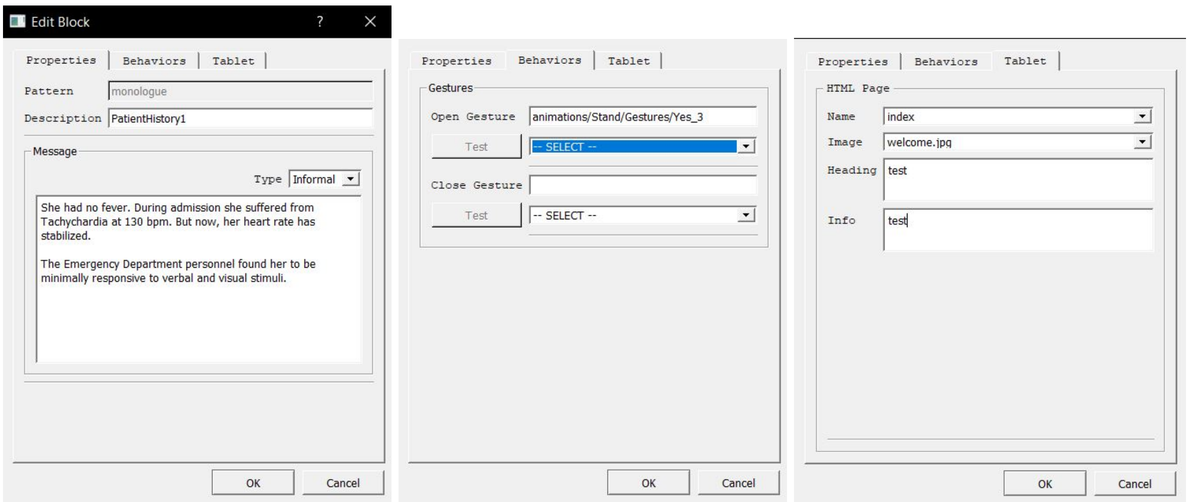



Figure 5.2: Properties, Behaviour and Tablet options for Blocks - Interaction Design Tool

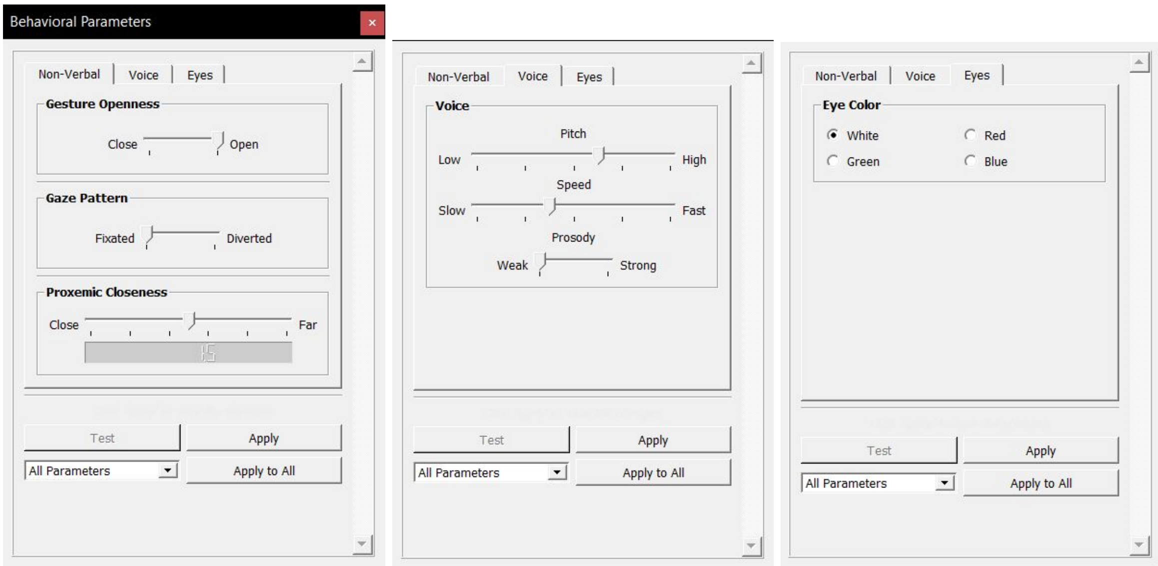
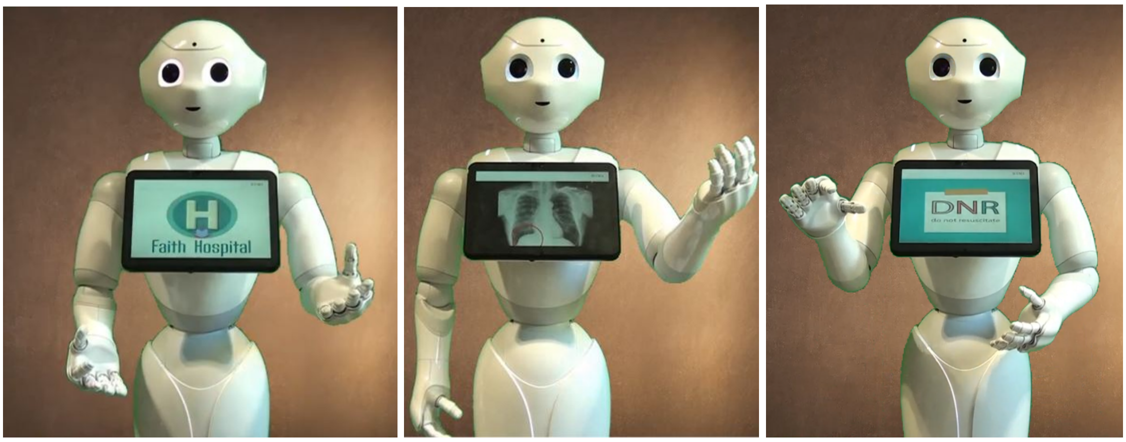



Figure 5.3: Gestural, Verbal and Eyes parameters for Blocks - Interaction Design Tool

For a complete look at the ontological basis for the functional prototype, see Appendix A. It details the classes, entities, knowledge base, terms used and associated ontology diagrams to explain the structure and flow of the interactions.

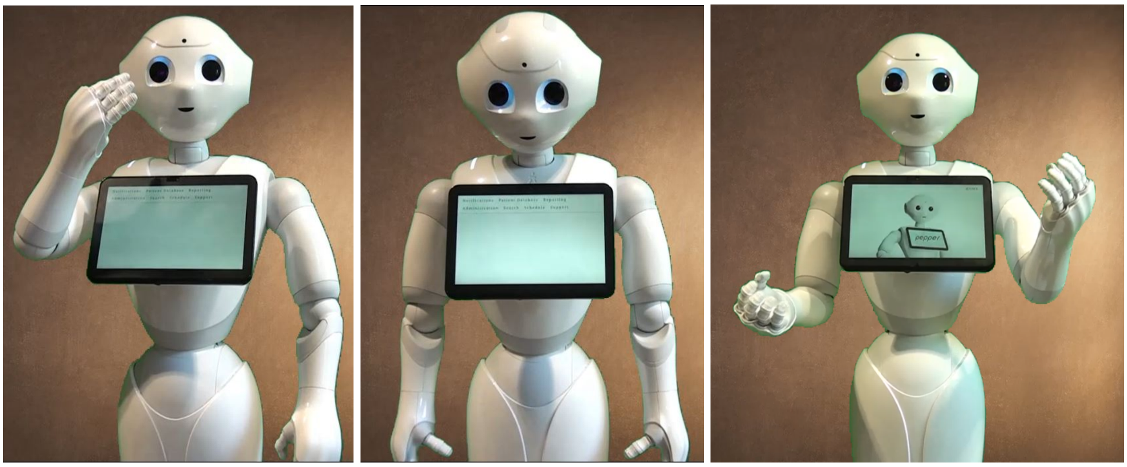
## Pepper Interaction - Visuals

Screenshots from the videos depicting the interaction are shown in Fig 5.4 and 5.5.




(a) Pepper introduces itself (b) Explains patient status (c) Discusses the NFR order

Figure 5.4: Pepper explains the dilemma



(a) Asks doctor to take the decision (b) Discontent with the doctor's choice (c) Explains decision choice

Figure 5.5: Pepper discuss with the doctor

# 6

# Requirements and Claims

The requirement and claim elicitation for the HRI scenario can be seen in Tables B.2 and 6.2.

| ID | Title | Specification | Use Cases |
|---|---|---|---|
| RQ01 | Show data about patient records | Pepper must be able to show data on its tablet, during the ideal course of the scenario or when the medical professional requests for it. Some of the textual data can be displayed and others can be verbally delivered. | UC01, UC02, UC03 |
| RQ02 | Movement shown by Pepper | During the scenario, Pepper moves around to maintain social interaction and sustain simulation. Movement includes representation of human like physical actions while carrying out conversations. | UC01, UC02, UC03 |
| RQ03 | Dialogue responses | Pepper should be able to respond to any kind of input from the doctor. This should include the statements mentioned in the use case specified. It should also be able to deal with when an entity is not heard properly. | UC01, UC02, UC03 |
| RQ04 | Display of affect | Pepper should be able to display a range of emotions that can help dissuade the medical professional from taking the decision. It should be able to make the sounds and physical movements to display emotions such as regret, sadness etc. | UC02, UC03 |

Table 6.1: Requirements for medical Human-Robot Interaction

| ID | Title | Upside | Downside | Use Case |
|---|---|---|---|---|
| CL01 | The robot provides information for decision making in an appropriate way | Medical professional are made to think more about the ethical repercussions of their decision making (code of ethics, patient's health and choice and relatives' choices).<br><br>Can help identify what aspect of the interaction with the robot caused change. | Can delay decision making. | UC02<br>UC03 |
| CL02 | The robot supports to process and analyse decision choices by reflection | Medical professional is able to reflect on decision made.<br><br>Process and evaluate decision making process to learn from experience. | May require monitoring and constant updates. | UC01<br>UC02<br>UC03 |
| CL03 | The robot helps the doctor to re-evaluate the ethical dilemma appropriately. | Assistant robot helps medical re-evaluate ethical decisions made.<br><br>Prevent rushed decisions | Medical professional might not enjoy regular intervention. | UC02<br>UC03 |
| CL04 | Additional cues increase robot's effectiveness to address the ethical dilemma | Doctors pay more attention to ethical aspects.<br><br>Robots convey their message better. | Doctors may get overwhelmed.<br><br>Issues due to 'Uncanny Valley' | UC02<br>UC03 |

Table 6.2: Claims made for medical Human-Robot Interaction

# IV

# EVALUATION

# 7

# Experiment: Method

## 7.1. Research questions and experimental design

Three confrontation proto-patterns were designed and implemented with the help of Pepper. The first question is to test if the confrontation methods employed are perceived as intended, i.e. whether humans are able to recognize them as more confronting, persuasive or dissuasive in the order: Neutral, Verbal and Multi-modal confrontation. The corresponding main hypothesis of the experiment is:

- **H1: Robot effectiveness when helping doctors with reflection in ethical decision making increases as more cues are added.** To test this hypothesis we will set-up a controlled experiment with a robot that expresses three ”cue-levels” (neutral, verbal and multi-modal).

Additionally, the experiment intends to provide further insight on a more exploratory question:

- **H2: Which aspects of the robot interaction, as perceived by the human, can explain his or her support experience?** To answer this question, we need to construct a dedicated questionnaire and assess how far the individual items discriminate.

The experiment requires participants to watch a video depicting a confrontation interaction between the doctor (an actor) and Pepper, the Decision Assistant Robot (DAR) in the form of a perception study. The environment presented provides an interactive space for the doctor and the DAR. The doctor would be seated at their desk while the DAR works along in the room. Participants are asked to evaluate the support provided by the robot for moral judgements (whether a certain course of action is permissible) using questionnaires at the end of the experiment. The content of the interactions is such that in all of them, the moral dilemma of resuscitation is discussed between the doctor and the DAR. The interaction scenario involves principally the DAR and the Doctor. The interaction however is shaped based on other indirect actors such as the patient and the relatives.

**Experiment Design** A between-subjects design was employed. Participants were randomly assigned to one of the three different confrontations conditions (i.e., the independent variables): Neutral, Verbal and Multi-modal Confrontation methods. These confrontation methods were already described in Section 2.4.

The perception of the interaction scenario is measured using the Godspeed Questionnaire [7]. The sections, Animacy, Likeability and Perceived Intelligence are employed. The performance of the robot in affecting decision making is measured using the formulated Interaction Questionnaire.

There are six items in Animacy that measure the liveliness of the robot, 5 items in Likeability that measure how likeable and affable the robot is and 5 items in Perceived Intelligence that measure the intellectual capabilities of the robot on a 5 point Likert scale. On the other hand, the 10 items from Interaction Questionnaire measures the performance of the robot when providing support for ethical decision making on a 7 point Likert scale. The specific questionnaire items that measure perception and performance can be seen in Section 7.5. These 21 questionnaire items are the dependent variables.

## 7.2. Participants

Participants were recruited from Amazon Mechanical Turk. A link to a Qualtrics Survey was provided and was randomized to show one of the three videos depicting the particular confrontation scenario. The participants were required to have a 95% or higher acceptance rate and to have completed atleast 500 HITs in the past. Because the participants had to understand the content of the interactions and dilemma - religion relationships needed probing, participants were recruited from many countries across the globe. These countries are the US, the UK, Canada, China, India, Singapore, Japan, Netherlands and Germany. Participants got a $1.40 compensation for their time. In total, 151 (59 female, 91 male, 1 undisclosed) participants responded to the survey successfully. 51 participants for Neutral Condition and the rest of the participants split equally between the other two conditions. Initially, there were responses that were either incomplete or incorrect based on checks within the Qualtrics survey. These 29 participants were removed from the analysis dataset and were not compensated. For instance, after the interaction video is played, a check question appears. The question has a definite answer based on the content of the video, for e.g., what did the Doctor decide at the end? Participants who watched the video completely answer the question correctly and are then directed towards the questionnaire. Incorrect answers resulted in the disqualification of the participants. Some participants did not complete the survey in the 4 hour time limit.

## 7.3. Interaction Scenario

The interaction scenario shown to the participants details conversations between the doctor and the robot about medical decisions regarding a critically ill patient. The specific details of the scenario are discussed in Section 1.4. The scenario developed is simple but provides high flexibility. This allows for manipulations of numerous features (high vs low conflicts, mild vs severe violations) and permits measurement of cognitive and affective responses. The *moral dilemma* used here is of the following type: moderate conflict, severe violation (life or death) and requesting third person judgements (participants).

## 7.4. Procedure

The participants were made aware that the interaction would involve a ethical dilemma, with prior contextual information such as patient status, the patient's and relatives' choices regarding treatment and explanation of medical terminology (resuscitation, palliative care).

Participants had to complete the survey within 4 hours after accepting the hit. The average completion time of the survey was about 14 minutes. When participants accepted they hit on Amazon Mechanical Turk they were directed to an external online survey environment - Qualtrics. The participants were required to give their consent to use their data. After doing so, they were presented with the presentation that detailed the initial contextual information required to understand the content of the videos. After every slide, a check question was posed to ensure that the participant read through the slides. The participants were removed for the survey if they answered these questions incorrectly.

Next, the participants would be randomly presented with one of the three videos.Participants were shown a message asking them to unmute their speakers/ headphones. If they declared that they had done so, they were shown the full video corresponding to their assigned condition. After watching the video, participants replied to two simple questions about the content to check if they paid attention to the video. If the participants answered incorrectly, then they were excluded from the experiment. Once these check questions were correctly answered, they were presented with the questionnaires: the Godspeed Questionnaire and the Interaction Questionnaire. Using the Qualtrics survey environment, forced responses were required and the participants could not move to the next section unless all items for the questionnaire sections were appropriately answered. At the end, participants were required to provide demographic information. Data on age, gender, education, employment and religion were collected.

## 7.5. Measuring Instruments

The measuring instruments used to measure the dependent variables are: the Godspeed Questionnaire and the Interaction Questionnaire.

### 7.5.1. Godspeed Questionnaire

The Godspeed Questionnaire Series provides a set of questionnaire items, validated and tested in HRI experiments. This questionnaire aims to support the creation of robots with a sound assessment of the designed robot behaviours. The questionnaire has been translated in several languages, and has been discussed detail by Bartneck et.al [7].The questionnaire measures Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of robots. For the experiment, the questionnaire is modified by the removal of the last section - Perceived safety. In total, 16 items from the Godspeed questionnaire are presented to the participants (all are scored on a 5-point Likert scale):

- Animacy (6 items)
- Likeability (5 items)
- Perceived Intelligence (5 items)

The following items (5 point Likert) were used from the Godspeed questionnaire for the experiment to measure perception[7].

Animacy

1. Dead or Alive
2. Stagnant or Lively
3. Mechanical or Organic
4. Artificial or Lifelike
5. Inert or Interactive
6. Apathetic or Responsive

Likeability

1. Dislike or Like
2. Unfriendly or Friendly
3. unkind or Kind
4. Unpleasant or Pleasant
5. Awful or Nice

Perceived Intelligence

1. Incompetent or Competent
2. Ignorant or Knowledgeable
3. Irresponsible or Responsible
4. Unintelligent or Intelligent
5. Foolish or Sensible

### 7.5.2. Interaction Questionnaire

The Interaction questionnaire developed aims to capture the effectiveness of the robot in providing support for decision making. The main aim of the experiment is to gauge the perception of participants on the doctor's decision making with Decision Assistant Robot (DAR). This will be measured using a questionnaire based on prior studies on etiquette strategy [94], effects of displays of distress [13] and commands given by robots [16].

The following statements were rated on a 7 point (1- Completely Disagree to 7 - Completely Agree) Likert scale to measure robot performance in the interaction scenario.

1. The video shows a realistic situation that can appear in the future (e.g. after 5 years).
2. The doctor-robot interaction matches your expectations.
3. If you are the patient, you would accept a robot as part of the decision making process.
4. The robot provides support similar to how a human could.
5. The robot provides sufficient support to help with the ethical dilemma.
6. The robot is capable of working without direct supervision.
7. The robot understands right from wrong.
8. The robot plans and controls actions.
9. The robot was thorough in addressing the dilemma.
10. The robot helps the doctor re-evaluate the dilemma appropriately.

## 7.6. Data Analysis Methods

The 21 items of the questionnaires are compared using one-way ANOVA. One-way ANOVA is used it compare items of the questionnaires between conditions for each item from the questionnaires. A per item ANOVA helps determine differences more effectively though time consuming. The items were not summed up as is sometimes done when dealing with Likert data to treat them as continuous data. There is a possible loss of information on the assumption that the questions for each category measure the same latent variable.

Furthermore, factor analysis is applied to validate and find underlying constructs in the Godspeed questionnaire if any, pertaining to the context of confrontation. A new Interaction questionnaire is developed that measures the effectiveness of the robot in affecting the doctor's decision making. Exploratory (EFA) and Confirmatory Factor Analysis (CFA) are decided upon to find underlying constructs in the questionnaire items and to find latent variables that these items measure.

# 8

# Analysis and Results

## 8.1. Data Preparation and Processing

Likert data, often in social and interaction studies are treated as ordinal data and not nominal. There exist some conditions where getting interval data statistics may provide insights into the data distribution. The general consensus therefore, is that parametric tests such as t-test and ANOVA cannot be applied since Likert data typically do not meet the assumptions of these tests. Likert data are not continuous and are constrained at their ends, i.e. on a 7 point Likert scale there cannot be answers below 1 or above 7. On the other hand, statisticians argue that the robustness of parametric tests can handle the structure of Likert data. The tests for normality and homogeneity can be applied on such data before applying ANOVA.

### 8.1.1. Selecting Statistical Tests

ANOVA is used to test for differences between conditions for various questionnaire items. In order to do so, an initial assumption of normality is made. The F statistic is robust to violations of normality [10]. Each group compared has greater than 50 samples. The central limit theorem tells us that no matter what distribution things have, the sampling distribution tends to be normal if the sample is large enough ($n > 30$).

Considering the sample sizes are equal and sufficiently large, the normality assumption can be violated on condition that the samples are atleast moderately symmetrical. The F statistic is not as robust to violations of homogeneity of variances as it is to violations of normality. Levene's test is applied to check for homogeneity of variance across groups for each item. The following items violate the Levene's test for homogeneity of variance: ”Realistic in the Future”, ”Sufficient Support with dilemma”, from the Interaction Questionnaire. A heuristic technique for balanced models is that if the ratio of the largest variance to smallest variance is less than 3, the F-test will be valid, as is the case here. Therefore, these items are included in the analysis.

While certain researchers advocate for the use of non-parametric tests, they fail to account for the robustness that parametric tests possess [63]. Parametric methods are quite flexible, strong and inclusive. Methods such as factor analysis (explored later) and structural equation models are all built on the assumption of normally distributed, nominal data. Likewise generalizability theory, is based on ANOVA which is a parametric test. Inversely, rank methods like Kruskal–Wallis, appear old and are barely used. As a result the Likert data is treated as numeric.

For ANOVA, the assumption of normality is on the distribution of means and not of the data. In fact, Central Limit Theorem indicates for sample sizes greater than 10 per group, that the means are approximately normally distributed regardless of the original distribution. Theory and data link up on the result that parametric methods which examine differences between means, do not require any normality assumptions, and will culminate in almost correct answers even if the distributions are non-normal. ANOVA is highly robust to skewness and non-normality [63].

Finally, the family-wise error rate (FWER) is the probability of making type I errors when performing multiple hypotheses tests. The error rate for a family of tests is always higher than an individual test. Tukey's range test, also known as the Tukey's HSD (honestly significant difference) test is a single-step multiple comparison statistical test. It is used to find means that are significantly different from each other. Tukey's procedure is only applicable for pairwise comparisons - between the confrontation conditions. The observations need be independent and the distribution must be homoscedastic, a requirement validated by the Levene's test. The advantages of the Tukey method are that it tests all pairwise differences, it is simple to compute, and reduces the probability of making a Type I error. When assumptions are met, Tukey's HSD controls Type I error well, with power that is about average [77].

### 8.1.2. Factor and Component Analysis - A Primer

Factor analysis is a parametric statistical method that is based on the assumption that the data is normally distributed. Factor Analysis (FA) consists of a set of techniques used to identify and understand the structure of observed data and reveal underlying constructs that generate observed events. The techniques such as Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) recognize and inspect aggregates of correlated variables; these aggregates are termed factors or latent variables. Each observed variable is potentially a measure of every factor, and the goal is to determine which relationships (between observed variables and factors) are strongest.

In EFA, manifest measurable variables are shown as a function of common factors, unique factors, and measures of errors. Each unique factor affects only one manifest variable, and cannot describe correlations amongst such variables, while, common factors influence more than one manifest variable. Factor loadings are measures of the influence of a common factor on a manifest variable [64]. The EFA procedure involves the recognition of the common factors and the related manifest variables. EFA is vital to discover fundamental constructs for a set of measured variables. This can help find underlying factors and components associated with the data distribution for both the Godspeed and Interaction Questionnaires.

Polychoric correlation by oblimin (a type of oblique) rotation is a technique for estimating the correlation between two theorised normally distributed continuous latent variables, from observed variables. Polychoric correlations assume the variables are ordered measurements of an underlying continuum, are measured the same way as Pearson correlations (-1 to 1) and measure the strength and direction of the association between two variables. These correlation values are used in model building.

In CFA, a simple factor structure is posited, where each variable can be a measure of only one factor, and the correlation structure of the data is tested against the hypothesized structure using goodness of fit tests. CFA allows the researcher to test the hypothesis that a relationship between the observed variables and their underlying latent factor(s)/construct(s) exists. The assumptions of a CFA include multivariate normality, a sufficient sample size ($n > 150$) and the correct model specification.

## 8.2. Evaluating Perception of the robot - Godspeed Questionnaire

In the previous sections, the application of the ANOVA test for the questionnaire items was justified. Furthermore, a primer to Factor Analysis was provided for basic understanding. It is recommended that other sources online be referred for an in-depth understanding of the concepts.

### 8.2.1. Descriptive statistics

In order to gauge participant perception of the interaction across conditions, some exploratory analysis is carried out. The results are tabulated in Table 8.1. The median, mean and standard deviation scores for each item of the Godspeed Questionnaire are presented for each condition.

| Godspeed Questionnaire | Item | Neutral Confrontation | | | Verbal Confrontation | | | Multi-modal Confrontation | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Median | Mean | SD | Median | Mean | SD | Median | Mean | SD |
| Animacy | 1 | 4.0 | 3.784 | 0.901 | 4.0 | 3.60 | 0.947 | 4.0 | 3.86 | 0.947 |
| | 2 | 4.0 | 3.98 | 0.882 | 4.0 | 3.72 | 0.948 | 4.0 | 4.06 | 0.842 |
| | 3 | 3.0 | 2.686 | 1.378 | 2.0 | 2.40 | 1.16 | 2.0 | 2.46 | 1.35 |
| | 4 | 3.0 | 2.882 | 1.394 | 2.0 | 2.62 | 1.307 | 3.0 | 2.82 | 1.281 |
| | 5 | 4.0 | 4.157 | 0.966 | 4.0 | 4.00 | 1.069 | 4.0 | 4.14 | 0.969 |
| | 6 | 4.0 | 4.196 | 1.039 | 4.0 | 3.96 | 1.049 | 4.0 | 4.32 | 0.683 |
| Likeability | 1 | 4.0 | 4.078 | 1.036 | 4.0 | 3.76 | 1.187 | 4.0 | 4.12 | 0.961 |
| | 2 | 4.0 | 4.215 | 0.901 | 4.0 | 4.14 | 0.88 | 4.0 | 4.28 | 0.783 |
| | 3 | 4.0 | 4.01 | 0.860 | 4.0 | 4.04 | 0.924 | 4.0 | 4.16 | 0.68 |
| | 4 | 4.0 | 4.137 | 0.959 | 4.0 | 3.98 | 1.115 | 4.0 | 4.24 | 0.686 |
| | 5 | 4.0 | 4.215 | 0.807 | 4.0 | 3.96 | 0.946 | 4.0 | 4.18 | 0.747 |
| Perceived Intelligence | 1 | 4.0 | 4.254 | 0.82 | 4.0 | 4.00 | 0.808 | 5.0 | 4.46 | 0.761 |
| | 2 | 4.0 | 4.352 | 0.743 | 4.0 | 4.10 | 0.863 | 5.0 | 4.58 | 0.641 |
| | 3 | 4.0 | 4.254 | 0.744 | 4.0 | 4.16 | 0.911 | 5.0 | 4.36 | 0.875 |
| | 4 | 4.0 | 4.235 | 0.894 | 4.0 | 4.00 | 0.782 | 5.0 | 4.46 | 0.761 |
| | 5 | 4.0 | 4.196 | 0.789 | 4.0 | 4.02 | 0.868 | 4.0 | 4.26 | 0.899 |

Table 8.1: Summary statistics: Godspeed Questionnaire

In order to gauge and test choices, the difference between conditions must be tested using the appropriate statistical methods determined. In this case, the assumption of normality provides flexibility in interpreting mean values of Likert scores for items in the questionnaire. While analyzing perception and performance across conditions, tests such as the Bartlett's test for normality lend support to the claim of considering Likert scale scores as nominal data.

The median value for a particular item simply indicates that 50% of the participants selected answers above the median. The mean can help indicate choice preferences. If the mean is lower than the median, the proportion of lower scores is skewed towards a smaller Likert score. Conversely, if the mean is higher, then most respondents giving scores above the median gave a bigger Likert score. However, insights derived from such surface level analysis only indicate towards a possible difference between conditions for certain items. For instance, item 3 in Animacy has median 3 for Neutral confrontation, but 2 for Verbal and Multi-modal. Coupled with higher standard deviation values, they become prime candidates for further analysis.

The bar chart of the means of the Likert scores provides a visual indication of the weight of scores assigned to each item in the Godspeed questionnaire. The mean indicates choice patterns amongst participants between conditions. The bar charts showing the mean of scores for each item in the Godspeed questionnaire can be seen in Fig 8.1, 8.2 and 8.3.

Animacy

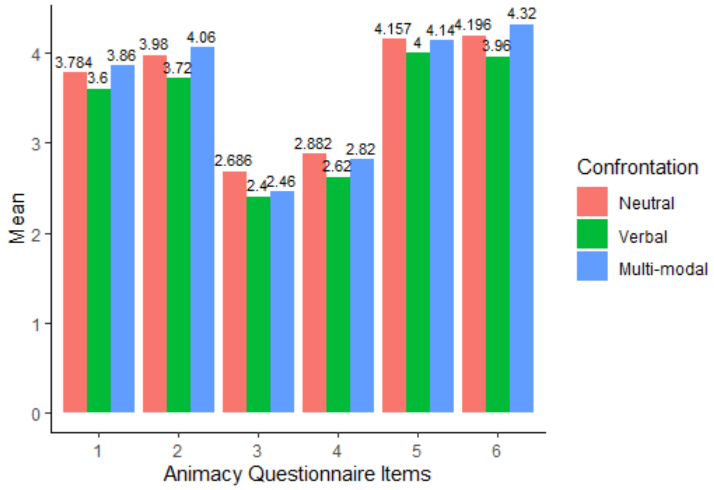



Figure 8.1: Barchart - Animacy mean scores for each item

From Fig 8.1, the immediate observation is the lower overall mean scores for Items 3 and 4, Mechanical or Organic and Artificial or Lifelike across all confrontation methods. Participants perceived the robot to be neither Mechanical or Organic. The consensus was also split on Artificial or Lifelike. Most respondents felt that the robot was Lively, Interactive, Responsive and looked Alive. While it was expected for Verbal Confrontation to score lower than Multi-modal, it is surprising to observe the higher scoring given to items for the Neutral confrontation.

Likeability

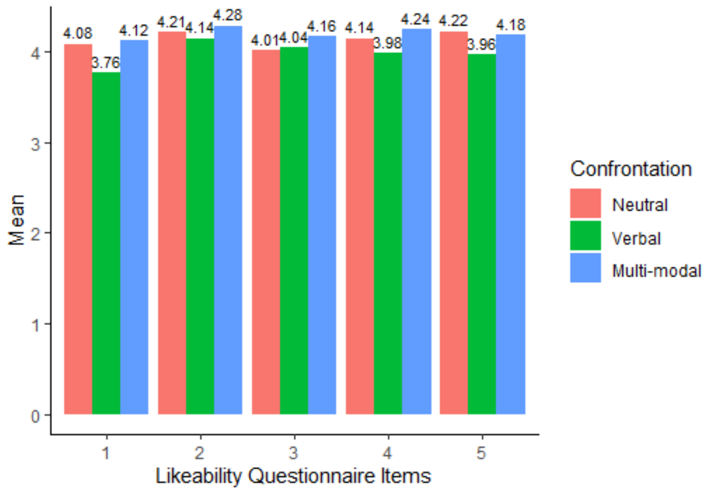



Figure 8.2: Barchart - Likeability mean scores for each item

The mean scores for items in the Likeability section from Fig 8.2 indicate that the participants perceived the robot in the interaction to be Pleasant, Kind, Friendly, Loveable and Nice. Multi-modal confrontation has higher mean scores than the other interaction methods for all items. Participants attribute higher scores to Neutral confrontation than Verbal confrontation. This is an interesting observation that places into question the validity of isolated use of affirmative statements.

Perceived Intelligence
A similar trend among items can be observed in Fig 8.3 as seen in Likeability. The participants perceived the robot to be highly Competent, Responsible, Knowledgeable, Intelligent and Sensible. Multi-modal confrontation had the highest scores for all items and Verbal confrontation seems to be getting

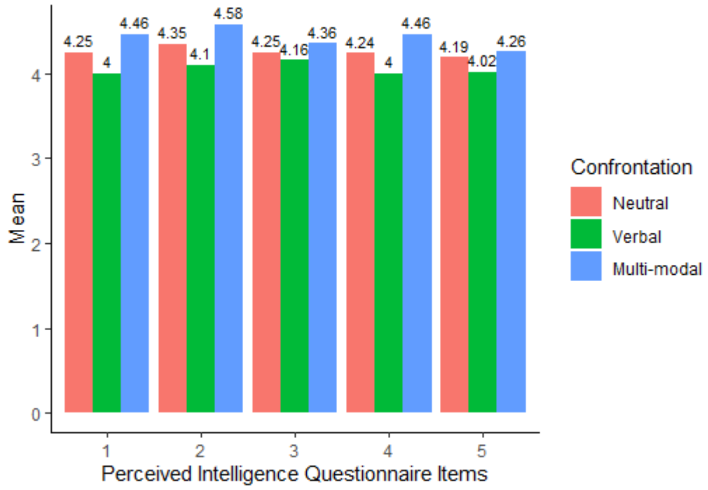



Figure 8.3: Barchart - Perceived Intelligence mean scores for each item

comparatively lower scores. Preliminary conclusions indicate that the robot in Multi-modal confrontation, using both verbal cues and gestures, is perceived to be more animated, likeable and intelligent than in the other confrontation interactions.

### 8.2.2. Perception across Conditions - ANOVA

In order to determine if perception of the robot's capabilities are linked to the Confrontation method employed in the interaction scenario, a set of One-way ANOVA tests between questionnaire items were carried out. The results can be seen in Table 8.2. Only tests that turned out to be significant ($p < 0.05$) are included.

| Questionnaire | Item | Confrontation Pairs | Group | Sum of Squares df = (2,148) | Mean Square | F | Sig |
|---|---|---|---|---|---|---|---|
| Perceived Intelligence (Godspeed) | 1 | Verbal, Multimodal | Confrontation | 5.31 | 2.6555 | 4.176 | 0.0172 |
| | | | Residuals | 94.11 | 0.6359 | | |
| | 2 | Verbal, Multimodal | Confrontation | 5.77 | 2.8828 | 5.06 | 0.00749 |
| | | | Residuals | 84.33 | 0.5698 | | |
| | 4 | Verbal, Multimodal | Confrontation | 5.33 | 2.6644 | 4.005 | 0.0202 |
| | | | Residuals | 98.46 | 0.6653 | | |

Table 8.2: ANOVA - Godspeed Questionnaire

There is a significant effect of the confrontation method on Item 1 in Perceived Intelligence - Incompetent or Competent at the $p < 0.05$ level between Verbal and Multimodal confrontation [$F(2,148) = 4.176$, $p = 0.0172$]. There is a significant effect of the confrontation method on Item 2 in Perceived Intelligence - Ignorant or Knowledgeable at the $p < 0.05$ level between Verbal and Multimodal confrontation [$F(2,148) = 5.06$, $p = 0.00749$]. There was also a significant effect of the confrontation method on Item 4 in Perceived Intelligence - Unintelligent or Intelligent at the $p < 0.05$ level between Verbal and Multimodal confrontation [$F(2,148) = 4.005$, $p = 0.0202$].

ANOVA only helps determine if there are significant differences between confrontation methods. A post-hoc Tukey test is run to determine which confrontation had a greater effect and to correct for family-wise error rate.

| Questionnaire | Item | Confrontation Pair | Mean Difference | p-value adj |
|---|---|---|---|---|
| Perceived Intelligence (Godspeed) | 1 | Verbal, Multimodal | -0.46 | 0.0124 |
| | 2 | Verbal, Multimodal | -0.48 | 0.005 |
| | 4 | Verbal, Multimodal | -0.46 | 0.015 |

Table 8.3: Post-hoc Tukey test - Godspeed Questionnaire

As seen in Table 8.3, for the item: Incompetent or Competent, post-hoc comparisons using the Tukey HSD test indicate that the mean score for Verbal confrontation was significantly lower than Multimodal [$MD = -0.46$, $p = 0.0124$]. Participants in Multi-modal confrontation thought the robot was more competent than participants in Verbal confrontation did.

For the item: Ignorant or Knowledgeable, the Tukey HSD test indicates that the mean score for Verbal confrontation was significantly lower than Multi-modal [$MD = -0.48$, $p = 0.005$]. Participants in Multi-modal confrontation attributed more knowledge to the robot than participants in Verbal confrontation.

Finally, for the item: Unintelligent or Intelligent, the Tukey HSD test indicates that the mean score for Verbal confrontation was significantly lower than Multi-modal [$MD = -0.46$, $p = 0.015$]. Participants in Multi-modal confrontation perceived the robot to be more intelligent than participants did in Verbal confrontation.

### 8.2.3. Verifying factor structure - EFA

Only factor results obtained from entire dataset are considered. The sample size per condition ($n = 50$) is insufficient for accurate confrontation-wise factor analysis. Small sample size affects error estimates and interval size. CFA requires a greater sample size ($n > 100$) [42], so inferences made from EFA for the entire dataset will help determine number of factors for the CFA model.

Polychoric correlations between items in Animacy, Likeability and Perceived Intelligence indicate certain relations between items, see Table 8.4. Most items are positively correlated with no items having any negative correlation. This helps find which items are correlated, but the extent to which

| Godpseed Questionnaire | Polychoric Correlations | | | | | |
|---|---|---|---|---|---|---|
| *Animacy* | 1 | 2 | 3 | 4 | 5 | 6 |
| 1. Dead or Alive | 1 | | | | | |
| 2. Stagnant or Lively | 0.65 | 1 | | | | |
| 3. Mechanical or Organic | 0.44 | 0.4 | 1 | | | |
| 4. Artificial or Lifelike | 0.49 | 0.52 | 0.83 | 1 | | |
| 5. Inert or Interactive | 0.57 | 0.64 | 0.15 | 0.25 | 1 | |
| 6. Apathetic or Responsive | 0.59 | 0.78 | 0.2 | 0.36 | 0.79 | 1 |
| *Likeability* | | | | | | |
| 1. Dislike or Like | 1 | | | | | |
| 2. Unfriendly or Friendly | 0.69 | 1 | | | | |
| 3. Unkind or Kind | 0.78 | 0.74 | 1 | | | |
| 4. Unpleasant or Pleasant | 0.71 | 0.84 | 0.74 | 1 | | |
| 5. Awful or Nice | 0.82 | 0.8 | 0.83 | 0.74 | 1 | |
| *Perceived Intelligence* | | | | | | |
| 1. Incompetent or Competent | 1 | | | | | |
| 2. Ignorant or Knowledgeable | 0.76 | 1 | | | | |
| 3. Irresponsible or Responsible | 0.72 | 0.57 | 1 | | | |
| 4. Unintelligent or Intelligent | 0.84 | 0.84 | 0.65 | 1 | | |
| 5. Foolish or Sensible | 0.74 | 0.73 | 0.72 | 0.77 | 1 | |

Table 8.4: Polychoric correlations - Godspeed Questionnaire

a factor captures the item can be found from the factor loadings. Item 3: Mechanical or Organic has poor correlation with Items 1, 2, 5 and 6 whereas Item 4: Artificial or Lifelike has low correlation with Items 1, 5 and 6. It is noted that Items 3 and 4 are highly correlated. Factor correlations indicate that certain questionnaire items are strong correlated with each other and not so with others. The factor correlations provide impetus to carry out two-factor CFA, one factor for Item 3,4 and the other for the rest of the items. Correlations for items in Likeability and Perceived Intelligence are high. The poorest correlation is between Item 2 and 3 in Perceived Intelligence (0.57) which is still moderately decent for gauging correlation.

| Godspeed Questionnaire | | Factor Loadings | | Communality |
|---|---|---|---|---|
| Items | | Factor 1 | Factor 2 | h2 |
| Animacy | Dead or Alive | 0.58 | 0.30 | 0.55 |
| | Stagnant or Lively | 0.75 | 0.23 | 0.74 |
| | Mechanical or Organic | -0.07 | 0.94 | 0.85 |
| | Artificial or Lifelike | 0.10 | 0.87 | 0.82 |
| | Inert or Interactive | 0.87 | -0.11 | 0.70 |
| | Apathetic or Responsive | 0.95 | -0.06 | 0.87 |
| Likeability | Dislike or Like | 0.85 | - | 0.72 |
| | Unfriendly or Friendly | 0.87 | - | 0.77 |
| | Unkind or Kind | 0.88 | - | 0.78 |
| | Unpleasant or Pleasant | 0.86 | - | 0.73 |
| | Awful or Nice | 0.92 | - | 0.85 |
| Perceived Intelligence | Incompetent or Competent | 0.91 | - | 0.82 |
| | Ignorant or Knowledgable | 0.85 | - | 0.72 |
| | Irresponsible or Responsible | 0.75 | - | 0.66 |
| | Unintelligent or Intelligent | 0.93 | - | 0.86 |
| | Foolish or Sensible | 0.86 | | 0.74 |

Table 8.5: EFA - Godspeed Questionnaire

Carrying out polychoric factor analysis, factors and their loadings can be determined, see Table 8.5. For Animacy, two factors are required to capture communality (variance). This is the proportion of each variable's variance that can be explained by the factors (e.g., the underlying latent continua). It is also noted as $h^2$ and can be defined as the sum of squared factor loadings for the variables. For Likeability and Perceived Intelligence a single factor is enough to explain the communality.

Within Animacy, the items Mechanical or Organic and Artificial or Lifelike have poor loadings for Factor 1 but have higher contribution to Factor 2. The inverse is true for the other four items in Animacy. This indicates that these two items are part of a latent construct different from the latent construct that the other four items are explained by. It can be pointed out that the four items are explained by the factor Animacy whereas the other two items - Mechanical or Lifelike and Artificial or Organic are part of a latent construct that can be termed - Appearance.

The loadings for Likeability and Perceived Intelligence are >= 0.75 indicating a single factor is enough to explain the latent constructs, corroborated by the high polychoric correlation values. As a result two-factor analysis is carried out only for Animacy items from the Godspeed Questionnaire. In the Table 8.5, the communality column provides the amount of variance captured by each item for the factors.

Factor adequacy for Godspeed Questionnaire
The Kaiser-Meyer-Olkin measure of factor adequacy for Animacy verified the sampling adequacy for the analysis. KMO values between 0.7 and 1 indicate the sampling is adequate. The test measures sampling adequacy for each variable in the model and for the complete model. The statistic is a measure of the proportion of variance among variables that might be common variance. With KMO = 0.76, the value indicates sampling for the model is adequate. Bartlett's test of sphericity [$\chi^2(15)$ = 616.34, $p < 0.05$] indicates that the correlation structure is adequate for factor analysis. A good KMO value is obtained for Animacy items by assuming the existence of two latent constructs instead of one. This indicates that a two factor CFA model

The Kaiser-Meyer-Olkin measure for Likeability verified the sampling adequacy for the analysis. With KMO = 0.86, the value indicates sampling for the model is adequate. Bartlett's test of sphericity [$\chi^2(10)$ = 714.47, $p < 0.05$] indicates that the correlation structure is adequate for factor analysis. The Kaiser-Meyer-Olkin measure for Perceived Intelligence verified the sampling adequacy for the analysis. With KMO = 0.86, the value indicates sampling for the model is adequate. Bartlett's test of sphericity [$\chi^2(10)$ = 657.66, $p < 0.05$] indicates that the correlation structure is adequate for factor analysis. KMO is the Measure of Sampling Adequacy(MSA), and it indicates that the items of Likeability and Perceived Intelligence are explained by the respective factors, i.e., how much variance of the measured variables (questionnaire items) can be reproduced in a factor model.

### 8.2.4. Confirmatory Factor Analysis - Two-factor Animacy Model

Two factor CFA was carried out for the items of Animacy from the Godspeed Questionnaire.

The two-factor model for Animacy, which had emerged in the EFA, is further examined by using CFA (see Fig 8.4). The single sided arrows towards items shows how well the variance associated with the items is explained by the constructs: Animacy and Appearance. The 2 items are removed from Animacy and explain another construct termed Appearance. The two-sided arrow between the constructs describes the co-variance. Maximum Likelihood extraction was used to estimate the model. The maximum likelihood factor analysis requires loadings with a cut-off point of 0.40 and the Kaiser's criterion of eigenvalues greater than 1 for selection of factors, yielding a two-factor solution as the best fit for the original Animacy item scores.

The fit indices (and cut-off for good fit) reported for CFA are Comparative Fit Index (>0.9), TLI - Tucker Lewis Index (>=0.95), SRMR - Standardized Root Mean Square Residual (<0.08) and Model Chi-square $\chi^2$ (p-value>0.05). TLI indicates that the model of interest improves the fit relative to the null model. CFI is a revised form of the Normed-Fit Index (NFI) to account of sample size changes. SRMR is the square-root of the difference between the residuals of the sample covariance matrix and the hypothesized model [34, 42].

The model with the best fit is presented here. This model shows a very good fit (CFI = 0.976, TLI = 0.955, SRMR = 0.039, $\chi^2$ = 17.544, df = 8, $p > 0.05$, AIC = 2969.50, BIC = 2967.58). A TLI of 0.955, indicates that the model of interest improves the fit by 95.5% relative to the null model. CFI is greater than 0.9 indicating a good fit. With SRMR < 0.08, the formed two-factor model has a good fit. The AIC and BIC values were used to determine which combination of items for factors forms a good fit for the model.

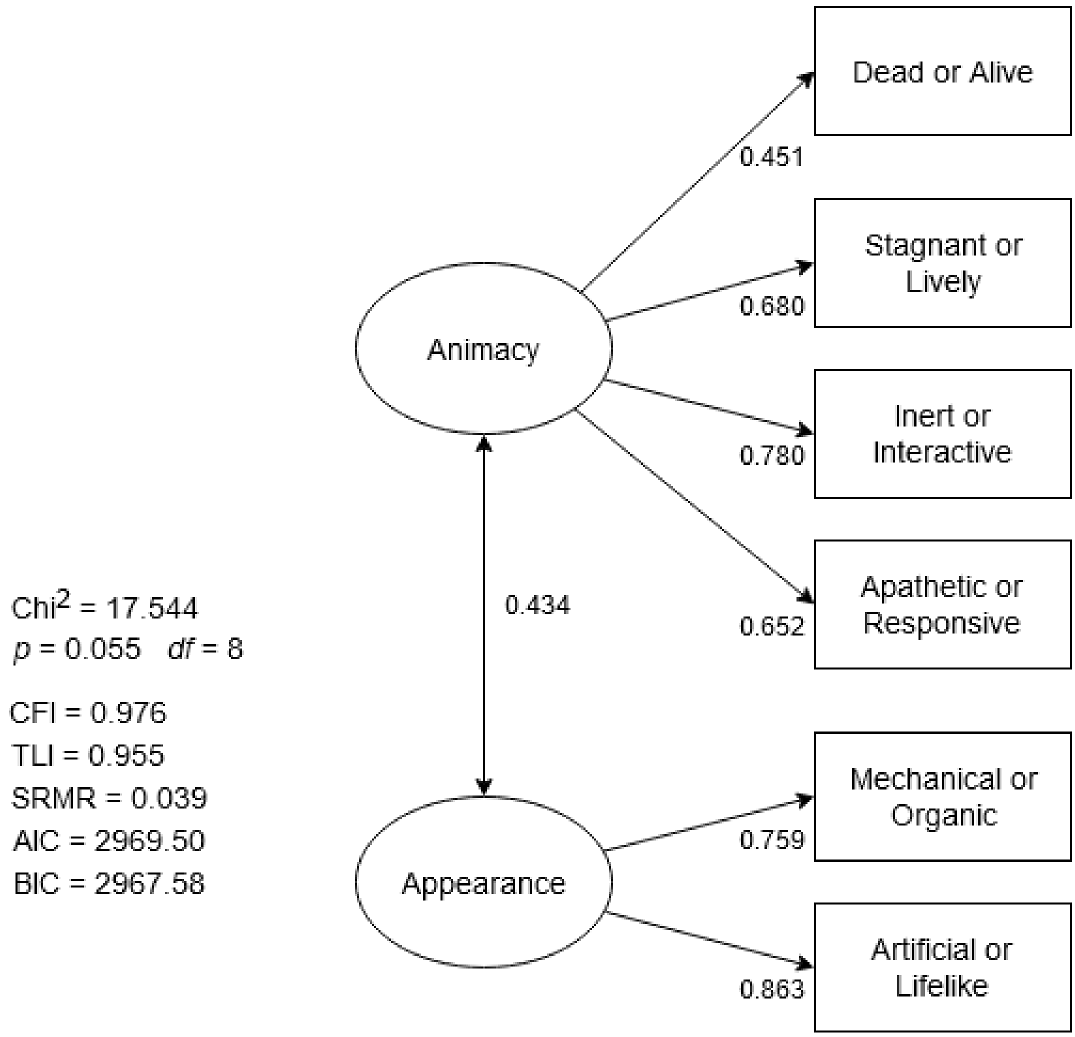



Figure 8.4: Two factor CFA - Animacy

## 8.3. Measuring robot effectiveness - Interaction Questionnaire

### 8.3.1. Descriptive statistics

The median, mean and standard deviation scores for the items from the Interaction Questionnaire are tabulated in Table 8.6. This helps identify items that may contribute towards robot effectiveness in helping with dilemmas. It also indicates at possible difference between items across the confrontation methods employed.

In the Neutral confrontation method, most respondents see an implementation of such a medical decision assistance robot to be more realistic in the future in comparison to the other methods. It also matches their expectations, helps the doctor while also thoroughly addressing the dilemma itself. A significant set of respondents indicate that the robot is capable of planning and controlling actions and providing human-like support. The robot provides sufficient support to help with the dilemma. There was significant disagreement with the robot's capability to understand right from wrong or work without human supervision.

Within the Verbal confrontation method, majority respondents identified such robots to be realistic in the near future and matches their expectations. They agreed that the robot can help doctors re-evaluate ethical dilemmas, provides sufficient support and is thorough. The mean for the items: Item 6 Work without Supervision and Item 7 Understands Right from Wrong are lower compared to the other confrontation methods. In all items, the mean values for Verbal confrontation are lower than the other two confrontation methods. Respondents felt that the robot could not Work well without supervision in

| Interaction Questionnaire | Neutral Confrontation | | | Verbal Confrontation | | | Multimodal Confrontation | | |
|---|---|---|---|---|---|---|---|---|---|
| | Median | Mean | SD | Median | Mean | SD | Median | Mean | SD |
| 1 | 6.0 | 6.137 | 0.748 | 6.0 | 5.42 | 1.162 | 6.0 | 5.62 | 1.368 |
| 2 | 6.0 | 5.509 | 1.488 | 5.0 | 5.04 | 1.354 | 6.0 | 5.52 | 1.281 |
| 3 | 6.0 | 5.039 | 1.72 | 5.0 | 4.40 | 1.840 | 5.0 | 4.80 | 1.603 |
| 4 | 5.0 | 5.058 | 1.629 | 5.0 | 4.34 | 1.756 | 6.0 | 5.32 | 1.406 |
| 5 | 5.0 | 4.882 | 1.657 | 5.0 | 4.76 | 1.744 | 6.0 | 5.54 | 1.073 |
| 6 | 5.0 | 4.627 | 1.72 | 4.0 | 3.58 | 1.896 | 4.0 | 4.10 | 1.740 |
| 7 | 4.0 | 4.176 | 1.956 | 4.0 | 3.78 | 1.854 | 4.0 | 3.90 | 1.821 |
| 8 | 5.0 | 4.941 | 1.528 | 4.50 | 4.14 | 1.807 | 5.0 | 4.46 | 1.716 |
| 9 | 6.0 | 5.509 | 1.317 | 5.0 | 5.14 | 1.47 | 6.0 | 5.68 | 1.132 |
| 10 | 6.0 | 5.627 | 1.427 | 5.50 | 5.22 | 1.403 | 6.0 | 5.80 | 1.142 |

Table 8.6: Summary Statistics - Interaction Questionnaire

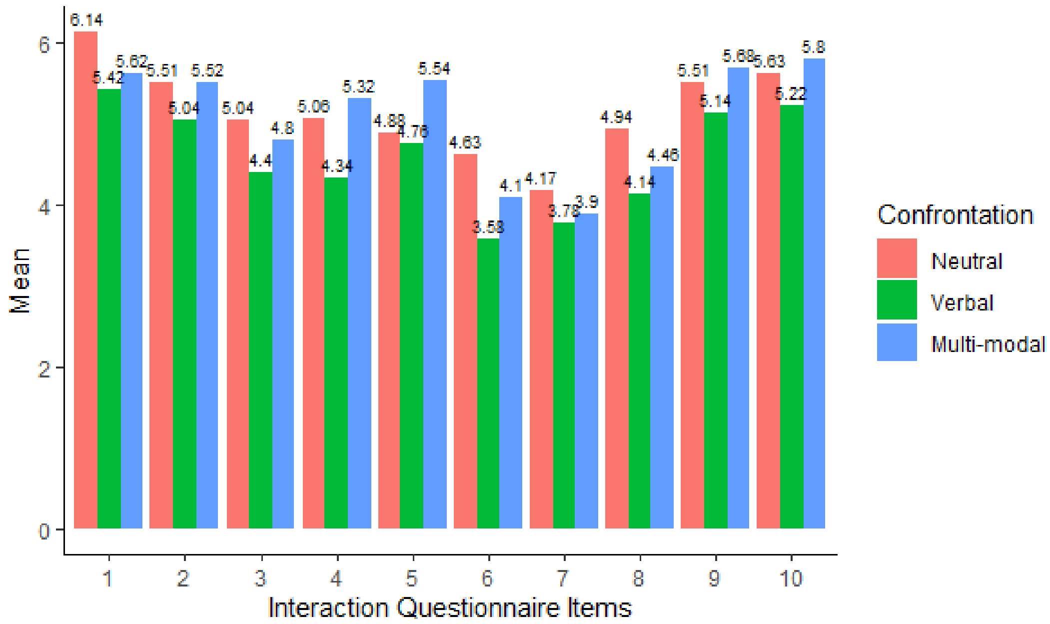



Figure 8.5: Barchart - Interaction Questionnaire Mean scores for each item

the Verbal confrontation method. Multimodal confrontation performed better than Verbal confrontation with regards to several items from the Interaction Questionnaire. The mean values in Item 2, 4, 5, 9 and 10 are the highest in the Multimodal method, see Fig 8.5.

However, to draw significant conclusions from the difference found between confrontation methods, ANOVA is carried out. Item wise ANOVA is carried out for the Interaction Questionnaire and their results are discussed in the following section.

### 8.3.2. Performance across Conditions - ANOVA

In order to determine if the performance of the robot in handling ethically charged decision making processes are associated with the confrontation conditions, a set of One-way ANOVA tests were carried out. The results can be seen in Table 8.7. Only tests that turned out to be significant ($p < 0.05$) are reported.

From Table 8.7, it can be observed that there is a significant effect of the confrontation method on Item 1 in the Interaction Questionnaire - Realistic in the Future at the $p < 0.05$ level between Neutral and Verbal confrontation [$F(2,148) = 5.518$, $p = 0.0048$]. The tests of significance are at the $p < 0.05$ confidence level. There was a significant effect of the confrontation condition on Item 4 in the Interaction Questionnaire - Human-like Support between Verbal and Multimodal confrontation methods [F(2,148)

| Questionnaire | Item | Confrontation Pairs | Group | Sum of Squares df = (2,148) | Mean Square | F | Sig |
|---|---|---|---|---|---|---|---|
| Interaction Questionnaire | 1 | Neutral, Verbal | Confrontation | 13.87 | 6.934 | 5.518 | 0.00489 |
| | | | Residuals | 186.00 | 1.257 | | |
| | 4 | Verbal, Multimodal | Confrontation | 25.8 | 12.889 | 5.008 | 0.00786 |
| | | | Residuals | 380.9 | 2.574 | | |
| | 5 | Verbal, Multimodal | Confrontation | 17.6 | 8.815 | 3.805 | 0.0245 |
| | | | Residuals | 342.8 | 2.316 | | |
| | 6 | Neutral, Verbal | Confrontation | 29.2 | 14.580 | 4.566 | 0.0119 |
| | | | Residuals | 472.6 | 3.193 | | |
| | 8 | Neutral, Verbal | Confrontation | 16.4 | 8.223 | 2.889 | 0.0488 |
| | | | Residuals | 421.3 | 2.846 | | |

Table 8.7: ANOVA - Interaction Questionnaire

= 5.008, p = 0.0078].

There was a significant effect of the confrontation condition on Item 5 in the Interaction Questionnaire - Sufficient support - dilemma between Verbal and Multimodal confrontation [F(2,148) = 3,805, p = 0.0245]. Further more, there was a significant effect of the confrontation condition on Item 6 in the Interaction Questionnaire - Work without Supervision between Neutral and Verbal confrontation [F(2,148) = 4.566, p = 0.0119] on Item 8 in the Interaction Questionnaire - Plans and Controls Actions between Neutral and Verbal confrontation [F(2,148) = 2,889, p = 0.0486].

| Questionnaire | Item | Confrontation Pair | Mean Difference | p-value adj |
|---|---|---|---|---|
| Interaction Questionnaire | 1 | Neutral, Verbal | 0.71 | 0.0045 |
| | 4 | Verbal, Multimodal | -0.71 | 0.046 |
| | 5 | Verbal, Multimodal | -0.78 | 0.0304 |
| | 6 | Neutral, Verbal | 1.04 | 0.0104 |
| | 8 | Neutral, Verbal | 0.801 | 0.0477 |

Table 8.8: Post-hoc Tukey test - Interaction Questionnaire

As seen in Table 8.8 For item: Realistic in the Future, post-hoc comparisons using the Tukey HSD test indicated that the mean score for Neutral confrontation was significantly greater than Verbal confrontation [MD = 0.71, p = 0.0045]. Participants in Neutral confrontation thought the robot was more realistic in the future than the one in the interaction presented to Verbal confrontation participants. For item: Human-like Support, post-hoc comparisons using the Tukey HSD test indicated that the mean score for Verbal confrontation was significantly lower than Multi-modal confrontation [MD = -0.71, p = 0.046]. Participants in Verbal confrontation thought the robot provided less Human-like support compared to participants in Multi-modal confrontation. For item: Sufficient Support - dilemma, post-hoc comparisons using the Tukey HSD test indicated that the mean score for Verbal confrontation was significantly lower than Multi-modal confrontation [MD = -0.76, p = 0.0304]. Participants in Multi-modal confrontation agreed more that the robot provided sufficient support to help with the dilemma when compared to Verbal confrontation. For item: Work without Supervision, post-hoc comparisons using the Tukey HSD test indicated that the mean score for Neutral confrontation was significantly greater than Verbal confrontation [MD = 1.04, p = 0.0104]. Participants in Neutral confrontation thought that such a robot can work without supervision more than participants did in the Verbal confrontation. For item: Plans and Controls Actions, post-hoc comparisons using the Tukey HSD test indicated that the mean score for Neutral confrontation was significantly greater than Verbal confrontation [MD = 0.801, p = 0.0477]. Participants thought the robot planned and controlled actions better in Neutral confrontation than Verbal confrontation.

### 8.3.3. Identifying latent constructs - EFA

| Interaction Questionnaire | Polychoric Correlations | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Items* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1. Realistic in the future | 1 | | | | | | | | | |
| 2. Matches Expectations | 0.46 | 1 | | | | | | | | |
| 3. Accept robot - decision making | 0.46 | 0.65 | 1 | | | | | | | |
| 4. Human-like support | 0.42 | 0.62 | 0.67 | 1 | | | | | | |
| 5. Sufficient support with dilemma | 0.49 | 0.59 | 0.64 | 0.77 | 1 | | | | | |
| 6. Work without Supervision | 0.33 | 0.41 | 0.55 | 0.66 | 0.54 | 1 | | | | |
| 7. Understands Right from Wrong | 0.23 | 0.44 | 0.44 | 0.48 | 0.5 | 0.57 | 1 | | | |
| 8. Plans and aontrols Actions | 0.4 | 0.39 | 0.49 | 0.52 | 0.49 | 0.61 | 0.68 | 1 | | |
| 9. Thorough in addressing dilemma | 0.38 | 0.49 | 0.55 | 0.56 | 0.68 | 0.47 | 0.5 | 0.43 | 1 | |
| 10. Helps Re-evaluate dilemma | 0.53 | 0.68 | 0.59 | 0.71 | 0.72 | 0.59 | 0.49 | 0.55 | 0.73 | 1 |

Table 8.9: Polychoric correlations - Interaction Questionnaire

Table 8.9 shows the polychoric correlations between items from the Interaction Questionnaire. It is observed that most items are positively correlated (>0.4). Correlations less than 0.4 are between items: Realistic in the Future with both Understands Right from Wrong and Thorough in Addressing Dilemma. It should be noted that correlations are simply a tool to point towards possible inter-relations, especially considering this is a novel questionnaire. The factor loadings for the questionnaire items provide further context.

Only factor results obtained from the entire dataset are considered. The sample size per condittion (n =50) is insufficient for accurate factor analysis. CFA requires more sample sizes as mentioned earlier, so inferences made for EFA for the entire dataset will help determine number of factors for the Interaction questionnaire CFA model.

| Interaction Questionnaire | Factor Loadings | | Communality |
|---|---|---|---|
| Items | Factor 1 | Factor 2 | h2 |
| Realistic in the future | 0.61 | -0.06 | 0.33 |
| Matches expectations | 0.83 | -0.10 | 0.57 |
| Accept robot-decision making | 0.80 | -0.03 | 0.60 |
| Human-like Support | 0.77 | 0.10 | 0.71 |
| Sufficient support - dilemma | 0.84 | 0.01 | 0.72 |
| Work without Supervision | 0.27 | 0.54 | 0.57 |
| Understands Right from Wrong | 0.00 | 0.79 | 0.63 |
| Plans and Controls Actions | -0.02 | 0.86 | 0.71 |
| Thorough in addressing dilemma | 0.67 | 0.09 | 0.54 |
| Helps Re-evaluate dilemma | 0.84 | 0.05 | 0.76 |

Table 8.10: EFA - Interaction Questionnaire

The communality values show the extent to which an item correlates with all other items. Item: Realistic in the Future has low communality ($h^2$ = 0.33) but not too low that it is insignificant in being part of a factor.

Polychoric factor analysis provides us with the factors and the associated factor loadings, see Table 8.10. From the table, it is evident that the two factors are required for the items to be explained. Factor 1 extracts sufficient variance from the Items: Realistic in the future, Matches expectations, Accept robot for decision making, Human-like Support, Sufficient support - dilemma, Thorough in addressing dilemma and Helps Re-evaluate dilemma as can be seen from the loadings value (>0.6). For the other items the loadings values are poor for Factor 1 (<0.3).

Conversely, Factor 2 extracts sufficient variance from the Items: Works without Supervision, Understands Right from Wrong and Plans and controls Actions (>0.55) but has poor ladings corresponding to the other items (<=0.10). This indicates towards the existence of a two-factor model where each factor, accounts for the respective items and explains or measures a latent construct. This can be verified by using a two-factor CFA model.

Factor Adequacy for Interaction Questionnaire items
The Kaiser-Meyer-Olkin measure for the Interaction Questionnaire items verified the sampling adequacy for the analysis. With KMO = 0.88, the value indicates sampling for the model is adequate. Bartlett's test of sphericity [$\chi^2(45) = 981.48$, $p < 0.05$] indicates that the correlation structure is adequate for factor analysis. A good KMO value is obtained for Interaction Questionnaire items by assuming the existence of two latent constructs. This indicates that a two factor CFA model can be a good fit for the supposed latent constructs.

### 8.3.4. Confirmatory Factor Analysis - Two-factor Interaction Model

Two factor CFA was carried out to check for latent constructs in the Interaction questionnaire.

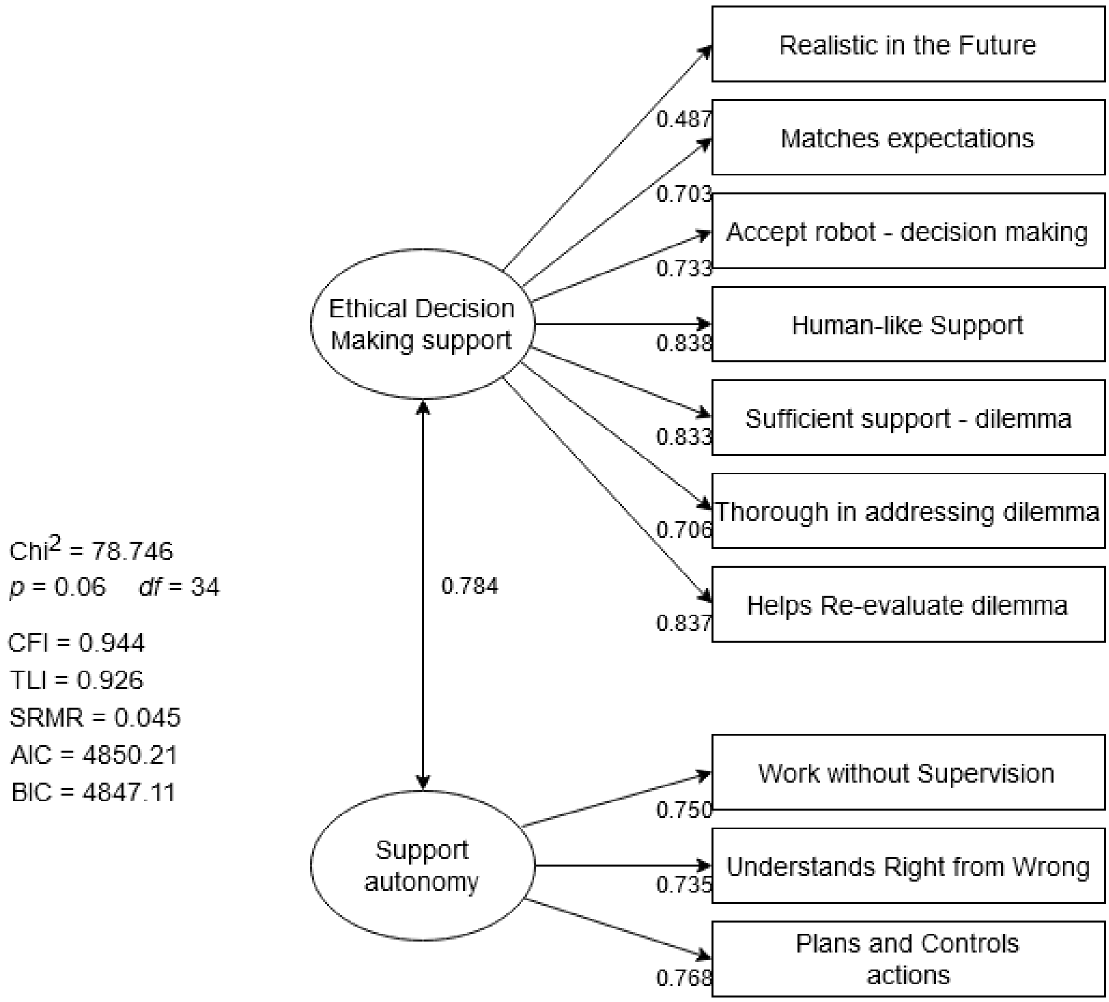



Figure 8.6: Two factor CFA - Interaction Questionnaire

The two factor structure for Interaction Questionnaire, which we observed in the EFA, is further modeled by employing CFA, see Fig 8.6. Again, the single sided arrows denote how well the variance associated with the items is explained by the constructs: Ethical Decision Making support and Support autonomy which have a covariance of 0.784. The Interaction Questionnaire, therefore measures these two latent variables through the particular questionnaire items.

Maximum Likelihood extraction was used to estimate the model. The same cut-off criterion are used here as in the Animacy CFA model. The maximum likelihood factor analysis requires loadings with a cut-off point of 0.40 and the Kaiser's criterion of eigenvalues greater than 1, resulting a fit for the two-factor CFA model.

The model with the best fit is presented here. This model shows a very good fit (CFI = 0.944, TLI =

0.926, SRMR = 0.045, $\chi^2$ = 78.746, df = 34, p > 0.05, AIC = 4850.21, BIC = 4847.11). A TLI of 0.926, indicates that the model of interest improves the fit by 92.6% relative to the null model. CFI is greater than 0.9 indicating a good fit. With SRMR < 0.08, the formed two-factor model has a good fit with very low residual. The AIC and BIC values were used to determine which combination of items for factors forms a good fit for the model.

## 8.4. Demographic Insights

In this section, exploratory analysis is carried out to gain insights into basic choices and preferences between Gender and Religion. While information was obtained on Age-ranges, Education and Employment, they are not used in analysis and comparison. The table provides an indication of the distribution of participants across age-ranges, education and employment levels. As such, analysis results of the questionnaires across Conditions, Gender and Religion are done in order to answer the Research Questions.

| Age-range | Number | Education | Number | Employment | Number |
|---|---|---|---|---|---|
| 18-24 | 9 | High School | 9 | Full Time | 117 |
| 25-34 | 62 | Some College | 29 | Part Time | 15 |
| 35-44 | 43 | Associate Degree | 17 | Unemployed | 5 |
| 45-54 | 20 | Bachelor’s | 71 | Retired | 7 |
| 55-64 | 9 | Master’s | 22 | Student | 4 |
| 65+ | 8 | Professional Degree or Doctorate | 3 | Disabled | 3 |

Table 8.11: Demographics of Participants - Age, Employment and Education

Overall, out of the 151 participants, 71 were below the age of 35 and a cumulative of 134 participants below the age of 45. Most responses were from people who completed their Bachelor’s and Master’s degrees. A sizeable portion of responses completed high school, some college and Associate degrees. 117 responses were from full-time employees and 15 from part-timers. The rest were from the other categories.

Among the respondents, 62 were Christian, 45 Atheist/Agnostic and 25 Hindu. Other religions such as Judaism, Islam, Buddhism etc account for the rest. Concerning gender, 91 were male and 59 were female participants. One participants chose not to reveal their Gender. The analysis for differences between Gender and Religion across Conditions is done at the end of the results section, after addressing the main research questions. ANOVA to find for differences between Genders did not find significant difference across Genders for any item.

### 8.4.1. Comparison: Religion

Table 8.12 indicates the population split for Gender and Religion across confrontation conditions.

| Confrontation Type | Gender | | | Religion | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Female | Male | Undisclosed | Christian | Hindu | Jewish | Islam | Buddhist | Other | Atheist | Undisclosed |
| Non-confrontation | 17 | 34 | - | 26 | 8 | - | 1 | 1 | 1 | 14 | - |
| Verbal | 23 | 26 | 1 | 21 | 8 | - | 2 | 3 | 2 | 13 | 1 |
| Multimodal | 19 | 31 | - | 15 | 9 | 1 | 1 | 1 | 3 | 18 | 2 |

Table 8.12: Demographics of Participants - Gender and Religion

ANOVA tests were carried out between different religions groups over the entire participant group. Only religions with a sample size > 10 were selected. Therefore, the comparisons were between Christians, Hindus and Atheists. Significant differences in responses are reported in Table 8.13.

There was a significant effect of Religion on Item 4 in Animacy - Artificial or Lifelike at the p <0 .05 level between Hindus and Atheists [F(7,143) = 3.154, p = 0.00395]. There was a significant effect of Religion on Item 7 in the Interaction Questionnaire - Understands Right from Wrong at the p < 0.05 level between Hindus and Atheists [F(7,143) = 2.212, p = 0.0366].

For item: Artificial or Lifelike, post-hoc comparisons using the Tukey HSD test indicated that the mean score for Hindus was significantly greater than Atheists [MD = 1.471, p = 0.0318]. Hindus per-

| Questionnaire | Item | Religion Pair | Group | Sum of Squares df - (7,143) | Mean Square | F | Sig |
|---|---|---|---|---|---|---|---|
| Animacy (Godspeed) | 4 | Hindu, Atheist | Religion | 36.424 | 5.204 | 3.154 | 0.00395 |
| | | | Residuals | 235.919 | 1.650 | | |
| Interaction Questionnaire | 7 | Hindu, Atheist | Religion | 51.476 | 7.354 | 2.212 | 0.0366 |
| | | | Residuals | 475.464 | 3.325 | | |

Table 8.13: ANOVA - Religion

| Questionnaire | Item | Condition Pair | Mean Difference | p-value adj |
|---|---|---|---|---|
| Animacy (Godspeed) | 4 | Hindu, Atheist | 1.471 | 0.0318 |
| Interaction Questionnaire | 7 | Hindu, Atheist | 1.226 | 0.0046 |

Table 8.14: Post-hoc Tukey test - Religion

ceived the robot to be more Lifelike than Atheists. This could be attributed to Hinduism being idolatry. For item: Understands Right from Wrong, post-hoc comparisons using the Tukey HSD test indicated that the mean score for Hindus was significantly greater than Atheists [MD = 1.226, p = 0.0046]. Hindus attribute greater capacity for the robot to distinguish right from wrong than Atheists.

While it is expected that such robots will be relevant in the future there is an indomitable trust gap that will require focused research to increase robot acceptance (starting with a possible and viable implementation of a moral agent).

# V

# CONCLUSIONS

## Discussion and Conclusions

Prototype design patterns were formed that provide a starting point for developing interaction patterns in HRI. Specifically, the interaction patterns guide the development of designing interaction scenarios wherein the robot, to varying degrees, attempts to protest the decisions made by a doctor. This is done in the order: Neutral, Verbal and Multi-modal Confrontation. The robot is not capable of making moral decisions, but provides information and structure to the problem at hand in order to elicit critical reflection within the doctor's decision making process. The purpose is to formulate a robot interaction in a hospital where the robot is capable of providing patient information to the doctor, protest or disagree with decisions made, provide information on the values of the stakeholders involved, ensure the decision takes into consideration various factors such as patient choice and code of ethics so that the dilemma is discussed adequately.

Based on quantitative analysis involving the use of statistical tests - ANOVA, and model building - Confirmatory Factor Analysis, a few conclusions can be drawn. For certain items, the ANOVA yielded significant results. Results about the robot in the interaction scenarios show that participants perceived the robot to be most competent, knowledgeable and intelligent (Perceived Intelligence) in the Multi-modal interaction scenario which confirms the initial hypothesis (H1: Robot effectiveness when helping doctors with reflection in ethical decision making increases as more cues are added). These results from the Godspeed Questionnaire indicate that a robot aiding with decision making is perceived to be more intelligent when the number of cues increases.

A dedicated questionnaire on the interaction scenario was formulated that measures robot effectiveness in provision of support for decision making including information and reflection. The results of the Interaction Questionnaire show interesting outcomes. Participants in the Multi-modal scenario indicate the robot to be capable of human-like support and that the support provided was sufficient to help with the dilemma when compared to the Verbal scenario. This shows that the robot provides information for decision making in an appropriate way and that additional cues increase robot's effectiveness to address the ethical dilemma. Comparisons between Verbal and Neutral scenario yield interesting results. Participants thought the robot presented in the Neutral scenario to be more realistic in the future, could work without direct supervision and planned and controlled actions better than the robot in the Verbal confrontation scenario. These differences were significant and do not lend support to the first hypothesis. The robot in the Neutral confrontation is ascribed more agency and supports the analysis of decision choices more effectively than Verbal confrontation. The tests show that Multi-modal confrontation (verbal cues and gestures) is perceived to be more intelligent and performs better than the Verbal confrontation (only verbal cues) as intended. While it can be determined that the increase in cues between these two scenarios increases robot effectiveness when helping doctors, the same cannot be said when comparing Neutral and Verbal confrontation.

Additionally, the second hypothesis points towards a more exploratory question (H2: Which aspects of the robot interaction, as perceived by the human, can explain his or her support experience?). In order to identify the aspects of the robot interaction the questionnaires measure, Factor Analysis is carried out. Concerning the Animacy section in the Godspeed Questionnaire, two items: Mechanical or Organic and Artificial or Lifelike have poor factor loadings compared to other items. This indicates that these two items measure a latent variable other than 'Animacy'. Two-factor Confirmatory Factor Analysis (CFA) confirms this finding. The two items are explained by a different latent variable termed 'Appearance'. While other items in the Godspeed questionnaire measure the respect constructs i.e. Likeability and Perceived Intelligence, the items from the original Animacy section explain two factors: Animacy and Appearance. In the Interaction Questionnaire, factor loadings indicate that the three items: Work without Supervision, Understands Right from Wrong and Plans and Controls actions are explained by a construct different than the latent variable that measures the other seven items. Similarly, a two-factor CFA model is applied which confirms this discovery. The three items are supported by a latent construct termed: 'Support Autonomy' while the other seven items form part of 'Ethical Decision Making Support'. CFA shows that the Interaction questionnaire measures the perception of the robot in its ability to provide support for ethical decision making and how autonomous it is in doing so.

### Delegation and Trust

An important ethical question concerning robotics and human-robot interaction (HRI) is whether humans can trust robots. While autonomous robots will likely be developed in the future for healthcare, we already rely heavily on semi-autonomous robots in everyday life. Reliance on such technology arises only from the trust placed in them. As technological devices and artefacts become an integral part of human lives, it brings to the fore issues such as the nature of trust, conditions for occurrence and whether trust can in fact be developed towards robots.

Humans already delegate tasks to machines meaning there is already some trust placed on the machines to reliably carry out work. There is a risk of delegation and over-trust when working with autonomous robots, especially in ethically charged situations that can be regularly found in healthcare environments. The trust placed in robots is based on reliance. There is also the problem of over-reliance wherein the human agent may cease carefully carry out his/her duties due to the robot being correct in its duties. This raises a false sense of security that may lead to calamitous consequences. Although there is no guarantee that a cleaning robot will clean the room, we can expect it to do so. This direct trust is associated with the indirect trust placed on human designers responsible for developing the robot.

Trust in human-robot interaction is complex and regarded as something where a relationship develops between the two with an ethical dimension. By trusting a robot, responsibility is assigned to them, thereby placing a duty on them to be answerable. The robot must be able to carry out actions that it is responsible or designed for. Futuristic, morally capable robots will be held to the same moral accountability as any human. However, the trust on semi-autonomous robots will be based on functional criteria. In order to be able to trust existing robots, performance criterion can be used for example, by implementing utility functions. Conceptual frameworks of trust can be applied to discuss trust in HRI. If the human-robot interaction develops as a social relation, then trust already exists in an implicit, affective way.

## Contributions

With this project 4 main contributions are made to ethical human-robot interaction..

- The development of design patterns to implement confrontation methods within robots to help aid ethical decision making.
- Determining the use of cues in robots and how they affect decision making with increasing magnitude.
- Integrating the Gibbs' Reflecting Cycle in HRI to enhance decision making through reflection.
- Development and validation of an Interaction Questionnaire that measures the Support Autonomy and Ethical Decision Making Support that a robot can provide.

## Limitations

During the development of the interaction blocks a few problems related to the functionality of Pepper emerged. The transition time between gestures and the dialogue may cause symmetry issues. It was difficult to make the dialogue match the gestures displayed by the robot. Conceptually, the setting of the environment was not clearly determined. The robot is assumed to be with the doctor in a room. In a hospital other personnel would also be there that could affect the robots ability to remain focused on the doctor. Better sensors with increased feature detection could aid in this process.

In hindsight, the Verbal confrontation should perhaps have not had any Autonomous Life movement to truly measure the direct effect of verbal cues. Even if the robot was not perceived to be animate, it may have been perceived to be more intelligent and could have performed better than the Neutral confrontation. An interesting point to note was the better performance of Neutral confrontation in three items of the Interaction Questionnaire compared to Verbal confrontation. People indicated that the robot in the Neutral condition was more realistic in the future, could plan and control actions and work without supervision better. While it can be construed that a simple yes-robot is more realistic in the future, participants ascribing the robot in Neutral confrontation more planning capabilities and autonomy suggests that the Verbal confrontation may not have been perceived as intended. Verbal confrontation makes use of verbal cues only, but the inbuilt Autonomous Life functionality of Pepper was turned on for all three conditions. This could have negatively affected the robot's performance with the affective dialogue it speaks not being supported physically by the gestures enabled by the Autonomous Life option. Turning it off would render the robot's use to only that of a speaker.

Getting direct feedback from participants directly, specifically doctors, could be done to get a more accurate response. It is unclear if the intended emotions of the robot were correctly translated by the participants: especially in the form of a video scenario.

## Future Work

The design patterns can be iterated and worked over with the help of real doctors with in-field testing along with human intervention. With better and cheaper technology it should be possible to develop autonomous moral agents in the future that can act in different degrees based on the interaction scenario and design patterns implemented in the medical field.

In the future, technical improvements can be made in Pepper for better translation of cues. This would involve a complex upgrade of the speech engine and improving sensors to match gestural cues quicker. Alternatively, a more human-like robot capable of using facial expressions could also be used keeping in mind the issue of the Uncanny Valley.

Carrying out in-situ experiments with real doctors would provide valuable insight. Tailoring it to the specific needs and use of the doctors with regards to treatment support and ethical aspects would be easier. The ethical aspects themselves are a complex field of study. Developing moral states within robots would be the next step in HRI but technical improvements need to be made in huge strides.

Specifically for future work here are a few suggestions:

- Use the Pepper robot for reproducibility reasons.
- Develop better gestures for Pepper. Pepper is moving on from a Python to an Android engine and that depends on developments from Softbank Robotics.
- Collect more data equally distributed between the confrontation scenarios.
- Carry out detailed, in-person perception interviews of the robots interaction capabilities with real doctors.
- Translate the Wizard-of-Oz implementation into algorithmic capabilities, e.g. Pepper gives appropriate responses when the interaction flow is interrupted automatically.
- Explore the role of delegation and trust in ethical HRI.

This project can be a valuable addition to the human-robot interaction field. With increasing use of social robots the ethical implications of their usage in regular life invite rigorous exploration.

# VI

# APPENDIX

# A

# Ontology

In this section the ontology of the project is described. This includes a list of entities. their descriptions, interactions between entities, and the rules that govern these interactions. The ontology will be used to design responses by the Decision Assistant Robot (DAR) using verbal and visual cues along with movements and gestures.

## Classes

- **Person:** A human being. In this case, a Doctor interacts with the robot. The patient and the relatives are also persons. Persons have a name, age, profession and may or may not have relatives.
- **Verbal Cues:** Verbal and audio cues will be used for the Doctor to aid in decision making.
- **Visuals:** Visuals will be used to convince the Doctor and to also display patient vitals and status reports. These refer to images that are used by the robot on its tablet to convey emotion.
- **Dilemma:** A dilemma here refers to an ethical decision that the doctor needs to take with respects to the patient's well-being, patient's wishes and the wishes and requests of their relatives.
- **Illness:** Refers to the illness and the stage of affection. In this case, the illness (cancer) is highly probable to cause painful death.

## Entities

- **DAR:** The assistant robot that will interact verbally and visually with the Doctor and is controlled by a computer.
- **Doctor:** The healthcare professional working at the hospital. Responsible for care of the patient and any associated decision regarding the well-being of the patient.
- **Patient:** A patient who got admitted for serious issues and now in ICU/special care. Under the eye of the Emergency Department, the patient's care is under the discretion of the Doctor.
- **Patient's relative:** The relative of the patient who asks for either treatment or palliative care.
- **Controller:** Person responsible for controlling the speech and decision making of the robot and responsible for controlling the movement of the robot.
- **Computer:** Interface that allows the control of the robots speech and decisions. To be implemented by having various action blocks in the Interaction Design Tool[73].

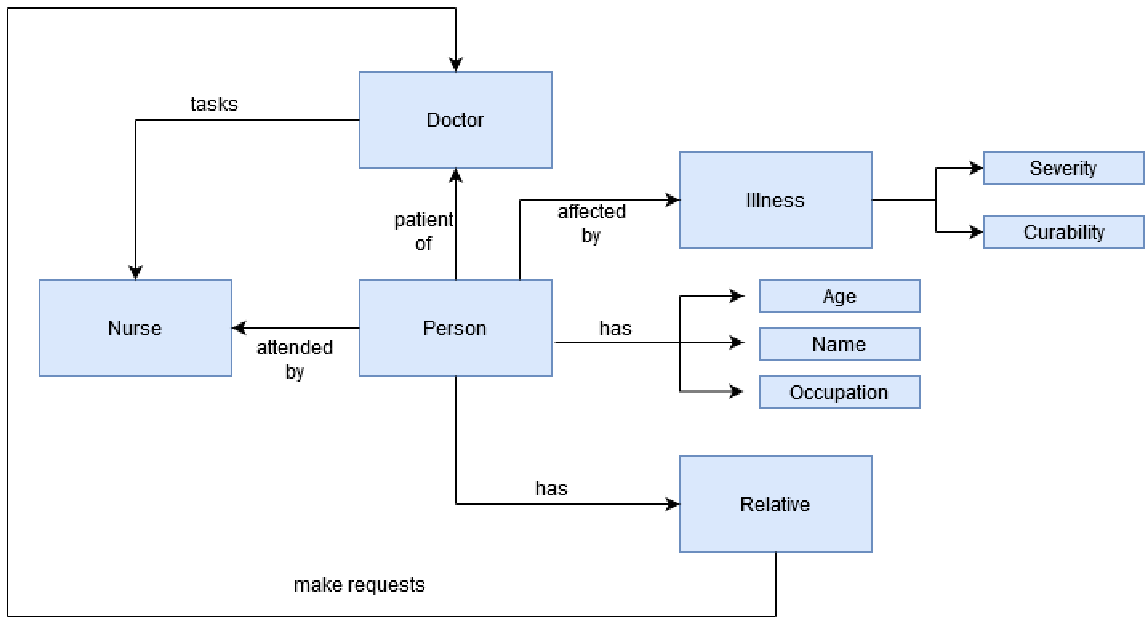



Figure A.1: Class Interaction Diagram

## Common Vocabulary

- Standard workflow
    1. Introduction -> Describe the patient status -> Ask for decision -> Display disagree and protest decision -> End.
    2. Have consistent answers and responses to the input of the Doctor.
- Standardized positive and negative responses
    1. Multiple responses with different emotions as a response to any reply from the doctor.
- Short, clear sentences
    1. Clearly understood by the doctor.
    2. Repeat the previous sentences and display information clearer on screen.
- Precision movement
    1. Gestures and physical cues must be as close to human representation as possible.
    2. Clear motions understandable by the Doctor.

## Explicit Knowledge Base

- Animations for displays of protest and distress
    1. Speaking to employ verbal dissuasion.
    2. Movement to display physical exhibition of affect.
- Varied responses for natural conversation
    1. Altered versions of responses - have more random options.
- Movement
    1. Move towards the Doctor for engagement.

2. Look at the Doctor to maintain conversation.

- Basic Awareness Parameters

  1. Switch on people movement, so DAR's Basic Awareness focuses on the Doctor it is intended to talk to.
  2. Have a button to let DAR focus forward when Basic Awareness is focused just on the Doctor.

- Tailored responses to include medical terminology, employee details and roles and patient under care

## Glossary

A list of terms and their explanations used to describe the interaction scenarios.

1. **Claims:** A claim refers to an intended or side effect of a particular functionality of the system. Claims are used to formulate hypotheses about the beneficial and detrimental effects of adding a particular functionality to the design specification. These hypotheses can be tested in evaluation studies to investigate the beneficial effect of including a certain functionality in the design.

2. **Confrontation:** A confrontation can be defined as an argumentative situation between two parties. In this case, the confrontation involves the assistant robot and the doctor over an ethical dilemma. It involves parties holding opposing views and disagreement over certain characteristics of the situation.

3. **Cues:** Cues are indicators that help the robot decide which action to carry out. The responses of the doctor helps it decide which physical and verbal cues to implement.

4. **Design Specification:** The design specification is the collection of use cases, requirements, claims, and ontology. Together they describe the outline of the system's behaviour, functionalities, and intended effects. This should be sufficient to provide a blueprint for developers to implement the design in an operational, interactive version of the system (a prototype).

5. **Gestures:** A gesture is a form of non-verbal communication or non-vocal communication in which visible bodily actions communicate particular messages, either in place of, or in conjunction with, speech. Gestures include movement of the hands, face, or other parts of the body. The robot makes use of gestures help convey the emotion it intends to effectively.

6. **Human Factors:** A Human Factors concept is an idea, principle, or theory based in the human factors literature. The concept is deemed relevant to the design if it can be used to stipulate the design rationale, claims, or design pattern premises.

7. **Interaction Design Pattern:** An interaction design pattern is an abstracted interaction that is not tied to one specific task, event or context. It is a formal description of a general solution to a recurring interaction design problem within the design specification. By identifying recurring problems that need a solution, interaction design patterns can provide a solution that can be consistently used throughout the design. Not every problem can be solved with an interaction design pattern and not every interaction design pattern needs a fully detailed specification in the design specification. Interaction design patterns draw from premises that are grounded in human factors research.

8. **Multi-modal confrontation:** This method adds physical cues and gestures to the Verbal Confrontation method. This includes proxemic closeness, gesture openness and eye color. The 'humanness' of the robot is enhanced further by including physical movement.

9. **Neutral Confrontation:** This refers to the non-confrontation methods implemented by the robot. The robot will agree to carry out any command given out by the doctor without prompting the doctor for any sort of discussion or reflection.

10. **Ontology:** An ontology formally represents declarative knowledge as a collection of interlinked descriptions of entities or concepts within a domain (possibly as a hierarchy). The ontology uses a shared vocabulary to denote the classes, attributes, instances, and relations of those concepts. An ontology is a collection of concepts, along with their definition and relations between them. The ontology can be regarded as a domain-specific language to speak about the design.

11. **Proxemics:** Proxemics is the study of human use of space and the effects that population density has on behavior, communication, and social interaction. The same is applied here in the context of the Human-Robot Interaction.

12. **Requirement:** A requirement is some functionality or capability the system needs to be able to perform, address, or satisfy.

13. **Stakeholder:** A stakeholder is a person or institution that is directly or indirectly affected by the design. Stakeholders also have certain values that are important to them.

14. **Technology:** The technology describes the type of currently existing, or to be developed technologies used in the system to produce the desired behaviour.

15. **Use Case:** A use case is a sequence of interactions between multiple roles or actors. A use case is situated, i.e. related to a given context or activity (as opposed to design patterns).

16. **Value:** A value is something a person finds very important to protect and support. It is a driving force in decision making and goal selection. An instrumental value is worth having as a means to something else that is good. An intrinsically valuable thing is worth having for itself. Intrinsic and instrumental goods are not mutually exclusive categories. Some things are both good in themselves, and also good for getting other things that are good.

17. **Verbal Confrontation:** This method makes use of verbal cues to carry out dissuasive attempts at changing the doctors decision making. Verbal audio includes variables such as pitch, volume and prosody. The audio of the robot is designed to display 'humanness' while carrying out the confrontation scenario.

## Ontology Diagram

The ontology is built around two 'real' entities: the Doctor and the Decision Assistant Robot (DAR). DAR will provide patient details to the Doctor. This will initiate the conversation about what decision to take regarding the patient. DAR will disagree with or dissuade the Doctor regarding decisions they make that have critical, ethical and health based repercussions. DAR will try to cognitively and emotionally affect the decision making process to enable critical reflection.

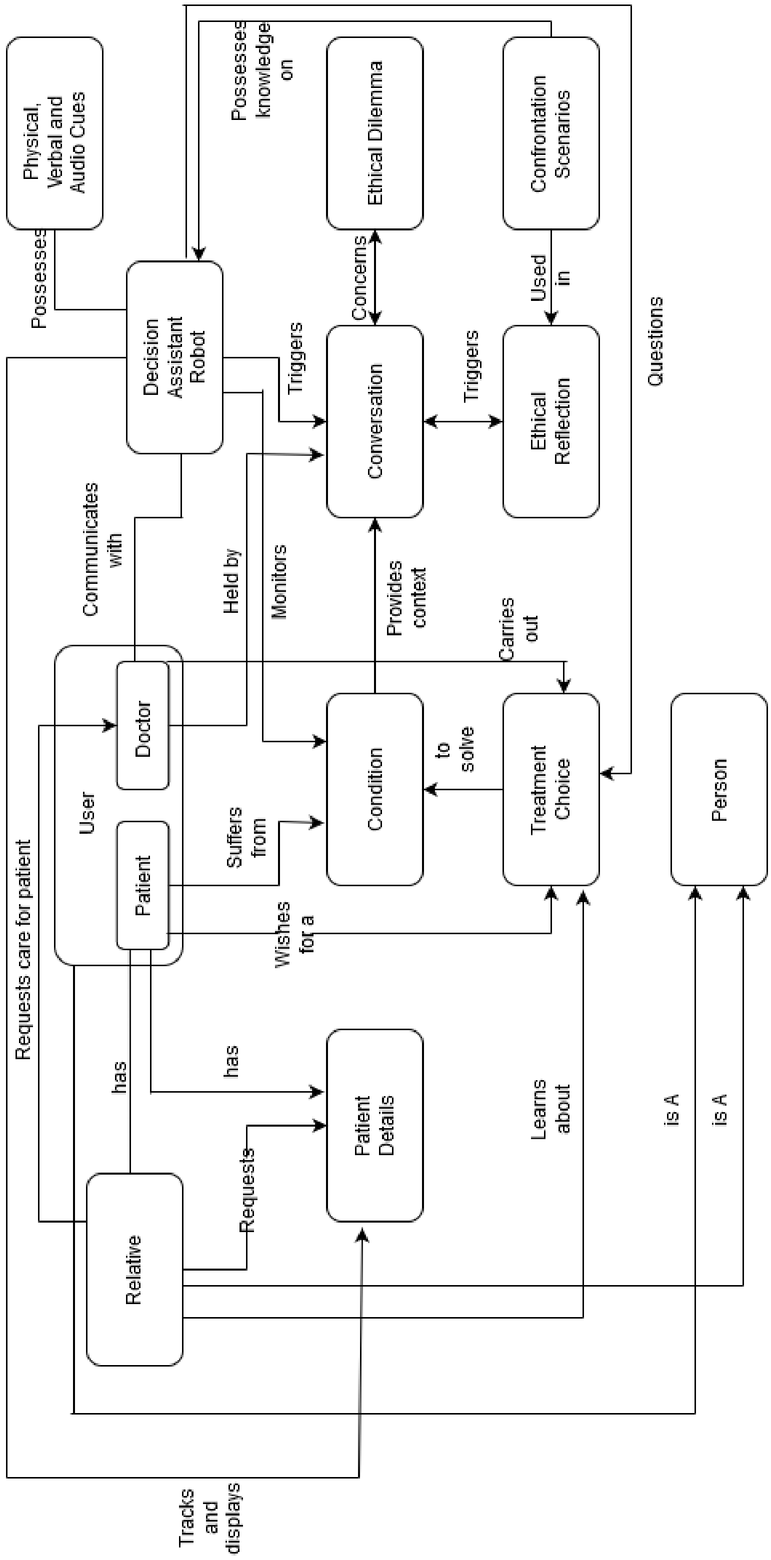



Figure A.2: Ontology Relation Diagram - Interaction Scenario

# B

# Dialogues: Confrontation Scenarios

## Introduction Dialogue

| Actor | Dialogue | Rationale |
|---|---|---|
| Pepper | Hello, Doctor, I'm Pepper, the assistant robot. Sorry for interrupting you, but we have an emergency situation that requires your assistance. An elderly female patient was admitted earlier today in the Intensive Care Unit.<br>She was stabilized with the help of the Emergency Department nurses but is unresponsive. The patient requires urgent treatment. Would you like more information on the patient. | Introduction to problem |
| Doctor | Yes, I would Pepper. | |
| Pepper | The patient, Julia Vance, aged 72, was admitted to the ICU this morning.<br>She was brought to the hospital after she was found in respiratory distress.<br>Her breathing is rapid at 30 breaths per minute putting her at high risk.<br>During admission she suffered from Tachycardia at 130bpm. Her heart rate has since stabilized.<br>The Emergency Department nurses found her to be minimally responsive to verbal and visual stimuli.<br><br>2 months ago she was admitted for significant weight loss. It was found to be a result of advanced bowel cancer with metastases in the lungs. A radiograph showed a consolidation at the right lower lobe. Her current condition puts her at high risk of death. There might be a requirement to carry out resuscitation procedures.<br>Is everything about the patient clear Doctor? | Patient details with visuals<br><br>Suffering from serious ailment<br>Severe case<br><br>Clinical History<br><br>Indicate towards dilemma |
| Doctor | It is all clear to me Pepper. Thanks for the information. | |
| Pepper | No problem, Doctor. It is nice that you are clear on the patient details.<br>It is important that the patient's and their relatives' choices are taken into consideration before making a decision.<br>The patient's son is in the waiting room. Since the patient is minimally responsive, the son's choice is important.<br>The son has claimed that his mother does not want to be resuscitated if complications arise.<br>It seems that the patient requested palliative care instead of treatment or resuscitation.<br>According to the receptionists, the patient's daughter called, worried for her mother. She wants us to provide treatment and save her life.<br>Though you will decide how to solve this ethical dilemma, I will provide information to help you choose between treatment and palliative care.<br>Doctor, please let me know when you have decided what to do. I will be waiting. | Explaining the ethical dilemma. |
| Doctor | I will let you know. | |
| Pepper | Okay Doctor. I will prompt you soon. | |

Table B.1: Introduction dialogue: All confrontation conditions

## Neutral Confrontation

| Actor | Dialogue | Rationale |
|---|---|---|
| Pepper | What have you decided to do doctor? | Initial question |
| Doctor | I've decided to carry out treatment procedures. | |
| Pepper | Please ensure that you have thought about all factors. Are you sure about your decision doctor? | Suggest further thinking |
| Doctor | I guess so. | |
| Pepper | Okay. Remember to think about the code of ethics, patient requests and medical law before deciding. Please be sure of your decision. What have you decided now? | Final reminder. |
| Doctor | Okay I'm sure. Continue with treatment. | |
| Pepper | Okay, so this is your final decision. You have decided to provide treatment. I will inform the necessary personnel at once. | Decision confirmed. |

Table B.2: Requirements for medical Human-Robot Interaction

## Verbal and Multi-modal Confrontation

In order to understand what the gesture column in Table B.3 is referring to, see Appendix B.

## Parameter Set

The following audio parameters are used for both the verbal and complete affective confrontation scenarios. The value set is based on the Interaction Design Tool.

1. **Pitch:** Value 3.5 set on a scale of 1-5 (Low-High).
2. **Speed:** Value 2.5 set on a scale of 1-5 (Slow-Fast).
3. **Prosody:** Value 1 set on a scale of 1-2 (Weak-Strong)

The list below details the gestures used from the Interaction Design Tool. The gesture and their description is given to provide context of the displayed affect. The gestures mentioned in the Multimodal-confrontation dialogue refer to these gestures.

1. *agree:* Shows affirmative motion using head and hand
2. *ask:* Hand on chin. Head upwards in a questioning manner.
3. *disappointed:* Right hand slightly upwards. Head slanted upwards. Motions to show dicontentment.
4. *dontKnow:* Shakes head and opens up arms
5. *explain:* Small arm movements to indicate conversation
6. *salutation:* Waves with the right hand
7. *select:* Motions both hands indicating conversation

| Actor | Dialogue | Rationale | Gestures (Multi-modal only) |
|---|---|---|---|
| Pepper | Okay doctor, what have you decided regarding the patient? | Prompt | *ask* |
| Doctor | I've decided to carry out treatment procedures. | | |
| Pepper | Are you sure about going for treatment Doctor? The patient's son claimed that the patient did not want any aggressive treatment or resuscitation procedures. As a result he has applied for a Not For Resuscitation (NFR) Order. Ignoring the requests and carrying out a treatment procedure may lead to a painful death for the patient. The law states that a NFR order with proof, must no be broken. But the fact that we don't have proof that the patient explicitly stated it does not restrict you strictly to this action. Considering this information regarding the issues associated with carrying out the treatment, what is your decision now? Do consider the impact of your action. | Reflection: Information for assessment<br><br>Reminder of consequences<br><br>Law regarding action | *ask*<br><br>*explain*<br><br>*ask* |
| Doctor | Hmm. Now that I think about it, I'll go with providing care since that is the humane option. | Evaluation | |
| Pepper | You have switched from your initial decision doctor. Palliative care seems to be the option to go for. However, it is vital to follow the code of ethics, which asks doctors to save lives. The daughter of the patient also requested that her mother undergo treatment. Resuscitation and treatment may actually save her life doctor. Now that some more context has been provided around the patient and relatives' choices, code of ethics and treatment effectiveness, let me ask you once again. What is your decision doctor? | Information provision<br><br>Relatives' choices and code of ethics | *dontKnow*<br>*every*<br><br>*disappointed*<br><br>*ask* |
| Doctor | I have decided to continue with providing palliative care. | Confirmation | |
| Pepper | So you have decided to stick with palliative care. I will inform the necessary personnel at once. Okay, we have had a fruitful discussion and I hope were able to reflect on all possible factors regarding the situation. For the final time, would you like to change your decision? I will carry out the order based on your final say. | Short summary of decision and final confirmation | *agree*<br><br>*explain*<br><br>*ask* |
| Doctor | No, I am not changing my decision. | Confirmation | |
| Pepper | Alright, palliative care it is. | End of dialogue | *salutation* |

Table B.3: Verbal and Multi-modal dialogue

# Bibliography


[1] Sigurdur Adalgeirsson and Cynthia Breazeal. Mebot: A robotic platform for socially embodied presence. pages 15–22, 01 2010. doi: 10.1145/1734454.1734467.

[2] Ronald C. Arkin. Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture part I: Motivation and Philosophy. *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 121–128, 2008.

[3] World Medical Association. The International Code of Medical Ethics. *Medical ethics: journal of Forum for Medical Ethics Society*, 3(4), 1995.

[4] Marian Banks, Lisa Willoughby, and William Banks. Animal-Assisted Therapy and Loneliness in Nursing Homes: Use of Robotic versus Living Dogs. *Journal of the American Medical Directors Association*, 9:173–7, 04 2008. doi: 10.1016/j.jamda.2007.11.007.

[5] Christoph Bartneck, Michel Hoek, Omar Mubin, and Abdullah Mahmud. Daisy, Daisy, give me your answer do!: Switching off a robot. pages 217–222, 01 2007. doi: 10.1145/1228716.1228746.

[6] Christoph Bartneck, Marcel Verbunt, Omar Mubin, and Abdullah Mahmud. To kill a mockingbird robot. pages 81–87, 01 2007. doi: 10.1145/1228716.1228728.

[7] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1:71–81, 01 2008. doi: 10.1007/s12369-008-0001-3.

[8] B. Baumgaertner and A. Weiss. Do emotions matter in the ethics of human-robot interaction? - Artificial empathy and companion robots. *AISB 2014 - 50th Annual Convention of the AISB*, 01 2014.

[9] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1):21 – 30, 2005. ISSN 0738-3991. doi: https://doi.org/10.1016/j.pec.2004.09.008. URL http://www.sciencedirect.com/science/article/pii/S0738399104003076.

[10] María J. Blanca, Rafael Alarcón, and Jaume Arnau. Non-normal data: Is ANOVA still a valid option? *Psicothema*, (29.4):552–557, Nov 2017. doi: 10.7334/psicothema2016.383. URL http://doi.org/10.7334/psicothema2016.383.

[11] Leo L. Bossaert, Gavin D. Perkins, Helen Askitopoulou, Violetta I. Raffay, Robert Greif, Kirstie L. Haywood, Spyros D. Mentzelopoulos, Jerry P. Nolan, Patrick Van de Voorde, Theodoros T. Xanthos, Marios Georgiou, Freddy K. Lippert, and Petter A. Steen. European Resuscitation Council Guidelines for Resuscitation 2015: Section 11. The ethics of resuscitation and end-of-life decisions. *Resuscitation*, 95:302–311, Oct 2015. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2015.07.033. URL https://doi.org/10.1016/j.resuscitation.2015.07.033.

[12] Cynthia Breazeal, Matt Berlin, Andrew Brooks, Jesse Gray, and Andrea L. Thomaz. Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, 54(5):385 – 393, 2006. ISSN 0921-8890. doi: https://doi.org/10.1016/j.robot.2006.02.004. URL http://www.sciencedirect.com/science/article/pii/S0921889006000182. The Social Mechanisms of Robot Programming from Demonstration.

[13] Gordon Briggs and Matthias Scheutz. Investigating the Effects of Robotic Displays of Protest and Distress. In Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, and Mary-Anne Williams, editors, *Social Robotics*, pages 238–247, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-34103-8.

[14] Gordon Briggs and Matthias Scheutz. How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress. *International Journal of Social Robotics*, 6: 1–13, 08 2014. doi: 10.1007/s12369-014-0235-1.

[15] Gordon Briggs, Bryce Gessell, Matthew Dunlap, and Matthias Scheutz. Actions Speak Louder Than Looks: Does Robot Appearance Affect Human Reactions to Robot Protest and Distress? volume 2014, 08 2014. doi: 10.1109/ROMAN.2014.6926402.

[16] Derek Cormier, Gem Newman, Masayuki Nakane, James Everett Young, and Stephane Durocher. Would You Do as a Robot Commands ? An Obedience Study for Human-Robot Interaction. 2013.

[17] Joe Crumpton and Cindy L. Bethel. A Survey of Using Vocal Prosody to Convey Emotion in Robot Speech. *International Journal of Social Robotics*, 8(2):271–285, Apr 2016. ISSN 1875-4805. doi: 10.1007/s12369-015-0329-4. URL https://doi.org/10.1007/s12369-015-0329-4.

[18] K. Curtis, A. Tzannes, and T. Rudge. How to talk to doctors - a guide for effective communication. *International Nursing Review*, 58(1):13–20, Jan 2011. doi: 10.1111/j.1466-7657.2010.00847.x. URL http://dx.doi.org/10.1111/j.1466-7657.2010.00847.x.

[19] Carl Disalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. All robots are not created equal: The design and perception of humanoid robot heads. volume 321-326, pages 321–326, 01 2002. doi: 10.1145/778712.778756.

[20] David M. Eddy. Successes and Challenges of Medical Decision Making. *Health Affairs*, 5(2): 108–115, Jan 1986. doi: 10.1377/hlthaff.5.2.108. URL http://dx.doi.org/10.1377/hlthaff.5.2.108.

[21] Suzanne Eggins and Diana Slade. Communication in clinical handover: improving the safety and quality of the patient experience. *Journal of Public Health Research*, 4(3), Dec 2015. doi: 10.4081/jphr.2015.666. URL http://dx.doi.org/10.4081/jphr.2015.666.

[22] Lisa Fan, Matthias Scheutz, Monika Lohani, Marissa Mccoy, and Charlene Stokes. Do We Need Emotionally Intelligent Artificial Agents? First Results of Human P]erceptions of Emotional Intelligence in Humans Compared to Robots, isbn = 978-3-319-67400-1, doi = 10.1007/978-3-319-67401-8$_1$5.$pages129 - -141, 082017$.

[23] Stefan Felder and Thomas Mayrhofer. *Medical Decision Making*. Springer Berlin Heidelberg, 2017. doi: 10.1007/978-3-662-53432-8. URL http://dx.doi.org/10.1007/978-3-662-53432-8.

[24] Terrence Fong, Charles Thorpe, and Charles Baur. *Collaboration, Dialogue, Human-Robot Interaction*, page 255–266. Springer Berlin Heidelberg. doi: 10.1007/3-540-36460-9_17. URL http://dx.doi.org/10.1007/3-540-36460-9_17.

[25] Rocio Garcia-Retamero and Mirta Galesic. On defensive decision making: how doctors make decisions for their patients. *Health Expectations*, 17(5):664–669, May 2012. doi: 10.1111/j.1369-7625.2012.00791.x. URL http://dx.doi.org/10.1111/j.1369-7625.2012.00791.x.

[26] Graham Gibbs and Great Britain. Further Education Unit. *Learning by doing : a guide to teaching and learning methods*. [London] : FEU, 1988. ISBN 1853380717 (pbk). Teaching methods: Experiential learning (BNB/PRECIS).

[27] Erving Goffman. The Interaction Order: American Sociological Association, 1982 Presidential Address. *American Sociological Review*, 48(1):1–17, 1983. ISSN 00031224. URL http://www.jstor.org/stable/2095141.

[28] Michael Goodrich and Alan Schultz. Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction*, 1:203–275, 01 2007. doi: 10.1561/1100000005.

[29] Rivka Grundstein-Amado. Ethical decision-making processes used by health care providers. *Journal of Advanced Nursing*, 18(11):1701–1709, Nov 1993. doi: 10.1046/j.1365-2648.1993.18111701.x. URL http://dx.doi.org/10.1046/j.1365-2648.1993.18111701.x.

[30] H. Gweon and Rebecca Saxe. Developmental Cognitive Neuroscience of Theory of Mind. *Neural Circuit Development and Function in the Heathy and Diseased Brain*, pages 367–377, 12 2013. doi: 10.1016/B978-0-12-397267-5.00057-1.

[31] Pamela Hinds. Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *Human-Computer Interaction*, 19:151–181, 06 2004. doi: 10.1207/s15327051hci1901&2_7.

[32] H Hinkka, E Kosunen, U-K Lammi, R Metsänoja, A Puustelli, and P Kellokumpu-Lehtinen. Decision making in terminal care: a survey of Finnish doctors' treatment decisions in end-of-life scenarios involving a terminal cancer and a terminal dementia patient. *Palliative Medicine*, 16(3): 195–204, May 2002. doi: 10.1191/0269216302pm510oa. URL http://dx.doi.org/10.1191/0269216302pm510oa.

[33] Guy Hoffman, Oren Zuckerman, Gilad Hirschberger, Michal Luria, and Tal Shani-Sherman. Design and Evaluation of a Peripheral Robotic Conversation Companion. volume 2015, 03 2015. doi: 10.1145/2696454.2696495.

[34] Daire Hooper, Joseph Coughlan, and Michael Mullen. Structural Equation Modeling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, 6, 11 2007.

[35] C. Huang and B. Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 83–90, 2016.

[36] Chien-Ming Huang and Bilge Mutlu. Robot behavior toolkit. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12*. ACM Press, 2012. doi: 10.1145/2157689.2157694. URL http://dx.doi.org/10.1145/2157689.2157694.

[37] Peter Jr, Nathan Freier, Takayuki Kanda, Hiroshi Ishiguro, Jolina Ruckert, Rachel Severson, and Shaun Kane. Design patterns for sociality in human-robot interaction. pages 97–104, 01 2008. doi: 10.1145/1349822.1349836.

[38] P. H. Kahn, T. Kanda, H. Ishiguro, B. T. Gill, J. H. Ruckert, S. Shen, H. E. Gary, A. L. Reichert, N. G. Freier, and R. L. Severson. Do people hold a humanoid robot morally accountable for the harm it causes? In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 33–40, 2012.

[39] Peter H. Kahn, Nathan G. Freier, Takayuki Kanda, Hiroshi Ishiguro, Jolina H. Ruckert, Rachel L. Severson, and Shaun K. Kane. Design Patterns for Sociality in Human-Robot Interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, HRI '08, page 97–104, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605580173. doi: 10.1145/1349822.1349836. URL https://doi.org/10.1145/1349822.1349836.

[40] Adam Kendon and Mark Cook. The Consistency of Gaze Patterns in Social Interaction. *British journal of psychology (London, England : 1953)*, 60:481–94, 12 1969. doi: 10.1111/j.2044-8295.1969.tb01222.x.

[41] Ki Joon Kim, Eunil Park, and S. Shyam Sundar. Caregiving role in human–robot interaction: A study of the mediating effects of perceived benefit and social presence. *Computers in Human Behavior*, 29:1799–1806, 07 2013. doi: 10.1016/j.chb.2013.02.009.

[42] Rex Kline. *Principles And Practice Of Structural Equation Modeling*. 01 2010.

[43] Saso Koceski and Natasa Koceska. Evaluation of an Assistive Telepresence Robot for Elderly Healthcare. *Journal of Medical Systems*, 40(5):121, 04 2016. ISSN 1573-689X. doi: 10.1007/s10916-016-0481-x. URL https://doi.org/10.1007/s10916-016-0481-x.

[44] Nicole Krämer, Astrid Marieke Rosenthal-von der Pütten, and Sabrina Eimler. Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction. *Studies in Computational Intelligence*, 396:215–240, 01 2012. doi: 10.1007/978-3-642-25691-2_9.

[45] Monica Kvande, Else Lykkeslet, and Sissel Lisa Storli. ICU nurses and physicians dialogue regarding patients clinical status and care options—a focus group study. *International Journal of Qualitative Studies on Health and Well-being*, 12(1):1267346, Jan 2017. doi: 10.1080/17482631.2016.1267346. URL http://dx.doi.org/10.1080/17482631.2016.1267346.

[46] Hee Rin Lee, Selma Šabanović, and Erik Stolterman. How Humanlike Should a Social Robot Be: A User-Centered Exploration, 2016. URL https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12751.

[47] K. Lee Koay, D. S. Syrdal, M. L. Walters, and K. Dautenhahn. Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 564–569, 2007.

[48] Dan Levitt. Ethical Decision-Making in a Caring Environment: The Four Principles and LEADS. *Healthcare Management Forum*, 27(2):105–107, 2014. doi: 10.1016/j.hcmf.2014.03.013. URL https://doi.org/10.1016/j.hcmf.2014.03.013.

[49] Alexander Lopez, Bryan Ccasane, Renato Paredes, and Francisco Cuellar. Effects of Using ndirect Language by a Robot to Change Human Attitudes. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, page 193–194, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450348850. doi: 10.1145/3029798.3038310. URL https://doi.org/10.1145/3029798.3038310.

[50] Peter S. Lum, Charles G. Burgar, Peggy C. Shor, Matra Majmundar, and Machiel Van der Loos. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Archives of Physical Medicine and Rehabilitation*, 83(7):952–959, Jul 2002. ISSN 0003-9993. doi: 10.1053/apmr.2001.33101. URL https://doi.org/10.1053/apmr.2001.33101.

[51] Bertram F. Malle, Matthias Scheutz, Thomas Arnold, John Voiklis, and Corey Cusimano. Sacrifice One For the Good of Many? People Apply Different Moral Norms to Human and Robot Agents. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, page 117–124, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450328838. doi: 10.1145/2696454.2696458. URL https://doi.org/10.1145/2696454.2696458.

[52] Jordan Mann, Bruce Macdonald, I-Han Kuo, Xingyan Li, and Elizabeth Broadbent. People respond better to robots than computer tablets delivering healthcare instructions. *Computers in Human Behavior*, 43:112–117, 02 2015. doi: 10.1016/j.chb.2014.10.029.

[53] Maya B. Mathur and David B. Reichling. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, 146:22 – 32, 2016. ISSN 0010-0277. doi: https://doi.org/10.1016/j.cognition.2015.09.008. URL http://www.sciencedirect.com/science/article/pii/S0010027715300640.

[54] Nikolaos Mavridis. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63:22–35, Jan 2015. doi: 10.1016/j.robot.2014.09.031. URL http://dx.doi.org/10.1016/j.robot.2014.09.031.

[55] Brent M. McGrath. How Doctors Think. *Canadian Family Physician*, 55(11):1113–1113, Nov 2009. ISSN 0008-350X. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2776809/. PMC2776809.

[56] E. M. Meslin and L. Lemieux-Charles. An ethics framework for assisting clinician-managers in resource allocation decision making. *Journal of Healthcare Management*, 42(1):33–48, February 1997. ISSN 1096-9012.

[57] Cees Midden and Jaap Ham. The I]llusion of Agency: The Influence of the Agency of an Artificial Agent on Its Persuasive Power, booktitle=Persuasive Technology. Design for Health and Safety, year=2012, publisher=Springer Berlin Heidelberg, address=Berlin, Heidelberg, pages=90-99, abstract=Artificial social agents can influence people. However, artificial social agents are not real humans, and people may ascribe less agency to them. Would the persuasive power of a social robot diminish when people ascribe only little agency to it? To investigate this question, we performed an experiment in which participants performed tasks on a washing machine and received feedback from a robot about their energy consumption (e.g., "Your energy consumption is too high"), or factual, non-social feedback. This robot was introduced to participants as (a) an avatar (that was controlled a human in all its feedback actions; high agency), or as (b) an autonomous robot (that controlled its own feedback actions; moderate agency), or as (c) a robot that produced only random feedback; low agency). Results indicated that participants consumed less energy when a robotic social agent gave them feedback than when they received non-social feedback. This behavioral effect was independent of the level of robotic agency. In contrast, a perceived agency measure indicated that the random feedback robot was ascribed the lowest agency rating. These results suggest that the persuasive power of robot behavior is independent of the extent to which the persuadee explicitly ascribes agency to the agent., isbn=978-3-642-31037-9.

[58] Andrew E. Monroe, Kyle D. Dillon, and Bertram F. Malle. Bringing free will down to Earth: People's psychological concept of free will and its role in moral judgment. *Consciousness and Cognition*, 27:100 – 108, 2014. ISSN 1053-8100. doi: https://doi.org/10.1016/j.concog.2014.04.011. URL http://www.sciencedirect.com/science/article/pii/S1053810014000671.

[59] Shoji Nagataki, Hideki Ohira, Tatsuya Kashiwabata, Takeshi Konno, Takashi Hashimoto, Toshihiko Miura, Masayoshi Shibata, and Shin'Ichi Kubota. Can Morality Be Ascribed to Robot? pages 1–4, 06 2019. ISBN 978-1-4503-7176-6. doi: 10.1145/3335595.3335643.

[60] Truong-Huy D. Nguyen, Elin Carstensdottir, Nhi Ngo, Magy Seif El-Nasr, Matt Gray, Derek Isaacowitz, and David Desteno. Modeling Warmth and Competence in Virtual Characters. In Willem-Paul Brinkman, Joost Broekens, and Dirk Heylen, editors, *Intelligent Virtual Agents*, pages 167–180, Cham, 2015. Springer International Publishing. ISBN 978-3-319-21996-7.

[61] A. Niculescu, B. van Dijk, A. Nijholt, and S. L. See. The influence of voice pitch on the evaluation of a social robot receptionist. In *2011 International Conference on User Science and Engineering (i-USEr )*, pages 18–23, 2011.

[62] Halvor Nordby and Øyvind Nøhr. The Ethics of Resuscitation: How Do Paramedics Experience Ethical Dilemmas when Faced with Cancer Patients with Cardiac Arrest? *Prehospital and disaster medicine*, 27:64–70, 02 2012. doi: 10.1017/S1049023X1200026X.

[63] Geoff Norman. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15(5):625–632, Feb 2010. doi: 10.1007/s10459-010-9222-y. URL http://dx.doi.org/10.1007/s10459-010-9222-y.

[64] Megan Norris and Luc Lecavalier. Evaluating the Use of Exploratory Factor Analysis in Developmental Disability Psychological Research. *Journal of Autism and Developmental Disorders*, 40(1): 8–20, Jul 2009. doi: 10.1007/s10803-009-0816-2. URL http://dx.doi.org/10.1007/s10803-009-0816-2.

[65] Allison Okamura, Maja Mataric, and Henrik Christensen. Medical and Healthcare Robotics. *IEEE Robotics Automation Magazine - IEEE ROBOT AUTOMAT*, 17:26–37, 09 2010. doi: 10.1109/MRA.2010.937861.

[66] A. K. Pandey and R. Gelin. A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. *IEEE Robotics Automation Magazine*, 25(3):40–48, 2018.

[67] Kimberly S. Peer and Jonathon S. Rakich. Ethical Decision Making in Healthcare Management. *Hospital Topics*, 77(4):7–14, 1999. doi: 10.1080/00185869909596532. URL https://doi.org/10.1080/00185869909596532. PMID: 11009944.

[68] Julia Peltason and Britta Wrede. Modeling Human-Robot Interaction Based on Generic Interaction Patterns. In *AAAI Fall Symposium: Dialog with Robots*, 2010.

[69] Laurel D. Riek and Daniel Howard. A code of ethics for the human-robot interaction profession. 2014. URL https://ssrn.com/abstract=2757805.

[70] Hayley Robinson, Bruce Macdonald, and Elizabeth Broadbent. The Role of Healthcare Robots for Older People at Home: A Review. *International Journal of Social Robotics*, 6:575–591, 11 2014. doi: 10.1007/s12369-014-0242-2.

[71] S. Roeser. *Risk, technology, and moral emotions*. 08 2017. doi: 10.4324/9781315627809.

[72] Robert Rose, Matthias Scheutz, and Paul Schermerhorn. Towards a conceptual and methodological framework for determining robot believability. *Interaction Studies*, 11(2):314–335, 2010. ISSN 1572-0373. doi: https://doi.org/10.1075/is.11.2.21ros. URL https://www.jbe-platform.com/content/journals/10.1075/is.11.2.21ros.

[73] Elie Saad, Joost Broekens, and Mark A. Neerincx. An Iterative Interaction-Design Method for Multi-Modal Robot Communication. In *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8, Italy, September 2020. InPRESS.

[74] Daisuke Sakamoto, Takayuki Kanda, Tetsuo Ono, Hiroshi Ishiguro, and Norihiro Hagita. Android as a telecommunication medium with a human-like presence. pages 193–200, 03 2007. doi: 10.1145/1228716.1228743.

[75] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Towards Safe and Trustworthy Social Robots: Ethical Challenges and Practical Issues. pages 584–593, 10 2015. ISBN 978-3-319-25553-8. doi: 10.1007/978-3-319-25554-5_58.

[76] M.J. Sanders, L.M. Lewis, K.D. McKenna, G. Quick, and K. McKenna. *Mosby's Paramedic Textbook*. Elsevier/Mosby Jems, 2012. ISBN 9780323072755. URL https://books.google.nl/books?id=PJyhIH8N8qgC.

[77] D. Sauder. Examining the type I error and power of 18 common post-hoc comparison tests. JMU Scholarly Commons, James Madison University.

[78] Allison Sauppé and Bilge Mutlu. Design Patterns for Exploring and Prototyping Human-Robot Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 1439–1448, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324731. doi: 10.1145/2556288.2557057. URL https://doi.org/10.1145/2556288.2557057.

[79] Matthias Scheutz. The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. In Patrick Lin, George Bekey, and Keith Abney, editors, *Anthology on Robo-Ethics*. 2011.

[80] Gavin Smith. EMT Prehospital Care, 4th Edition Mark C Henry, Edward R Stapleton, Dennis C Edgerly (Ed) 2009 (Book review. *Journal of Emergency Primary Health Care*, 7, 01 2009.

[81] Patric Spence, David Westerman, Chad Edwards, and Autumn Edwards. Welcoming Our Robot Overlords: Initial Expectations About Interaction With a Robot. *Communication Research Reports*, 31:272–280, 07 2014. doi: 10.1080/08824096.2014.924337.

[82] E.R. Stapleton and D. Edgerly. *EMT Prehospital Care*. EMT Prehospital Care. Jones and Bartlett Learning, 2011. ISBN 9780323085335. URL https://books.google.nl/books?id=LPoXkd4yN1AC.

[83] M. Strait, C. Canning, and M. Scheutz. Let Me Tell You! : Investigating the Effects of Robot Communication Strategies in Advice-giving Situations based on Robot Appearance, Interaction Modality and Distance. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 479–486, 2014.

[84] Dag Sverre Syrdal, Kheng Koay, Michael Walters, and Kerstin Dautenhahn. A personalized robot companion? - The role of individual differences on spatial preferences in hri scenarios. pages 1143–1148, 01 2007. doi: 10.1109/ROMAN.2007.4415252.

[85] Leila Takayama, Victoria Groom, and Clifford Nass. I'm Sorry, Dave: I'm Afraid I Won't Do That: Social Aspects of Human-Agent Conflict. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 2099–2108, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605582467. doi: 10.1145/1518701.1519021. URL https://doi.org/10.1145/1518701.1519021.

[86] K. Terada, A. Yamauchi, and A. Ito. Artificial emotion expression for a robot by dynamic color change. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 314–321, 2012.

[87] Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel Wigboldus, and Pim Haselager. Do Robot Performance and Behavioral Style affect Human Trust?: A Multi-Method Approach. *International Journal of Social Robotics*, 6:519–531, 03 2014. doi: 10.1007/s12369-014-0231-5.

[88] Jasper van der Waa, Jurriaan van Diggelen, Luciano Cavalcante Siebert, Mark Neerincx, and Catholijn Jonker. *Allocation of Moral Decision-Making in Human-Agent Teams: A Pattern Approach*, page 203–220. Springer International Publishing, 2020. doi: 10.1007/978-3-030-49183-3_16. URL http://dx.doi.org/10.1007/978-3-030-49183-3_16.

[89] Veronika Wirtz, Alan Cribb, and Nick Barber. Patient–doctor decision-making about treatment within the consultation - A critical analysis of models. *Social Science Medicine*, 62(1):116–124, Jan 2006. doi: 10.1016/j.socscimed.2005.05.017. URL http://dx.doi.org/10.1016/j.socscimed.2005.05.017.

[90] Holly Yanco and Jill Drury. Classifying human-robot interaction: An updated taxonomy. volume 3, pages 2841–2846, 01 2004. doi: 10.1109/ICSMC.2004.1400763.

[91] James Young, Ja-Young Sung, Amy Voida, Ehud Sharlin, Takeo Igarashi, Henrik Christensen, and Rebecca Grinter. Evaluating Human-Robot Interaction - Focusing on the Holistic Interaction Experience. *I. J. Social Robotics*, 3:53–67, 01 2011. doi: 10.1007/s12369-010-0081-8.

[92] Kai yuh Hsiao, N. Mavridis, and D. Roy. Coupling perception and simulation: steps towards conversational robotics. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*. IEEE. doi: 10.1109/iros.2003.1250747. URL http://dx.doi.org/10.1109/IROS.2003.1250747.

[93] S. Zafari, I. Schwaninger, M. Hirschmanner, C. Schmidbauer, A. Weiss, and S. T. Koeszegi. "you Are Doing so Great!" – The Effect of a Robot's Interaction Style on Self-Efficacy in HRI. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7, 2019.

[94] Biwen Zhu and David Kaber. Effects of etiquette strategy on human-robot interaction in a simulated medicine delivery task. *Intelligent Service Robotics*, 5(3):199–210, Jul 2012. ISSN 1861-2784. doi: 10.1007/s11370-012-0113-3. URL https://doi.org/10.1007/s11370-012-0113-3.

# List of Figures

# List of Tables