Enhancing robot legibility in hospital environments through shape-changing

Tom Ossendrijver

Interaction Technology, University of Twente

Supervisors: Daniel Davison, Khiet Truong, Bob Schadenberg

Thursday 15$^{\text{th}}$ May, 2025

Keywords: Human-robot interaction, Shape-changing robot, Hospital robot, Legibility, Social robotics, Robot navigation

Abstract

As robots increasingly navigate social environments like hospital hallways, effective communication of their intentions is of importance. This paper investigates shape-changing, the ability of an entity to intentionally manipulate its torso to undergo deformations with the aim of conveying messages, intentions, or information, as a novel communication modality to improve robot legibility for hospital navigation.

While existing non-verbal modalities, such as lights and gestures, are already used to enhance robot legibility, they can fall short in meeting the complex demands of hospital environments. To address the limitations, this study explores shape-changing as a novel modality. Inspired by natural behaviors and animation principles, shape-changing offers visually clear and intuitive cues that may better align with the crowded and diverse nature of hospital environments.

We conducted a 2x2 within-subjects experiment in a virtual reality hospital environment, comparing shape-changing with blinkers. Participants interacted with a robot using gaze, shape-changing, blinkers, or a combination of these modalities. We measured response time, accuracy, confidence, and surprise ratings, which are supported by qualitative findings. Additionally, we used the Robotic Social Attributes Scale (ROSAS) to examine how shape-changing influences the perceived social attributes of the robot.

Results suggest that shape-changing is a promising modality for improving legibility, but its effectiveness depends on design and integration with other modalities. While shape-changing was perceived as intuitive, it could also be intimidating or shocking. Combining shape-changing with blinkers improved user confidence and reduced surprise, though inconsistencies between modalities sometimes caused mixed signals.

The main contribution of this paper is to demonstrate that shape-changing shows promise as a novel modality for improving robot legibility in hospital environments. By testing shape-changing in a real robot prototype, we ensure the findings are applicable in practice. Additionally, this study provides valuable insights for the design and integration of shape-changing behavior to improve human-robot interaction, leading to more legible communication in social robotics.

Acknowledgments

First of all, I would like to thank my supervisors, Daniel Davison, Khiet Truong, and Bob Schadenberg, for guiding me through my thesis, providing invaluable feedback, and offering continuous support throughout the process. I am deeply grateful for the knowledge I gained from them during this journey.

Secondly, I want to thank the Interaction Lab and team at the University of Twente. Without their support, resources, and the environment they provided for conducting my studies, this work would not have been possible.

Furthermore, I am thankful to all the participants who took part in the studies, as well as the technical staff who greatly assisted with the experimental setup. Their contributions were essential to the success of this project.

Finally, on a personal note, I would like to thank my family and the friends I have made throughout this journey for their support and encouragement, which helped me through challenging times.

List of Figures

Contents

1 Introduction

In today's world, the presence of social robots is growing exponentially [39], particularly in areas such as Human-Robot Interaction (HRI) [10], with the goal of improving human-robot communication [27]. Robots are increasingly being integrated into social environments, including hospitals, where they assist healthcare personnel by reducing patient stress and loneliness, thereby improving patient care [40, 45, 44]. Furthermore, effective communication between humans and robots is crucial, especially in diverse and crowded spaces like hospital hallways, where robots must navigate safely and perform tasks such as delivering medications or guiding patients [12].

However, in the complex landscape of human-robot interaction in hospital navigation, ensuring effective communication is a challenge due to the diversity among individuals encountered. This diversity includes variations in language, culture, and individual needs, influencing the effectiveness and in turn acceptance of social robots [43].

To improve the acceptance of these robots, trust should be built, which can be increased by making robots more legible [41]. Robot behavior is considered legible if "a human observer or interactor is able to understand its intentions, and the behavior met the expectations of the human observer or interactor" [28].

To make robots more legible, various modalities, such as light indicators, auditory signals, screen displays and gestures, have been employed. However, these modalities have limitations in visibility, design constraints, and information transfer [25, 50, 36, 31, 22]. For instance, light indicators, may offer limited expressiveness in certain contexts, as they often rely on simple color or blinking patterns that can lack nuance, due to their binary nature. While usefully for conveying basic statuses, like: on, standby and error, communicating more complex subtle intentions may be challenging. Auditory signals, such as beeps or tones are not always straightforward, as audio cues for subtle changes can be difficult for people to interpret accurately, especially in noisy hospital environments where alarms and background noise are present, making the auditory signals less dominant. Screen displays, can be difficult to read in crowded or fast-paced hospital like settings. Gestures, such as pointing, waving or nodding, are a form of nonverbal communication that can help indicate a robot's intent or direct human attention. However, they often require human-like features, such as hands, arms or a head, which may not be feasible for all robot designs. Overall, while the modalities help improve the legibility, they have their limitations. Due to the diversity in a hospital it is a challenge to make the robots intentions understandable, while meeting the expectations as interpretation is subject to individual variation [1]. which can lead to misunderstandings and inefficiencies in communication [47]. In turn leading to difficulties understanding the robot behavior, causing irritation [31, 25].

Indicates a need for innovative approaches to enhance robot legibility, while keeping into account the diverse patients and medical professionals with various expectations, perceptions and communication styles. This needs to be considered for the acceptance of robotic technology [48].

To address these challenges, researchers have explored multi-modal interfaces, as they offer an approach addressing various needs and preferences of users, allowing them to engage with the robot through their preferred sensory channel. This has shown to improve the acceptance of robotic services for patient care [32]. Humans naturally make use of non-verbal cues such as gestures, postures, and facial expressions to enrich communication and convey meaning beyond words. Incorporating non-verbal communication strategies into robots has shown to significantly improve human-robot interaction in social environments [25]. Building on this, researchers have begun to explore more embodied and expressive forms of non-verbal communication [19].

One emerging approach from this line of research is shape-changing. In this context, we aim to complement existing modalities with shape-changing as a novel non-verbal channel to communicate a robot's intent. We define shape-changing as: "the ability of an entity to intentionally manipulate its torso to undergo deformations to convey messages, intentions, or information." Presenting a novel approach to enhance robot legibility by providing intuitive non-verbal cues.

Shape-changing draws inspiration from natural communication strategies observed in humans and animals, while making use of animation principles and taking into account the expressive movement parameters to shape the movements interpretation. While prior work has demonstrated the potential of shape-changing to convey emotions, as demonstrated in the sprout project [53], its use case for improving legibility, especially in social hallways remains to our knowledgeable unexplored. Our research is not focused on replacing the modalities but add a new modality, thereby widening the communication spectrum.

We propose that shape-changing can overcome the limitations of the previously mentioned modalities by enabling more expressive communication than blinkers and increasing information transfer through the use of frame sequences (frame-based progressions of physical transformations) that are intuitively interpretable and capable of conveying subtle nuances. This is interpreted intuitively as it is based on natural communication strategies. Additionally, shape-changing allows for more design flexibility as it does not require a human like design, while being attention grabbing and visually clear due to the deformation of the robot's entire torso as a physical cue (see Table 1). The shape-changing robot design used in this study is shown in Figure 1, adopting a deformed, left leaning pose.

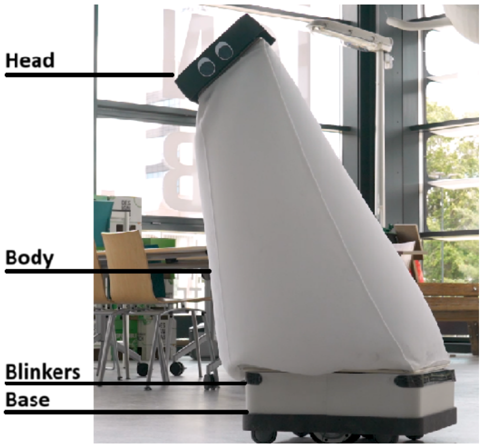


Figure 1: Shape-changing robot

To assess the legibility of shape-changing, we compare it with gaze, a naturally present and continuously available communication modality, given the robot's face. Since communication inherently occurs when an open channel exists [15], gaze is always active during interactions. Therefore, by comparing gaze alone with the combination of gaze and shape-changing, we aim to determine whether shape-changing contributes to the robot's legibility beyond what gaze already communicates. We further evaluate how shape-changing performs compared to blinkers, vehicle like indicators, widely used in daily life to communicate directional intent, to determine if shape-changing cues offer advantages over familiar standardized signals. We also explore if combining shape-changing with blinkers improves legibility, building on evidence that multi-modal approaches improve robot acceptance [32]. And determine if shape-changing influences the perception of robots. Providing insights, required for making robotic technologies more accepted [48]

Therefore, this study aims to explore the potential of shape-changing as a novel communication modality. By trying to answer the research question: *How can we use shape-changing to enhance robot legibility for navigation in social hallways?* This research question aims to bridge the gap in literature while laying a foundation for future developments in human-robot interaction. To answer this main research question, we came up with the following sub research questions:

- How legible is shape-changing in comparison to gaze as directional modality?
- How legible is shape-changing in comparison to blinkers as a directional modality?
- To what extent can shape-changing contribute to improving the legibility of other modalities, such as blinkers?
- How does shape-changing influence participants perceptions of the robot's likability, intelligence, and safety during navigation?

This study contributes to the field of human-robot interaction by demonstrating the potential of shape-changing as a novel communication modality to improve robot legibility. It also highlights the importance of multi-modal approaches, showing that combining shape-changing with modalities like blinkers can improve the information richness and predictability of the robot's intent. Additionally, the study bridges the gap between theoretical design and practical application by implementing shape-changing behaviors in a real-world robot prototype and translating them into a VR environment for testing.

Chapter 2 provides background on the current state of human-robot interaction, the challenges faced in hospital hallways, the limitations of traditional communication modalities, and the potential of shape-changing as a novel interaction method. Chapter 3 outlines the problem statement, research questions, and corresponding hypotheses. Chapter 4 details the design and implementation of the shape-changing behavior, including key design decisions informed by pilot testing. Chapter 5 describes the main study, covering the experimental scenarios and conditions, participant recruitment, setup, data collection procedures, and an analysis of the required sample size, along with an overview of the planned analysis approach. Chapter 6 presents the quantitative results, while Chapter 7 explores the thematic analysis of qualitative feedback. Chapter 8 discusses the findings in relation to the research questions, highlights fulfilled requirements, addresses limitations, and offers recommendations for future work. Finally, Chapter 9 concludes the study by summarizing the key findings.

This research aims to contribute to more legible human-robot interactions in social environments.

2 Background

In this background section, we outline the opportunities and challenges of designing robots that can legibly communicate their movement intentions in shared spaces, such as hospital hallways. First, we discuss the importance of non-verbal communication, highlighting its advantages and challenges in these environments. Next, we introduce shape-changing as a novel non-verbal communication modality that conveys intent through familiar human body language and animal-like behavior, enhanced by animation principles and movement parameters. [1]

2.1 Social robots in healthcare environments

The integration of social robots into healthcare settings has grown significantly [39], driven by their potential to improve patient care and assist medical staff [40, 44] and with the goal of improving human-robot interaction [27].

Robots in hospitals perform diverse tasks, such as delivering medications, folding laundry, helping with rehabilitation, doing administrative tasks, acting as mediator for nurses, helping people with motor impairments, keeping elderly company while navigating and more [12, 45]. Indicating the dual role hospital robots have, they must accomplish functional tasks (for example, navigation and object transport), while operating in a social complex environments where interactions with patients, staff, and visitors take place.

For example, a delivery robot navigating a crowded hospital hallway must not only avoid obstacles but also communicate its navigational intents clearly to humans. Whether it is yielding, changing direction, or taking priority, which requires a rich information transfer. Similarly, a companion robot interacting with patients must adapt its behavior to diverse cultural norms and individual preferences to avoid misunderstandings [48]. This combination between task efficiency and social competence makes social hallway navigation in hospitals an interesting domain to focus on for studying human-robot communication.

Despite the potential of social robots navigating these hospital environments, navigating such social spaces, bring challenges, which should be taken into account.

For example, hospitals hallways serve diverse users with variations in language, culture, sensory abilities (such as visual or hearing impairments), experience (first encounter, repeated exposure), cognitive states (such as stressed or distracted) and other individual needs, making it challenging to design universally understandable (intuitive) robot behaviors. This diversity influences both the acceptance and effectiveness of social robots [43], as these individual differences lead to varying interpretations of robotic intentions [1]. Which can lead to misunderstandings and inefficiencies in communication [47]. In turn leading to difficulties understanding the robots behavior, causing irritation [31, 25].

Further, hospital hallways are high-traffic environments where urgent scenarios (such as staff rushing to emergencies) require robot actions to be immediately visible to all nearby humans. This visibility is required for legibility, with legibility is meant that, "a human observer or interactor is able to understand its intentions, and the behavior met the expectations of the human observer or interactor" [28]. Where response time and accuracy are metrics for assessing legibility, as shown by [13, 21, 20]. Making robots more legible, it is believed to build trust between humans and robots, which in turn improves the acceptance of social robots [41].

2.2 Non-verbal communication for robot legibility

To take these challenges into account, multi-modal interfaces have being explored, as this allows humans to engage with the robot through their preferred sensory channel, which has shown to improve the acceptance of robotic services for patient care [32]. Human communication naturally relies on implicit non-verbal cues like altering body posture to convey confidence or doubt [9], or by making use of gaze and pointing gestures to draw attention towards a shared object or location [7]. Further, as highlighted by Tomasello, pantomiming, a form of iconic gesture that mimics real-world actions, plays a key role in natural communication by leveraging the recipient's ability to interpret intentional actions. This directs the imagination of others and relies on shared common ground for effective interpretation [11]. It is believed that humans making use of such implicit and non-verbal cues, enables seamless human navigation in crowds, however this kind of effective and efficient communication often falls short in human-robot encounters [25]. Highlighting the potential of improving social robot navigation in hospital hallways by leveraging more effective non-verbal communication.

[1]Parts of this background section are adapted from Shape_changing_research_proposal_Tom_Ossendrijver_24_10_2024, with adjustments for context.

Non-verbal communication modalities, such as visual light indicators, auditory signals, screen displays, and gestures, have already been employed to enhance robot legibility. However they have their strengths and weaknesses for conveying navigational intents (such as, yielding, turning, taking priority).

As the light indicators have demonstrated significant improvements in legibility and comfort in shared environments [25], it can be more challenging to communicate yielding, taking priority or the degree of turning, due to its limited information richness. Similarly, including color based feedback, though effective as a mechanism [33], fails to accommodate color blind users.

Further, Höfker studied the effectiveness of auditory signals to convey robot intentions [51]. While sounds can enhance user confidence, they are not always straightforwards and subtleties can be hard for people to interpret [50]. For navigational intents, noisy hospital environments can make it even more challenging.

Also, screen displays have demonstrated promise in improving communication efficiency [24]. But they can be difficult to read in crowded or fast-paced settings, requiring a line of sight, further as evidenced by Hebesberger, Koertner, Gisinger and Pripfl [31], screen usage and interface design are critical factors to make it work, as factors, such as lack of information or frequent changes can make it challenging.

Finally, Robots that incorporate expressive gestures, such as subtle body movements or arm gestures, have also shown promise. Hwang, Ahn, Macdonald and Ahn demonstrated that adding nuanced gestures to robot dialogue improved communication efficiency [42]. However, Salem, Eyssel, Rohlfing, Kopp and Joublin, emphasized that such gestures often require human-like features, which may not be feasible for all robot designs [22]. Moreover, expressive gestures can lead users to overestimate a robot's abilities and misinterpret its true capabilities or intentions [55].

Modality	Strengths	Weaknesses
Visual light indicators	Grabs attention, non-intrusive, useful for conveying basic statuses (red = stop, green = go).	Visibility issues (harder to notice in bright environments), limited information transfer (binary), requires line of sight.
Auditory signals	Grabs attention, no need for line of sight.	Noise pollution, hard to hear in noisy environments, beeps and boops are not always straightforward and more binary.
Screen displays	Detailed information, easy to convey different information, non intrusive.	Requires line of sight (visibility issues from certain angles or distances, especially in motion and crowds), requires user focus, physical size can make the robot bulky.
Gesturing	Natural and intuitive human like interaction, attention grabbing, rich information transfer (subtle nuances).	Complex to design, need for sophisticated hardware, requires line of sight, requires a human like design, lead to overestimating robot abilities.
Shape-changing	Expectations: Natural and intuitive animal and human like interaction, attention grabbing, rich information transfer (subtle nuances), design flexibility.	Expectations: requires line of sight, need for sophisticated hardware, complex to design.

Table 1: Modality overview for navigational cues in hospital environments

Together, these modalities and their limitations for communicating navigational cues in hospital like environments (summarized in Table 1 [2], based on our interpretation of expected strengths and weaknesses) indicate the need for innovative communication modalities that integrate legibility with intuitiveness, information richness, and clear visibility to meet the diverse demands of hospital environments.

2.3 Shape-changing as a novel non-verbal communication modality

To address these limitations and enabling seamless navigation in crowds like humans by means of implicit and non-verbal cues [25], we propose shape-changing as a novel communication modality for robotic navigation in hospitals. Inspired by the Sprout project demonstrating that soft actuators can effectively convey a robot's internal states and intent [54], we adapt this approach specifically for navigational communication. We define shape-changing as: "*the ability of an entity to intentionally manipulate its torso to undergo deformations to convey messages, intentions, or information*".

[2] Adapted and adjusted for clarity from Shape_changing_research_proposal_Tom_Ossendrijver_24_10_2024.

The word intentionally is added to separate purposeful communication from natural movements, such as minor passive chest expansion during breathing. Deformation refers to reversible, physical shape changes (such as, leaning and expanding) that keep the robots recognizability, unlike modular or reconfigurable transformations, which can alter the robots entire identity by rearranging parts [37, 6].

We focus on the torso, as the robot is specifically designed to deform this area, enabling expressive changes in orientation, form, and volume, three of the shape-changing dimensions described by Rasmussen [18]. This allows for a more focused exploration of torso deformation for legibly communicating intent in hospital environments.

Shape-changing movements will be based on pre-existing mental models by incorporating familiar and intuitive cues, such as human body language and animal-like behaviors, this to allow robots to communicate intent more naturally. Including the use of universally recognized shapes resembling objects, animals, or human forms, while taking into account expressive movement parameters described by Laban and Bogart [34, 56]. This naturalistic movement has already shown to increase robot predictability [23], where predictable motion is defined as movement that aligns with the observers expectations, thereby improving legibility.

The potential of shape-changing to enhance robot communication has been demonstrated in various research projects. For instance, the Sprout project explored how fiber-embedded actuators could enable robots to change their shape for emotional expressiveness, showing that deformation patterns could convey particular emotional cues [53]. Similarly, research on soft skin texture modulation has shown that robots can mimic natural behaviors, such as the appearance of a cat's fur, the display of spines by a blowfish or the manifestation of goosebumps in humans, to create relatable and intuitive interactions [35].

Moreover, shape-changing allows for the integration of animation principles, such as squash and stretch, anticipation, and exaggeration, which can further enhance the robot's ability to convey emotions and intentions [4, 38]. For example, exaggerated expansions can communicate urgency, while subtle contractions can express calmness or neutrality. The principle of anticipation, where small preparatory movements precede a larger action, can improve the predictability of robot behavior, potentially making it easier for humans to anticipate and respond appropriately.

2.4 Summary

In summary, the context of this research is the growing need for effective human-robot communication in hospital environments, where traditional modalities such as visual light, auditory signals, screen displays, and gestures may not be sufficient due to their limitations in visibility, design constraints, and information transfer. Shape-changing offers an alternative by enabling robots to dynamically adjust their torso to convey intuitive and legible cues. By drawing inspiration from natural human and animal behaviors and integrating animation principles, while taking into account the expressive movement parameters, shape-changing has the potential to enhance robot expressiveness and appeal, potentially improving legibility and acceptance in social hallways. Our research aims to explore the potential of shape-changing as a novel communication modality, addressing the current gap in the literature (mentioned in section 3) and providing a foundation for future developments in human-robot interaction.

3 Problem statement

Despite advancements in robotics, the integration of shape-changing behaviors to improve the legibility of robotic navigational intent remains an underexplored area of research. While shape-changing has been shown to support expressivity and convey internal and emotional states [53, 35], there is still a need for a framework that applies shape-changing technology to enhance robot legibility.

This gap is particularly relevant in diverse and crowded social environments, where robots must not only perform functional tasks but also clearly communicate their navigational intent to nearby humans. Misunderstandings in such settings can reduce both the effectiveness and acceptance of social robots [43].

This research aims to address this gap by investigating the potential of shape-changing to enhance or hinder the communication of robotic navigational intent in hospital hallways. By focusing on shape-changing, our aim is to provide new insights and perspectives that can inform the design of socially compatible robots, potentially improving legibility, fostering trust, and facilitating more effective human-robot interactions in social environments [41, 31]. To explore the potential of shape-changing, this study addresses the following research question:

"How can we use shape-changing to enhance robot legibility for navigation in social hallways?". To answer this question, we collect data as explained in subsubsection 5.4.1. To guide and structure the study, the following sub-research questions and associated hypotheses were formulated:

- **How legible is shape-changing in comparison to gaze as directional modality?**
  - *H0: Shape-changing does not enhance the legibility of the robot's intent. No measurable differences will be observed in participant responses (legibility score, confidence, or surprise) between conditions with and without shape-changing.*
  - *H1: Shape-changing enhances the legibility of the robot's intent, resulting in measurable improvements in participant responses, including higher legibility scores, greater confidence, and reduced surprise, compared to conditions without shape-changing.*

  This hypotheses is added, to establish a baseline comparison between shape-changing and gaze. The robot used in this study is integrated with a face (see Figure 1). As gaze naturally serves as a form of communication whenever an open channel is present [15], it is important to determine if the addition of shape-changing meaningfully improves the legibility of the robot compared to what gaze alone provides. This comparison helps to ensure that any observed effects on participant perception can be attributed to the shape-changing behavior itself rather than being influenced by the presence of gaze cues.

- **How legible is shape-changing in comparison to blinkers as directional modality?**
  - *H0: Shape-changing is not more legible than blinkers as a directional modality for robot navigation. There are no measurable differences in participant understanding of the robot's intended direction, legibility scores, confidence levels, or surprise when interpreting the robot's movements.*
  - *H1: Shape-changing will be more legible than blinkers as a directional modality for robot navigation, resulting in improved participant understanding of the robot's intended direction, higher legibility scores, greater confidence, and reduced surprise when interpreting the robot's movements.*

  This hypotheses is added to enable a comparison between a well known directional cue, blinkers, and the novel modality of shape-changing. Blinkers (or indicator lights) are used in everyday life to communicate directional intent, particularly in vehicles, thus likely familiar and easily interpreted by participants. By comparing shape-changing to this familiar modality, this study aims to evaluate how shape-changing compared against blinkers in term of legibility.

- **To what extend can shape-changing contribute to improving the legibility of other modalities, such as blinkers?**
  - *H0: Shape-changing does not enhance the legibility of blinkers when used in combination. There are no measurable differences in participant understanding of the robot's intent, legibility scores, confidence levels, or surprise compared to using blinkers alone.*
  - *H1: Shape-changing will enhance the legibility of blinkers when used in combination, leading to improved participant understanding of the robot's intent, higher legibility scores, greater confidence in interpreting the robot's movements, and reduced surprise compared to using blinkers alone.*

  This hypothesis is added to find out whether shape-changing can enhance the legibility of an existing directional cue, blinkers, when used in combination. Previous research has shown that multi-modal interfaces can improve the effectiveness and acceptance of robotic services in patient care [32]. This study aims to determine whether shape-changing contributes positively to a multi-modal setup by exploring its potential to improve participants understanding of the robot's intent.

- **How does shape-changing influence participants perceptions of the robot's likability, intelligence and safety during navigation?**
  - *H0: Shape-changing will not lead to higher ratings of warmth (likability), competence (perceived intelligence), and safety (reduced discomfort) during robot navigation compared to a baseline without shape-changing, nor compared to blinkers without shape-changing.*
  - *H1: Shape-changing will lead to higher ratings of warmth (likability), competence (perceived intelligence), and safety (reduced discomfort) during robot navigation, compared to a baseline without shape-changing and compared to blinkers without shape-changing.*

  This hypothesis is added to find out how shape-changing influences participants perceptions of the robot's legibility, specifically in terms of its likability (warmth), perceived intelligence (competence) and safety (reduced discomfort). These insights are required for making robotic technologies more accepted [48]. By evaluating these attributes, the study aims to provide a more detailed understanding of how shape-changing can make robots appear more socially compatible and trustworthy when navigating.

Based on existing research and anticipated findings, shape-changing is expected to enhance robot legibility for navigation in social hallways by providing an intuitive and information-rich communication modality. Shape-changing allows robots to communicate their intent in ways that are natural and easily interpretable, drawing inspiration from human and animal behaviors, enriched with animation principles to improve expressiveness and appeal. Shape-changing is expected to surpass or complement traditional modalities like blinkers by conveying richer information, enhancing legibility, and improving user perceptions of the robot's likability, intelligence, and safety.

This research aims to demonstrate that shape-changing can effectively communicate directional intent through legible interactions. While the primary focus is on navigation in hospital hallways, we believe that the application of shape-changing can potentially extend beyond this context. As shape-changing offers broader design possibilities for robots to communicate intent in diverse social environments and for a variety of tasks.

The next section outlines the design of the study and environment used for testing these hypotheses.

4 Design and implementation of shape-changing in a hospital robot: from prototype to VR testing

To explore how shape-changing can influence robot legibility for navigation in social hallways, we developed and implemented shape-changing behaviors in a hospital navigation context. This chapter outlines the final design choices and implementation steps that allowed us to experiment with the shape-changing modality in a controlled virtual reality (VR) environment.

The design process was iterative and informed by theoretical insights, real-world robot constraints, and participant feedback from pilot testing. While an extensive iterative process preceded the final study design, this chapter focuses on the core decisions that shaped the implementation of the experimental setup. The iterative design process and design reasoning can be found in Appendix H, with an overview of guiding principles used for decision making in Appendix I.

Our aim was to translate the abstract concept of shape-changing into practical, expressive robot movements that could legibly convey directional intent, allowing us to compare it with vehicle indicators. The robot's behaviors were based on literature about non-verbal communication of human and animal movement, animation principles and the movement parameters. In this chapter we provide the following: First, the theoretical foundations for the shape-changing behaviors are presented. Next, we explain the implementation of these behaviors in the physical robot prototype and translate it to the VR environment. Finally, we discuss the key design refinements made based on pilot testing.

4.1 Physical robot movement based on literature

To design shape-changing behaviors that are able to legibly convey navigational intent, we drew inspiration from a broad set of theoretical frameworks, including human and animal communication, animation principles and expressive movement parameters from Laban and Bogart. In this section we will outline how these frameworks informed the development of the robots shape-changing behavior. [3]

4.1.1 Human and animal shape-changing behavior

Shape-changing in nature can serve as a form of non-verbal communication. Where both humans and animals use changes in posture and form to convey internal states. By making use of known animal behavior or human forms, we try to reach common ground for effective communication

In humans posture can be used to communicate social cues. For example, expanding the chest and standing tall can signal confidence, while slouching, may convey doubt or submission [9].

Similarly, animals make use of shape-changing behavior to communicate intent or emotions. Animals like peacocks and frigatebirds use expansion to draw attention and communicate mate readiness [5, 8]. While for intimidation and enhancing appearance, cats puff up their fur, and frill-neck lizards extend their frills to appear larger deterring threats [49, 3]. Further, in social species like wolves, posture is used to expressing social hierarchy: dominant individuals stand tall and broad, while subordinates lower their bodies and shrink in size to signal submission [2, 26].

Humans and animals seem to have some overlapping patterns in this regards, where both make use of expansion to communicate dominance or urgency, and contraction for yielding or submission, which principle will be applied to the robots communication behavior.

4.1.2 Animation principles for expressiveness

Further to make the robots behavior believable while helping to improve the robot expressing its intent, animation principles are combined with the shape-changing behavior [38]. The animation principles, as outlined by Thomas [4], can help make the robots intent more clear, by improving:

- *Expressive movements:* the incorporation of principles such as squash and stretch (organic deformation) and exaggeration (increasing intensity) can facilitate the expression intensity of emotions in robotic movements, while making it more organic.
- *Predictable interactions*: the principle of anticipation (preparatory motions) help enhance the predictability and make the robot movement more understandable, making it easier for humans to anticipate and respond to robot behaviors.

[3] This section extends the shape-changing movement framework first outlined in our 2024 research proposal *Shape_changing_research_proposal_Tom_Ossendrijver_24_10_2024*.

- *Aesthetic appeal:* the aesthetic appeal of robot movements is enhanced by the incorporation of staging (action readability), follow-through (residual motion) and arcs (curved trajectories), which help to enhance interactions to be more enjoyable and engaging.

4.1.3 Shape associations

Shapes inherently carry associations that influence perception. When designing the shape-changing motion, we considered the shape associations described by [57]:

- *Rounded shapes*, are perceived as approachable and friendly, making them suitable for calm states of the robot, giving a warm and welcoming feeling.
- *Square or rectangular shapes*, can convey stability, strength, or authority, making them suitable to communicate the robot will stand it's ground.
- *Triangular shapes*, are associated with danger and directionality. Making them suitable to communicate urgency or directional intent.

4.1.4 Movement parameters

For further refinement of the robotic shape-changing behavior, we considered expressive movement dimensions as described by Laban [34], which concentrated on the categorization of movement into four classes:

- *Weight*: Heavier, more forceful movements signal power and urgency, while lighter movements suggest calmness or retreat.
- *Time*: Abrupt or sharp movement can feel urgent, while sustained, smooth motions communicate ease or patience.
- *Space*: Direct, goal oriented and efficient pathways convey decisiveness or focus, whereas indirect, curved paths communicate exploration or hesitation.
- *Flow*: Bound flow is controlled and communicates emotional restrained, while free flow expresses spontaneity or fluidity, communicating emotional release.

As well as combining it with insights from shape-changing studies from HRI, such as mentionings of the paper [29]. which stated:

- *Direction*: Upward movements are associated with positive emotions, high arousal, and high dominance, whereas downward movements are associated with negative emotions, low arousal, and low dominance.
- *Orientation*: Flat orientations are associated with positive valence, whereas leaning orientations (either towards or away) are linked to negative valence.
- *Velocity*: Faster movements are strongly correlated with higher ratings of valence, arousal, and dominance, with arousal being particularly influenced by movement speed.
- *Fluidity*: Although fluidity alone does not have a significant main effect on emotional dimensions, interactions with other variables (such as velocity and direction) can suggest more complex dynamics.

4.1.5 Navigational intents through shape-changing

Based on the frameworks described above, we developed four core expressive shape-changing behaviors to communicate the robot's internal state and navigational intent:

- *Expanding to indicate a change in state or assert dominance*

  Expanding will serve as a clear signal of authority or a shift in the robots internal state. This expansion will involve an abrupt, fast and heavier movement, towards a square shape, conveying power and dominance. Expanding upward and outward will not only make the robots actions noticeable but also increase its dominance, indicating that the robot has decided to take control of the space or situation. This behavior draws attention to the robot before a change of state, ensuring that others are aware of its intention to act.

- *Contracting to yield or show submission*

  Contracting will signal the robots intention to reduce its activity and yield to others. This behavior will involve slow, lighter, and smoother movements, indicating submission. The robot will contract downward and inward, while doing so, it will take on a more round shape, symbolizing softness, approachability, and reduced dominance. This creates a non-threatening presence, encouraging humans to pass or take priority. The downward contraction signifies yielding or respect, especially in crowded environments, allowing the robot to blend into the background and minimize its impact on those around it.

- *Breathing to indicate activity or emotional states*

  Breathing-like movements could be used to indicate the robots emotional or activity states. Slower, smoother expansions and contractions will convey calmness, relaxation, or idle states. While faster, sharper breathing will indicate urgency, arousal, or heightened activity, making the robot appear more alert, in a hurry or engaged.

- *Elongation for directional intent*

  This behavior involves a slight orientation change and slow forward stretch accompanied by a slight upward tilt, forming a triangular posture that indicates the intended direction of movement. This upward component enhances the perceived dominance of the motion, indication the robot's decision to proceed in the specified direction. Following this initial directional cue, the system enters a preparatory phase where the robot maintains its orientation while slowly contracting backward (squash phase). Followed with a rapid forwards stretch (stretch phase) towards the previously indicated direction of intent, immediately followed by execution of the turn.

In our study, we made use of expanding to indicate a change in state or assert dominance and elongation for directional intent. These behaviors were implemented in the shape-changing robot available at the University of Twente, despite not being able to take triangular shapes it is capable of growing, shrinking, leaning, and inflating. By integrating these movements, the robot can potentially communicate its intentions through legible shape-changing behaviors, improving its interaction with humans in a hospital like environment.

4.2 Real robot design and challenges

To translate the designed shape-changing behaviors into a physical robot capable of deforming its torso, prototyping was conducted on the Harmony project robot of the University of Twente, see the shape-changing robot in Figure 1. This marked the starting point for a more extensive development process.

The development of a shape-changing robot started with the design of an initial prototype created for the HARMONY project, an European initiative focusing on assistive robots for healthcare. This robot was shipped from Zurich to the University of Twente, with the objective to further develop and implement shape-changing behaviors.

To enable this functionality, a team at the University of Twente worked on programming the robot's integrated T-CAN485 development board and sought an intuitive tool for controlling its movement. For this purpose, Bottango, a visual robot control environment, was selected. By connecting the device running Bottango directly to the robot through a cable, the team was able to experiment with motion patterns and observe real-time feedback on the robot's physical behavior. This functionality allowed for quick prototyping and iterative improvements of the movement patterns.

In addition to real-time control, Bottango offered the ability to create and bake animations into a command stream which could be stored on the robot's microcontroller. This enabled the robot to execute pre-defined movement sequences.

Using this setup, a variety of motion patterns were tested based on the mentioned shape-changing behavior of subsection 4.1. These tests helped to validate if the designed shape-changing behaviors were physically feasible on the real robot and if they appeared appropriate in real life to convey navigational intent.

Overall, this prototyping phase focused on translating theoretical shape-changing movement concepts into real world robot actuated movements, bridging the gap between design theory and practical robotic behavior. The process demonstrated that shape-changing behaviors could be implemented in a real-world robot.

4.3 Design decisions

To evaluate the legibility of the shape-changing robot in a hospital hallway context, several key design decisions were made regarding the structure of the study and the participant interaction methodology. These decisions were informed by the guiding principles outlined in Appendix I and the technical limitations of the physical robot described in Appendix J. Each choice was made to balance feasibility, ecological validity, and experimental control, while ensuring the study could capture the necessary measurements to address the sub-research questions. The key design decisions focused on:

- The choice of the type of media format: video VS VR.
- The choice of first person: front view VS side view.
- The choice of movement: static VS freely.
- The choice of moment of pause: before intent VS participant decide to pause.

4.3.1 Media format: video VS VR

The initial study plan involved using the real-world robot and presenting its behavior through pre-recorded video in an online format. This approach was evaluated for its accessibility and potential to reach a large participant pool through survey platforms such as OpenSesame, a free tool commonly used for psychology experiments.

However, several concerns emerged with the video-based format. Online studies carry inherent risks, including participant distraction, uncontrolled testing environments, and limited engagement, which can compromise data quality. Additionally, offering sufficient incentives for online participation was beyond the available budget.

Simultaneously, a virtual reality (VR) environment simulating a hospital hallway was in development at the University of Twente. Considering the limitations of the physical robot (as explained in Appendix J) and the need for a controlled yet ecologically valid setting, the VR environment was selected as the appropriate medium for this study.

Also considering VR's ability to create a more realistic and immersive environment closely aligning with the real-world. providing better control over external distractions, leading to more focused and consistent participant interactions and valid comparisons.

Although developing and updating the VR environment would require more time and poses potential risks of discomfort, we could make use of the in development VR hospital hallway environment of the university of Twente and expected a minimal chance of VR sickness given the short duration of the study. Therefore, we determined that the benefits of VR outweighed the challenges it would bring along, making it the preferred choice for this research.

4.3.2 First-person front view VS side view

In designing the participant's viewpoint, two options were considered: a first-person front view, simulating a head-on encounter with the robot, and a first-person side view, representing a robot crossing the participants path sideways.

The front view was selected for its expected increase in engagement, as participants would have to respond to the robot coming their way, this seemed in line with a typical hallway interaction, where individuals encounter a service robot approaching head-on. This perspective allows the participant to directly observe the robot's shape-changing behavior and respond as they would in a natural, face-to-face setting.

In contrast, the side view provided a less direct experience of interaction. Because the robot would not actively engage with the participant from this angle, there was a greater risk of delayed or passive responses. Therefore, the front-facing perspective was chosen to support consistent, reliable, and meaningful participant responses.

4.3.3 Free movement VS static

Right now, in the VR environment, the robot was intended to walk towards participants from a first-person view. However, the question remained if participants should be allowed to move freely or remain in a fixed position.

Free movement was initially considered, as it was expected to encourage more naturalistic and engaged responses. To measure this, ultrasonic sensors were explored to track participant positioning and movement in response to the robots behavior. However, several limitations arose. Distinguishing between intentional movement and subtle hesitation or adjustment proved challenging, complicating the interpretation of reaction times. Additionally, real-time calibration issues between the ultrasonic sensors and the Unity VR environment introduced timing inaccuracies, making the collected data less reliable.

Prior studies, for example [28], made use of a static pause button to capture participant responses, providing a more easy and validated alternative method. Based on this a static seated setup was selected. This allowed for consistent, simplified data collection and analysis, while at the same time supported accessibility for a wider participant pool, including individuals with reduced mobility.

4.3.4 Pause button timing: Participant-controlled VS automated pauses

Another important design consideration involved the timing of the pause used to capture participant interpretation. Two options were considered: (1) pausing the video automatically just before the robot's intent became clear, or (2) allowing participants to manually pause the video when they believed they understood the robot's intention.

Both methods were considered valid ways to measure legibility [28]. For the automated pause method, the cognitive load is reduced by clearly defining the moment of prediction, in trade off for richness of the collected data. As it only captures response accuracy (correct/incorrect) and does not provide insights into the response time.

To allow for the collection of both response time and accuracy, the participant controlled pause method was chosen. Although this may increase the cognitive load, the minimal nature of the interaction (pressing a single button), was not expected to affect the participants performance or comfort, especially with the short duration of the study.

A summary of the key study design decisions and their trade offs can be found in Appendix A.

4.4 Unity implementation

Following the finalized design decisions, the study was implemented within a virtual reality (VR) hospital environment developed at the University of Twente. This environment was created to evaluate the shape-changing robot behaviors in a controlled, immersive, and ecologically valid context. The VR environment was developed in Unity, a game engine selected for its flexibility, ease of integration with VR platforms, and support for future modifications. To ensure broad device compatibility and future scalability, the application was configured using OpenXR, an open standard for VR/AR development. For this study, the environment was run on a Meta Quest 3 headset.

Although the base environment was pre-existing, some modifications were required to align it with the experimental requirements. These included:

- Adjusting the layout to have unobstructed visibility of the robot's behaviors.
- Adding and removing environmental objects to maintain focus on the robot and interpretation of its shape-changing movements.
- Implementing functionality to support data collection and participant interaction, such as integrating shape-changing behavior, capturing pause inputs, and logging response times.

These adjustments were made for the final design setup to provide a consistent, repeatable, and contextually appropriate testing environment for measuring the legibility of robot behavior.

4.4.1 Environmental setup and features

Initial setup

The base environment used in this study was a realistic VR simulation of a hospital hallway, developed at the University of Twente. Initially, this setup featured third-person camera perspectives, with a robot interacting with virtual humanoid actors. Ambient hospital sounds, along with environmental details such as plants, benches, and posters, were included to enhance realism.

To meet the requirements of the current study, several adjustments were made. These included implementing a first-person front-facing perspective, integrating a functional pause button, and modifying the participant setup to be static rather than free-moving, in accordance with the study's design decisions Appendix A.

Environment adjustments

To create the environment meeting the design decisions. The VR camera was positioned at the center of the hallway to provide a clear and unobstructed view of the robot's movements. Surrounding objects, such as plants and furniture, were removed to remove potential visual bias or obstructions that could influence participants perception of the robots intentions.

Additionally, the batch play mode and automatic video recording functions were disabled. Allowing for more control over the playback functionality, which made it possible to select and play our own playable assets, providing full control over the robot behavior during testing.

VR integration

For VR deployment, the OpenXR plugin was used instead of the Oculus integration plugin, which produced compatibility errors. Early testing on Linux systems using SteamVR led to performance issues, including crashes and latency. To have a stable experience, the environment was run on a Windows based system using a wired connection via Oculus Link. This setup minimized lag and lead to smooth performance with the Meta Quest 3 headset.

Robot movement path

The robot's movement was designed to follow a symmetrical path in both left and right turn conditions, as is visible in Figure 3. It approached the participant from the back of the hallway and executed its turn only at the last moment. This layout maximized the time participants had to observe and interpret the robots behavior.

Using Unity's Timeline system, the robot's movement was synchronized to start at frame 0, proceed straight to frame 465, and then turn either left or right. This ensured identical timing and angles across conditions.

Pause button

A pause button was integrated into the system to capture the participants response time. A new input action was added to the XRI Input Action Asset, mapping the pause function to the trigger button on the right-hand controller. The pause logic toggled the timeline into a paused state upon input detection, with the exact frame of response logged for later analysis.

Participants were instructed: "Observe the robot and press 'pause' when you believe you understand what it will do next." This method allowed for the collection of the reaction time.

Feedback canvases

For data collection as mentioned in subsubsection 5.4.1, participant feedback is collected. For feedback collection, three canvases were implemented. These were activated via keyboard commands during the post trial phase to prevent headset removal during testing.

- Pressing 1 displayed the question "What do you think is the robot's intent?".
- Pressing 2 displayed, "How confident are you in your choice?", which is a 5 point likert scale, from not at all confident to very confident.
- Pressing 3 displayed, "How surprising was the outcome?", which is a 5 point likert scale, from not at all surprising to very surprising as shown in Figure 2.

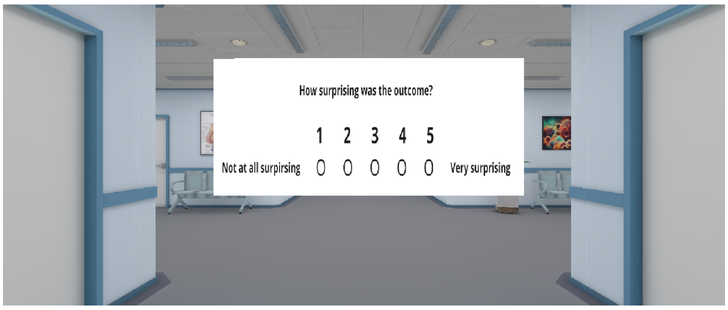


Figure 2: Surprise feedback canvas

The canvases were implemented tasking the participant not only verbally from outside the environment, but also from within the environment. These canvases were custom made images by means of the application Miro, this to improve the visual quality, ensuring the questions and Likert scales were visible and consistent in the environment as intended.

Blinkers

The initial robot design did not integrate blinkers, so to mimic vehicle turn signals, orange light objects were added to the robot as visual indicators of directional intent. These were implemented as spotlights with an intensity of 10, ensuring they were easily visible. The blinkers were positioned at the robot's base, as robots and cars often implement these blinkers at the base of their system. The blinkers could be activated and deactivated in turns, as needed for the experiment.

Timeline control

Unity's Timeline was used to manage the synchronization of robot animations, movements, gaze behavior, and environmental cues. The timeline allowed us to control parameters with key frames for parameters such as inflating, leaning, nodding, shrinking, and twisting, while allowing for the synchronization with the robot's movement. It made it easier for the shape-changing movement to be combined with the blinkers and gaze behavior as intended. As separate tracks allowed us to make adjustments to a part of the robotic behavior, such as the blinkers or the robotic movement without influencing the other parts of the sequence. While at the same time making it easier to switch between the scenario's on the spot. As due to the timeline with multiple tracks, it would only require one mouse button press to turn off the blinkers or shape-changing modality track from the robot's behavior timeline.

4.4.2 Implementation of physical robot movement into VR

The shape-changing behaviors developed for the physical robot were now implemented in the VR environment using Unity's timeline system, which allowed for exact control over the robots movement, gaze, blinkers, and shape-changing behaviors by means of separate animation tracks. Where all animations were structured at 60 frames per second, allowing consistent timing and synchronization between modalities across experimental conditions.

Robot movement

The robot's movement track was implemented together with the predetermined paths, as shown in Figure 3. It approached the participant head-on and executed a turn either left or right at a fixed point (frame 465). Between frames 250 and 450, the robot passed through a crossroad area, with its body positioned at the center by frame 350. This design provided participants with a clearly defined time window to interpret the robot's communicative cues.

The purpose of crossing the crossroad is to evaluate whether participants perceive the gaze, blinkers, and shape-changing behaviors differently. This crossroad allows us to explore whether shape-changing has the potential to communicate directional intent in a more analog way, thereby making the communication richer compared to a more binary approach like blinkers. Additionally, the crossroad simulates a realistic situation in which conveying directional intent, or simply indicating the robot's intention to pass, can be challenging.

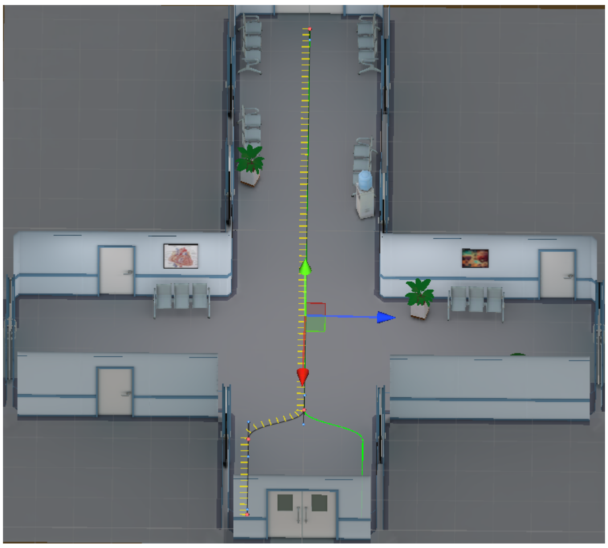


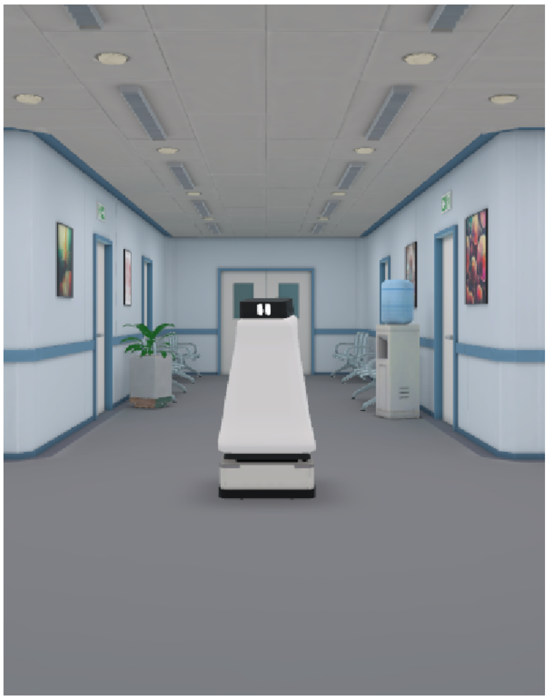
Figure 3: Hospital hallway with robot path

Gaze

To ensure consistency across all conditions, the robot's gaze behavior was standardized. From frame 0 to 117, the robot faced forward, performing a subtle nod to acknowledge the participant. Starting at frame 200, the robot turns its head in the intended direction of movement, followed by another nod to indicate its intention to move in that direction.

This directional gaze was then held until frame 465, when the robot started to turn. After the turn, the gaze returned to a neutral forward facing position. The gaze behavior remained consistent across all scenarios.

(a) Subtle nod



(b) Directional cue

(c) Conformational nod

Figure 4: Sequence of gaze behaviour for directional intent

Blinking

The blinkers were represented by animated orange spotlights attached to the base of the robot, simulating vehicle turn indicators. These were controlled through Unity's timeline to synchronize with the robot's gaze and movement behaviors.

The blinking sequence included four phases:

- *Anticipatory phase* (frames 55–180): Both blinkers rapidly flash three times to draw the participant's attention. Signaling to the participant that something is about to happen, while drawing their attention to the blinkers

- *Directional cue* (frames 200–260): The blinker on the side of the intended direction remained active for an extended period, mimicking a vehicle's turn signal, lasting one second (60 frames). Providing a clear and known directional cue.

- *Enhanced expressiveness* (frames 260–360): The blinking speed increased during the robot's dynamic squash and stretch movement, aligning with the shape-changing behavior, while enhancing the directional signal and visual visibility. The light blinked four times in this interval, at a rate of 10 frames on and 10 frames off.

- *Final phase* (frames 360–465): The blinking returns to a steady one-second interval, consistent with typical vehicle turning signals. This period concluded with the final frame which is the moment where the robot's intent should already be communicated to the participant.

(a) Anticipatory blinking

(b) Directional cue

Figure 5: Sequence of blinking behaviour for directional intent

Shape-changing movement

The shape-changing behavior was developed to communicate directional intent through a sequence of expressive torso movements. This included inflation, leaning (forward and to the side), nodding, shrinking, twisting, and elongation, each parameter is animated though the key frames, together creating a sequence of frames in Unity's timeline.

The behavior went as follows:

- *Attention phase* (starting at frame 55): The robot starts inflating and growing to draw attention as anticipatory cue.

- *Preparation phase* (frames 117–200): The robot leans slightly forwards, enhancing the nodding motion of the gaze track. From frame 160 to 200, the robot slowly settles with some bouncy movement into a more forwards-leaning posture, while focusing the gaze on the participant, to communicate the robot sees the participants and is preparing to communicate its intent.

- *Directional cue* (frames 200–360): At frame 200, the robot initiated its directional behavior. It twisted and leaned toward the intended direction, then transitioned into a squash-and-stretch sequence:

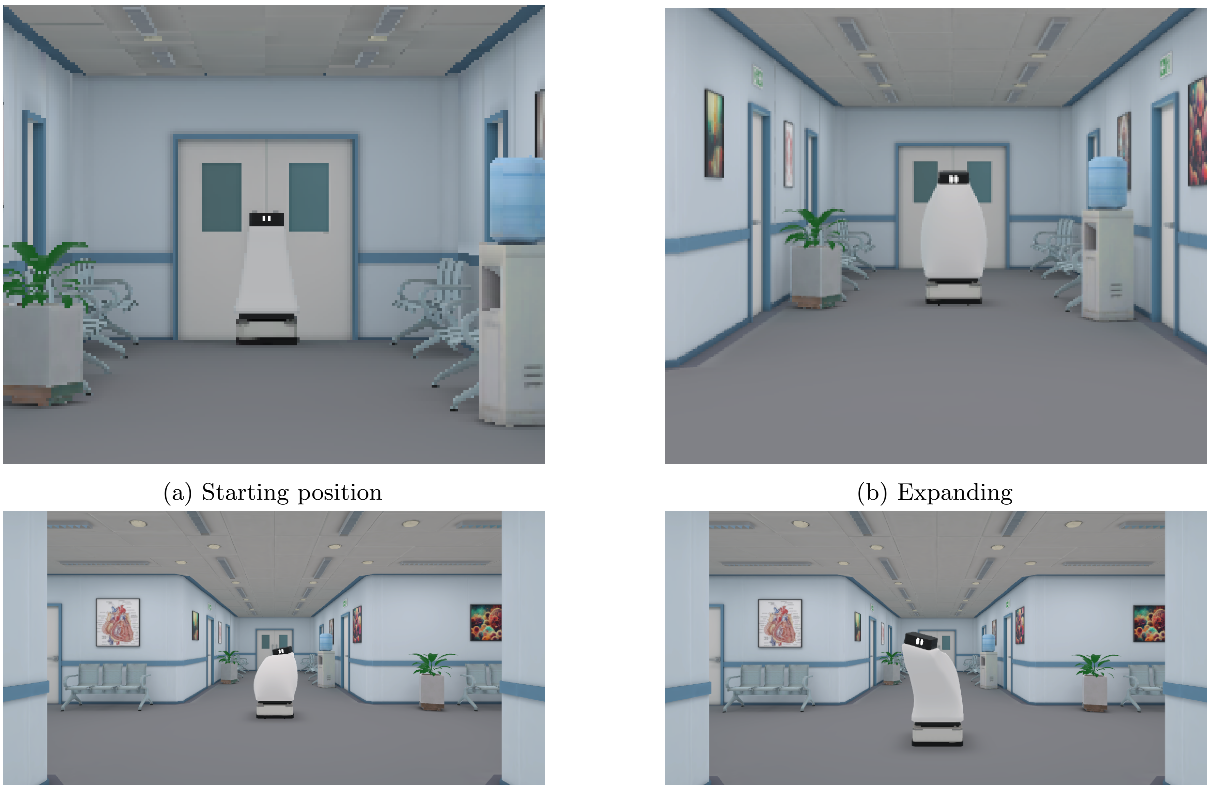
  – From frame 260 to 300, it contracted backward, shrinking and deflating preparing for an exaggerated stretch forwards.

  – From frame 300 to 340, it squashed to center, being inflated, appearing more compact.

  – Then, from frame 340 to 360, it rapidly stretched its entire body forward, extending fully and leaning in the intended direction while deflating, creating a distinct "shooting" gesture with its entire body.

- *Hold phase* (frames 360–465): The robot maintained this extended pose with slight natural oscillation (bouncing motion), to make the fast movement seem more organic and lifelike instead of static.

The sequence was mirrored for left and right turns, ensuring a symmetrical implementation and fair comparison between conditions. Although the movement might appear too exaggerated, this is in line with the animation principles, which suggest exaggeration to increase expressiveness. This design is used as an exploratory prototype to determine if shape-changes can be used to communicate directional intent and to what extent. While probably not fully optimized, as there are other ways the robot could use shape-changing to communicate intent, this prototype is based on existing literature as outlined in subsection 4.1 and aims to determine the potential of shape-changing independently and in combination with other modalities.



(a) Starting position (b) Expanding

(c) Preparing (d) Directional cue

Figure 6: Sequence of shape-changing behaviour for directional intent

4.5 Iterative design for participants

Before conducting the main study, pilot tests were conducted to identify potential challenges, improve the experimental setup, and enhance the robot's movement for improved communication of intent. This process is intertwined with the design decisions as these tests provided insights into participant behavior, allowing us to collect feedback and make required adjustments on the design. Key improvements based on the iterative design study included slightly reducing exaggeration of the robot's movement design, introducing a warm-up phase, and adjusting the environment to determine if participants could make a distinction between left and right.

Warm-up phase and VR familiarization

During the pilot test it became clear the environment was quite overwhelming for some participants, especially for those unfamiliar with virtual reality. This occasionally resulted in delayed engagement with the task and inconsistent use of the pause button, potentially due to hesitation in an unfamiliar setting. Informing us that participants required some time to get used to the environment before they could focus on interaction with the robot.

During this warm-up phase, participants were placed in the VR environment and observed the robot moving in a looped, non-experimental pattern. They were encouraged to freely test the pause button, allowing them to understand how it worked and what to expect once triggered. This helped participants become comfortable with the VR headset, controls, and visual context before being asked to interpret the robot's behavior under time pressure.

Left and right orientation clarity

During early trials, it became evident that some participants experienced difficulty distinguishing between left and right turns, which could affect the accuracy of directional responses.

To address this, the VR camera was repositioned next to an eye test poster mounted on the participant's right-hand side. At the beginning of the experiment (during the warm-up phase), participants were asked to conduct the eye test on their right hand-side.

This poster served two purposes: first it helped to determine if a participant might struggle with the distinction between left and right, requiring us to validate at the end of a round if the participants directional response matched the intended choice. Secondly, it ensures participants had sufficient vision, with or without glasses, to clearly experience the environment. If the VR display would appear too blurry, adjustments could still be made to the VR headset to improve visibility.

Refinement of shape-changing behavior

Initial implementations of the shape-changing movement were perceived as overly exaggerated by some participants, particularly the backward and sidewards motion, which were sometimes interpreted as feinting or hesitation rather than clear directional intent. This unintended effect caused confusion rather than improving the legibility of the robot. To tackle this, the backwards and sidewards movements were slightly reduced, ensuring the robot's intent was easier to interpret.

Question adjustments

Finally, the pilot tests also highlighted the need to improve the clarity and specificity of post-trial interview questions. Some questions were too broad, requiring adjustments, to help participants provide meaningful responses. For example, the question: "Can you describe how you interpreted the robot's intent?", was changed to, "Can you explain the steps or thought process you used to understand what the robot was trying to do or communicate?". This adjustment helped participants to give a more structured and detailed answer. Why the interview is added is explained in the study consideration subsection 4.6, the part "Exploratory study" and further details on the specific interview questions can be found in step 6 of the procedure subsection 5.5.

4.6 study considerations

To ensure the robustness of the main study and the validity of the data collected, several important considerations were taken into account. These refinements aim to strengthen the study design, support valid participant responses, and preserve the exploratory character of the early testing phases. The following key aspects were addressed:

- Participant grouping:

  As scenarios could be switched with a few mouse clicks thanks to Unity's timeline, participants were grouped once recruited. This made it possible to quickly recruit participants in hallways and involve them with the tests. The participants were grouped based on:

  - Experience with robots: low, medium or high
  - Possession of driver license: yes or no.

  These criteria were selected based on the assumption that previous experience with robots and having driving experience could potentially influence the participants interpretation of robot behavior and their perception of blinkers, as it could potentially be interpreted faster due to their experience with road vehicles. To have a balanced distribution between the scenario's, an Excel sheet was used to manage participants assignment to a scenario. The sheet recorded each participants robot's experience level and driver license status, tracking how often these features appeared in each scenario. Participants were then assigned to the scenario with the lowest count, leading to a fair and even distribution of the participants based on these features.

- Participants learning effect:

  Based on initial testing, it was determined that conducting the experiment over three rounds provides the optimal balance between participant learning and reliable data collection. Only the third round will be used for analysis, as rounds 1 and 2 serve as learning and exploratory rounds, allowing participants to naturally interpret the robot's shape-changing behavior without explicit instructions on directional intent.

  - Rounds 1 and 2 as learning phase: In these early rounds, participants often found it unclear what the robot's movements indicated. However, these rounds helped with the familiarization with the robot's behavior, while offering exploratory insights into how participants interpret shape-changing movements beyond directional intent. Providing insights on other intents shape-changing could be used for to communicate.
  - Round 3 for analysis: By the third round, participants often have an understanding of the robot's behavior and the expected task responses. This round provides stable and potentially the most insightful data, as participants are still wondering what is going to happen, while being confident in their interpretation without the test becoming repetitive.
  - Exclusion of round 4: A fourth round was deliberately excluded to avoid the development of response strategies or pattern recognition, which could bias results. Extending the session further risked producing repetitive or overly confident responses based on familiarity rather than real-time interpretation.

  We believe that by focusing the analysis on round 3, we can collect freely interpretable data from previous rounds, while analyzing the natural response in which the study intent is clear and preventing the inclusion of redundant data.

- Participant exclusion

  To keep the study's exploratory nature, individuals with prior knowledge of the research's focus on directional intent were excluded. As this potentially removes the exploratory effect as the participant would know what to expect, this could indirectly also bias the participant to guess 50% correctly without truly knowing the robots intent.

- Exploratory study:

  For significant results, a large number of participants will likely be required, as explained in subsubsection 5.6.1. However, for this thesis, we expect to obtain a limited number of participants, likely too few for statistically significant findings. Because of this, the focus shifts towards a more exploratory approach, leading to the inclusion of a post-task interview to explore participants perceptions of the robot's gaining qualitative insights into whether participants perceive higher legibility. With this qualitative approach we hope to provide deeper insights into participant experiences and complement the quantitative data. So the focus shifts towards identifying trends and exploratory insights rather than achieving statistical significance.

- Drop rate:

  There is a potential chance participants drop out, so for the study we kept it simple to recruit participants as we anticipate a drop rate of 10–20% due to factors such as motion sickness, technical issues, or external disturbances during the study.

- Technical reliability:

  Technical issues can occur, due to hardware or software malfunctions. Such as loading time delays, which we considered as potential risks to the data accuracy. To prevent these risks, equipment is pre-tested before the experiments.

- Generalizability:

  Finally the findings from this study may be specific to VR and may not fully extend to real-world settings. This limitation is mitigated as much as possible by making use of a realistic VR environment that closely simulates a hospital hallway.

5 Main study

This study aimed to answer the research question: "How can we use shape-changing to enhance robot legibility for navigation in social hallways?". To achieve this, sub-research questions were addressed through a 2x2 within-subjects design experiment conducted in a virtual reality (VR) environment. In this experiment, participants experienced a first-person front view of the robot while being seated and were instructed to press the pause button when they believed they understood the robot's intent.

As explained in the design chapter Appendix A, the main study's design decisions were based on theoretical foundations, implementability, and iterative improvements based on participant feedback. Where the main study focused on evaluating legibility through a combination of, response time, accuracy, confidence and surprise ratings. Together, these measures provides insights on how legible and intuitively the robot communicated its navigational intent.

In this section details will be provided considering the main studies experimental scenarios and conditions, participants criteria, recruitment process, experimental setup, procedure, data collection, G*power test of participant amount and analytical approach used to interpret results.

5.1 Scenarios and conditions

To investigate the sub-research questions, a 2x2 within-subjects design was implemented, resulting in four experimental conditions. These scenarios were designed to isolate and evaluate the impact of shape-changing, blinkers, and their combination, while gaze was included consistently across all conditions due to the robot being integrated with a face.

The scenario's are as follows:

Blinkers Shape-changing	- Blinkers (-B)	+ Blinkers (+B)
- Shape-changing (-S)	-B-S (1)	+B-S (3)
+ Shape-changing (+S)	-B+S (2)	+B+S (4)

Figure 7: 2x2 study with gaze behavior as base

1. **Baseline: Gaze only (-B-S)**

   Serves as the control condition. Gaze cues are inherently present due to the robot's facial design, as communication occurs whenever an open channel exists [15]. By evaluating legibility for this setup, we try to obtain a reference point against which the effectiveness of other modalities can be compared.

2. **Shape-changing: Shape-changing with gaze (-B+S)**

   Shape-changing, which is inspired by the natural human and animal movement and integrated with animation principles, is tested here in combination with gaze to determine how it contributed to the robots legibility. This scenario helps to determine if shape-changing can improve the legibility and the ability of participants to interpret the robots intent.

3. **Blinkers: Blinkers with gaze (+B-S)**

   Blinkers are an often used modality in daily life (such as vehicle indicators), which we want to compare to shape-changing. Combining blinkers with gaze, we want to determine if this modality scores better considering its legibility in comparison to the more novel shape-changing modality.

4. **Combined modality: Shape-changing with blinkers and gaze (+B+S)**

   By combining shape-changing with blinkers and gaze we want to create a multi-modal approach. This scenario is used to determine if the integration of multiple modalities improves the legibility compared to each modality individually, potentially it provides insights on if shape-changing can add to other modalities.

These scenario are selected to isolate the modalities, making it possible for us to compare their effect on legibility, individually and in combination. The aim is to compare the legibility of shape-changing with that of blinkers and explore if it can potentially be used to enhance a multi-modal communication approach. By looking at these scenario's we are able to evaluate how these modalities impact the legibility of a robot communicating its directional intent.

The baseline includes gaze, mostly as the robot that is used for this study has a face, which naturally leads to the inclusion of gaze cues. As the robot has a face, we will consider gaze as a considered part of the shape-changing behavior. For example, when the robot rotates its body or leans towards a specific direction, the head will naturally follow, indirectly simulating gaze. So to determine the effectiveness of shape-changing and the blinkers in this project, gaze was included in all the scenario's. This mitigates the potential that findings from the comparison are due to the effect of gaze instead of due to the shape-changing behavior.

Further, each scenario contributes to the sub-research questions:

- **How legible is shape-changing in comparison to gaze as directional modality?** For this a comparison between the Baseline (Gaze only) and the Shape-changing (with gaze), will take place.

  By comparing the Baseline with the Shape-changing scenario we determine whether the introduction of shape-changing makes the robot's intended movement more legible compared to only relying on gaze. Thereby we hope to discover if shape-changing can improve overall legibility of a robot with a face.

- **How legible is shape-changing in comparison to blinkers as directional modality?** For this a comparison between Blinkers (with gaze) and Shape-changing (with gaze), will take place.

  By comparing Shape-changing and Blinkers we will determine if shape-changing offers a more legible way to convey directional intent compared to the commonly used blinkers. This comparison will help us understand how shape-changing compares to blinkers and whether it could be adopted for better robot legibility in everyday situations.

- **To what extend can shape-changing contribute to improving the legibility of other modalities, such as blinkers?** For this a comparison between Combined modality (gaze, blinkers and shape-changing) and Blinkers (with gaze), will take place.

  Comparing Blinkers and Combined modality can help to find out if shape-changing can improve legibility of blinkers. This can help us understand if shape-changing can be seen as an additional modality that enhances other directional cues, such as blinkers.

However to answer the research question, "*How does shape-changing influence participants perceptions of the robot's likability, intelligence and safety during navigation?*", We require more data to be obtained besides the response time, accuracy, confidence and surprise rate. To be able to answer this sub research question the Robotic Social Attributes Scale (ROSAS) questionnaire is added at the end of the test for each scenario. After which the following scenarios will be compared with each other:

- Baseline vs Shape-changing
- Shape-changing vs Blinkers
- Blinkers vs Combined modality
- Shape-changing vs Combined modality
- Baseline vs Blinkers
- Baseline vs Combined modality

Through these comparisons, the study aims to determine whether shape-changing behavior enhances participants perceptions of the robot's likability, intelligence, and safety. Given the number of pairwise comparisons, a Benjamini-Hochberg False Discovery Rate (FDR) correction will be applied to control for potential Type 1 and Type 2 errors, as explained in subsubsection 5.6.2

Overall, to evaluate the unique contribution of shape-changing, we compare its effects on legibility to those of blinkers and gaze individually, as well as against a combined modality that integrates all three cues. Further, we assess the effect of shape-changing on shaping perceptions of warmth, competence and safety. These findings aim to inform the design of socially accepted robots by clarifying how shape-changing influences perceptions.

To answer the hypotheses, both quantitative and qualitative data will be collected. Quantitative data includes response time, accuracy, confidence ratings, surprise ratings, and ROSAS questionnaire responses. Qualitative data will be obtained from structured interviews to capture participants experiences and perceptions, as explained in subsubsection 5.4.1.

By designing the study with this approach, we aim to answer the sub-research questions and evaluate the potential of shape-changing behavior to enhance the legibility and social perception of robots with a head during navigation in social hallways.

5.2 Participants

5.2.1 Requirement

Participants needed to be at least 18 years old, capable of following instructions, and proficient in basic English or Dutch. These criteria include as many as possible participant leading to inclusivity, reflecting the diversity typical in hospital environments, while also accommodating the researcher's language capabilities. Informed consent was mandatory for all participants. Further,

to maintain the study's exploratory character of the early testing phases, individuals with prior knowledge of the research's focus on directional intent were excluded.

5.2.2 Recruitment process

The recruitment process includes walking up to various persons at the University of Twente and politely asking them if they might be interested to participate in a study on social robot behavior. Further, friends and family members are asked which are unknowing about the ongoing research to prevent biasing the data. Care was taken to avoid recruiting multiple individuals from the same working study group to prevent cross-participant learning about the study. Also in case we still wanted to ask another individual from the participants study group, the participant was politely asked not to share details with others at the table who might also participate. To enhance engagement, participants were offered a small incentive in the form of snacks, distributed beforehand to ensure that participation was not purely incentive-driven.

5.3 Experimental setup

5.3.1 Participant setup

Participants were seated next to the researcher at a large table. The table was prepared with: an information sheets and consent forms placed alongside a pencil; a secondary computer screen positioned closer to the participant, allowing them to read and respond to questions using a mouse; and a VR headset placed on the right hand-side of the participant within reach, as displayed in Figure 8. To ensure unbiased responses, no scenario-related information was displayed on any screen until after the participant wears the VR headset. This measure minimized the change to provide additional unintended cues about the study to the participant.

The experiments were conducted in various environments, including silent dedicated testing rooms, living rooms of participants and more noisy university tinkering environments. This is not a problem as the headset created environmental noise on purpose to simulate a more noisy hospital environment.

The equipment includes a meta quest 3 headset with a right hand controller, of which the index trigger button is used to pause the simulation during the task. This VR is connected by a link cable to have a better connection to the computer for better performance, thus a reduced change of frame drops. This cable is connected to a Lenovo Legion Y540-15IRH laptop, as this laptop seemed suitable for connecting a VR headset and run experimental tasks without lag.

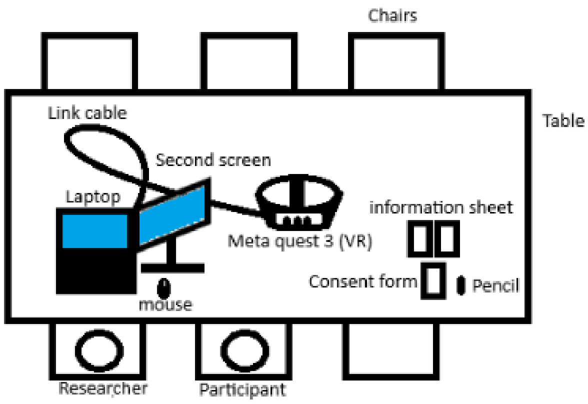


Figure 8: Participant setup

5.3.2 Virtual environment

The Virtual Reality (VR) environment was designed to simulate a hospital hallway. In this setting, a robot navigates toward the participant, who is tasked with interpreting the robot's intentions. This environment is build in Unity and made to look like a realistic hospital environment, including hospital environmental sounds. In this environment participants observe the robot driving in the VR environment and respond when they believe they understand its intent by pressing the index trigger button to pause the environment.

5.4 Data collection

5.4.1 Data Types

The data collection consist of a mix between qualitative and quantitative data. Since the aim of this thesis is not to obtain statistical significance but to identify trends and patterns (see subsubsection 5.6.1), a mixed-methods approach is used. This allows qualitative data to complement and improve the interpretation of quantitative findings.

For measuring the legibility the response time and accuracy is combined with confidence and surprise ratings. Where response time and accuracy are metrics for assessing legibility, as they reflect how quickly and correctly participants can predict the robot's intentions, these measures have being combined before into one measure, as shown by [13, 21, 20]. While confidence and surprise is added to capture additional dimensions of participants perception, which is inspired by [14], explaining that the confidence offers insights into how clearly and predictably the robot communicates its intent. Also surprise is included based on findings from [21, 17, 16]. As unexpected behavior can negatively affect legibility by failing to meet participants expectations. Meaning surprise allows us to measure how well the robots behavior aligns with the anticipation, potentially improving the evaluation of legibility. By combining these measures we build on established ways to measure legibility, combining subjective and objective measures. Leading to the understanding of both performance and perception of the shape-changing modality. We found these references thanks to the paper on legibility of robot behavior [28]. Overall, the quantitative data includes:

The response time, which is the frame in which the participant presses the pause button, meaning the frame in which the participant believes to understand what the robot's intent is in the environment. This frame is recorded in the unity environment and noted by the researcher in Qualtrics.

Accuracy, which indicates whether the participant correctly understood the robot's intent. This is verbally provided by the participant and noted by the researcher in Qualtrics.

Legibility score, as response time and accuracy are established metrics for assessing legibility. Response time and accuracy will be combined into one measure for the legibility score. Where guessing wrong gets a score of 0 and guessing right gets a higher score the faster it happens. For this we implemented the formula of Figure 9.

$$\text{legibilityScore} = \begin{cases} 0, & \text{if guess is wrong} \\ \text{maxScore} \times \left(1 - \frac{\text{frame}-\text{startFrame}}{\text{endFrame}-\text{startFrame}}\right), & \text{if guess is correct and frame} \leq 465 \\ 0, & \text{if guess is correct and frame} > 465 \end{cases}$$

Figure 9: Formula for legibility score

Confidence rating, which scale is shown in the environment as a 5 point likert scale, where the participant verbally states the assigned number, which is noted by the researcher in Qualtrics.

Surprise rating, which is just like the confidence rating shown in the environment as a 5 point likert scale and the participant again asked to verbally rate the surprise. The number is noted down by the researcher in Qualtrics.

ROSAS questionnare [30], this is a 9 point likert scale questionnaire assessing the robot's social attributes (warmth, competence, and discomfort) which is completed by participants in Qualtrics after the experiment and interview.

The qualitative data includes a structured interview, conducted immediately after testing, as we consider this information crucial. To ensure participants have the test and their observations fresh in mind, the interview takes place before the ROSAS questionnaire. The questions focused on asking about the participants experience, thought process, improvement suggestions, etc. These responses are written down real-time during the interview.

5.4.2 Data collection

Considering the data collection, Unity is used to measure the response time in frames, here we can take a look in the timeline and read out at which time frame a participant presses the pause button. This is written down and stored just like all the other quantitative data in Qualtrics. Considering the interview, this data which consist of the verbal responses from participants is transcribed in OneDrive word, as this is accepted for data storing. The data will be retained for the duration of the thesis and securely deleted afterwards the completion of the study.

5.4.3 Ethical considerations

Participants are required to sign a consent form before participation, and these forms will be securely stored in a safe at the University of Twente. All data collected during the study will be kept confidential, and no personally identifiable information will be included in the published results. Data will be made anonymous immediately and stored only for the duration of the thesis, as it is relevant solely to this research.

Participation in the study is entirely voluntary, and participants can withdraw at any time without consequences. They are informed of their right to withdraw before their data is made anonymous. After it is made anonymous, withdrawal is not possible to maintain the integrity of the dataset.

The study includes the potential risks of motion sickness or discomfort due to the virtual reality environment. To mitigate these risks, participants can take breaks or stop their participation entirely. At the end of the study, participants will be thoroughly debriefed and informed about the study's specific insights.

This study has been reviewed and approved by the ethics committee at the University of Twente. Access to the anonymous data will be restricted to authorized researchers only.

5.5 Procedure

The procedure of the experiment consisted of a sequence of steps to ensure consistency, starting with:

Step 1: introduction and consent form.

The participant is informed by means of the information sheet, which includes the details of the study, its purpose and the potential risks it includes. Afterwards the study is verbally explained, ensuring the participant understands the objectives and steps of the study. Afterwards the participant is tasked to sign the consent form to confirm their understanding and acceptance to participate in the study. in this study the time the participant spends in the Virtual Reality world is kept rather short, to minimize the VR-induced fatigue, however the participant is also told to take the headset off and take a break or stop the study if needed to prevent the risk of becoming motion sick.

Step 2: Pre-experiment demographics and experience survey

After signing the consent form the participant is asked to fill in demographic questions, including: gender identification (Woman, man, nonbinary/genderqueer, prefer not to answer, or other) and age group (8–24, 25–34, 35–44, 45–54, 55–64, or 65+). Further the familiarity level of participant is categorized into low, medium or high, based on the ticked box (”I have no prior experience with robots”; ”I have seen robots in videos or media but never interacted with one”; ”I know what robots are, but never interacted with one”; ”I have interacted with robots before”; ”I have used robots in a professional setting (e.g., manufacturing, healthcare)”, ”I have built, programmed, or worked extensively with robots” or ”other”). Finally the participant is asked if they possess a driver's license, to assess potential biases in interpreting blinkers.

Step 3: Scenario assignment

Participants were assigned to one of four experimental scenarios based on their survey responses. This is done by means of a sorting mechanism based on driver license and their experience with robots.

Scenario (1). Baseline with gaze only.

Scenario (2). Shape-changing and gaze.

Scenario (3). Blinkers and gaze.

Scenario (4). Shape-changing with blinkers and gaze.

The assigning process takes less than a minute. Afterwards participants were informed that their scenario would be explained at the end of the study to avoid influencing their responses.

Step 4: Virtual reality introduction

After being sorted in one of the scenario's the researcher will explain to the participant how to use the Oculus quest 3 VR headset and explains how the controller works. Here it is explained to the participant how to use the index trigger as it serves as a pause button of the robots movement, which they can press once they think to understand what the robot will do next. Afterwards the participant is tasked to put on the VR headset, while the test environment is loading. Here the headset is adjusted for comfort and after putting on the headset, the participant is asked to read the letters on the right hand side on the wall, this to see if participants are able to distinguish left and right from each other and are in the procession of sufficient vision to observe the robot. Afterwards the participant is tasked to look at the robot driving around and press the trigger, this shows the participant how the trigger works, while also validating the trigger its functionality. From here on it is time for the experiment, which is based on previous ways of conducting research on robot legibility as reviewed by [28].

Step 5: The experiment

Once it is validated everything works and the participant is ready for the experiment, the participant is briefed on the main task, by repeatedly (every round once) reminding the participant before the start: "remember to observe the robot and press 'pause' when you think to understand what it will do next." A verbal countdown is started from 3 to 0 after which the experiment begins in which the robot walks towards the participant. Once the participant thinks to understand the robot intent, it presses pause and a question pops up on the screen "what do you think is the robot's intent?". The participant answers verbally, which is written down by the researcher in combination with the response time which is shown in frames in the Unity environment. After answering this question, the following pop-up appears with the question: "How confident are you in your choice?". The participant verbally rates their confidence on a 1-5 likert scale, which answer is noted down by the researcher. Afterwards the participant watches the entire video of the robot's movement and is asked: "How surprising was the outcome?" Which the participant is asked to rate on a 1-5 likert scale and again noted down by the researcher. This procedure is repeated for three rounds in which the robot moves to the left in the first round, the right in the second round and left again in the third round.

Step 6: Post-experiment structured interview

After the 3 rounds the participant is asked to take off the VR headset, to join in on the interview. The following questions were asked (dutch translation between brackets):

- How did you experience the robot? (Hoe heb je de robot ervaren?).
- Can you explain the steps or thought process you used to understand what the robot was trying to do or communicate? (Kun je uitleggen welke stappen of gedachten in je opkwamen om te begrijpen wat de robot probeerde te doen of te communiceren?).
- What changed for you between rounds in how you understood the robot? (Wat veranderde er voor jou tussen de rondes in hoe je de robot begreep?).
- Was there anything that helped you understand its intent? (Was er iets dat jou helpte met het begrijpen van de robot zijn intentie?).
- What did you find confusing or unclear about the robot's intent? (Wat vond je verwarrend of onduidelijk aan de robot zijn intentie?).
- How natural did you experience the robot communicating its intent? (Hoe natuurlijk voelde de communicatie van de robot zijn intentie?).
- If you were to redesign the robot's intent to make them clearer, what would you change? (Als je iets mocht aanpassen zodat de robot zijn intentie duidelijker communiceert, wat zou je aanpassen?).
- How safe did you feel interacting with the robot? (Hoe veilig voelde de interactie met de robot?).

Step 7: ROSAS questionaire

After the participant answered all the questions of the structured interview in this order. The participant is tasked to fill out a ROSAS (Robot Social Attributes Scale) questionnaire on warmth, competence and safety in the Qualtrics environment.

Step 8: Finalize study

Finally after filling in the ROSAS, the participant is thanked for participation and explained its scenario and what the study exactly entails. Afterwards, the participant is asked if questions still remain to be answered, After answering all the participants questions the research is finished, overall taking approximately 20 minutes per participant.

5.6 Analysis

5.6.1 Pre-analyses G*power test participant amount

To determine the required amount of participants for statistical significance, a G*Power analysis was conducted. The analysis considered scenario's across measures such as legibility score, confidence, and surprise ratings. Below is a breakdown of the estimated requirements:

- **Sub research question 1:** To compare two scenarios using MANOVA (Multivariate Analysis of Variance) with an expected low effect size (0.1), a significance level of 5% ($\alpha = 0.05$), a power of 80% (1-$\beta = 0.8$), and three response variables, a total sample size of 114 participants is required. If an ANOVA is conducted for separate variables under the same parameters, 788 participants would be needed.

- **Sub research question 2:** Similar requirements to Sub-Research Question 1, with MANOVA needing 114 participants and ANOVA requiring 788 participants.

- **Sub research question 3:** Using the same parameters for a MANOVA with two groups and three variables also required 114 participants, while ANOVA would need 788 participants.

- **Sub research question 4:** For sub research question 4 we expect to require a comparison between the shape-changing included and not included scenario's. Including the 3 dependent variables, warmth, competence and discomfort. Requiring a MANOVA test. Again we expect an effect size of 0.1, α error of 0.05, a 1-β error prob of 0.8, with 3 variables, but now with a group size of 4, as all the scenario's will be included. Requiring a total sample size of 56 participants. When an ANOVA is conducted on these findings a total sample size of 1096 participants is required.

Given the constraints of time and resources, achieving the required sample sizes for statistical significance is unrealistic for a thesis. Due to this a mixed-methods approach has been chosen, combining quantitative data to identify trends and patterns with qualitative data to provide deeper insights. This shift makes the study more exploratory in nature.

Due to the exploratory nature the focus will be on the recruitment of a small sample size of in total **20 participants**. This small sample size made statistical methods like MANOVA and ANOVA unsuitable due to the increased likelihood of violating assumptions such as homogeneity of variances (similar variability) and normality of residuals (normal distributed variabled). By reading discovering statistics using SPSS, by Andy Fields [52], we found the need for a Non-parametric test, such as the kruskal-Wallis test or the Mann-Whitney U test. These tests compare medians rather than means, do not assume a normal distribution and are more robust to unequal variances and are therefore more appropriate for smaller sample sizes. However, these tests can only compare one dependent variable a time, so multiple tests need to be conducted, which means there is a need for a correction to determine if there is a true statistical significant score, as each test improves the change for a type 1 error, however too much correction can also cause a potential type 2 error.

Given these limitations, the primary focus will shift from significance testing to exploratory trend analysis. Descriptive statistics and effect sizes will be used to identify meaningful patterns in the data, while qualitative data, analyzed thematically, will serve to validate, nuance, or challenge the interpretations drawn from the quantitative results. This approach enables the study to contribute valuable insights despite the limitations in sample size and statistical power.

5.6.2 Hypothesis testing and analytical approach

This section outlines how the hypotheses will be tested and either confirmed or declined through specific scenario comparisons and analyses.

Initial analytical step

As an initial analytical step expected for a 2x2 study design, we perform a two-way ANOVA to explore main and interaction effects. This provides an overview and illustrates how the data would be analyzed under ideal conditions with a sufficiently large sample size. However, given the small sample size (N = 20) and the exploratory nature of the study (as mentioned in subsubsection 5.6.1), the assumptions of ANOVA, such as normality and homogeneity of variances, are likely to be violated. Therefore, we complement this initial analysis with non-parametric Mann–Whitney U tests, which are more appropriate for small samples and do not rely on these assumptions. These tests form the basis of our statistical interpretation.

How legible is shape-changing in comparison to gaze as directional modality?

A comparison will take place between Baseline and Shape-changing. The dependent variables would include the legibility score, confidence, and surprise ratings. The legibility score will be calculated for all scenarios. In each scenario, the robot starts moving at frame 0, but its directional intent is only communicated from frame 200 onward. Therefore, we use frame 200 as the start frame, as this isolates the directional intent cue from earlier anticipatory, attention-drawing movements. The end frame is defined as frame 465, as at this frame the intent is already shown to the participant. The formula for calculating the legibility score is shown in Figure 9. The max score will be defined as 10 and rounded off at the fourth decimal. Once all the data points are collected (see Appendix F, Figure 39), descriptive statistics will be calculated using Excel, and the IBM SPSS Statistics application will be used for further analysis.

We will make use of the Mann-Whitney U test, conducting one for each dependent variable to find out if one of the variables contribute to a difference between the scenario's. As multiple tests are performed the chance of a type 1 error increases, meaning a Bonferroni correction is required, as there are 3 dependent variables we divide the alpha level (required P-value) by 3. Afterwards we calculate the effect size, where a larger effect size indicates a larger effect, by means of r (Rank-biserial correlation), this is calculated by deviding Z-value (standart test statistic from the test), by $\sqrt{N}$ (N = total sample size of both groups) and make use of descriptive statistics, such as the mean, median, the interquartile range (IQR) and range, creating a boxplot for visual representation of the data and to provide additional context for the observed trends.

To complement the quantitative findings, we will perform a thematic analysis on qualitative feedback collected from participant interviews. This analysis will focus on identifying key themes related to participant perceptions of shape-changing. Specific areas of interest include participants descriptions of their confidence and surprise, as well as their reasoning on why shape-changing might improve or fail to improve legibility.

This hypothesis will be supported if Shape-changing shows an improvement in the legibility score and confidence, with reduced surprise ratings compared to the Baseline. However, since this study is exploratory, the goal is to identify trends rather than draw definitive conclusions.

How legible is shape-changing in comparison to blinkers as directional modality?

A comparison will take place between Shape-changing and Blinkers. The dependent variables include the legibility score, confidence, and surprise ratings. The same analysis steps will be followed as for the first sub-research question, including the Mann-Whitney U test, effect size calculation, and descriptive statistics.

Also here qualitative data is added to confirm the findings. The focus of the qualitative data will be on the participants perceptions of shape-changing and blinkers, exploring their confidence, surprise, and reasoning about why one modality might be more legible in comparison to the other.

This hypothesis will be supported if Shape-changing shows a higher legibility score, greater confidence, and less surprise compared to Blinkers. As with the previous question, this study is exploratory and focuses on identifying trends rather than drawing definitive conclusions.

To what extend can shape-changing contribute to improving the legibility of other modalities, such as blinkers?

A comparison will take place between Blinkers and the Combined modality scenario. The analysis process will be the same as in the previous questions, with the Mann-Whitney U test, effect size calculations, and descriptive statistics. The qualitative data will help validate the findings, focusing on whether participants perceive the Combined modality as more legible than blinkers alone, and their reasoning for this.

This hypothesis will be supported if the Combined modality shows higher legibility scores, greater confidence, and lower surprise ratings compared to Blinkers alone. As with the previous sub-research questions, the study aims to identify trends rather than draw definitive conclusions.

How does shape-changing influence participants perceptions of the robot's likability, intelligence and safety during navigation?

As mentioned before in subsection 5.1, comparisons will take place between multiple different scenario's. For these comparisons, ratings on the dependent variables warmth, competence, and discomfort will be collected using the ROSAS questionnaire. Descriptive statistics (mean, median, IQR, and range) will be calculated for each dependent variable in each scenario, and boxplots will visualize the data distribution.

The 6 data points for each dependent variable will be combined into one mean value per participant, as done in the traditional ROSAS [30]. These mean values will be analyzed using the Mann-Whitney U test for each dependent variable across the scenarios. In this test a Bonnferroni correction can be too strict, which could lead to a type 2 error (failing to detect a significant difference) as we have 6 comparisons and 3 dependent variables. Instead, the Benjamini-Hochberg False Discovery Rate (FDR) procedure will be applied to control for multiple comparisons. This method is more appropriate as it balances the risk of Type 1 and Type 2 errors. The Benjamini-Hochberg FDR is applied to the alpha level at the end of the analyses to determine if the findings are significant. This procedure is applied separately for each dependent variable to ensure it is not overly strict, as each variable addresses a distinct part of the sub-research question. The adjusted P-values, used to determine the significance of the findings, are provided in Appendix E (Figure 38). Afterwards, effect size will be calculated to assess the strength of the observed differences.

Further, qualitative data will be analyzed. The tags obtained from the analysis, will be matched to the attributes they aligned with most closely. This will be accomplished through the use of pre trained word embeddings (Word2Vec) to measure semantic similarity between the tags and predefined seed words representing each attribute. The seed words for warmth, competence, and safety were derived from the ROSAS framework, and tags from the thematic analysis are evaluated to determine which best fit each attribute. The code used for this analysis can be found in Appendix G. This is done to validate the findings, focusing on participants perceptions of the robot's warmth, competence, and safety. The analysis will help determine whether, and how, the Shape-changing and Combined modality scenarios score higher on these traits compared to the Baseline and Blinkers scenarios.

This hypothesis will be supported if Shape-changing and Combined modality score higher on warmth, competence, and safety compared to the Baseline and Blinkers. As with other research questions, the study is exploratory, aiming to identify trends rather than establish significant findings.

Finally as mentioned before the interview data will be analyzed thematically, meaning that the data will be reviewed to identify, organize and interpret patterns or themes within the responses. This included becoming familiar with the data re-reading the transcripts to understand the content and context, highlighting and labeling specific pieces of data, searching for themes by means of grouping the labels into categories, reviewing the themes and defining the final themes, to clearly represent the content of the data. Additionally, the connection to the scenarios is maintained, allowing for a comparison of how often a theme is mentioned and how many participants in each scenario referenced towards it.

6 Quantitative results

To answer the research questions, statistical analyses were conducted based on the dataset presented in Appendix F. This section presents the results for each research question, including statistical findings. Descriptive statistics, such as the mean, median, interquartile range, and range, are visualized using boxplots (see Appendix C), along with effect size calculations. These quantitative results are complemented by qualitative findings from participant feedback. Since this study is exploratory, the focus is on identifying trends rather than finding statistical significances. However, statistical analyses are included to determine potential differences between scenarios. The analysis begins with a two-way ANOVA. While the number of participants is limited, as discussed in subsubsection 5.6.1, the ANOVA provides initial insights into possible main and interaction effects and illustrates how the data would be analyzed under ideal conditions with a sufficient sample size.

6.1 Initial analytical step

Source	**Dependent Variable**	**F**	**Sig.**	η_p^2
Shape-changing	Legibility	0.051	.824	.003
	Surprise	1.316	.268	.076
	Confidence	0.508	.486	.031
	Warmth	12.001	*.003	*.429
	Competence	2.437	.138	.132
	Discomfort	7.491	*.015	*.319
Blinkers	Legibility	0.029	.867	.002
	Surprise	0.053	.821	.003
	Confidence	0.000	1.000	.000
	Warmth	4.429	.051	.217
	Competence	1.012	.329	.059
	Discomfort	1.176	.294	.068
Interaction (S $\times$ B)	Legibility	0.321	.579	.020
	Surprise	2.579	.128	.139
	Confidence	1.143	.301	.067
	Warmth	0.085	.775	.005
	Competence	6.413	*.022	*.286
	Discomfort	4.930	*.041	*.236

Note. Significant values ($p < .05$) are marked with *.

Table 2: Tests of between-subjects effects (2x2 study MANOVA)

Shape-changing significantly influenced perceptions of warmth and discomfort ($p < .05$). Interaction effects between Shape-changing and Blinkers were significant for competence and discomfort. There was a trend toward significance for the effect of Blinkers on warmth ($p = .051$).

Now that the overall between subject effects have been explored, we will proceed by addressing each of the research questions individually. Given the small sample size and potential violations of normality and homogeneity of variances, we will use the Mann-Whitney U test to compare conditions and gain a more clear understanding of how shape-changing influenced the participant responses.

6.2 RQ1: How legible is shape-changing in comparison to gaze as directional modality?

As mentioned in the analysis subsubsection 5.6.2, this sub-research question compares the Baseline and Shape-changing scenario's using the dependent variables legibility score, confidence ratings, and surprise ratings. A Bonferroni correction was applied to the alpha level, requiring a P-value of 0.0167 ($p = \frac{0.05}{3}$) for statistical significance.

Table 3 summarizes the results of the Mann-Whitney U tests for legibility, confidence, and surprise ratings, including descriptive statistics (mean, median, IQR, range), test statistics (U, Z, p), and effect sizes (r).

Measure	Scenario	Mean	Median	IQR	U	Z	p	Effect Size (r)
Legibility	Baseline	3.22	2.50	4.55	9.00	-0.73	0.465	-0.231
	Shape-changing	4.26	3.95	4.62				
Confidence	Baseline	4.20	4.00	1.50	11.00	-0.33	0.740	-0.105
	Shape-changing	4.00	4.00	2.00				
Surprise	Baseline	1.40	1.00	1.00	10.00	-0.60	0.549	-0.190
	Shape-changing	1.60	2.00	1.00				

Note. Significant values ($p < .05$) are marked with *; no significant values were found in this instance.

Table 3: Results for RQ1: Comparison of Baseline and Shape-changing

None of the comparisons reached statistical significance ($p < 0.0167$). However, visual inspection of the descriptive statistics suggests that Baseline scored slightly lower in legibility, higher in confidence, and similar in surprise. Boxplots visualizing the distribution of legibility, confidence, and surprise ratings can be found below in Figure 10, Figure 11, and Figure 12), where X indicates the mean. Or in Appendix C (Figure 26, Figure 27, and Figure 28).

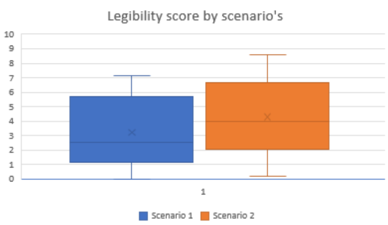


Figure 10: RQ1 boxplot legibility

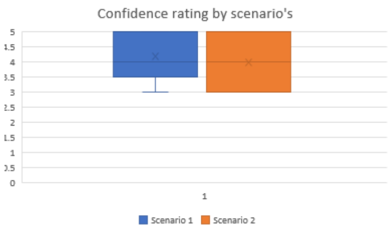


Figure 11: RQ1 boxplot confidence

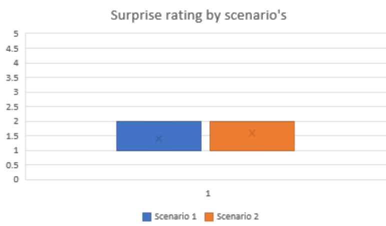


Figure 12: RQ1 boxplot surprise

6.3 RQ2: How legible is shape-changing in comparison to blinkers as directional modality?

As mentioned in the analysis subsubsection 5.6.2, this sub-research question compares Shape-changing and Blinkers using the dependent variables legibility score, confidence ratings, and surprise ratings. A Bonferroni correction was applied to the alpha level, requiring a P-value of 0.0167 ($p = \frac{0.05}{3}$) for statistical significance.

Table 4 summarizes the results of the Mann-Whitney U tests for legibility, confidence, and surprise ratings, including descriptive statistics (mean, median, IQR, range), test statistics (U, Z, p), and effect sizes (r).

Measure	Scenario	Mean	Median	IQR	U	Z	p	Effect Size (r)
Legibility	Shape-changing	4.26	3.95	4.62	12.00	-0.104	0.917	-0.033
	Blinkers	4.19	4.23	3.80				
Confidence	Shape-changing	4.00	4.00	2.00	12.00	-0.112	0.911	-0.035
	Blinkers	3.60	5.00	3.50				
Surprise	Shape-changing	1.60	2.00	1.00	12.00	-0.113	0.910	-0.036
	Blinkers	2.20	1.00	3.00				

Note. Significant values ($p < .05$) are marked with *; no significant values were found in this instance.

Table 4: Results for RQ2: Comparison of Shape-changing and Blinkers

None of the comparisons reached statistical significance ($p < 0.0167$). However, visual inspection of the descriptive statistics suggests that Shape-changing scored slightly higher in legibility, higher in confidence and lower in surprise. Boxplots visualizing the distribution of legibility, confidence, and surprise ratings can be found below in Figure 13, Figure 14, and Figure 15), where X indicates the mean. Or in Appendix C (Figure 26, Figure 27, and Figure 28).

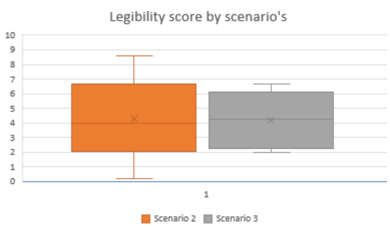


Figure 13: RQ2 boxplot legibility

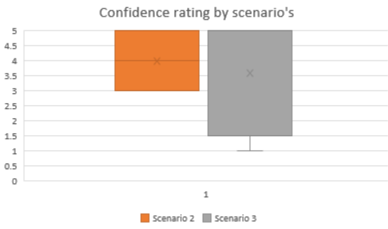


Figure 14: RQ2 boxplot confidence

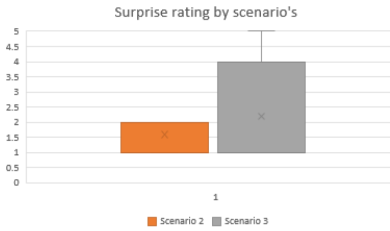


Figure 15: RQ2 boxplot surprise

6.4 RQ3: To what extend can shape-changing contribute to improving the legibility of other modalities, such as blinkers?

As mentioned in the analysis subsubsection 5.6.2, this sub-research question compares Blinkers and the Combined modality scenario using the dependent variables legibility score, confidence ratings, and surprise ratings. A Bonferroni correction was applied to the alpha level, requiring a P-value of 0.0167 ($p = \frac{0.05}{3}$) for statistical significance.

Table 5 summarizes the results of the Mann-Whitney U tests for legibility, confidence, and surprise ratings, including descriptive statistics (mean, median, IQR, range), test statistics (U, Z, p), and effect sizes (r).

Measure	**Scenario**	**Mean**	**Median**	**IQR**	U	Z	p	**Effect Size** (r)
Legibility	Blinkers	4.19	4.23	3.80	10.00	-0.522	0.602	-0.165
	Combined modality	3.74	1.75	7.14				
Confidence	Blinkers	3.60	5.00	3.50	9.00	-0.900	0.368	-0.285
	Combined modality	4.60	5.00	1.00				
Surprise	Blinkers	2.20	1.00	3.00	7.50	-1.491	0.136	-0.472
	Combined modality	1.00	1.00	0.00				

Note. Significant values ($p < .05$) are marked with *; no significant values were found in this instance.

Table 5: Results for RQ3: Comparison of Blinkers and Combined modality

None of the comparisons reached statistical significance ($p < 0.0167$). However, visual inspection of the descriptive statistics suggests that Blinkers scored more stable and higher in legibility, but lower in confidence and higher surprise. Boxplots visualizing the distribution of legibility, confidence, and surprise ratings can be found below in Figure 16, Figure 17, and Figure 18), where X indicates the mean. Or in Appendix C (Figure 26, Figure 27, and Figure 28).

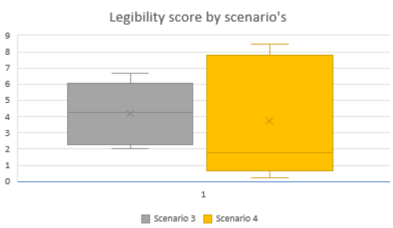


Figure 16: RQ3 boxplot legibility

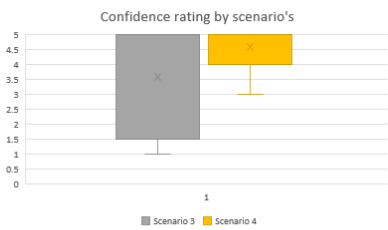


Figure 17: RQ3 boxplot confidence

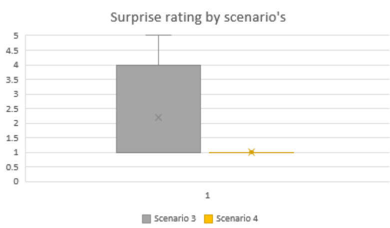


Figure 18: RQ3 boxplot surprise

6.5 RQ4: How does shape-changing influence participants perceptions of the robot's likability, intelligence and safety during navigation?

As mentioned before in the analysis subsubsection 5.6.2, to answer this sub-research question 6 comparisons will be conducted. These comparisons will take place for the 3 dependent variables warmth, competence and discomfort. A Benjamini-Hochberg correction was applied to adjust for multiple comparisons.

The results of the Mann-Whitney U tests for warmth, competence, and discomfort ratings are summarized in Table 7, Table 9, and Table 11, respectively. Descriptive statistics for each scenario are provided in Table 6, Table 8, and Table 10.

6.5.1 ROSAS warmth analyses

The descriptive statistic for the ROSAS warmth analyses are displayed below in Table 6 and displayed in boxplot format in Figure 19, where X indicated the mean.

Scenario	Mean	Median	IQR	Range
Baseline	2.8333	3	2.92	1.33–5.33
Shape-changing	4.7666	4.8330	1.5	3.5–5.5
Blinkers	4.0666	4.3333	2.17	2.33–5.83
Combined modality	5.7000	5.5	0.83	5.33–6.5

Table 6: Descriptive statistics for warmth ratings.

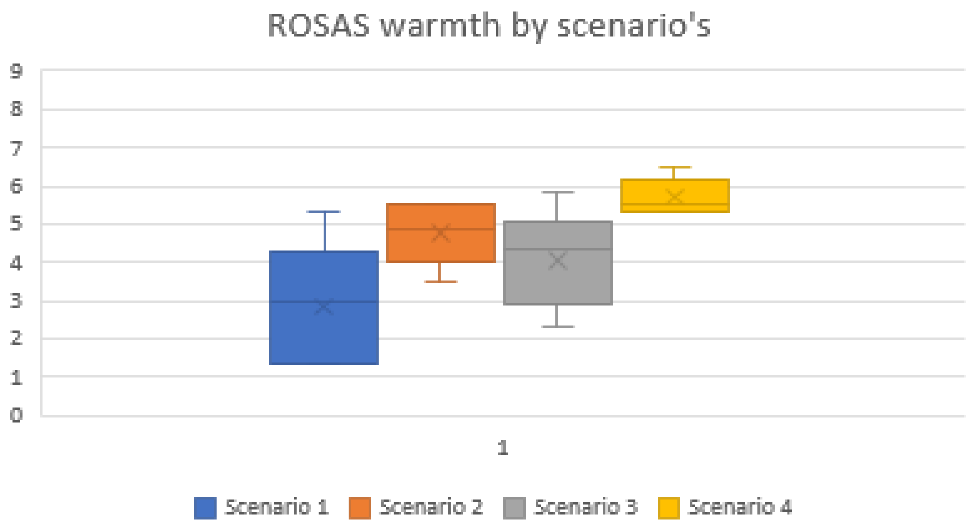


Figure 19: Warmth ratings

A series of Mann-Whitney U tests were conducted to compare warmth ratings across six different scenario pairs. The results for each comparison are summarized below in Table 7.

Comparison	U	Z	p-value	p-adj (BH)	Effect Size (r)
1. Baseline vs Shape-changing	3	-1.996	0.045866*	0.137598	-0.6312
2. Baseline vs Blinkers	6	-1.366	0.176	0.2112	-0.4320
3. Baseline vs Combined modality	1	-2.440	0.015*	0.09	-0.7717
4. Shape-changing vs Blinkers	7.5	-1.054	0.292	0.292	-0.3333
5. Shape-changing vs Combined modality	5	-1.591	0.112	0.168	-0.5032
6. Blinkers vs Combined modality	3.5	-1.897	0.058	0.116	-0.5999

Note. Significant values ($p < .05$) are marked with *.

Table 7: Mann-Whitney U test results and effect sizes for warmth ratings.

None of the comparisons reached statistical significance ($p - adj(BH) < 0.05$). However, comparison 1 and 3 indicates a trend. While the descriptive statistics show a clear visual trend as Shape-changing and especially Combined modality, scored highest on warmth.

6.5.2 ROSAS competence analyses

The descriptive statistic for the ROSAS competence analyses are displayed below in Table 8 and displayed in boxplot format in Figure 20, where X indicated the mean.

Scenario	Mean	Median	IQR	Range
Baseline	4.7333	4.3333	1	4.33–5.83
Shape-changing	6.7000	6.6667	1.25	5.83–7.83
Blinkers	5.4667	4.6667	2.17	4.50–7.17
Combined modality	5.0000	4.5000	2.58	3.67–7.33

Table 8: Descriptive statistics for competence ratings.

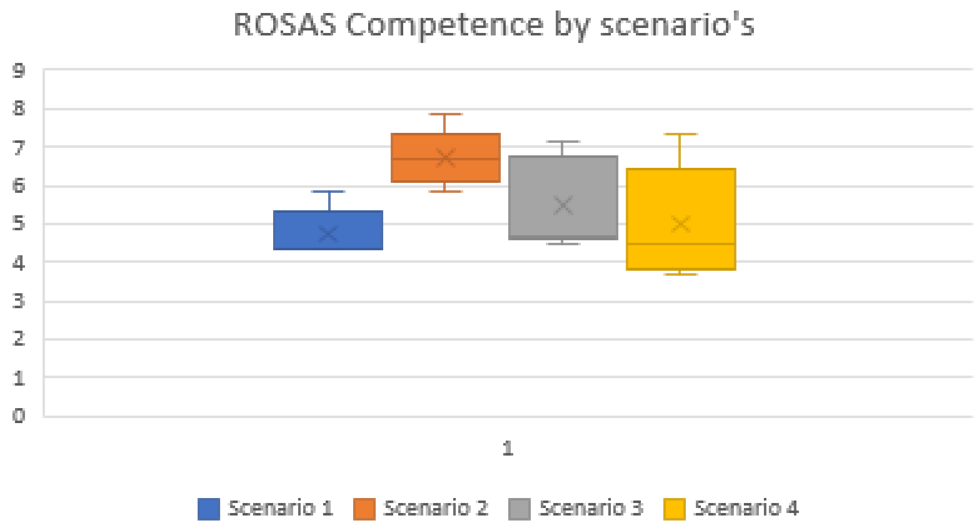


Figure 20: Competence ratings

A series of Mann-Whitney U tests were conducted to compare competence ratings across six different scenario pairs. The results for each comparison are summarized below in Table 9.

Comparison	U	Z	p-value	p-adj (BH)	Effect Size (r)
1. Baseline vs Shape-changing	0.5	-2.546	0.011*	0.066	-0.8052
2. Baseline vs Blinkers	6	-1.379	0.168	0.252	-0.4361
3. Baseline vs Combined modality	12	-0.106	0.916	0.916	-0.0335
4. Shape-changing vs Blinkers	5.5	-1.471	0.141	0.252	-0.4652
5. Shape-changing vs Combined modality	4	-1.776	0.076	0.228	-0.5617
6. Blinkers vs Combined modality	8.5	-0.841	0.401	0.4812	-0.2660

Note. Significant values ($p < .05$) are marked with *.

Table 9: Mann-Whitney U test results and effect sizes for competence ratings.

None of the comparisons reached statistical significance ($p - adj(BH) < 0.05$). However, comparison 1 indicates a trend. While the descriptive statistics show a clear visual trend as Shape-changing scored highest on competence.

6.5.3 ROSAS discomfort analyses

The descriptive statistic for the ROSAS warmth analyses are displayed below in Table 10 and displayed in boxplot format in Figure 21, where X indicated the mean.

Scenario	Mean	Median	IQR	Range
Baseline	1.4667	1.1667	0.75	1.17–2.33
Shape-changing	3.0667	2.6667	1.5	2.33–5.17
Blinkers	1.8333	1.6667	2.17	1.33–2.33
Combined modality	2.0000	2.0000	0.83	1.50–2.67

Table 10: Descriptive statistics for discomfort ratings.

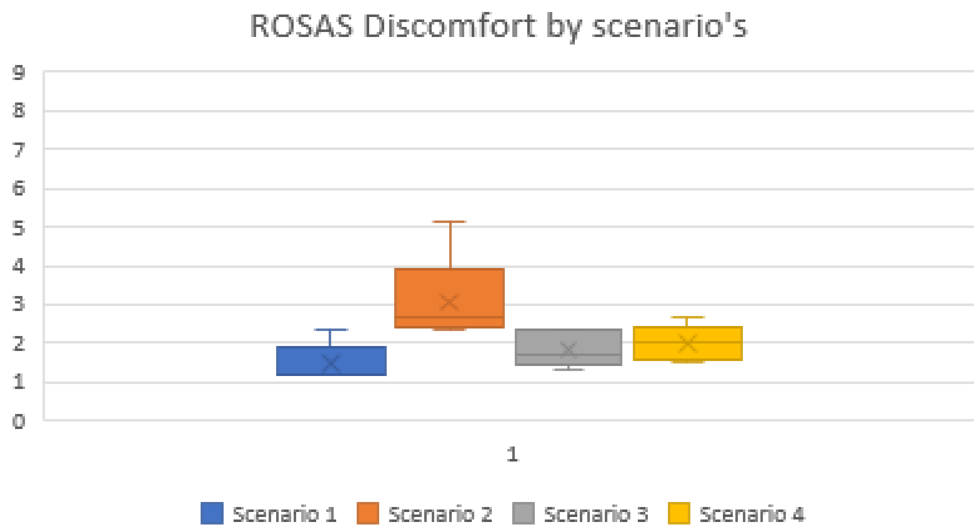


Figure 21: Discomfort ratings

A series of Mann-Whitney U tests were conducted to compare competence ratings across six different scenario pairs. The results for each comparison are summarized below in Table 11.

Comparison	U	Z	p-value	p-adj (BH)	Effect Size (r)
1. Baseline vs Shape-changing	0.5	-2.554	0.011*	0.045*	-0.8077
2. Baseline vs Blinkers	5.5	-1.504	0.133	0.1596	-0.4756
3. Baseline vs Combined modality	4.5	-1.697	0.090	0.135	-0.5367
4. Shape-changing vs Blinkers	1	-2.440	0.015*	0.045*	-0.4652
5. Shape-changing vs Combined modality	3	-2.009	0.045*	0.09	-0.7717
6. Blinkers vs Combined modality	10	-0.527	0.598	0.598	-0.1667

Note. Significant values ($p < .05$) are marked with *.

Table 11: Mann-Whitney U test results and effect sizes for discomfort ratings.

Comparison 1 and 4 reached statistical significance ($p{-}adj(BH) < 0.05$). While, comparison 5 indicates a trend. Also the descriptive statistic visually show that Shape-changing, scored highest on discomfort.

7 Thematic analyses

In the thematic analysis, a combination of bottom-up and top-down approaches was used. Initial tags were defined based on observation during the study and the intended design features of the robot, in order to determine if participants recognized or noted these elements. Other tags were based on recurring themes in participant responses.

The following tags were defined based on the intended design and initial observations: safe, smart, positive, scared, unsafe, unlogical, attention grabbing, animal like, human like, modality helpful, unnatural, not human like, unclear, expected to turn into the hallway, potential improvements, too extreme, driving, learning over rounds, blinkers too digital, context mismatch and overload of modalities. While the other tags displayed in Appendix D emerged from recurring themes in the participant interviews.

All interviews were read multiple times, and tags were assigned to parts of the text that reflected these themes. After defining the initial tags, the interviews were reviewed twice more to ensure consistency in tagging and to include any new tags that emerged.

Once all tags were assigned, related tags were grouped together to form broader themes. To analyze the data, the number of times each tag was mentioned was counted, and the number of participants who mentioned each tag was noted. Each participant could mention a tag multiple times within a scenario. We measured two measures: the total amount of tag mentions across all participants, and the number of unique participants who mentioned an unique tag (ranging from 0-5 for each tag, as we had five participants each scenario).

Finally, these numbers were added up to see how often each theme appeared across all scenarios. The results of this analysis, are shown in Appendix D.

The final structure includes six themes. Themes 1 and 2 focus on emotional perceptions, capturing positive and negative emotions, respectively. These are analyzed further in subsubsection 7.2.2, where insights into the research questions are presented based on these perceptions. Specifically, tags associated with these themes were mapped using semantic similarity analysis to match the tag to the most representative attribute. Themes 3 and 4 address the clarity of communication, distinguishing between clear communication and communication issues. Related findings and research question insights are provided in subsubsection 7.4.2. The final two themes cover future improvements and general observations, providing additional context and suggestions derived from participant feedback. These thematic findings contribute to a deeper understanding and explanation of the quantitative results.

7.1 Theme 1: positive emotions/perceptions

Scenario	**Total mentions across all tags**	**Unique tag–participant mentions**	**Total possible unique mentions 11 tags x 5 participants**
Baseline	17	14	55
Shape-changing	17	10	55
Blinkers	11	9	55
Combined modality	23	17	55

Table 12: Descriptive statistics thematic analyses on positive emotions/perceptions.

This subsection highlights the key tags that emerged within the theme, along with overall observations from the scenario's.

Baseline Key themes in the Baseline include: friendly (3), cute (3), safe (4), not scary (4), and nice (3). The gaze-only scenario was perceived as safe, friendly, and not scary. For example, Participant 19 stated, "It felt safe, I did not feel intimidated, it looked friendly." However, the overall positive emotions were moderate, as the same participant noted that the participant and the robot "did not interact with each other." Communication issues will be discussed further in subsection 7.4.

Shape-changing Key themes in Shape-changing include: friendly (6), smart (5), and positive (2). The gaze with shape-changing scenario was described as smart and friendly. Participant 4 commented, "I found it a friendly robot, as I found it lively." Participant 5 noted that the robot was designed smartly, as the growing motion made it clear that the robot was trying to communicate that it saw the participant. However, this interaction was also perceived as somewhat intimidating. Further discussion on negative associations can be found in negative emotions/perceptions subsection 7.2, while clear and helpful communication is explored in Clear/helpful communication subsection 7.3.

Blinkers Key themes in Blinkers include: safe (2) and not scary (3). The gaze with blinker scenario was perceived as safe and not scary. Participant 7 mentioned, "I didn't feel anxious or scared, no weird feeling about it." Participant 10 added, "Safe, the first time I thought you're going to run into me, but that didn't happen," which also indicates some uncertainty. This uncertainty is further discussed in communication issues subsection 7.4. Overall, the gaze with blinker scenario had the lowest number of positive emotions.

Combined modality Key themes in Combined modality include: cute (4), safe (6), surprising (5), nice (2), and fluffy (3). The combination of gaze, blinkers, and shape-changing was perceived as safe, surprising, and cute. Participant 12 noted, "One of the things that came to my mind is that it's cute and a little unrealistic, with reality, it wasn't making possible movements." This is an interesting observation, as the robot was developed based on a real robot. Overall, this scenario received the most positive reactions from participants.

7.2 Theme 2: negative emotions/perceptions

Scenario	Total mentions across all tags	Unique tag–participant mentions	Total possible unique mentions 8 tags x 5 participants
Baseline	6	6	40
Shape-changing	18	13	40
Blinkers	9	8	40
Combined modality	4	4	40

Table 13: Descriptive statistics thematic analyses on negative emotions/perceptions.

7.2.1 Observations

This subsubsection highlights the key tags that emerged within the theme, along with overall observations from the scenario's.

Baseline Key themes in the Baseline include: unpredictable (2), unsafe (1), scared (1), unlogical (1), and strange (1).

The gaze-only scenario was perceived as somewhat unpredictable and slightly unsafe. Participant 19 noted, "It was like robot behavior of what I would have seen, it did its own thing. If it was a human, it would just look at me and turn its head, but now it just turns in its own way without any form of interaction." This suggests that the interaction was not clear enough, making it difficult to predict the robot's intent. The communication issues will be discussed further in subsection 7.4. Despite this, the overall negative emotions were relatively low, as the scenario was not considered intimidating, shocking, or intense.

Shape-changing Key themes in Shape-changing include: intimidating (5), unsafe (4), unpredictable (2), and shocking (3). The gaze with shape-changing scenario was perceived as intimidating. Participant 3 mentioned, "If he is taller than me, I find that intimidating," but also added, "When he widens, he communicates, 'Look out, I'm coming, move out of the way.'" This indicates that while the robot's movements were intimidating, they communicated its intent. The scenario was also described as shocking. Participant 3 stated, "He made himself really big, and I thought, 'Oh shit,' like some kind of visual cue that he's onto me. The way he was moving forward seemed dangerous to me." Despite the shock, the robot's intent was communicated clearly, as further explained in Theme 3, Clear/helpful communication subsection 7.3. Some participants, like Participant 5, experienced this as positive: "Because of the blowing up effect, people are going to stop because they might be a little bit shocked, you jump up a little bit. It's positive, like a tap, making you think, 'What do you want to do?'" However, most negative emotional responses were associated with this scenario.

Blinkers Key themes in Blinkers include: unpredictable (3), unsafe (2), and scared (1).

The gaze with blinkers scenario was perceived as unpredictable and unsafe. Participant 7 commented, "Now he runs me over, and then at the last moment, he goes the other way," explaining that the robot's unpredictability until the last moment made them feel unsafe. Despite this, Blinkers had a moderate number of negative emotional responses overall.

Combined modality Key themes in Combined modality include: intimidating (1), intense (1), and unlogical (1). The combination of gaze, blinkers, and shape-changing was associated with the lowest number of negative emotions. Overall, participants felt safe. An interesting observation came from Participant 12: "I felt reasonably safe, but I can imagine that someone who has never seen it before might find it an intimidating move, especially the sizing up, it's animalistic, like when a cat puts its hair up, and you know something is wrong." This perception of the puffing as animal-like aligns with the intended design, communicating danger or that something is about to happen. While the participant did not feel intimidated, they acknowledged that others might, as seen in Shape-changing.

7.2.2 RQ insights positive and negative emotions/perceptions

This section, discusses some insights for **RQ4: "How does shape-changing influence participants perceptions of the robot's likability, intelligence, and safety during navigation?"**. By focusing on the positive and negative attributes related to warmth (likability), competence (intelligence), and discomfort (safety). Tags were mapped to ROSAS attributes (warmth, competence, discomfort) using semantic similarity analysis as explained in subsubsection 5.6.2.

Warmth (Likability)

Positive tags mapped to warmth included fluffy and happy, while no negative tags were found to indicate a lack of likability. Based on these associations, Combined modality was perceived as the most likable (3 tags from 2 participants), followed by scenarios 2 and 3, each with 1 tag. Baseline received no warmth-related tags, suggesting it was viewed as the least likable scenario.

Competence (Intelligence)

Positive tags reflecting competence included smart and friendly. In contrast, unlogical was considered a negative competence-related tag. Shape-changing stood out as the most intelligent scenario, with 11 positive tags (4 participants) and no negative associations. Baseline followed with 3 positive tags (2 participants), while scenarios 3 and 4 had lower intelligence ratings. All three of those scenarios also received one negative competence tag, suggesting some participants found the robot's behavior lacking in intelligence.

Safety (Discomfort)

Safety was the most frequently discussed attribute. Positive safety-related tags included safe, cute, fine, not scary, and positive. Conversely, negative tags included intimidating, unsafe, scared, strange, shocking, unpredictable, and intense. Combined modality was perceived as the safest, with 18 positive tags (14 participants) and only 3 negative associations. Baseline followed with 14 positive tags (12 participants), but also had 5 negative safety-related tags. Blinkers received 8 positive and 8 negative tags, indicating mixed perceptions. Shape-changing stood out as the least safe scenario, with just 5 positive safety tags and 18 negative tags (from 13 participants), suggesting that the robot in this condition was perceived as the least safe or most intimidating.

Scenario	Likability (Warmth)	Intelligence (Competence)	Safety (Discomfort)
Baseline	0 (0/10)	3 (2/10) / 1 (1/5) neg	14 pos (12/35) / 5 (5/35)
Shape-changing	1 (1/10)	11 (4/10) / 0 (0/5) neg	5 pos (5/35) / 18 (13/35)
Blinkers	1 (1/10)	1 (1/10) / 1 (1/5) neg	8 pos (7/35) / 8 (7/35)
Combined modality	3 (2/10)	2 (1/10) / 1 (1/5) neg	18 pos (14/35) / 3 (3/35)

Table 14: Combined positive and negative ROSAS related tags across scenarios for likability, intelligence, and safety.

7.3 Theme 3: Clear/helpful communication

Scenario	Total mentions across all tags	Unique tag–participant mentions	Total possible unique mentions 9 tags x 5 participants
Baseline	6	5	45
Shape-changing	46	17	45
Blinkers	18	9	45
Combined modality	46	25	45

Table 15: Descriptive statistics thematic analyses on clear/helpful communication.

This subsection highlights the key tags that emerged within the theme, along with overall observations from the scenario's.

Baseline Key themes in the Baseline include: clear (2) and natural (3). The gaze-only scenario was perceived as somewhat natural and smooth. Participant 9 noted, "Well, reasonably natural for a robot it's better than just a light, I guess. It was a fluid movement, but it was only a small part of the robot." This suggests that while the robot's movements were natural and smooth, the communication could have been clearer, potentially by making the gaze more pronounced or attention-grabbing (as mentioned in theme 5, Modality improvements/change subsection 7.5, which highlights the use of the entire body to communicate). This scenario scored the lowest in clear/helpful communication, as no participants mentioned the gaze as truly helpful, indicating that the robot did not interact with participants in a meaningful enough way.

Shape-changing Key themes in Shape-changing include: shape-changing helped (16), clear (10), attention-grabbing (8), animal-like (4), and human-like (4). The shape-changing behavior was described as attention-grabbing and clear. Participant 5 stated, "Becoming big, the intention already became clear to me that he was showing he is going to do something," indicating that the growing motion helped convey the robot's intent, such as signaling an upcoming movement or change in direction. For some, it communicated the need to create space. Participant 3 mentioned, "I see him getting thick and wide, then I think, 'He's coming, can you get out of the way?'" Participants also described the movement as more animal-like than human-like. Participant 6 commented, "It was fluid, but that's not how a human moves, it was more like an octopus," suggesting an expectation of more human-like movement. Overall, shape-changing effectively communicated intent, though it also introduced some trade-offs, such as being perceived as shocking or intimidating, as discussed in Theme 2, negative emotions/perceptions subsection 7.2. Overall, this scenario scored among the highest in clear/helpful communication.

Blinkers Key themes in Blinkers include: blinkers helpful (12), attention-grabbing (3), and clear (1). The blinker scenario was perceived as helpful. Participant 10 stated, "At first, the robot came towards me, and I thought he wants to greet me or needs me. Then the lights started flashing as an indicator, and I thought he is going to slow down. But then a light came on, and I thought he is going to take a turn." This indicates that the blinkers not only communicated intent but also grabbed participants attention. However, fewer participants mentioned tags related to clear/helpful communication in this scenario. Participant 7 noted, "I did see the head turn a little, but it didn't tell me as much as when I saw the light, and I hadn't seen it at first." This suggests that while the blinkers were helpful, they were not always noticed immediately. Further discussion on this can be found in Theme 4: Communication Issues/Errors subsection 7.4. The lower number of responses may also be due to blinkers being a more familiar modality, making them less impressive.

Combined modality Key themes in Combined modality include: shape-changing helped (14) and attention grabbing (11). The combination of gaze, blinkers, and shape-changing was frequently described as clear and helpful, with more participants mentioning various tags related to this theme. The communication was perceived as natural and animal-like. Participant 15 noted, "It felt quite natural, kind of like a bird mating dance." Participants also highlighted that the blinkers, combined with shape-changing, added additional information. Participant 12 mentioned, "Then the light went on, just like with autos he turns right immediately. I expected him to turn right into that path, but the movement in combination made it seem like I was in the way, so it sounds logical that he drove around me." However, some participants found the modalities somewhat contradictory. For example, the shape-changing communicated that the participant was in the way, while the blinkers signaled a turn. Notably, shape-changing was mentioned as helpful more often than the blinkers, potentially because it drew attention away from the blinkers. This is further discussed in Theme 6, General Observations subsection 7.6.

7.4 Theme 4: Communication issues/errors

Scenario	Total mentions across all tags	Unique tag–participant mentions	Total possible unique mentions 7 tags x 5 participants
Baseline	31	19	35
Shape-changing	24	17	35
Blinkers	28	15	35
Combined modality	25	17	35

Table 16: Descriptive statistics thematic analyses on communication issues/errors.

7.4.1 Observations

This subsubsection highlights the key tags that emerged within the theme, along with overall observations from the scenario's.

Baseline Key themes in the Baseline include: unclear (8), timing off (8), uncertainty (4), and not humanlike (4). The gaze-only scenario was perceived as unclear, with timing that seemed off. Participant 14 noted, "I knew he was looking at something, but did this mean anything, or was he just looking where he was going? You didn't know this beforehand, and he did this very early on." Due to the timing issue, participants also felt the robot did not seem humanlike. Participant 14 added, "You don't do this with people you don't look before you walk somewhere. Like in a hospital, I already see that door, but I don't look to the left 1100 meters in advance." This suggests that the early gaze did not clearly communicate intent to participants. This scenario scores highest on communication issues/errors.

Shape-changing Key themes in Shape-changing include: timing off (11), unclear (4), confusing (3), and unnatural (3). The gaze with shape-changing scenario was described as having timing issues, making the robot's intent somewhat confusing. Participant 4 mentioned, "It was confusing that he was coming straight at me, as if he was going to greet me, and then at the last minute, he went right or left. I didn't understand why he was blowing himself up." This indicates that the anticipatory movement was unclear. Participant 13 added, "A person would adjust their path earlier. Right now, he does this last minute." Participant 5 mentioned, "If I am in a hurry I should be able to read its intent immediately, but if an object is coming at you quickly it should act faster. So then a less extended animation would be better," indicating that one participant would prioritize a faster way of communication. Despite this, the timing issue was primarily related to the time it took for the robot to turn its body, with participants noting that the robot seemed to hold its movement for too long and that the animation took too much time to communicate. Further discussion on this can be found in Theme 5, Potential Improvements subsection 7.5.

Blinkers Key themes in Blinkers include: unclear (8), timing off (8), expected to turn into hallway (7), and uncertainty (4). The gaze with blinker scenario was perceived as unclear, as participants expected the robot to turn into the hallway. Participant 17 stated, "Generally, it was a little unclear. I thought he was going to turn into the corridor because of the flashing light, just like with a car, and not that he was passing by someone." This uncertainty may have been caused by timing issues. Participant 11 noted, "He would turn into the corridor earlier at the intersection. Perhaps turning on the blinker light later would have helped avoid this expectation, though I might still have wondered if he was going to turn into the door or not at all." This suggests that the timing issue led participants to expect the robot to turn 90 degrees into the hallway, later into the door, or not at all. This may indicate that blinkers communicate directional intent faster or in a more digital manner, making the communicated intent less information rich. Further discussion on this can be found in Theme 6, General Observations subsection 7.6.

Combined modality Key themes in Combined modality include: unclear (10), unnatural (4), confusing (3), and uncertainty (3). The combination of gaze, blinkers, and shape-changing was perceived as unclear and somewhat unnatural. Participant 18 commented, "It's not very natural, as I had no idea what it was trying to communicate. It was just a robot." However, it later became clear that this participant did not notice the blinkers on the robot. This suggests that shape-changing may be more dominant or distracting, causing the blinkers to be overlooked. This could lead to an overload of modalities, which is further explained in Theme 6, General Observations subsection 7.6.

7.4.2 RQ insights on clear communication and communication issues

This section provides insights into the comparison between scenarios, focusing on the clarity of communication and its impact on the legibility of the robot's intent for each sub research question.

RQ1: How legible is shape-changing in comparison to gaze as directional modality?

When comparing the Baseline (gaze-only) with Shape-changing, it is observed that Shape-changing elicited more responses related to clear and helpful communication. Participants in the Baseline found the gaze to be natural and smooth but noted that it lacked meaningful interaction. In contrast, Shape-changing was described as more attention-grabbing and clear. Some participants interpreted the growing motion as a signal to make space, while others saw it as a way to draw attention. This aligns indirectly with the intended design, as the growing motion was meant to communicate the robot taking space. Overall, it can be suggested that shape-changing enhances the legibility of the robot's intent, as it made the gaze more noticeable and was mentioned more often as clear and helpful compared to gaze alone.

At the same time, when comparing the Baseline with Shape-changing regarding communication issues, the Baseline elicited more theme-related responses from participants. The gaze modality alone was often described as unclear and not humanlike, as the interaction occurred too early, leading to uncertainty. Participants felt the gaze did not communicate intent clearly or richly enough. In contrast, Shape-changing was more frequently associated with timing issues, particularly the time the robot took before changing direction after finishing its animation. While this was desired behavior to provide participants with more time to decide on the robots intent, it caused some confusion. One participant explicitly noted that in situations requiring urgency, a quicker form of communication would be preferred, stating that a less extended animation would be more appropriate when the robot is approaching quickly. Despite this, shape-changing was still noted to communicate intent more clearly and more humanlike. Overall, it can be suggested that shape-changing reduces communication errors by making the robot's intent clearer and more humanlike, even though there are time-related problems due to speed and the holding of the intention for the purpose of this study.

RQ2: How legible is shape-changing in comparison to blinkers as a directional modality?

When comparing Shape-changing with Blinkers, Shape-changing elicited more positive responses related to clear and helpful communication. While Blinkers was described as helpful and attention-grabbing, it appeared to draw attention less effectively compared to shape-changing. Shape-changing was perceived as more intuitive, as participants described it more often as animal-like or human-like. Blinkers, although effective, relied on timing; when the timing was off, participants expressed that the robot's exact intent was unclear (as explained in Theme 4, Communication issues/errors subsection 7.4). This issue occurred less frequently in the shape-changing scenario. This difference may be due to the digital nature of blinkers (as discussed in Theme 6, General observations subsection 7.6) or the fact that blinkers communicate intent faster, as they involve turning on and off at a defined frame rate, whereas shape-changing is a sequence of movements. Based on participant responses, it can be suggested that shape-changing was perceived to be more clear and information-rich in communicating intent compared to blinkers.

At the same time, when comparing Shape-changing with Blinkers regarding communication issues, Blinkers elicited more theme-related responses from participants, despite participants noting a wider variety of issues for Shape-changing. The blinker modality alone often led participants to expect the robot to turn 90 degrees into the hallway, creating uncertainty as the robot continued driving toward the participant until the last moment. In contrast, shape-changing was more often described as confusing and unnatural. Overall, it can be suggested that shape-changing and blinkers perform similarly, however, shape-changing may reduce communication errors by more richly communicating the intent to pass rather than turn 90 degrees. However, blinkers communicate faster and are less confusing.

RQ3: To what extent can shape-changing contribute to improving the legibility of other modalities, such as blinkers?

When comparing Blinkers with Combined modality, Combined modality elicited more positive theme related responses from participants. The combination of blinkers and shape-changing was described as more attention-grabbing and natural, allowing participants to rely on their preferred modality. Participants also noted that the combination made the communication more information rich. However, some participants found the modalities somewhat contradictory. For example, shape-changing communicated one thing (such as, "you are in the way"), while the blinkers communicated another (such as, "I am going to turn"). This suggests that while combining modalities can enhance communication, consistency between the modalities is required to avoid providing mixed signals.

At the same time, when comparing Blinkers with Combined modality regarding communication issues, Blinkers elicited more theme-related responses from participants, despite participants noting a wider variety of issues for Combined modality. The blinker modality alone was frequently associated with timing issues and expectations for the robot to turn into the hallway. Combined modality, on the other hand, was often described as more unclear, unnatural, and confusing. However, the addition of shape-changing to the blinkers reduced the number of tags related to timing issues and expectations of turning into the hallway. Potentially indicating that the addition of shape-changing can help make the communication more rich, despite more unclear, as it is perceived as more confusing and unnatural.

7.5 Theme 5: Modality improvements/change

Scenario	**Total mentions across all tags**	**Unique tag–participant mentions**	**Total possible unique mentions 7 tags x 5 participants**
Baseline	13	10	35
Shape-changing	38	17	35
Blinkers	15	11	35
Combined modality	12	8	35

Table 17: Descriptive statistics thematic analyses on potential improvements.

This theme focuses on potential improvements suggested by participants, which can be considered for future developments.

Baseline

For the gaze-only scenario, participants noted 4 potential improvements to make the communication clearer. Key feedback included:

- Use the entire body to communicate, making it more human-like.
- Make the robot drive slower and turn earlier.
- Incorporate traffic rules, as participants expected the robot to follow standard rules.
- Incorporate blinkers or use the eyes as blinkers.

The scenario was perceived as needing improvements in driving behavior and movement speed, as the robot was described as driving too fast and turning too late. Participants also suggested adding modalities, such as blinkers, to enhance clarity.

Shape-changing

For the gaze with shape-changing scenario, participants noted 12 potential improvements to make the communication clearer. Key feedback included:

- Apply the shape-changing movement later to address timing issues.
- Make the shape-changing movement less extreme and more proportional to the intended turn.
- Incorporate additional modalities such as speech, sounds, or blinkers.
- Make the animated shape-changing movement faster
- Make the animated shape-changing movement more direct, turn intermediately after the animation.
- Let the robot give instructions to the participant
- Make the head bounce in the direction of intent
- Make the eyes blinker arrows
- Incorporate a warm up period, to make humans used to the robot its way of interaction, increasing the robots interaction speed weekly.
- Add a projector displaying a trajectory on the ground, encircling the objects that are visible for the robot.

- Incorporate social rules, such as always moving to the left in the Netherlands.
- Make the robot drive slower and turn earlier.

Much of the feedback focused on refining the shape-changing behavior to make it less extreme and more clear. Participants also emphasized improvements to the robot's driving behavior, including movement speed, and suggested integrating multiple modalities. One participant noted that adding blinkers might be too vague, as they would only communicate left or right without indicating the degree of the turn.

Blinkers

For the gaze with blinkers scenario, participants noted 3 potential improvements to make the communication clearer. Key feedback included:

- Make use of colors to differentiate between braking, accelerating, and turning.
- Implement sounds or speech to complement the blinkers.
- Make the robot turn earlier

Most of the feedback focused on the robot's driving behavior, particularly turning too late. Participants also suggested adding modalities, such as sounds or speech, and using different colors for the blinkers to indicate braking, accelerating, or turning.

Combined modality

For the gaze with blinkers and shape-changing scenario, participants noted 9 potential improvements to make the communication clearer. Key feedback included:

- Make more use of the eyes and incorporate colors.
- Make the robot drive slower or pause at times
- Incorporate arms to indicate turning or change the blinkers to arrows.
- Use a curved display with animations to indicate slight turns.
- Incorporate sounds.
- Make the robot indicate more clearly when the participant is an obstacle.
- Make use of the gaze less intense.
- Turn earlier.
- Maybe incorporate blinkers.

Feedback focused on improving the robot's driving behavior and movement speed, as well as enhancing the blinkers or adding more modalities. One notable suggestion was to incorporate blinkers, which revealed that the participant was unaware blinkers were already implemented. This is further discussed in Theme 6, General observations subsection 7.6.

7.6 Theme 6: General observations

Scenario	Total mentions across all tags	Unique tag–participant mentions	Total possible unique mentions 10 tags x 5 participants
Baseline	29	12	50
Shape-changing	61	15	50
Blinkers	43	22	50
Combined modality	58	20	50

Table 18: Descriptive statistics thematic analyses on general observations.

This theme highlights some of the general observations and interesting insights encountered during the study and interviews. These observations were tagged to capture notable findings that emerged during the interviews.

Baseline
In the gaze only scenario the general observations included:

- **Gaze modality:** All participants mentioned the gaze, indicating that the modality was noticeable enough for participants to recognize it as a form of communication. However, one participant suggested incorporating eyes for future improvements, indicating that part of the modality was overlooked. Participant 19 noted, "I think I noticed the head, but forgot about it, it was not indicating enough for a rotation," suggesting uncertainty about whether the gaze was used to communicate intent.

- **Learning over rounds:** Four out of five participants mentioned learning over rounds. They often noted that the gaze modality was unclear initially, but by the second round, they began to understand how the robot communicated, and by the third round, it was clear that the robot used gaze for directional intent.

- **Context mismatch:** One participant expected the robot to interact with them, but instead, the robot simply avoided them, creating a mismatch between expectations and reality.

Shape-changing
In the gaze and shape-changing scenario, the general observations included:

- **Shape-changing modality:** All participants noted something related to the shape-changing movements, indicating that the modality effectively communicated the robot's intent.

- **Gaze modality:** Four out of five participants mentioned the gaze, suggesting that participants focused their attention on the robot's head during the shape-changing movements.

- **Blinkers as an addition:** One participant noted that adding blinkers might be too vague, as they would only communicate left or right without indicating the degree of the turn.

- **Learning over rounds:** All participants mentioned learning over rounds. In the first round, the scenario was unclear, and participants were unsure how to respond. By the second round, the communication became clearer, and by the third round, it was evident that the robot used shape-changing for directional intent.

Blinkers
In the gaze and blinkers scenario, the general observations included:

- **Gaze modality:** All participants mentioned the gaze, suggesting that despite the blinkers being the primary modality for directional intent, participants focused their attention on the robot's head.

- **Blinkers modality:** All participants mentioned the blinkers, indicating that the modality effectively communicated directional intent. However, three participants described the blinkers as too digital, as they communicated a full turn like a vehicle would, which did not align with the robot's actual movement.

- **Delayed noticing of modalities:** Four out of five participants mentioned noticing one of the modalities only in later rounds. Two participants noted seeing the blinkers only in later rounds, as they initially focused on the gaze, while two others noticed the gaze later.

- **Overload of signals:** One participant felt overwhelmed by the number of signals being communicated simultaneously.

- **Learning over rounds:** Four out of five participants mentioned learning over rounds. The first round was unclear, but by the second round, participants understood what to expect, and by the third round, the communication was obvious, with the second modality also being noticed.

Combined modality
In the gaze, blinkers, and shape-changing scenario, the general observations included:

- **Shape-changing modality:** All participants noted something related to the shape-changing movements, indicating that the modality effectively communicated the robot's intent.

- **Gaze modality:** Only two participants mentioned the gaze, suggesting that participants were less focused on the robot's face compared to the other modalities.
- **Blinkers modality:** Only two participants mentioned the blinkers, while three participants did not mention them at all. Post interview, these three participants confirmed they had not noticed the blinkers, indicating that the shape-changing behavior may have been more distracting or dominant.
- **Contradicting modalities:** Participant 12 noted that the blinkers felt too digital and contradicted the shape-changing movement. The shape-changing communicated "you are in the way," while the blinkers signaled a turn.
- **Context mismatch:** Two participants expected the robot to interact with them, and one mentioned that in a hospital setting, they expected the robot to engage with them rather than simply avoiding them.
- **Learning over rounds:** Four out of five participants mentioned learning over rounds. The first round was unclear, but by the second round, it was considered to be clear what to expect and how to respond. The third round was more uncertain, as participants did not expect the robot to repeat the same behavior, but after observing it again, the communication became clear.

8 Discussion

In this section, we discuss the findings of the study in relation to the research questions, combining both quantitative and qualitative results. We explore how shape-changing impacts the legibility of the robot's intent, compare shape-changing to blinkers as a directional modality, examine the potential of shape-changing to enhance other modalities, and analyze how shape-changing influences participants perceptions of the robot's likability, intelligence, and safety during navigation. Additionally, we reflect on how the design met the initial requirements, outline the study's limitations, and suggest directions for future research.

8.1 RQ 1: How legible is shape-changing in comparison to gaze as directional modality?

In this subsection we will discuss research question 1, comparing the results of Shape-changing with those of gaze, to investigate whether the introduction of shape-changing enhances the legibility of the robot's navigational intent. While the quantitative results did not reveal statistically significant differences between the conditions, the descriptive statistics indicated a slight improvement in legibility scores when shape-changing was used. This suggests a potential benefit, although with a small effect size.

The qualitative findings further support this interpretation, as indicated in subsubsection 7.4.2, participants found shape-changing more intuitive and attention-grabbing than gaze alone. These insights imply that, although not conclusively proven by statistical significance, shape-changing may offer an added value in improving the legibility and interpretability of robot behavior during navigation.

8.1.1 Interpretation of quantitative and qualitative findings

The Mann-Whitney U tests revealed no statistically significant differences between Baseline and Shape-changing across legibility, confidence, or surprise ratings. While shape-changing showed slightly higher mean and median legibility scores, the effect sizes were small. This suggests a minimal, non-significant positive trend, with little impact on confidence or perceived surprise.

The qualitative findings provide deeper insights into how shape-changing influenced the participants perception. In the Baseline, participants found the gaze modality to be natural and smooth but often described it as unclear and lacking meaningful interaction. The gaze alone was perceived as insufficient to communicate the robot's intent clearly, with participants noting that the interaction occurred too early, leading to uncertainty. For example, one participant expressed uncertainty about whether the robot's gaze was meaningful or simply directional. This lack of clarity potentially contributed to the lower legibility scores in the Baseline.

In contrast, Shape-changing lead to more responses related to clear and helpful communication. Participants described the shape-changing movements as attention-grabbing and intuitive, with some interpreting the growing motion as a signal to make space or draw attention. This aligns with the intended design outlined in subsection 4.1. The behavior was based on non-verbal communication strategies observed in humans and animals, such as expansion signaling dominance or readiness [9, 5, 3] and implemented using expressive movement parameters like weight, time, and direction drawn from Laban's framework [34]. Additionally, participants interpretation of the expansion as a cue to communicate it is going to do something or asking to make way confirms the relevance of applying these naturalistic motion movements in the robots shape-changing behaviors.

The upward and outward deformation was designed to increase visibility and convey a shift in internal state, leveraging Rasmussen's dimensions of shape-changing [18], particularly "form" and "volume" changes. In the qualitative findings we saw that this growing motion was either perceived as clear or as somewhat intimidating. However, this likely made the robots intent more noticeable. As noted by [55], "large, noticeable shape changes were particularly effective in capturing attention and facilitation navigation in shared spaces." This attention-grabbing behavior may have contributed to the higher mean and median legibility scores in Shape-changing. Additionally, the incorporation of animation principles, such as anticipation and exaggeration may have further supported legibility by making the robot's actions more predictable and visually appealing [4].

While the results are not statistically significant or indicative of a large advantage, they do suggest that shape-changing can help enhance the legibility of intent. Especially as it performs closely to gaze, which is shown to improve legibility [25, 46]. This aligns with the findings of the sprout project, which states that deformations enable robots to express their internal states and potentially can be used to convey intent [54]. Future work could further investigate how refinements in movement parameters, animation principles, or other aspects of the robot's physical movement (as outlined in subsection 4.1) influence the results, to determine how each aspect influences the robot legibility. Also shape-changing could be investigated in real hospital settings or other crowded environments, which may also reveal stronger benefits where legible signals are more critical.

8.1.2 Hypotheses evaluation

- *H0: Shape-changing does not enhance the legibility of the robot's intent. No measurable differences will be observed in participant responses (legibility score, confidence, or surprise) between conditions with and without shape-changing.*
- *H1: Shape-changing enhances the legibility of the robot's intent, resulting in measurable improvements in participant responses, including higher legibility scores, greater confidence, and reduced surprise, compared to conditions without shape-changing.*

Although the quantitative analysis did not reveal statistically significant differences in legibility scores, confidence, or surprise between the Baseline and Shape-changing, the descriptive statistics and qualitative findings suggest that shape-changing may offer benefits for enhancing the legibility of the robot's intent. Specifically, Shape-changing demonstrated higher mean and median legibility scores, and participants more frequently described the shape-changing movements as clear, helpful, attention-grabbing, and intuitive.

However, as mentioned in subsubsection 7.2.2, shape-changing also introduced some negative emotional responses, such as intimidation and shocking, which may have offset its potential benefits, by increasing its surprise rate.

Overall, while shape-changing shows promise in improving legibility and confidence, its effectiveness depends on the design of the movement and mitigation of negative emotional responses, which potentially increase the surprise rate. Meaning, H0 is partially disconfirmed, and H1 is partially supported.

8.2 RQ2: How legible is shape-changing in comparison to blinkers as a directional modality?

In this subsection we will discuss research question 2, comparing the results of Shape-changing with those of Blinkers, to determine how shape-changing competes with known directional modalities like the blinkers. While the quantitative results did not show statistically significant differences between Shape-changing and Blinkers, the descriptive statistics and qualitative insights suggest that shape-changing and blinkers score alike, however shape-changing may enhance the richness of communication but has its own trade-offs compared to blinkers.

8.2.1 Interpretation of quantitative and qualitative findings

The Mann-Whitney U tests revealed no significant differences between shape-changing and blinkers across legibility, confidence, or surprise ratings. Interestingly, the descriptive statistics often contradicted each other across mean and median values. Shape-changing scored higher in mean legibility and confidence, but lower in mean surprise, whereas blinkers scored higher in median legibility and confidence, but lower in median surprise. These opposing trends, combined with small effect sizes, suggest that both modalities performed similarly overall, with no clear advantage.

As discussed in subsubsection 8.1.1, participants in Shape-changing found the shape-changing movements to be attention-grabbing and intuitive. In contrast, participants in Blinkers described the blinkers as helpful but less attention-grabbing, with some uncertainty arising from timing issues making it somewhat unclear.

A key difference between the two modalities lies in how participants interpreted the robot's intent. While shape-changing was often described as more natural and clear, with participants linking the movements to animal-like or human-like behaviors, blinkers were seen as more familiar but also caused more uncertainty due to the timing being off. For example, one participant noted that the blinkers made them think the robot would turn 90 degrees into the hallway, but the robot continued driving straight until the last moment. Suggesting that blinkers, while effective in signaling direction, may lack the richness of shape-changing, which communicates intent through a sequence of movements rather than a binary signal. However, shape-changing can also be more confusing due to its complexity.

Additionally, the qualitative feedback highlighted that shape-changing was more likely to draw attention and create a sense of anticipation. Participants described the growing motion of the robot as a signal, drawing their attention, indicating it is going to do something. Which aligned with the animation principle of anticipation, where the prepatory movement improves action predictability, by indicating upcoming behavior [4]. In contrast, blinkers, despite including anticipatory behavior, were sometimes noted later or misinterpreted, particularly when participants focused on the robot's gaze or when the timing of the blinkers did not align with the robot's movement. This aligns with the quantitative findings, where Blinkers had a higher IQR in surprise rating, potentially due to that some participants expected the robot to turn into the hallway. This expectation may come from the more immediate and digital nature of blinkers compared to shape-changing. Whatsoever, it could also be due that blinkers communicate the intent faster as shape-changing requires a sequence of frames to communicate its intent. This could have caused participants to wait for the interaction to finish before pausing, by which time the robot had already passed the hallway somewhat further. This suggests that pausing later may have made the intent clearer to participants, or that shape-changing communicates intent more richly at the cost of requiring more time to do so.

These differences highlight that while both modalities can be effective, they serve different strengths. Blinkers appear more immediate and straightforward, which can be beneficial in hospitals fast-paced environments where quick decisions are needed. while, Shape-changing on the other hand, may offer a richer and more expressive form of communication by signaling both direction and timing in a more embodied way, even if it takes slightly longer to interpret.

Overall, the comparison aligns with our expectations, as participants largely interpreted shape-changing as intended, and the more consistent performance of blinkers reflects their familiarity and simplicity. However, it was somewhat unexpected that blinkers were often only noticed in one of the later testing rounds, despite having a direct line of sight. While flashing lights like blinkers are generally considered attention grabbing cues [25]. Suggesting that even well known modalities may be overlooked depending on placement or context, indicating the importance of clearly visible cues.

Future work could explore alternative placements, such as positioning blinkers higher on the robot, for example on the head, to improve visibility and ensure they are noticed earlier. Additionally, future studies could look into the timing of blinker activation, as mismatches between signal and actual behavior led to uncertainty in some cases. Also future studies could look into the long term use case of shape-changing, to see how they compare against blinkers when the modality is more known. Lastly conducting the test without direct line of sight to uncover if shape-changing communicates more visibly compared to blinkers.

8.2.2 Hypotheses evaluation

- *H0: Shape-changing is not more legible than blinkers as a directional modality for robot navigation. There are no measurable differences in participant understanding of the robot's intended direction, legibility scores, confidence levels, or surprise when interpreting the robot's movements.*
- *H1: Shape-changing will be more legible than blinkers as a directional modality for robot navigation, resulting in improved participant understanding of the robot's intended direction, higher legibility scores, greater confidence, and reduced surprise when interpreting the robot's movements.*

Although the quantitative analysis did not show statistically significant differences in legibility scores, confidence, or surprise rate, between Shape-changing and Blinkers. The descriptive statistics revealed that Blinkers had a lower mean, but higher median legibility score and lower variability, suggesting blinkers may be more consistent in communicating intent. Qualitatively, shape-changing was described as more clear and natural, with participants linking the movements to animal-like or human-like behaviors. However, it also introduced more negative emotional responses, such as intimidation and shocking, which may have offset its potential benefits. In contrast, blinkers were perceived as familiar and straightforward but sometimes led to uncertainty due to timing issues, such as participants expecting the robot to turn into the hallway.

Overall, both modalities performed similarly, but shape-changing may offer richer communication, allowing to communicate various intents, at the cost of requiring more time to communicate, increased complexity and negative emotional responses. However, while shape-changing offers richer communication, it does not clearly outperform blinkers in terms of legibility or confidence, and it introduces higher emotional responses. Meaning, H0 is supported, and H1 is disconfirmed.

8.3 RQ3: To what extent can shape-changing contribute to improving the legibility of other modalities, such as blinkers?

In this subsection, we will discuss research question 3, comparing the results of Blinkers with Combined modality to determine whether the addition of shape-changing enhances the legibility of blinkers. While the quantitative results did not show statistically significant differences between the two scenarios, the descriptive statistics reveal some notable trends. Combined modality had a lower legibility score compared to Blinkers, but it also showed higher confidence ratings and lower surprise ratings. These findings suggest that while shape-changing may not significantly improve the speed of understanding the robot's intent, it can add richness to the communication of intent, making it feel more predictable, though it may also introduce some complexity.

8.3.1 Interpretation of quantitative and qualitative findings

The Mann-Whitney U tests revealed no significant differences between blinkers and the combined modality in legibility, confidence, or surprise. Descriptive statistics, however, showed conflicting trends: blinkers had higher mean and median legibility scores, while the combined modality showed higher mean confidence and lower mean surprise. Median surprise remained equal, but variability increased in Combined modality for legibility and decreased for surprise and confidence. These opposing patterns, paired with small to medium effect sizes, suggest that while the addition of shape-changing did not improve legibility, it may have made the robot's behavior feel more predictable and less surprising to participants.

As discussed in subsubsection 8.2.1, participants in Blinkers found the blinkers helpful but sometimes experienced uncertainty due to timing issues, particularly when the robot's movement did not align with their expectations, such as when it continued straight after the blinkers indicated a turn, leading participants to expect it would turn into the hallway. In contrast, participants in Combined modality described the combination of modalities as more attention-grabbing and natural, though some found the communication somewhat unclear or even contradictory.

A key difference between the two scenarios lies in how participants perceived the richness of the communication. While blinkers alone were effective in signaling a directional change, the addition of shape-changing provided a more natural and expressive way to communicate intent. In Combined modality, the shape-changing introduced a sense of anticipation and buildup, as intended by integrating animation principles, specifically the anticipation principle [4]. Several participants interpreted this buildup as the robot preparing to act, and often waited for the entire animation sequence to finish before reacting. This may have helped reduce surprise and increased confidence, as seen in the quantitative results for Combined modality.

However, not all participants interpreted the combined cues as working together. One participant, for instance, felt that the shape-changing signaled that they were in the way, while the blinkers suggested the robot intended to turn, leading to uncertainty. This highlights a challenge in the multi-modal approach, that if the perceived intent, direction or timing of signals do not align, the combination may reduce rather than enhance legibility.

Particularly surprising was that three participants did not mention the blinkers at all in Combined modality, despite having a direct line of sight. Normally, blinkers are considered visual cues that can capture attention effectively, especially when flashing [25], which was integrated in the behavior design. The fact that they were overlooked in the multi modal scenario suggests that the shape-changing behavior may have been too dominant or distracting. While it met the intended design goal, using naturalistic and exaggerated movement to draw attention, this may have being overdone and need to be dosed more carefully. Future implementations could consider reducing the strength of certain movement parameters or simplifying some of the expressive elements from the animation principles (as discussed in subsection 4.1).

Interestingly, despite this overshadowing effect, Combined modality still showed lower surprise ratings and more consistent confidence scores compared to blinkers alone. This suggests that while shape-changing may have dominated the visual channel, it still contributed to making the robot's intent feel more predictable and aligned with participants expectations. In that sense, the combination still improved the overall richness of the interaction. However, designers must ensure consistency between modalities, and carefully consider the balance of visual strength. If not well coordinated, expressive modalities like shape-changing can unintentionally compete with simpler but essential cues like from blinkers.

While multi modal modalities have shown to improve the acceptance of robotic services in patient care settings [32], this result suggests that modality combinations do not automatically improve legibility. In this case, the addition of shape-changing did not lead to a clear improvement over blinkers alone which was somewhat unexpected. Future work could investigate how to better coordinate timing, direction, and visual balance across modalities, and explore the impact of environmental context. Testing with adjusted parameters, such as slower shape-changing or improved blinker placement, may help discover how to fully incorporate shape-changing in a multi modal design, that enhances legibility without introducing conflicting cues.

8.3.2 Hypotheses evaluation

- *H0: Shape-changing does not enhance the legibility of blinkers when used in combination. There are no measurable differences in participant understanding of the robot's intent, legibility scores, confidence levels, or surprise compared to using blinkers alone.*

- *H1: Shape-changing will enhance the legibility of blinkers when used in combination, leading to improved participant understanding of the robot's intent, higher legibility scores, greater confidence in interpreting the robot's movements, and reduced surprise compared to using blinkers alone.*

Although the quantitative analysis did not show statistically significant differences in legibility scores, confidence, or surprise rate, between Blinkers and Combined modality. The descriptive statistics revealed that Combined modality had a lower legibility score, potentially due to the time required for shape-changing to communicate intent. In contrast, Combined modality had higher confidence ratings and lower surprise ratings, suggesting that the combination of modalities made the robot's intent feel more predictable. Qualitatively, participants described the combination of blinkers and shape-changing as more attention-grabbing and natural, with some noting that the shape-changing movements made the robot more expressive. However, some participants found the combination of modalities somewhat contradictory or unclear, as the shape-changing communicated one thing (e.g., "you are in the way"), while the blinkers signaled another (e.g., "I am going to turn"). This suggests that while combining modalities can enhance communication, consistency between the modalities is required to avoid mixed signals. Further, despite shape-changing may overshadow the blinkers by drawing the attention away from the base, it still contributed to lowering the surprise rate and improving consistency in confidence. Indicating that the addition of shape-changing to blinkers can potentially enrich communication, but it requires careful design to ensure clarity and avoid overwhelming participants.

Overall, while the combination of shape-changing and blinkers adds richness to communication, it does not improve legibility, however it does enhance confidence and surprise rate to some extend. Meaning, H0 is partially disconfirmed and H1 is partially supported.

8.4 RQ4: How does shape-changing influence participants perceptions of the robot's likability, intelligence, and safety during navigation?

The ROSAS questionnaire results reveal notable differences in how participants perceived the robot's warmth (likability), competence (intelligence), and discomfort (safety) across the scenarios. Below, we discuss each attribute in turn, highlighting key trends and significant findings.

8.4.1 Interpretation of quantitative findings

Warmth (likability)

While no statistically significant differences were found after correction, descriptive data showed clear trends: the highest warmth ratings occurred in the Combined modality scenario, followed by Shape-changing alone. Effect sizes between several conditions were large, particularly when comparing Shape-changing or Combined modality with the Baseline, suggesting that shape-changing improves perceived likability.

Competence (intelligence)

Participants rated Shape-changing alone as the most competent, with a notably higher median score than all other scenarios. Although not statistically significant, the large effect sizes, especially when comparing Shape-changing to the Baseline, suggest that shape-changing movement may enhance perceived intelligence. However, combining shape-changing with blinkers did not maintain this advantage, possibly due to conflicting or unclear cues.

Discomfort (safety)

Discomfort was the only attribute where statistically significant differences were found. Participants reported significantly more discomfort in the Shape-changing condition, compared to both the Baseline and Blinkers conditions:

- Baseline vs Shape-changing: $U = 0.5$, $Z = -2.554$, $p = 0.045$, large effect size ($r = -0.81$)

- Blinkers vs Shape-changing: $U = 1.0$, $Z = -2.440$, $p = 0.045$, medium effect size ($r = -0.47$)

These results indicate that shape-changing alone significantly increased feelings of discomfort and reduced perceived safety. However, combining it with blinkers in Combined modality appeared to mitigate this effect somewhat. Although not statistically significant, the discomfort ratings in Combined modality were lower than in Shape-changing, and the effect size for the comparison between Blinkers and Combined modality was small, suggesting the addition of blinkers helped reduce the intensity of negative emotional responses.

Summarized

These findings suggest that Shape-changing enhances perceived warmth and competence but significantly increases discomfort when used alone. The combination of blinkers and shape-changing may reaches a better balance, improving likability while keeping discomfort at more acceptable levels.

8.4.2 Interpretation of qualitative findings

Warmth (Likability)

Qualitatively, Combined modality was associated with tags like "fluffy." This aligns with its highest median warmth rating in the quantitative results. This suggests that the added expressiveness through both motion and signaling increased the robot's likability. These results are in line with the principles outlined in subsection 4.1 (Design of the shape-changing movement), which state that rounded shapes, combined with increased aesthetic appeal through arced movements, are associated with warmth and approachability [57, 4]. Similarly, shape-changing alone also received a warmth related tag, reinforcing the notion that expressive behavior can help improve likability. In contrast, the baseline condition lacked warmth related tags, reflecting its more neutral behavior.

Competence (Intelligence)

Shape-changing was also associated with descriptors like "smart" and "friendly," suggesting a positive influence on perceived intelligence. This complements findings that abrupt, direct, goal-oriented movements and expressive expansions (such as elongation to signal direction) can increase perceived decisiveness and confidence [34, 4], which participants may have interpreted as signs of competence. Interestingly, competence scored lower when shape-changing was combined with blinkers. Some participants described the robot as "unlogical," possibly due to the appearance of inconsistent or competing cues, which may have affected the interpretation of the robot's intelligence.

Discomfort (Safety)

The most significant and somewhat unexpected finding came from the emotional reaction of discomfort in the shape-changing condition. Tags like "intimidating," "shocking," and "unsafe" were frequently used. This condition also scored lowest on tags like "cute," "safe," and "nice," which matches the quantitative results. It also aligns with the design intent in someway, as the used animal behaviors, are in nature used not only to draw attention, but also to ward off threats or assert dominance, such as cats puffing up or the frill neck lizards extending their frills appear more imposing [49, 3]. Human posture, like standing tall or chest expansion, also communicates confidence or dominance [9]. Additionally, sudden, upward, and forceful expansions can further signal dominance or urgency [34, 29]. These combined cues may have unintentionally mimicked animalistic threat displays, leading to subconscious associations with territorial or intimidating behavior, leading to a feeling of intimidation, explaining the higher discomfort ratings. These findings are also in line with the SPROUT project, where participants similarly described expansion behaviors with terms like "defensive" and "intimidation" [54]. This could also have being expected in our study as it ties back to the shape-changing design principle (subsubsection 4.1.5) that expansion asserts dominance and contracting signal submission.

However, the exaggerated interpretation, leading to intimidation came unexpected for us. In contrast, prior work has shown that shape-changing can lead to a sense of awareness and psychological safety [55]. When shape-changing is combined with blinkers, these cues appeared to be perceived as less extreme. More participants described the robot as "safe" and "cute," suggesting that the addition of blinkers may have softened the emotional impact of the shape-changing behavior. This highlights how combining multiple modalities can influence emotional responses, potentially by providing additional context or clarity to the robot's intent.

Summarized

Overall we see that the qualitative findings hint that shape-changing can enhance the perceived likability and intelligence, however with a trade off in emotional comfort, due to dominant or overly exaggerated signals. Interestingly, combining shape-changing with blinkers, seem to retain some benefits while reducing the discomfort. This supports literature suggesting that combining modalities can enrich interaction and improve acceptance [32], however, only if well integrated. These insights align with the quantitative results, highlighting the trade-offs between likability, intelligence, and comfort. Based on these findings, future research could explore less dominant shape-changing patterns, possibly inspired by submissive animal cues, such as yielding, or by making use of less dominant design parameters from subsection 4.1 to avoid unintended intimidation. This could allow shape-changing to enhance warmth and competence without sacrificing user comfort. As highlighted by [55], rhythmic up-and-down motions were perceived as cheerful and lively, reinforcing the importance of thoughtful design. Additionally, studying long-term exposure and learning effects may reveal how user perceptions evolve as they become more familiar with such cues.

8.4.3 Hypotheses evaluation

- *H0: Shape-changing will not lead to higher ratings of warmth (likability), competence (perceived intelligence), and safety (reduced discomfort) during robot navigation compared to a baseline without shape-changing, nor compared to blinkers without shape-changing.*
- *H1: Shape-changing will lead to higher ratings of warmth (likability), competence (perceived intelligence), and safety (reduced discomfort) during robot navigation, compared to a baseline without shape-changing and compared to blinkers without shape-changing.*

Warmth (Likability)

Combined modality scored the highest in warmth, with a median rating of 5.5, followed by Shape-changing with a median of 4.833. Blinkers and the Baseline scored lower, with medians of 4.333 and 3.0, respectively. Qualitatively, participants described Combined modality as "fluffy", while Shape-changing was associated with tags like "happy." This suggests that shape-changing, especially when combined with blinkers, enhances the robot's likability. However, the Baseline lacked positive warmth related tags, potentially explaining its low warmth score.

Competence (Intelligence)

Shape-changing scored the highest in competence, with a median rating of 6.6667, indicating that participants perceived the robot as more intelligent when shape-changing was used alone. Blinkers followed with a median of 4.6667, while Combined modality and the Baseline scored lower, with medians of 4.5 and 4.3333, respectively. Qualitatively, participants described Shape-changing as "smart" and "friendly," while Combined modality was sometimes perceived as "unlogical." This suggests that while shape-changing alone enhances perceptions of intelligence, combining it with blinkers does not further improve this perception and may even slightly reduce it.

Discomfort (Safety)

The Baseline had the lowest discomfort rating (median = 1.1667), suggesting that participants felt safest in this scenario. Blinkers followed with a median of 1.6667, while Combined modality and Shape-changing had higher discomfort ratings, with medians of 2.0 and 2.6667, respectively. Qualitatively, Shape-changing was associated less with tags like "cute" and "not scary" and more with tags like "intimidating" and "shocking," while Combined modality was described as "safe" and "cute." This indicates that shape-changing alone increases discomfort, but combining it with blinkers reduces this effect, making the interaction feel safer and more predictable.

Summarized

Overall, when comparing the two scenario pairs, Shape-changing and Combined modality outperformed Baseline and Blinkers in 7 out of 12 measured dimensions. Specifically:

- Warmth: Combined modality and Shape-changing scored highest (both qualitatively and quantitatively).
- Competence: Shape-changing led (both measures), while Combined modality scored lower.
- Discomfort: Combined modality ranked highest qualitatively, whereas Shape-changing scored lower on both measures.

Indicating, H0 is partially disconfirmed, and H1 is partially supported.

8.5 Requirements accomplished

This subsection explains how well the study met the requirements as mentioned in Appendix I. While most requirements were met, some limitations arose during implementation.

- Robot legibility and communication effectiveness
  - The study successfully measured the legibility of the robot's intent through shape-changing using a combination of quantitative metrics (response time, accuracy, confidence and surprise rating) and qualitative feedback (interviews). Fully meeting the requirement.
- Modality isolation, comparison and integration
  - The 2x2 experimental design allowed for the the isolation of the shape-changing condition and comparison of shape-changing with traditional modalities like blinkers. However the gaze modality is not completely isolated as the gaze modality is always present in the conditions, due to the robot having a face, leading to a partially met requirement.
- Iterative design and feedback
  - The study incorporated iterative design and participant feedback, leading to improvements in the robot's movements and the experimental setup. leading to a fully met requirement. However, despite being able to communicate intent, additional iterations could further improve the modalities.
- Ecological validity (hospital environment)
  - The VR environment successfully simulated a hospital hallway, with hospital-like features such as crosswalks, hospital like objects and ambient sounds, fully meeting the requirement. However, future studies could explore ways to improve the environment even further, with for example tactile feedback or smell.
- Participant understanding and perception
  - The study successfully measured participants perceptions of the robot's likability, intelligence, and safety using the ROSAS questionnaire and thematic analysis of interview data. Fully meeting the requirement.
- Data analysis and statistical power
  - The study collected data from 20 participants, using a combination of quantitative metrics (response time, accuracy, confidence and surprise ratings) and qualitative feedback (interviews). Non-parametric statistical tests (e.g., Mann-Whitney U test) were used to analyze the data, and corrections (e.g., Bonferroni) were applied to control for multiple comparisons. However, the requirement was partially met as the study provided some exploratory insights, but lacked the statistical power for definitive conclusions.
- Technical feasibility of robotic movement design
  - The study demonstrated that shape-changing movements could be implemented in a real-world robot prototype and translated into a VR environment. The requirement is partially met, as the shape-changing was integrated. However, further hardware improvements are required for real-world applicability.
- Participant diversity and inclusivity
  - The study included a wide participant pool, with participants of varying ages, genders, and levels of experience with robots. However, the requirement was partially met as the participant pool was small, it was not inclusive enough to meet a full range of potential users.
- Ethical and safety considerations
  - The study followed the ethical guidelines from the University's ethics committee, with all participants providing informed consent and data being anonymized. Participants were informed of their right to withdraw at any time, and the VR environment was pre-tested and kept as short as possible to minimize the risk of motion sickness or discomfort. Fully meeting the requirement.

8.6 Limitations

This study has several limitations that should be mentioned. First, the small sample size limited the ability to generalize results, as the study did not include all cultural and demographic backgrounds. Additionally, the study was exploratory rather than confirmatory, meaning it aimed to explore potential effects rather than establish definitive conclusions.

The use of a VR environment provided a controlled and immersive setting but did not fully replicate real-world interactions, which may had an impact on the ecological validity of the results. Similarly, limited contextual testing was conducted, as the study focused only on a hospital setting. Testing in other environments could provide additional insights into the effectiveness of shape-changing cues.

Technical constraints further impacted the study. The physical robot used in prototyping could either drive or shape-change, but not both simultaneously, limiting the ability to test real-world interactions. Additionally, a learning effect may have influenced participants performance, as they experienced multiple rounds of testing, potentially improving their performance over time. Moreover, the study only compared and combined gaze, shape-changing, and blinkers, leaving out other modalities like sound or speech, which could have provided additional insights.

Feedback from participants, as seen in the theme modality improvements, also suggested that further iterations are needed to optimize design choices. Participants proposed various improvements or changes to enhance clarity, timing, and movement integration, indicating that the current implementation could be further improved.

Finally, time constraints limited the depth and scope of the research, and the study focused on short-term interactions, which may not fully capture how people would engage with the robot over extended periods in real-world settings. These limitations, provide areas for future research to build upon.

8.7 Future work

While this study provided insights into shape-changing as a communication modality for robots navigating hospitals, several areas require further exploration.

First, future studies should aim to include a larger and more diverse sample size to improve the generalizability of the findings. Expanding the participant pool to include different cultural and demographic backgrounds would help to understand potential differences in perception and interpretation. Although the study was exploratory, a bigger sample size could help draw more definitive conclusions.

Secondly, future work should explore the use of shape-changing in environments beyond hospitals, such as offices, restaurants, or homes, to assess its effectiveness in different settings. This would help determine whether the findings from this study are context-specific or applicable to a broader range of scenarios.

The technical capabilities of the robot should also be improved. The current hardware constrained the ability to drive and shape-change simultaneously, limiting applicability. Future studies could implement more advanced prototypes capable of performing both. Additionally, despite requiring more time, incorporating dynamic participant movement rather than a static perspective in VR could potentially provide a more realistic and engaging setting. Considering the hardware, future studies could also explore the complete isolation of the shape-changing modality by integrating the modality on a faceless robot and adding a shape-changing only condition.

The study also highlighted the need for further exploration of modality combinations. While this research focused on gaze, shape-changing, and blinkers, future work could investigate the integration of additional modalities, such as sound, speech, or other visual cues. This could provide a more rich understanding of how multi-modal communication can enhance robot legibility and user experience.

Moreover, the feedback from participants suggested that the current design could be further optimized. Future iterations should focus on refining the timing, movement patterns, and blinker placement to improve visibility and legibility. For instance, blinkers could be placed higher on the robot (such as near the head) to increase appearance. Adjusting the timing of blinker activation is also important, as mismatches between signaling and movement led to uncertainty in some cases.

Future work could explore how to better coordinate the timing, direction, and visual balance across modalities. For example, adjusting parameters such as slower shape-changing, clearer anticipation phases, or synchronized blinkers may help shape-changing fit better within a multi-modal design, enhancing legibility without introducing conflicting cues. Additionally, conducting tests where the robot is not in the direct line of sight may reveal whether shape-changing is more visible and legible compared to blinkers alone.

Long-term interaction studies are another area for future research. While this study focused on short-term interactions, understanding how users engage with shape-changing robots over extended periods would provide insights into how familiarity with shape-changing impacts trust and legibility over time. This could include exploring how repeated exposure affects user comfort, expectations, and interpretation of intent compared to more familiar cues like blinkers.

The learning effect observed in this study also suggests that shape-changing may have a steeper learning curve than blinkers, but may offer richer, more nuanced communication once understood. Future research should investigate how quickly users adapt to shape-changing cues and whether improvements in movement design can help make shape-changing more intuitive.

Finally, the impact of specific shape-changing movement characteristics requires further research. In particular, examining how the frameworks outlined in subsection 4.1, influences the interpretation, which could help to understand how each framework shapes the interpretation of shape-changing movements. For example future work could investigate how refinements in timing, weight, fluidity, or exaggeration, influences the legibility. In particular, less dominant shape-changing patterns inspired by submissive animal cues (such as, contraction, yielding, soft movements) may help reduce unintended intimidation and improve perceived warmth and competence.

Key points for future researchers developing shape-changing robots

Based on the findings of this study and the frameworks applied in the design process, the following design-oriented guidelines are offered for researchers and developers working on shape-changing behaviors in robots.

A valuable starting point for future researchers are the frameworks presented in subsection 4.1 which provide a valuable foundation for developing shape-changing behaviors. These frameworks build upon natural human and animal behaviors, animation principles, shape associations, and expressive movement parameters. While it remains unclear which of these frameworks most strongly informed the legibility of the robot and participants perceptions, this layered method offers a rich starting point to develop shape-changing behavior.

We believe that grounding intended communication in natural, familiar behaviors, such as humans and animal behaviors can help establish a common ground with users. This natural base can be enhanced using animation principles (such as, anticipation or exaggeration) to make the intent more visually clear. Shape associations (such as, round = safe, sharp = urgent) influence perception, and movement parameters (such as, speed, direction, weight) can shape perception and emotional tone. Different combinations of these elements may lead to varied interpretations, emphasizing the need to align design expression with communicative goals.

However, a key takeaway from our study is the risk of overdoing shape-changing. While expressive movements can attract attention, overly large, sudden, or intense deformations may be interpreted as intimidating. Researchers should be careful not to prioritize expressiveness at the cost of user comfort or perceived safety. Fine tuning movement through early user feedback is important to meet a closer balance between legibility and comfort.

Additionally, if shape-changing is combined with other modalities (such as, blinkers, gaze, or sound), synchronization is important. Our findings showed that mismatched or contradictory cues reduced confidence and caused uncertainty. To improve legibility and trust, multi-modal directional cues must reinforce one another through synchronized timing, direction, and intensity. For instance, in some conditions of our study, participants were uncertain when blinkers indicated a turn before the shape-change had fully finished. This suggests that asynchronous timing between modalities may weaken legibility. Also in our multi modal scenario shape-changing was considered too intens, drawing attention away from the blinkers, thus instead of the modalities complementing each other, they would compete with each other.

By addressing these areas, future research can potentially further enhance shape-changing for robot navigation, improving its effectiveness as a communication tool for human-robot interaction.

9 Conclusion

This thesis explored the potential of shape-changing as a modality to enhance robot legibility in social navigation, particularly in hospital environments. Through a combination of quantitative and qualitative analyses, the study investigated the research question, "How can we use shape-changing to enhance robot legibility for navigation in social hallways?" This included examining how shape-changing impacts the legibility of a robot's intent, how it compares to traditional modalities like blinkers, how it functions in multi-modal combinations, and how it influences user perceptions of the robot's likability, intelligence, and safety.

The findings suggest that shape-changing has the potential to improve the legibility of a robot communicating its directional intent. Participants described shape-changing as clear and helpful, often linking it to animal-like or human-like behaviors. However, shape-changing also introduced some trade-offs, such as increased feelings of intimidation and shocking, which could offset its benefits. While the quantitative results did not show statistically significant differences, the qualitative insights and descriptive statistics indicate that shape-changing can enhance the richness of communication, and when combined with other modalities like blinkers, it also showed potential to increase user confidence and reduce the surprise rate.

When compared to blinkers, shape-changing performed similarly in terms of legibility but offered a more expressive and natural way to communicate intent. Blinkers, while familiar and straightforward, sometimes led to uncertainty due to timing issues, such as participants expecting the robot to turn into a hallway. Shape-changing, on the other hand, provided a more analog and information-rich communication method, though it required more time to convey intent and could be perceived as more complex.

The combination of shape-changing with blinkers showed promise in improving user confidence and reducing surprise rate, suggesting that multi-modal communication can make the robot's intent feel more predictable. However, the combination also introduced some uncertainty, as the modalities sometimes sent mixed signals. This underscores the importance of careful design to ensure consistency and clarity when integrating multiple communication modalities.

Considering user perceptions, shape-changing enhanced the robot's likability and intelligence when used with gaze alone. However, it also increased feelings of discomfort, as participants found the movements intimidating or shocking. Combining shape-changing with blinkers helped to minimize these negative effects, creating a more balanced interaction that was perceived as safer and more likable. However, the gaze-only scenario remained the safest, indicating that simpler modalities may be preferable for comfort.

Overall, this study contributes to the field of human-robot interaction by demonstrating the potential of shape-changing as a novel communication modality to enhance robot legibility. These findings also reflect the motivation outlined in the background, where the need was identified for innovative communication strategies that integrate legibility with intuitiveness, information richness, and clear visibility, particularly in complex environments like hospitals. Shape-changing appears to fulfill many of these criteria: it offers intuitive, information rich, and visually attention grabbing signals, although its effectiveness depends strongly on how it is designed and integrated. The study further touches upon the importance of multi-modal approaches, showing that combining shape-changing with modalities like blinkers can improve information richness and predictability of the robots intent. Additionally this study based the interaction on a real robot and made the transition to a VR environment, showing that nowadays, shape-changing is possible in real world robots, while providing an immersive setting for testing robot legibility.

As highlighted in the future work, upcoming research should build on this study by expanding participant diversity and testing shape-changing in more varied and dynamic environments, including real hospital contexts. Further exploration of additional modalities, such as sound or speech, and refined combinations of shape-changing with blinkers or gaze may help improve multi-modal coordination. This includes optimizing timing, directional consistency, and cue balance to prevent conflicting signals. In particular, refinements in movement parameters like speed, weight, and fluidity, as well as the use of less dominant, more yielding patterns inspired by animal behavior, could help reduce unintended intimidation while maintaining expressiveness. Investigating how these individual design frameworks, outlined in subsection 4.1, influence interpretation can help future developers understand which elements most directly improves legibility. Moreover, long-term interaction studies and repeated exposure will be of importance to assess how users adapt to shape-changing over time and whether it offers lasting benefits over more familiar modalities like blinkers.

In conclusion, shape-changing shows promise as a communication modality for improving robot legibility, but its effectiveness depends on its design and integration, especially when combined with other modalities. By addressing the trade-offs and limitations identified during this study, future research can potentially further advance the development of socially compatible robots that are legible in their communication, leading to a greater acceptance of robots.

References

[1] Michael Argyle. “Non-verbal communication in human social interaction”. In: *Non-verbal communication* 2.1 (1972).

[2] MW Fox. “Social dynamics of three captive wolf packs”. In: *Behaviour* 47.3-4 (1973), pp. 290–301.

[3] Richard Shine. “Function and evolution of the frill of the frillneck lizard, Chlamydosaurus kingii (Sauria: Agamidae)”. In: *Biological Journal of the Linnean Society* 40.1 (1990), pp. 11–20.

[4] Frank Thomas and Ollie Johnston. “The illusion of life: Disney animation”. In: *(No Title)* (1995).

[5] M Petrie, P Cotgreave, and I Stewart. “Variation in the train morphology of peacocks (Pavo cristatus)”. In: *Journal of Zoology* 238.2 (1996), pp. 365–371.

[6] B. Salemi, Wei-Min Shen, and P. Will. “Hormone-controlled metamorphic robots”. In: *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*. Vol. 4. 2001, 4194–4199 vol.4. DOI: 10.1109/ROBOT.2001.933273.

[7] Brian Scassellati. “Theory of mind for a humanoid robot”. In: *Autonomous Robots* 12 (2002), pp. 13–24.

[8] Vinni Madsen et al. “Bimodal signaling of a sexually selected trait: gular pouch drumming in the Magnificent Frigatebird”. In: *The Condor* 106.1 (2004), pp. 156–160.

[9] Pablo Briñol, Richard E Petty, and Benjamin Wagner. “Body posture effects on self-evaluation: A self-validation approach”. In: *European Journal of Social Psychology* 39.6 (2009), pp. 1053–1064.

[10] Cindy L Bethel and Robin R Murphy. “Review of human studies methods in HRI and recommendations”. In: *International Journal of Social Robotics* 2.4 (2010), pp. 347–359.

[11] Michael Tomasello. *Origins of human communication*. MIT press, 2010.

[12] Richard Bloss. “Mobile hospital robots cure numerous logistic needs”. In: *Industrial Robot: An International Journal* 38.6 (2011), pp. 567–571.

[13] Michael J. Gielniak and Andrea L. Thomaz. “Generating anticipation in robot motion”. In: *2011 RO-MAN*. 2011, pp. 449–454. DOI: 10.1109/ROMAN.2011.6005255.

[14] Leila Takayama, Doug Dooley, and Wendy Ju. “Expressing thought: improving robot readability with animation principles”. In: *Proceedings of the 6th International Conference on Human-Robot Interaction*. HRI ’11. Lausanne, Switzerland: Association for Computing Machinery, 2011, 69–76. ISBN: 9781450305617. DOI: 10.1145/1957656.1957674. URL: https://doi.org/10.1145/1957656.1957674.

[15] Paul Watzlawick, Janet Beavin Bavelas, and Don D Jackson. *Pragmatics of human communication: A study of interactional patterns, pathologies and paradoxes*. WW Norton & Company, 2011.

[16] Christina Lichtenthäler, Tamara Lorenzy, and Alexandra Kirsch. “Influence of legibility on perceived safety in a virtual human-robot path crossing task”. In: *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. 2012, pp. 676–681. DOI: 10.1109/ROMAN.2012.6343829.

[17] Christina Lichtenthäler et al. “Increasing perceived value between human and robots — Measuring legibility in human aware navigation”. In: *2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. 2012, pp. 89–94. DOI: 10.1109/ARSO.2012.6213405.

[18] Majken K Rasmussen et al. “Shape-changing interfaces: a review of the design space and open research questions”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, pp. 735–744.

[19] Tiago Ribeiro and Ana Paiva. “The illusion of robotic life: principles and practices of animation for robots”. In: *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI ’12. Boston, Massachusetts, USA: Association for Computing Machinery, 2012, 383–390. ISBN: 9781450310635. DOI: 10.1145/2157689.2157814. URL: https://doi.org/10.1145/2157689.2157814.

[20] Anca Dragan and Siddhartha Srinivasa. “Generating legible motion”. In: (2013).

[21] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. “Legibility and predictability of robot motion”. In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2013, pp. 301–308. DOI: 10.1109/HRI.2013.6483603.

[22] Maha Salem et al. “To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability”. In: *International Journal of Social Robotics* 5 (2013), pp. 313–323.

[23] Anca Dragan and Siddhartha Srinivasa. “Familiarization to robot motion”. In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 2014, pp. 366–373.

[24] Nicole Mirnig et al. “Screen feedback in human-robot interaction: How to enhance robot expressiveness”. In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE. 2014, pp. 224–230.

[25] Alyxander David May, Christian Dondrup, and Marc Hanheide. “Show me your moves! Conveying navigation intention of a mobile robot to humans”. In: *2015 European Conference on Mobile Robots (ECMR)*. IEEE. 2015, pp. 1–6.

[26] GEORGE B. RABB, JEROME H. WOOLPY, and BENSON E. GINSBURG. “Social Relationships in a Group of Captive Wolves”. In: *American Zoologist* 7.2 (Aug. 2015), pp. 305–311. ISSN: 0003-1569. DOI: 10.1093/icb/7.2.305. eprint: https://academic.oup.com/icb/article-pdf/7/2/305/247303/7-2-305.pdf. URL: https://doi.org/10.1093/icb/7.2.305.

[27] Minae Kwon, Malte F Jung, and Ross A Knepper. “Human expectations of social robots”. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2016, pp. 463–464.

[28] Christina Lichtenthäler and Alexandra Kirsch. “Legibility of robot behavior: A literature review”. In: (2016).

[29] Haodan Tan et al. “Happy Moves, Sad Grooves: Using Theories of Biological Motion and Affect to Design Shape-Changing Interfaces”. In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. DIS ’16. Brisbane, QLD, Australia: Association for Computing Machinery, 2016, 1282–1293. ISBN: 9781450340311. DOI: 10.1145/2901790.2901845. URL: https://doi.org/10.1145/2901790.2901845.

[30] Colleen M Carpinella et al. “The robotic social attributes scale (RoSAS) development and validation”. In: *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*. 2017, pp. 254–262.

[31] Denise Hebesberger et al. “A long-term autonomous robot at a care hospital: A mixed methods study on social acceptance and experiences of staff and older adults”. In: *International Journal of Social Robotics* 9.3 (2017), pp. 417–429.

[32] Alessandro Di Nuovo et al. “The multi-modal interface of Robot-Era multi-robot services tailored for the elderly”. In: *Intelligent Service Robotics* 11 (2018), pp. 109–126.

[33] Aljoscha Pörtner et al. “The power of color: A study on the effective use of colored light in human-robot interaction”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2018, pp. 3395–3402.

[34] Hamish Tennent, Dylan Moore, and Wendy Ju. “Character Actor: Design and Evaluation of Expressive Robot Car Seat Motion”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1 (Jan. 2018), pp. 1–23. DOI: 10.1145/3161407.

[35] Yuhan Hu and Guy Hoffman. “Using Skin Texture Change to Design Emotion Expression in Social Robots”. In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2019, pp. 2–10. DOI: 10.1109/HRI.2019.8673012.

[36] Carlos Lobato et al. “Human-robot dialogue and Collaboration for social navigation in crowded environments”. In: *2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE. 2019, pp. 1–6.

[37] John W Romanishin, John Mamish, and Daniela Rus. “Decentralized control for 3D M-blocks for path following, line formation, and light gradient aggregation”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 4862–4868.

[38] Trenton Schulz, Jim Torresen, and Jo Herstad. “Animation Techniques in Human-Robot Interaction User Studies: A Systematic Literature Review”. In: *J. Hum.-Robot Interact.* 8.2 (2019). DOI: 10.1145/3317325. URL: https://doi.org/10.1145/3317325.

[39] Andrea Bonarini. “Communication in human-robot interaction”. In: *Current Robotics Reports* 1 (2020), pp. 279–285.

[40] Carlos A Cifuentes et al. “Social robots in therapy and care”. In: *Current Robotics Reports* 1 (2020), pp. 59–74.

[41] Sara Cooper et al. “ARI: The social assistive robot and companion”. In: *2020 29th IEEE International conference on robot and human interactive communication (RO-MAN)*. IEEE. 2020, pp. 745–751.

[42] Eui Jun Hwang et al. “Demonstration of hospital receptionist robot with extended hybrid code network to select responses and gestures”. In: *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE. 2020, pp. 8013–8018.

[43] Thomas B Sheridan. “A review of recent research in social robotics”. In: *Current Opinion in Psychology* 36 (2020). Cyberpsychology, pp. 7–12. ISSN: 2352-250X. DOI: `https://doi.org/10.1016/j.copsyc.2020.01.003`. URL: `https://www.sciencedirect.com/science/article/pii/S2352250X2030004X`.

[44] Carina Soledad González-González, Verónica Violant-Holz, and Rosa Maria Gil-Iranzo. “Social robots in hospitals: a systematic review”. In: *Applied Sciences* 11.13 (2021), p. 5976.

[45] Carina Soledad González-González, Verónica Violant-Holz, and Rosa Maria Gil-Iranzo. “Social Robots in Hospitals: A Systematic Review”. In: *Applied Sciences* 11.13 (2021). ISSN: 2076-3417. DOI: `10.3390/app11135976`. URL: `https://www.mdpi.com/2076-3417/11/13/5976`.

[46] Blake Holman et al. “Watch where you’re going! gaze and head orientation as predictors for social robot navigation”. In: *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE. 2021, pp. 3553–3559.

[47] Mikaela Law et al. “Case studies on the usability, acceptability and functionality of autonomous mobile delivery robots in real-world healthcare settings”. In: *Intelligent Service Robotics* 14.3 (2021), pp. 387–398.

[48] Velvetina Lim, Maki Rooksby, and Emily S Cross. “Social robots on a global stage: establishing a role for culture during human–robot interaction”. In: *International Journal of Social Robotics* 13.6 (2021), pp. 1307–1333.

[49] Nicolae Sfetcu. “About Cats”. In: *Published by Nicolae Sfetcu* (2021).

[50] Akihiro Maehigashi, Takahiro Tsumura, and Seiji Yamada. “Effects of Beep-Sound Timings on Trust Dynamics in Human-Robot Interaction”. In: *International Conference on Social Robotics*. Springer. 2022, pp. 652–662.

[51] T Höfker. “Musical utterances to evoke empathy and prosocial behavior toward a hospital robot”. MA thesis. University of Twente, 2023.

[52] Andy Field. *Discovering statistics using IBM SPSS statistics*. Sage publications limited, 2024.

[53] Amy Koike, Michael Wehner, and Bilge Mutlu. “Sprout: Designing Expressivity for Robots Using Fiber-Embedded Actuator”. In: *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. HRI ’24. ¡conf-loc¿, ¡city¿Boulder¡/city¿, ¡state¿CO¡/state¿, ¡country¿USA¡/country¿, ¡/conf-loc¿: Association for Computing Machinery, 2024, 403–412. ISBN: 9798400703225. DOI: `10.1145/3610977.3634983`. URL: `https://doi.org/10.1145/3610977.3634983`.

[54] Amy Koike, Michael Wehner, and Bilge Mutlu. “Sprout: Designing Expressivity for Robots Using Fiber-Embedded Actuator”. In: *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 2024, pp. 403–412.

[55] Bob R. Schadenberg et al. “Goo-y: A Robot’s Shape-Changing Capability for Fast, Non-Anthropomorphic Communication with People (and Animals)”. In: *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA ’25. New York, NY, USA: Association for Computing Machinery, 2025. ISBN: 9798400713958. DOI: `10.1145/3706599.3719793`. URL: `https://doi.org/10.1145/3706599.3719793`.

[56] Anne Bogart and T Landauthe. “The Viewpoints Book: A Practical Guide to Viewpoints and Composition. 2005”. In: *New York: Theatre Communications Group* ().

[57] *Tips Techniques shape language*. `https://www.waltdisney.org`. Accessed: 2024-28-5.

10 Appendix

A Design decisions

	Pros	Cons
Media format: Video	Simple to develop, Minimal test setup	Less immersive, Distracting environmental cues, Lacks the 3D aspect, potentially reducing the legibility of the movement overall
Media format: Virtual reality (VR)	More realistic (easier to perceive robot intent naturally), Increased engagement, Stable environment	Harder to develop, Bigger test setup, Possible discomfort

Figure 22: Video VS VR

	Pros	Cons
First person front view	Realistic situation as the person actually is required to look and respond to the robot, Increased engagement (forcing the participant to pause in time)	Limited view (only the robots front view, potentially communicating something differently from what is to be expected, as debt can be missing)
First person side view	Complete view of the movement	Reduced immersion (as the participant is more detached from the situation), No pressure for the participant to respond as the robot will not directly interact with the participant itself

Figure 23: Front view VS side view

	Pros	Cons
Movement: Moving around	Participants feel more engaged leading to more natural movements allowing to detect small movements. Less bias as task is easier given	Tracking is more though, also excludes participants with a more limited freedom of movement. Harder to analyse. Can not easily measure response time
Movement: Pause button	Easier to develop, more precise timing, easier to analyse, available to many	Less engaged, can't detect doubt or small movements. Can bias research on making the participant focus on the robot

Figure 24: Static VS freely

	Pros	Cons
Pause video before intent	Lower cognitive load, Stopping the video before intent is clear the focus is more on predicting the intent	May miss individual difference in perceiving the movement (some need more time or more context while others less) Less high-resolution data as it is correct or incorrect
Participant pauses video	Richer data (providing higher resolution, combining accuracy and response time), Can measure individual differences in the understanding of movement.	Might increase cognitive load, requires calculation time for combining response time and accuracy in one score.

Figure 25: Before intent VS participant decide to pause

B AI usage disclaimer

This thesis was developed with the assistance of AI-based tools, including GPT, to help with rewriting text and providing structural and content related advice. The AI was used purely as a support tool for improving readability, while making sure that all research, analysis, and conclusions remain our original work. Any decisions considering the content, argumentation, and final wording come from the author.

C Boxplots

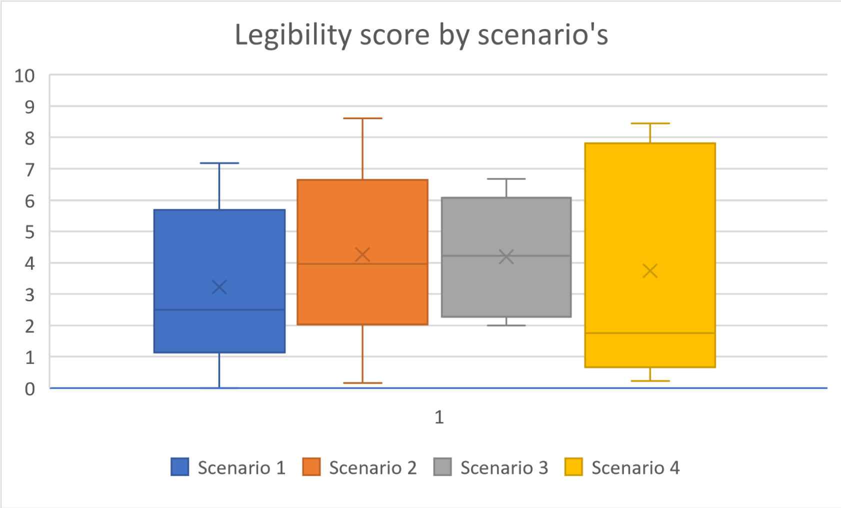


Figure 26: Boxplot of legibility scores for all scenarios.

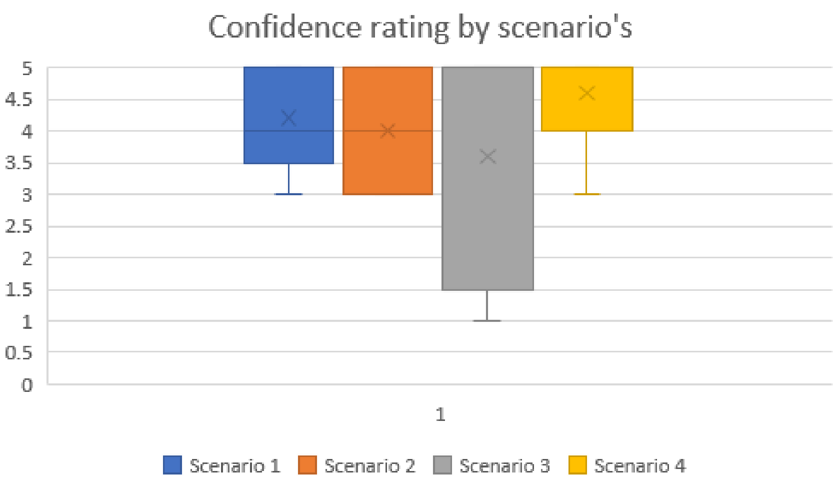


Figure 27: Boxplot of confidence ratings for all scenarios.

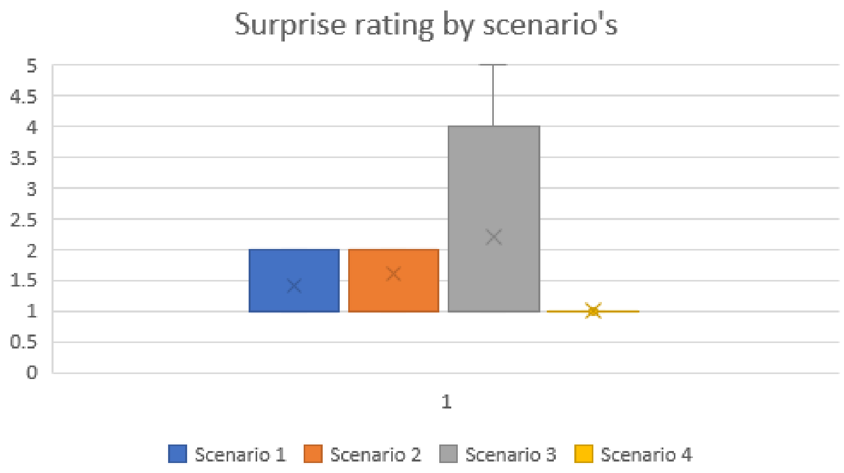


Figure 28: Boxplot of surprise ratings for all scenarios.

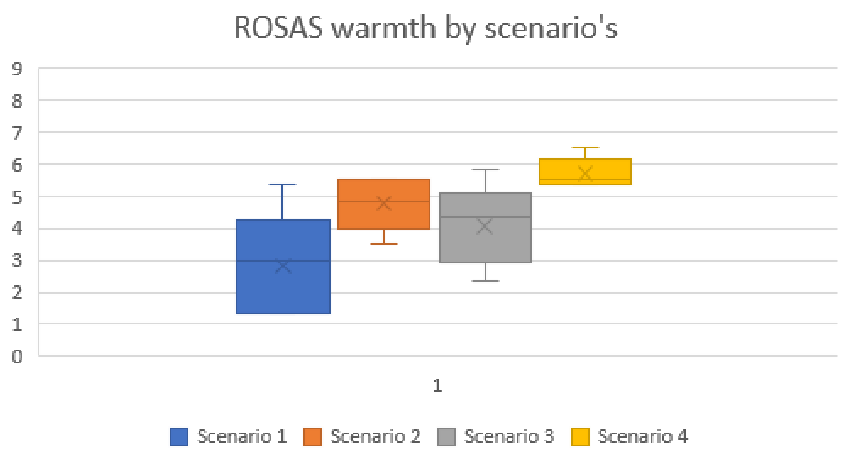


Figure 29: Boxplot of warmth ratings for all scenarios.

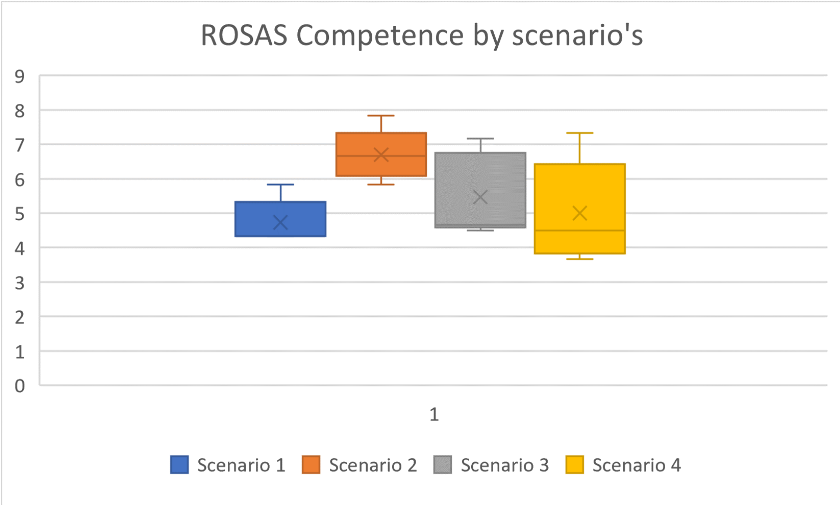


Figure 30: Boxplot of competence ratings for all scenarios.

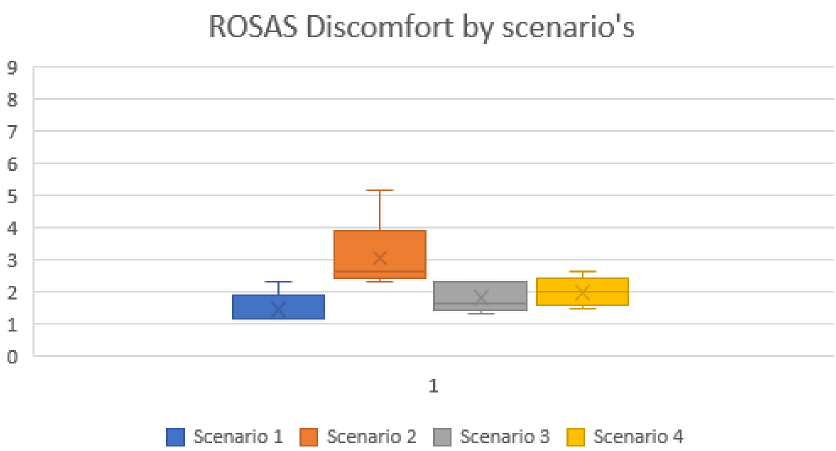


Figure 31: Boxplot of discomfort ratings for all scenarios.

D Thematic analyses

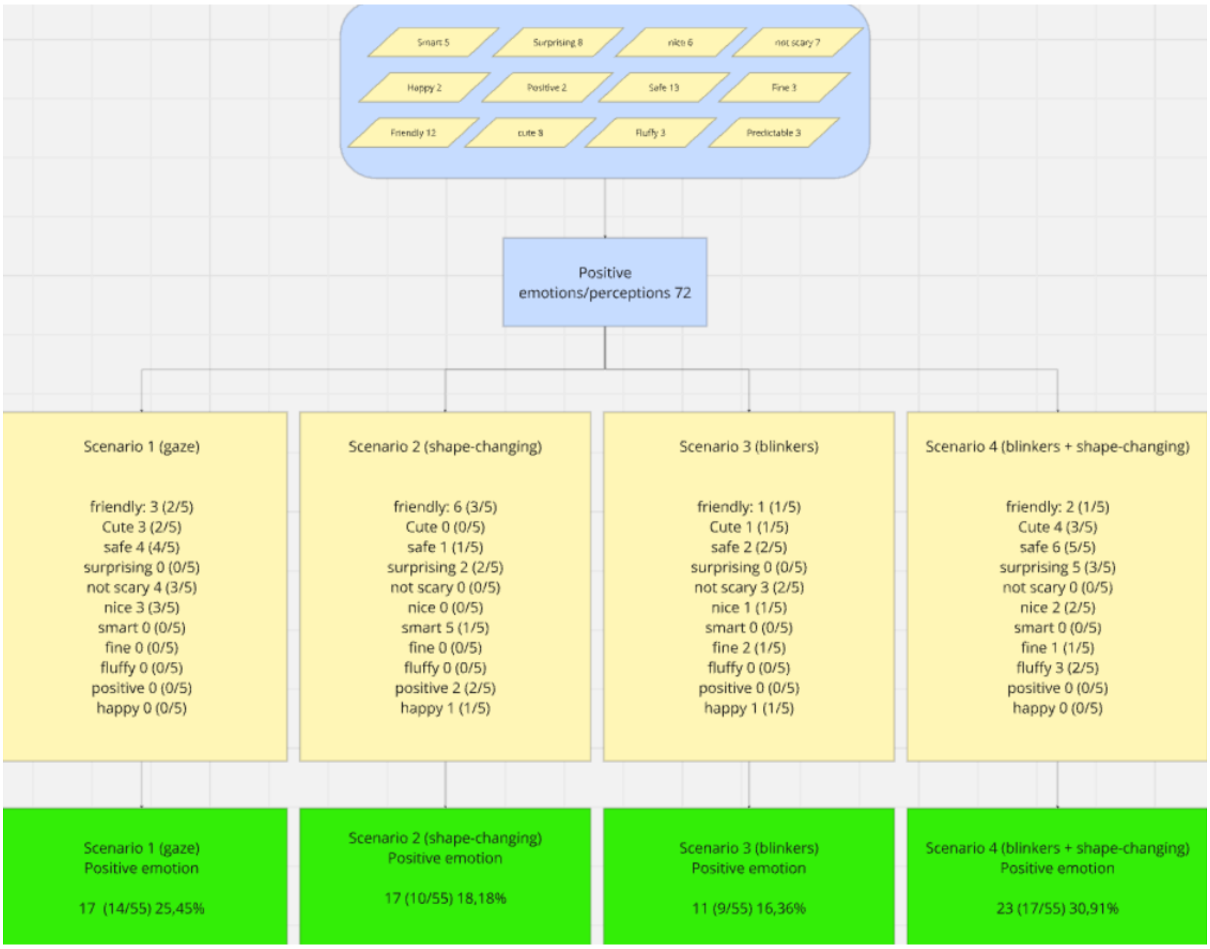


Figure 32: Positive emotions thematic analyses.

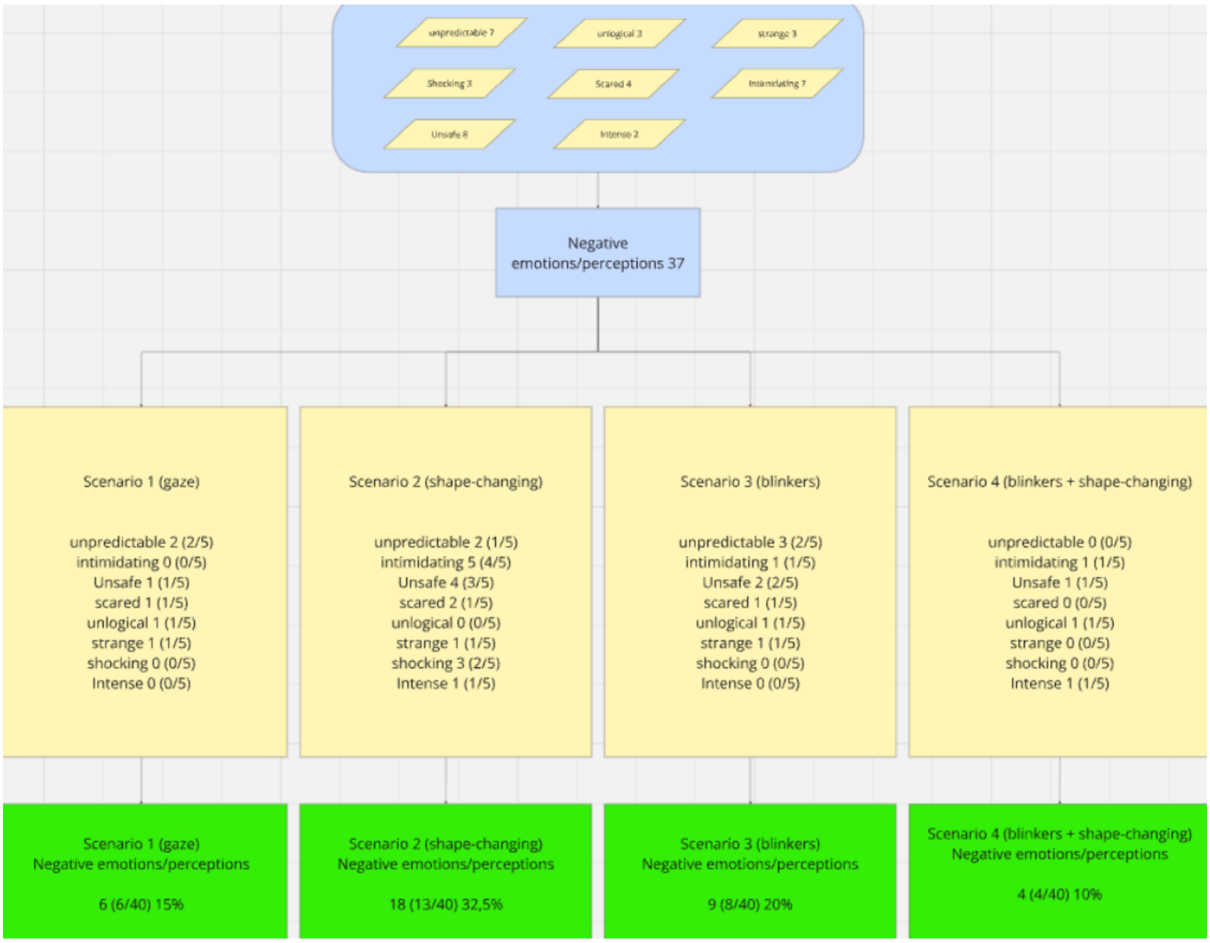


Figure 33: Negative emotions thematic analyses.

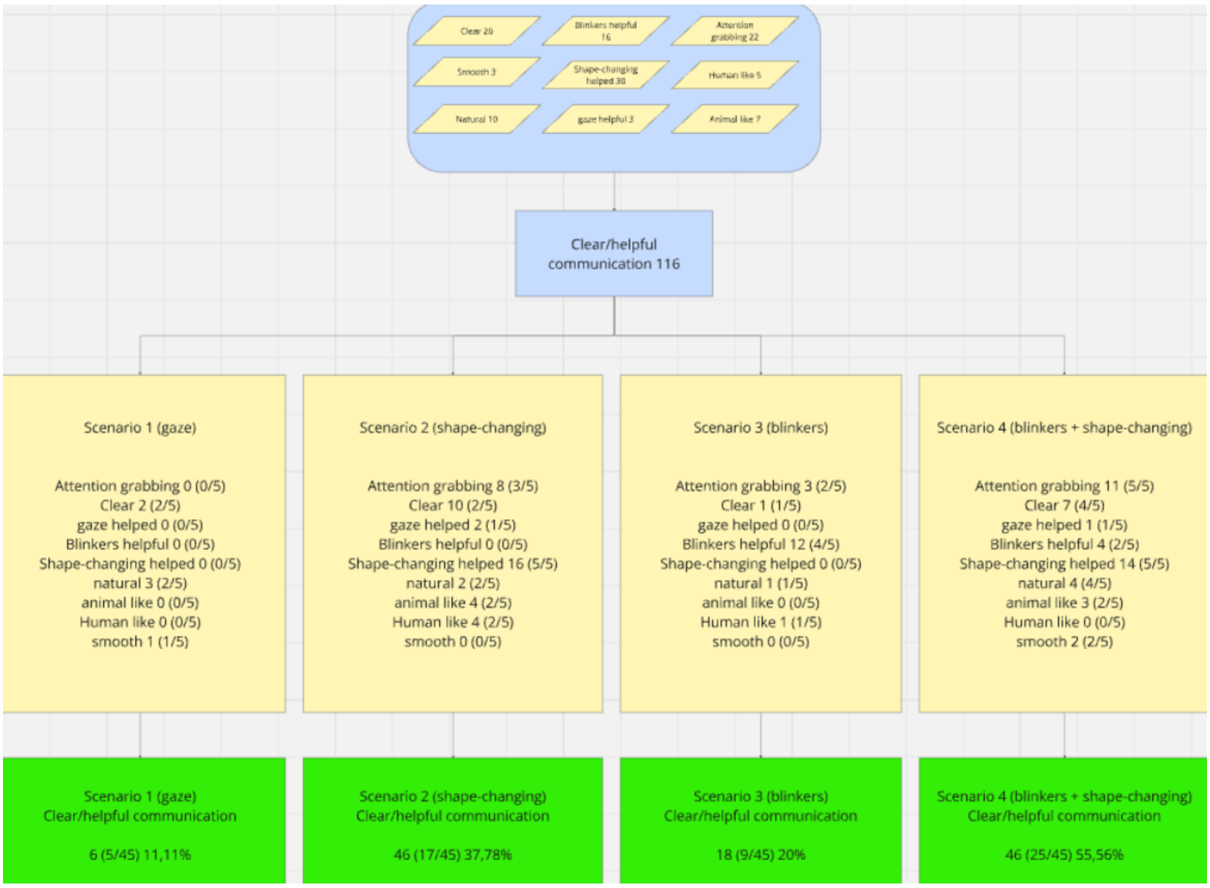


Figure 34: Clear communication thematic analyses.

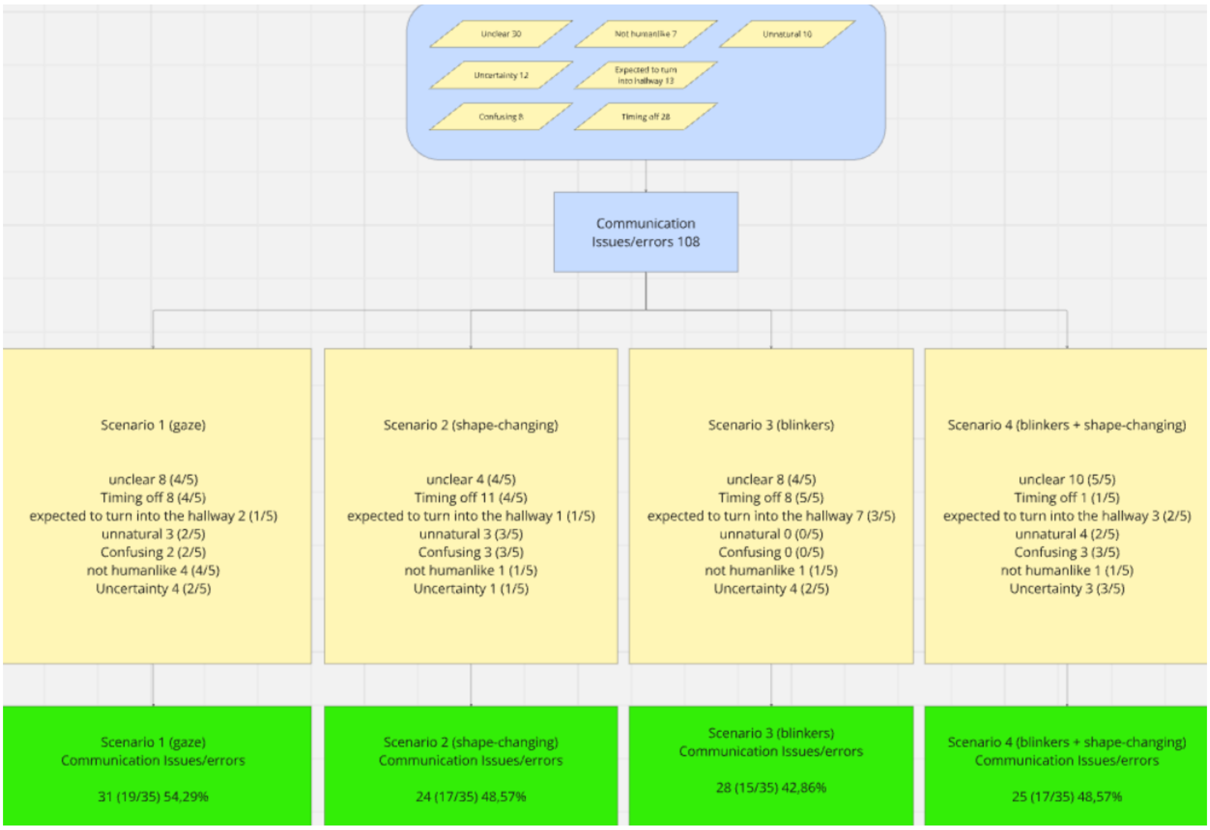


Figure 35: Unclear communication thematic analyses.

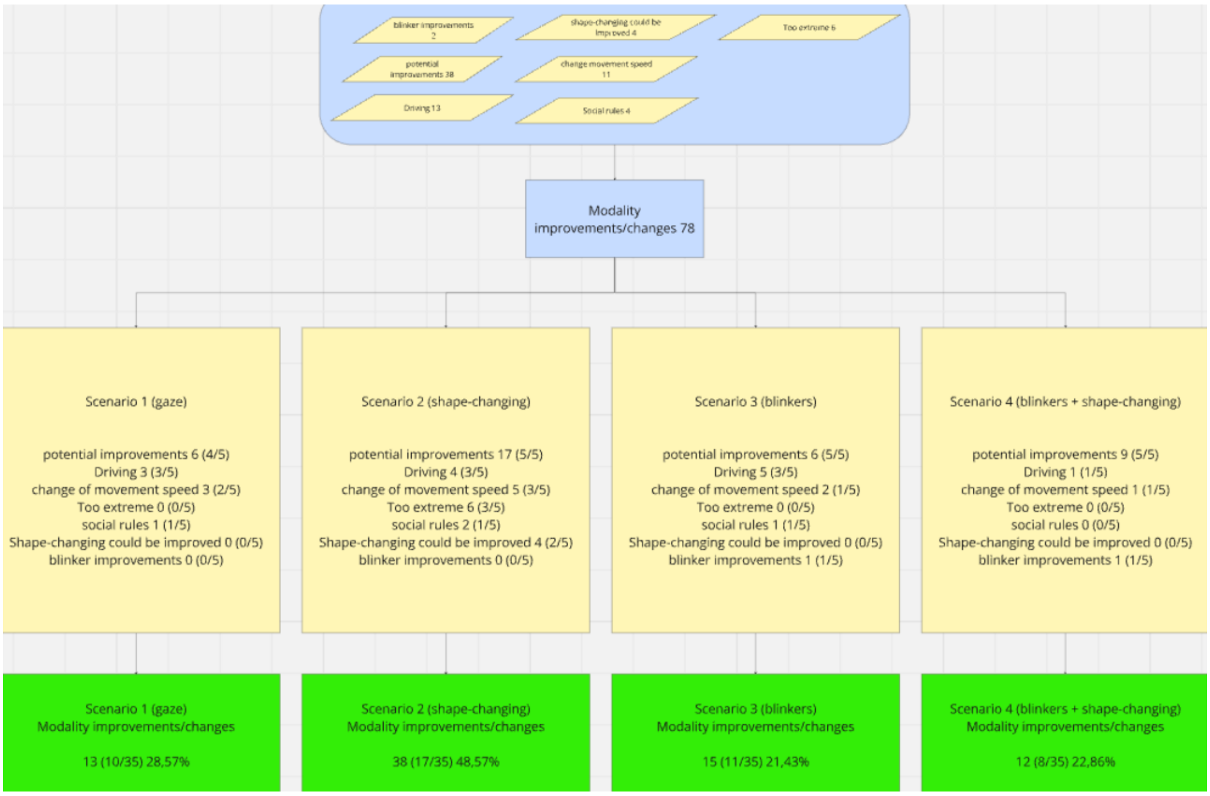


Figure 36: Scenario improvements thematic analyses.

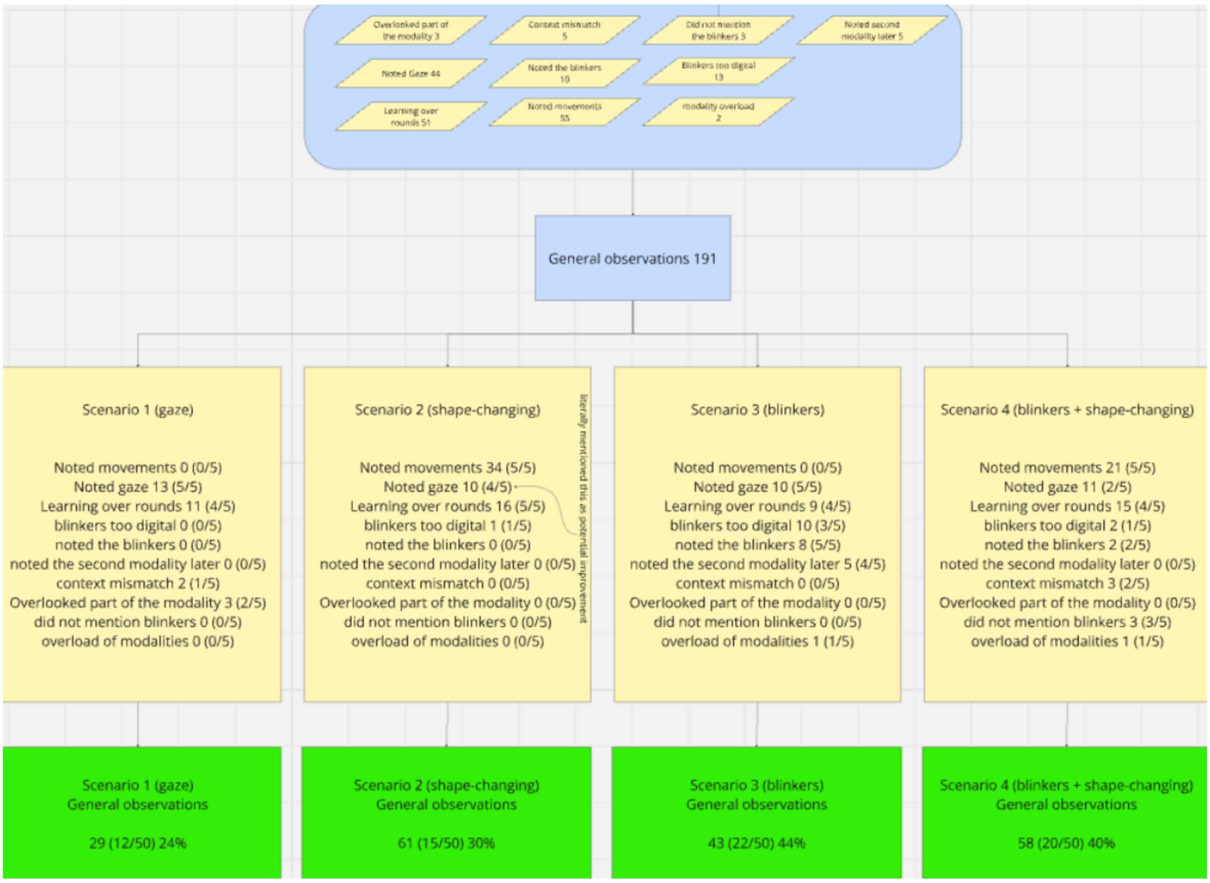


Figure 37: General observations thematic analyses.

E benjamini-Hochberg (BH)

The benjamini-Hochberg (BH) is calculated by ranking the P-values in order and calculating their BH-critical value $\frac{Rank}{P-value} \times a$, where a is set to 0,05, meaning the false discovery rate is set to 5%. This is followed by calculating the adjusted P-values that are compared to a common critical value. The adjusted P value can be calculated by:$Pvalue \times \frac{Amountofpvalues}{rank}$

P values	Warmth	0,045866	Rank	P values	(Pvalue/Rank)* a	calculated critical value (a=0.05)	BH-Adjusted P-Values	Largest to smallest cant increase
		0,176	1	0,015		0,00075	0,09	0,09
		0,015	2	0,045866		0,00114665	0,137598	0,137598
		0,292	3	0,058		0,000966667	0,116	0,116
		0,112	4	0,112		0,0014	0,168	0,168
		0,058	5	0,176		0,00176	0,2112	0,2112
			6	0,292		0,002433333	0,292	0,292
	Competence	0,011						
		0,168	1	0,011		0,00055	0,066	0,066
		0,916	2	0,076		0,0019	0,228	0,228
		0,141	3	0,141		0,00235	0,282	0,252
		0,076	4	0,168		0,0021	0,252	0,252
		0,401	5	0,401		0,00401	0,4812	0,4812
			6	0,916		0,007633333	0,916	0,916
	Discomfort	0,011						
		0,133	1	0,011		0,00055	0,066	0,045
		0,09	2	0,015		0,000375	0,045	0,045
		0,015	3	0,045		0,00075	0,09	0,09
		0,045	4	0,09		0,001125	0,135	0,135
		0,598	5	0,133		0,00133	0,1596	0,1596
			6	0,598		0,004983333	0,598	0,598

Figure 38: Benjamini-Hochberg

F Data points

Participant	Scenario	Accuracy	Response_Time	Legibility_Score	Confidence_Rating	Surprise_Rating	Warmth	Competence	Discomfort
8	1	1	398,8434	2,4965	5	1	3	4,3333	1,1667
9	1	1	354	4,1887	4	2	5,333333333	5,8333	2,3333
14	1	1	405	2,2642	4	1	3,166666667	4,3333	1,5
16	1	1	279	7,1698	5	1	1,3333	4,8333	1,1667
19	1	1	516	0	3	2	1,3333	4,3333	1,1667
3	2	1	460,6442	0,1644	3	2	3,5	7,8333	2,3333
4	2	1	340,6096	4,694	4	2	5,5	6,6667	2,6667
5	2	1	237	8,6038	5	1	4,5	5,8333	2,5
6	2	1	360,1992	3,9547	3	2	4,833	6,8333	2,6667
13	2	1	362	3,8868	5	1	5,5	6,3333	5,1667
7	3	1	397,32	2,554	1	5	5,8333	7,1667	1,6667
10	3	1	320	5,4717	5	1	4,3333	4,6667	2,3333
11	3	1	412	2	2	3	4,3333	6,3333	1,3333
17	3	1	288	6,6793	5	1	2,3333	4,5	2,3333
20	3	1	353	4,2264	5	1	3,5	4,6667	1,5
1	4	1	435,6204	1,1087	5	1	5,333333333	4	2,1667
2	4	1	418,5776	1,7518	5	1	6,5	7,3333	1,6667
12	4	1	241	8,4528	5	1	5,333333333	3,6667	1,5
15	4	1	459	0,2264	3	1	5,8333	4,5	2,6667
18	4	1	275	7,1698	5	1	5,5	5,5	2

Figure 39: Data points used for SPSS

G Word embedding mapping

```
from gensim.models import KeyedVectors
import numpy as np

# Load pre-trained Word2Vec model (e.g., Google News vectors)
# Download the model from: https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/
# Place it in your working directory and update the path below.
model_path = 'GoogleNews-vectors-negative300.bin'
model = KeyedVectors.load_word2vec_format(model_path, binary=True)

# Define words for each theme
warmth_seeds = ['Compassionate', 'Happy', 'Feeling', 'Social', 'Organic', 'Emotional', '
competence_seeds = ['Capable', 'Responsive', 'Interactive', 'Reliable', 'Competent', 'K
safety_seeds = ['Scary', 'Strange', 'Awkward', 'Dangerous', 'Awful', 'Aggressive', 'Safe

# Calculate average vectors for each theme
def get_average_vector(seed_words, model):
    vectors = [model[word] for word in seed_words if word in model]
    return np.mean(vectors, axis=0) if vectors else None

warmth_vector = get_average_vector(warmth_seeds, model)
competence_vector = get_average_vector(competence_seeds, model)
safety_vector = get_average_vector(safety_seeds, model)

# calculate similarity between a word and a theme vector
def get_similarity(word, theme_vector, model):
    if word in model:
        return np.dot(model[word], theme_vector) / (np.linalg.norm(model[word]) * np.lin
    else:
        return None

# Assign a word to the closest theme
def assign_theme(word, model, warmth_vector, competence_vector, safety_vector):
    # Check if the word is in the model
    if word not in model:
        return None  # Skip words not in the model

    # Compute similarities only if theme vectors are not None
    similarities = {}
    if warmth_vector is not None:
        similarities['warmth'] = get_similarity(word, warmth_vector, model)
    if competence_vector is not None:
        similarities['competence'] = get_similarity(word, competence_vector, model)
    if safety_vector is not None:
        similarities['safety'] = get_similarity(word, safety_vector, model)

    # If no valid similarities, return None
    if not similarities:
        return None

    # Return the theme with the highest similarity
    return max(similarities, key=similarities.get)

# Example usage
thesis_words = ['unpredictable', 'intimidating', 'Unsafe', 'scared', 'logical', 'strange
for word in thesis_words:
    theme = assign_theme(word, model, warmth_vector, competence_vector, safety_vector)
```

```
if theme is not None:
    print(f"Word: {word}, Theme: {theme}")
else:
    print(f"Word: {word} could not be assigned to a theme.")
```

H Iterative design process

After identifying the problem of robot legibility in hospital environments. The choice was made to evaluate the legibility of shape-changing robots in communicating navigational intent, specifically directional intent, allowing it to be compared with vehicle indicators, for this it is required to design a controlled and representative study environment. The steps for this design process, are as shown in figure 40, this went as follows:

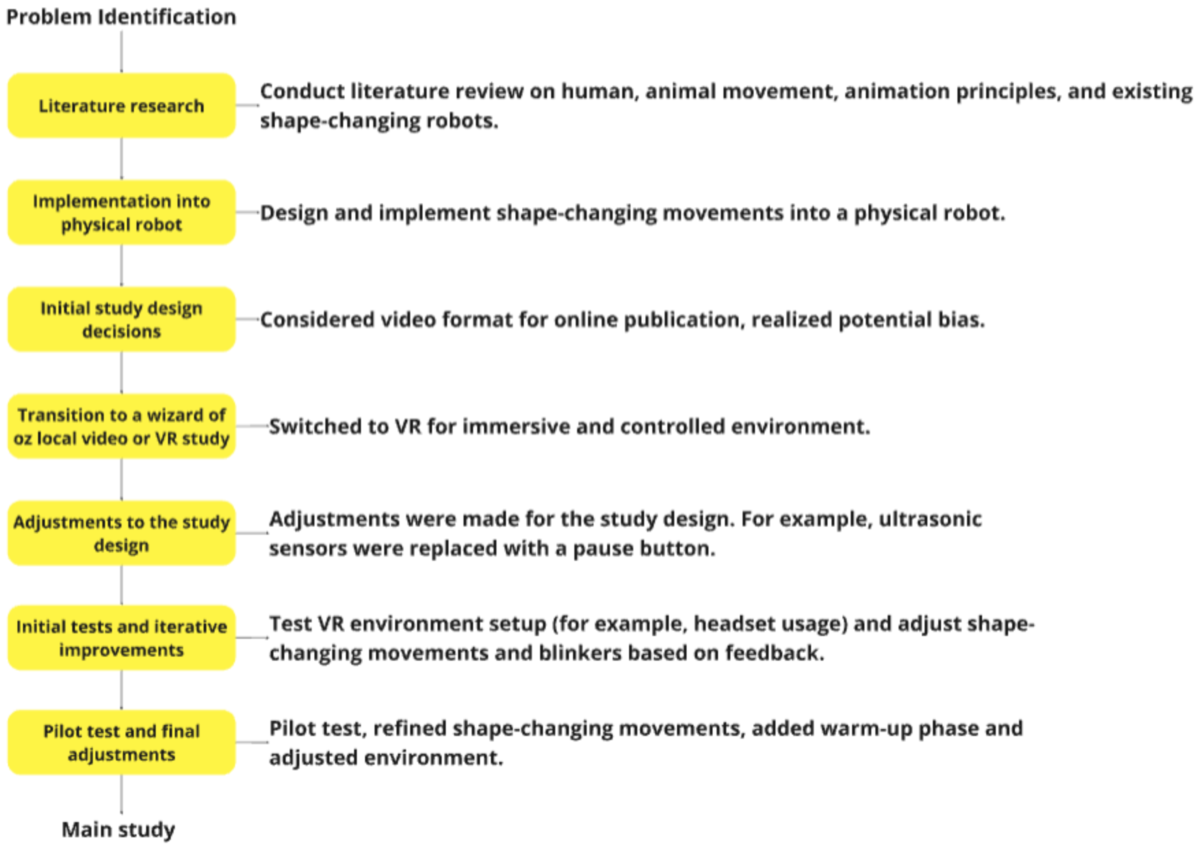


Figure 40: Design process

1. **Literature review**

   To address the legibility problem in hospitals, a literature review was conducted to explore human and animal movements, animation principles, and existing shape-changing robots. Based on this research, robotic movements were designed, which could be implemented into a real-world robot.

2. **Implementation into a physical robot**

   The designed shape-changing movements were then implemented into a physical robot. This step involved translating theoretical movement patterns into practical, real-world robotic behavior, possibly allowing the robot to communicate its intent through shape-changing.

3. **Initial study design decisions**

   After implementing the movements into the physical robot, initial study design decisions were made. The first idea was to use a video format to publish the robot's movements online, allowing for a broader audience. However, after further literature review and discussions with my supervisor, it became clear that an online video format might bias the results, especially since the study aimed to measure legibility, which required response time measurements. As a result, it was decided to conduct the study locally.

4. **Transition to a wizard of oz local video or VR study**

   Due to technical difficulties with the physical robot and other considerations, the decision was made to switch to VR-based study. This shift allowed for a more controlled environment and helped to guarantee more certainty in the measurement of legibility without the potential biases of an online format.

5. **Adjustments to the study design**

   With the switch to a virtual study, several adjustments were made to the study design. For example, it was decided to use a first-person front view in VR to simulate a realistic hospital hallway scenario. Additionally, ultrasonic sensors, initially planned for tracking participant movements, were replaced with a pause button to simplify data collection and analysis.

6. **Initial tests and iterative improvements**

   After adjusting the study design, initial tests were conducted in the VR environment to observe participant reactions to the robotic movements and determine the overall setup of the study. These tests aimed to determine practical aspects, such as whether participants would keep their VR headset on during the experiment or remove it, which led to the inclusion of cues in the environment to maintain engagement. Based on the feedback from these tests, iterative improvements were made to the robotic modalities. For example, the shape-changing movements were deemed too fast, and the blinkers were found to lack anticipatory cues. These adjustments were taken into account while ensuring that the physical robot could still execute the adjusted movements.

7. **Pilot test and final adjustments**

   A pilot test was then conducted to evaluate the study design and robotic movements. Feedback from the pilot test led to further improvements, such as reducing the exaggeration of the shape-changing movements, adding a warm-up phase for participants, and making environmental changes.

After all these improvements, the main study was conducted. This study was designed to evaluate the effectiveness of shape-changing as a communication modality for robot legibility in a hospital environment.

I Guiding principles

The study was designed to meet a set of requirements that helped determine the design decisions and implementation. These requirements were based on the need to create a realistic, controlled, and measurable environment for evaluating the legibility of shape-changing robots in hospital settings.

- Robot legibility and communication effectiveness
  - The study must measure the legibility of the robot's intent through shape-changing. The primary goal of the study is to determine how shape-changing impacts the robot's ability to communicate its intent clearly and intuitively. it is tried to reach this by using a combination of quantitative measures (response time, accuracy, confidence, surprise) and qualitative feedback (interviews) to measure legibility.
- Modality isolation, comparison and integration
  - The study must isolate shape-changing to determine the effectiveness, also compare shape-changing to traditional modalities, such as blinkers and explore how they can be integrated for multi-modal communication. This helps in understanding if shape-changing can enhance legibility, how shape-changing performs, compared to existing modalities and whether it can enhance them can potentially inspire future researchers to include this novel modality. It is tried to reach this by using a 2x2 experimental design to compare gaze-only, gaze + shape-changing, gaze + blinkers, and gaze + blinkers + shape-changing.
- Iterative design and feedback
  - The study should incorporate iterative design and participant feedback to improve the robot's movements and the experimental design. Iterative design helps to improve the robot's movements, to make it more legible and intuitive, while participant feedback helps to identify areas for improvement. To reach this pilot tests will be conducted to gather feedback on the robot's movements and the study setup. Based on this feedback, adjustments will be made to the robot's modalities and the VR environment.
- Ecological validity (hospital environment)
  - The study must simulate a realistic hospital environment to ensure ecological validity. This to determine how shape-changing impacts robot legibility in real-world hospital environments, the virtual environment must closely mimic a hospital setting. To reach this the VR environment will include hospital-like features such as hallways, crosswalks, hospital like objects and ambient hospital sounds.
- Participant understanding and perception
  - The study must determine how participants perceive the robot's likability, intelligence, and safety in the various scenarios. As user perception is important for the acceptance of robots in social environments like hospitals. Understanding how participants perceive the robot can inform future design improvements. To reach this use will be made of the Robotic Social Attributes Scale (ROSAS) questionnaire to measure participants perceptions of the robot's warmth (likability), competence (intelligence), and discomfort (safety). Additionally, conduct thematic analysis of interview data to gain deeper insights into participants experiences.
- Data analysis and statistical power
  - The study must include a sufficient number of participants to conduct a meaningful statistical analysis. While the study is exploratory, it should still aim to identify trends and patterns that can inform future research. To reach this data will be collected from at least 20 participants, by making it inclusive for as many participants as possible. Researching with a combination of quantitative metrics (response time, accuracy, confidence, surprise) and qualitative feedback (interviews). Use non-parametric statistical tests (for example, Mann-whitney U test) to analyze the data, as they are more robust for small sample sizes. Apply corrections (such as, Bonferroni or Benjamini-Hochberg) to control for multiple comparisons.

- Technical feasibility of robotic movement design
  - The study must ensure that the shape-changing robot design is technically feasible and can be implemented in a real-world robot. As the findings should be transferable to real-world robots, not just theoretical or virtual models. Technical feasibility helps that the design can be translated into practical applications. To reach this use is made of a real-world robot prototype for initial testing to validate the shape-changing movements. After initial testing the movements will be transferred to a VR environment for controlled experiments, this method can help to ensure that the movements are realistic and executable by real world robots.
- Participant diversity and inclusivity
  - The study must include a diverse participant pool to reflect the variety of individuals who might interact with robots in hospital environments. A diverse participant pool ensures that the findings are generalizable and applicable to a wide range of users, including those with different levels of experience with robots and cultural backgrounds. To reach this participants will be recruited from an university setting, helping to reach a mix of ages and genders. Also it will be tested at various home places, which can help reach different levels of experience with robots.
- Ethical and safety considerations
  - The study must meet the ethical guidelines, including informed consent, data privacy, and the right to withdraw, while offering a safe environment. These ethical considerations are required for participant well-being, while ensuring safety and comfort to minimize the risk of effects, such as motion sickness in VR. To reach this it is ensured to obtain ethical approval from the university's ethics committee. Ensure informed consent is obtained from all participants, and explain their right to withdraw at any time. Anonymize all data to protect participant privacy. Pre-test the VR environment to minimize the risk of motion sickness or discomfort, and allow participants to take breaks or stop the study if needed.

J Limitations of the robot

While the physical robot prototype allowed for the initial implementation and evaluation of shape-changing behaviors, several practical and technical constraints limited its suitability for use in the main study. These limitations, combined with the study requirements outlined in Appendix I, informed the decision to shift toward a video-based experimental setup in a virtual hospital environment.

- *Software limitations (Bottango)*

  Although Bottango was selected for its ease of use and real-time control features, it was not specifically designed for deformable shape-changing robots. The software assumes static body movement, which led to a difference between the on-screen visualization and the physical deformation of the robot. To address this, no virtual objects were added in the Bottango environment, and actuators were instead individually controlled. While this approach allowed for the creation of shape-changing movements, it reduced development efficiency and flexibility.

- *Communication through gaze*

  The prototype included a face element, introducing an additional communication channel through gaze. Since the presence of gaze can unintentionally influence interpretation during interaction [15], the robot's gaze direction was standardized across all conditions, using the shape-changing gaze direction as the primary reference point. This standardization increased the complexity of the visual design and setup but was necessary to isolate the effects of shape-changing behavior.

- *Hardware integration constraints*

  The prototype could not simultaneously support locomotion and shape-changing due to hardware limitations. During testing, the system was configured to prioritize the execution of shape-changing behaviors, which meant that the driving functionality was disabled. Additionally, components required for remote control, such as the router for triggering animations on command, were unavailable. As a result, the robot could only perform preloaded animations in a continuous loop from startup. While restoring full mobility and control was technically feasible, doing so would have required substantial additional integration work beyond the scope of the current study.

- *Robot availability*

  Due to interest from multiple research teams, access to the robot was limited. This introduced uncertainties regarding scheduling and long term availability for testing, making it challenging to plan a study using the physical robot.

- *Safety considerations*

  Testing the robot in a real world hospital environment would have introduced safety concerns, especially compared to the usage of video recordings. Additionally, conducting the study in uncontrolled environments would have made it more difficult to isolate the effects of shape-changing behaviors from environmental influences. As a result, we chose for a video based study design, which allowed for consistent presentation of robot behavior in a recorded hospital environment while ensuring participant safety.

These limitations, in combination with the study requirements described in Appendix I, led to the design decisions discussed in Appendix A.