

<https://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

Can You Trust What a Robot Says?
The Influence of Robot Morphology on Trust

A thesis
submitted in partial fulfilment
of the requirements for the degree
of
Master of Science (Research) in Computer Science
at
The University of Waikato
by
BEN JONES



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2026

Abstract

As social robots are increasingly used for sharing information, it is essential to determine whether people can trust them. Too little trust and people abandon a useful tool; too much and it may cause over-reliance. Morphology is part of communication, a channel that operates in parallel to spoken words, and it can affect perceptions before the interaction has even begun. Human-like appearances are often thought to make a robot more trustworthy, but our review of the literature shows that trust is not determined by embodiment type alone. Features such as resemblance, gaze, and expressive movement act as social heuristics, signalling what a robot is for and what it can do. Because trust arises only where a person is vulnerable and relying on the robot, information-seeking represents a meaningful context for studying trust even when the stakes are low. Within this context, trust is multidimensional, encompassing performance, relational, and moral judgements. Existing research has examined trust in information from websites, search engines, and disembodied agents, and trust in robots performing industrial or collaborative tasks, but little is known about how morphology affects trust when a robot is itself the information source.

In this thesis, we address this gap by examining how a robot’s physical form shapes the trust people place in it as an information source. Two studies were conducted using the same four devices: a disembodied smart speaker and zoomorphic, mechanomorphic, and humanoid robots. Each robot performed the same information-seeking task, allowing reported and enacted trust to be separated. First, in a survey, 252 participants watched a video of the robots and rated them using the Human-Robot Interaction Evaluation Scale and the Trust Perception Scale-HRI. Under passive observation, morphology had no significant effect on reported trust and only small effects on social perceptions; prior familiarity predicted perceptions instead, and the humanoid did not receive a trust benefit from anthropomorphic attributions.

An observational user study then brought the scenario into the real world, where 54 participants chose to approach a robot in person and ask their own questions. Once people sought information themselves, morphology produced large effects, but the dimensions of trust moved in different, sometimes opposing, directions. The

humanoid was rated as the most helpful and fit-for-purpose, yet participants were most likely to double-check what it said. The mechanomorphic robot was most relied upon because its form legibly signalled its communicative purpose. This appearance-function fit, rather than human-likeness, determined whether a robot was trusted appropriately. Cuteness and emotion encouraged people to approach, but cuteness alone did not signal credibility, as the zoomorphic robot was cute yet poorly received as an information source.

Morphology’s effect on trust is thus conditional on engagement, moving the facets of trust independently rather than uniformly. This thesis contributes an early empirical account of how morphology shapes trust in robots as information sources. Our findings indicate that a form that honestly signals what a robot is for, rather than one that is simply human-like, better supports users in calibrating trust. Systematically varying morphological features is a clear next step for tracing these effects to specific design choices.

Acknowledgements

I want to thank the University of Waikato for making studying in Tauranga the best choice, for developing my interest in research through the Summer Research Scholarship Programme, and for awarding me the Research Masters Scholarship to help me continue my growth.

Thank you to Lani Wreaks and Logan Shaw from the Waikato Students' Union for jumping straight on board with integrating the robots into Orientation Week. You made them much more engaging and useful than I ever could, and made this thesis feel like it has genuine value rather than just being theoretical.

Many thanks to my hype crew on level 4. It was always a delight coming to see you all. I would make trips to the main campus and come up just to chat, and you would always be happy and make time, and that made getting through the tedious parts so much better. You've been there alongside me, sharing the wins, and being genuinely interested in supporting what I do. I love helping out, but you are the reason why I would be free labour when I could. Not only do you do amazing work, but you also make the campus a great vibe. So thank you for just being awesome.

Thank you, Zane Hamilton, for being the trailblazer and the big brother that I could always turn to. You know you're a great teacher, but mostly you've been an even better mate. There could not be a better person to have around; you and I strive for the same things, and no one understands my situation better than you. I can never repay the kindness you've given me, and the number of messages, lunches, and disruptions in your office over the last few years. Whenever I needed help, I could rest easy knowing help is on its way. Believe it or not, I look up to you. I enjoy your company, and am honoured to call you my friend. Thank you, Dr Nicholas Vanderschantz, for making me realise where my true research passion lies. You have an affinity for making people consider the best way to do things for the people they are for. I have many different interests; you have allowed me to turn what I care about into a career I can pursue. No one better embodies and understands what is behind the whakataukī: He aha te mea nui o te ao? Thank you, Dr Jessica Turner, our fearless leader, without whom nothing would happen, constantly fighting to do what is best. I sincerely appreciate the investment you have put into me to allow my growth as a student, a researcher, and a professional

to happen. Right from the beginning, you have been a pleasure to learn from and work alongside; your passion for education is what made me care so deeply about our campus. Your commitment to doing what is in my best interests allowed me to put my faith in you wholeheartedly, even when I made things more difficult for myself. You are wise beyond your years, and I look forward to learning more from you.

Finally, and most importantly, thank you to my family. I cannot express how much gratitude I have. No words can adequately convey how much I appreciate each of you three. You have supported me relentlessly over the course of my life, and I am immensely grateful for you making me the man I am today, through how I was raised and the mind I inherited. You have never stopped me from trying to do a million different things, even when you strongly (and correctly) advise that it is too much. To my ever-loving mother, when my wheels fell off, you are why things kept on moving, and you were always doing what I could not, and somehow loved me through it all. To my fantastic father, you have always helped me when I needed it, had constant confidence in me, and unenviably tried to understand what goes on in my mind. To my inspirational sister, your determination is what I aspire to, and I appreciate you being interested in what I have to say. Even when I was not the nicest or did not have the capacity to give back, I always had all of your relentless support to fall back on, as I often did. You have always been understanding, probably more than I deserved, especially with all the time I couldn't spend with you. I could not have done this degree without you.

Thank you once again to all of my supporters: my friends, my family, those at University, and those at karate. You all have a place in my heart.

Contents

1	Introduction	1
1.1	Background	1
1.2	Thesis Statement & Scope	4
1.3	Research Questions	4
1.4	Structure of Thesis	5
2	Literature Review	7
2.1	Information-Seeking	8
2.1.1	Question-Asking Behaviour	8
2.1.2	Digital Information Sources	10
2.1.3	Trust in Information-Seeking	11
2.1.4	Robots as an Information Source	12
2.2	Robot Morphology	14
2.2.1	Physical Characteristics	14
2.2.2	Impact on Perceptions	15
2.2.3	Morphological Signals	17
2.3	Evaluation Frameworks & Methodologies	19
2.3.1	Standard Measures	19
2.3.2	Experiment Designs	23
2.4	Definitions of Trust	25
2.4.1	Trust Measurements	26
2.4.2	Trust in Automation	30
2.4.3	Trust in Human-Robot Interaction	33
2.5	Chapter Summary	37
3	Methodology	39
3.1	Research Design	39
3.2	Robot Platforms	42
3.3	Study Context	45
3.4	Measurement Approach	46
3.5	Chapter Summary	48

4	Survey	50
4.1	Methods	50
4.1.1	Survey Design	51
4.1.2	Participants & Recruitment	52
4.1.3	Analytical Approach	52
4.2	Results	53
4.2.1	Responses	54
4.2.2	Descriptive Statistics	55
4.2.3	Effects of Morphology	55
4.2.4	Effects of Participant Characteristics	57
4.3	Chapter Summary	58
5	User Study	59
5.1	Methods	59
5.1.1	Setting & Approach	59
5.1.2	Participants	60
5.1.3	Interaction Flow	61
5.1.4	Technical Implementation	61
5.1.5	Measurement and Data Analysis	65
5.2	Results	68
5.2.1	Participants & Overview	68
5.2.2	Descriptive Statistics	68
5.2.3	Inferential Statistics	69
5.2.4	Post-Hoc Pairwise Comparisons	70
5.2.5	Summary of Quantitative Findings	71
5.2.6	Qualitative Findings	72
5.2.7	Direct Perceptions of Trust in Appearance	74
5.3	Chapter Summary	75
6	Discussion	76
6.1	Survey	76
6.1.1	Small Effect of Morphology	77
6.1.2	Scale Dimensions	78
6.1.3	Prior Familiarity	79
6.1.4	Limitations	80
6.1.5	Future Work	81
6.1.6	Survey Summary	81
6.2	User Study	82
6.2.1	Interpretation of the Findings	82
6.2.2	Appearance-Function Fit	84
6.2.3	Cute Factor	85

6.2.4	Trust Calibration	87
6.2.5	Perceived Answer Quality	88
6.2.6	Limitations	89
6.2.7	Future Work	89
6.2.8	User Study Summary	91
6.3	Combined Study Findings	91
6.4	Relation of RQ2 to RQ1	92
6.5	Chapter Summary	94
7	Conclusion	95
7.1	Key Findings	96
7.1.1	Answer to RQ1	96
7.1.2	Answer to RQ2	97
7.2	Contributions	99
7.3	Concluding Remarks	100
	Bibliography	102
A	Interaction Taxonomy	122
B	Ethical Approval	131
B.1	Survey	131
B.2	User Study	134
C	Survey Content	137
D	Online Video Content	141

List of Figures

3.1	Devices Selected for Studies	42
3.2	METAMORPH Representation of Morphologies	44
4.1	HRIES Dimensions Morphology Bar Charts	56
4.2	TPS-HRI Morphology Bar Chart	56
5.1	Robot Conversation Process	62

List of Tables

3.1	Device Morphology Matrix	43
3.2	Robot Graph Edit Distance	44
3.3	Map of Trust Dimensions to Measures	48
4.1	Survey Participant Ages (N = 252)	54
4.2	Survey Participant Genders (N = 252)	54
4.3	Survey Participants Robot Familiarity (N = 252)	55
4.4	Survey HRIES Descriptive Statistics	55
4.5	Survey TPS-HRI Descriptive Statistics	57
4.6	Univariate ANOVA for HRIES Dimensions	57
5.1	User Study Descriptive Statistics	69
5.2	User Study Kruskal-Wallis Tests	70
5.3	Helpful Rating Morphology Pairwise Comparison	71
5.4	Useful Rating Morphology Pairwise Comparison	71
5.5	Believable Rating Morphology Pairwise Comparison	72

Chapter 1

Introduction

Before a robot says a word, its appearance has already spoken for it. Morphology refers to the combination of a robot’s physical appearance, movement, communication style, and contextual presentation that shapes how users perceive and interpret a robot [1]. Trust is one of these perceptions, and supporting users to place appropriate trust in a robot is important if those users are expected to rely on the information provided. Social robots are appearing in hospitals, schools, stores, and other public places [2, 3], and advances in conversational artificial intelligence (AI) mean that answering spoken questions is now one of the most plausible everyday use cases, as can be seen with the increasing adoption of AI in general [4]. A robot placed in a foyer can greet a visitor, listen to a question about directions, and produce a fluent and confident-sounding answer. Whether that answer should be believed is different from the perceived quality of the answer, which motivates the main questions we examine throughout this thesis. Specifically, can you trust what a robot says, and how does its appearance affect that?

1.1 Background

Information-seeking is a fundamental human behaviour, driven by the need to reduce uncertainty and solve problems [5]. The sources people consult for their information have evolved, as have the technologies that can be used as a source; from libraries, printed media, and human experts, to websites and search engines [6, 7]. Recently, conversational AI agents have become a common information-seeking strategy for people. A known issue with these systems for information-seeking is that they synthesise an answer rather than point to one, potentially taking away the source evaluation that users would previously have undertaken themselves [8]. Embodied agents are an extension of AI agents, adding a face to the answer. With their physical presence, they can be an always-available first point of contact,

taking some of the workload off human workers. Speech-to-speech AI models can now hold a natural spoken conversation inexpensively, which makes a robot that answers questions in a public setting a very real possibility today.

The evolution of information-seeking behaviour changes what users base their trust assessments on. When information comes from a document or a list of search results, a person can separate the content from the channel and evaluate each separately [6]. When information comes from a robot, this is no longer possible because the robot is both the interface and the source, so a judgement about the information provided and a judgement about the robot are intertwined. How content is presented is known to affect trust independent of accuracy [9]. For example, people report more trust in the structured and confident answers of a conversational agent than in a list of links [10]. A robot adds a physical body to an interaction, which helps users make judgements before, during, and after an answer is given.

Concern about misinformation and AI systems that produce confident but incorrect answers is prevalent [11, 12]. If robots are intended to be used to provide information, the implementation should consider whether the trust users place in them is appropriate to the level of reliance they warrant. The goal of robots that provide information should not be to make them seem as trustworthy as possible. Instead, the design of the robot should accurately reflect its purpose and capabilities to help the user calibrate appropriate reliance [13]. Too little trust can mean a genuinely useful tool is not used, referred to as disuse. Conversely, too much trust can mean a user relies on information that should have been checked, a form of misuse. These two directions are failure modes in automation research, and both can have consequences [13, 14].

Trust is only meaningful under particular conditions. Wagner and Arkin [15] identified two: risk in the potential for some loss, and interdependence on the system to play a role in managing that risk. Everyday information-seeking meets both of these conditions, even when the stakes of acting on that information are low. Acting on a wrong answer can have consequences, even if the consequence is just a little bit of time being wasted. When the robot is the source, it controls access to the answer, and the robot's response and other qualities can influence whether users will verify the answer independently [16]. Trust is a dynamic factor in this situation, and at low stakes, it is more than simply preference. A user calibrating their trust does not know everything about a system, such as error rates or the original source of the information. Instead, they must use the signals given to them, such as morphology which is unique to robots. Before any performance is known, the only direct observation they have is what the robot looks like.

Robot morphology is more than aesthetic appeal; it encompasses its structure, styling, visible features (e.g., cameras, speakers, antennas, and hands), and

movement. It is a distinct communication channel, separate from spoken word. Technical features and styling can communicate social heuristics that help a user understand a robot’s intent [17]. Eyes and cameras imply awareness of the world, hands imply the ability to manipulate physical objects, and a human-like body can imply human-like competence [18, 19]. A user infers these capabilities before an interaction begins, and it is an unconscious behaviour that may work against a designer’s intention [20]. In a meta-analysis of trust factors in human-robot interaction (HRI) by Hancock *et al.* [21], robot-related factors were among the strongest predictors of trust, with morphology in particular having a strong effect.

Morphological signals may also unintentionally mislead a user. There are two types of signals we refer to throughout this thesis. Role legibility is the degree to which an observer can infer, from the form alone, what a robot is intended to do. Appearance-function fit is the degree to which the role and capabilities the form implies match those the robot actually possesses. The two signals are similar but distinct. Legibility concerns whether the signal can be read; fit concerns whether what it communicates is true. A form can therefore be legible but misleading, projecting a clear role the robot cannot fulfil, or capable but illegible, offering no readable role at all.

For example, a highly human-like appearance implies cognitive and social capabilities that the software behind it may not possess. When those expectations are not met, trust can collapse [22, 23]. Design tropes used to make a robot appealing further complicate this. Cuteness, for example, is widely used to lower hesitation and invite engagement [24], but it remains to be seen whether this translates into confidence in the system and its credibility.

A common theme in robot morphology and trust research is that a more human-like appearance is more trustworthy [25]. People tend to trust people; this is an intuitive assumption at a surface level, and the intention is that anthropomorphic traits can inherit some of that trust. In this thesis, we examine whether that assumption holds true, specifically in the context of information-seeking.

Research on trust in information-seeking is concentrated on digital interfaces and disembodied agents, such as websites, search engines, and conversational assistants [26–28], where the source has no physical form for the user to interpret. Meanwhile, research on trust in HRI has focused largely on industrial [29, 30], collaborative [31, 32], and task-oriented contexts [33, 34], where trust concerns a robot’s physical actions rather than the information it provides. The intersection of the two, how a robot’s morphology shapes trust when the robot is the information source, has received little direct attention. Without an understanding of how morphology shapes trust, designers risk creating robots whose appearance encourages inappropriate trust, unwarranted mistrust, over-reliance, or a misunderstanding of what the robot can actually do.

1.2 Thesis Statement & Scope

As robots enter everyday public spaces, people will increasingly turn to them for simple information. Such an exchange is only successful when the person is willing to believe the answer, and only believing when appropriate. We assert that a robot’s morphology, meaning its physical form and expressive movement, affects perceived trustworthiness in the information it provides. These judgements are related but not interchangeable, as a robot may appear helpful or capable without its information being regarded as sufficiently credible to act upon without verification. More precisely, morphology influences epistemic trust.

Epistemic trust concerns whether a person regards the robot as a helpful source, whether they judge its answers to be good, and whether they are willing to act on those answers without independently checking them. Epistemic trust is distinct from trust in the robot’s physical safety and from the relational trust that builds over long-term use. The relationship matters because the aim of good design is not to maximise trust but to help calibrate it, so that a person relies on a robot’s answers only to the degree those answers merit it.

The types of interactions we study in this thesis are brief, voluntary spoken exchanges in which a person asks a robot a question and expects a factual reply, set in an everyday public place. These are first encounters rather than established relationships, and they are low-stakes, since the information is useful, but a wrong answer has little consequence. The robots we studied are four commercially available, desktop-scale devices selected to span a range of embodiment, with a smart speaker as a disembodied baseline, a zoomorphic robot, a mechanomorphic robot, and a humanoid. These devices appear in Figure 3.1, and Chapter 3 justifies their selection.

Several related research areas fall outside this scope. High-stakes settings, such as healthcare, where a wrong answer could cause real harm, are not examined. In this work, we examine a diverse range of morphologies rather than specific features. Highly realistic androids and large robot platforms are also not assessed, since their scale and appearance differ markedly from the desktop devices used here. Finally, because each study captures a first encounter, the way trust develops over repeated interactions is left to future work. Chapters 3 and 6 discuss the boundaries as both a limit on the findings and as a direction for later study.

1.3 Research Questions

The assertion above cannot be investigated without first establishing what is already known about the ways morphology shapes perceptions. Morphology, trust, and information-seeking have each been developed separately, and information-seeking

with robots, in particular, is an area with limited existing work. In the first research question, we ask what is already known in these areas, and what it means when brought together. In the second research question, we ask whether expectations about trust and morphology in the literature remain relevant when applied to a real-life information-seeking context. Thus, the research questions are as follows:

RQ1 Which morphological features influence perceived trust, and by what mechanisms?

RQ2 How does robot morphology affect epistemic trust during information-seeking scenarios?

RQ1 is answered by establishing what morphological features robots have and how they are classified, how those features affect perceptions and interaction expectations, and what trust means in both HRI and information-seeking. RQ1 is addressed through the literature review in Chapter 2, where we highlight the morphological factors relevant to trust and how trust is measured. By answering RQ1, we identify a gap in existing research to study.

Answering RQ2 is complicated by a well-documented gap between what people report about trust and what they do [35]. RQ2 is therefore addressed through two complementary studies that hold the robots, task, and topic constant, using morphology as the independent variable. One study is an online survey capturing the trust people report after passively observing a robot being used, and the other is a naturalistic user study measuring indicators of trust that people enact when they choose to approach and rely on the same robots. Throughout this thesis, reported trust refers to what a person states about the trust they place in a robot based on what they think they trust, and enacted trust to what they are observed to do in terms of what they actually trust. Together, the studies help answer not only whether morphology shifts trust, but also how different kinds of trust change within the same robot.

1.4 Structure of Thesis

In Chapter 2, we conduct a literature review to explore information-seeking behaviour, the classification and impact of robot morphology, and the various aspects of trust in HRI, addressing RQ1. The gap in answering RQ1 is the motivation behind Chapter 3, where we justify the approach taken to assess how perceptions are shaped by robot morphology, using an online survey and a real-world observational user study to answer RQ2. In Chapter 4, we cover the approach and results of the survey, and in Chapter 5, the approach and results of the user study. In Chapter 6, we provide an interpretation of the findings, including the impact of

users and robots on trust. Finally, Chapter 7 is the conclusion to the thesis, and we summarise what each of the findings mean for the overall research area.

Chapter 2

Literature Review

To understand how morphology shapes trust in information-seeking, it is first necessary to examine each of these areas in the existing literature. Trust in this research context sits at the intersection of three bodies of research: how people seek and evaluate information, how a robot’s physical form influences perceptions, and how trust is defined, measured, and affected in human-robot interaction. Each of these areas has developed separately. This literature review aims to bring them together and identify gaps in the research. We also examine how interactions are structured and how perceptions are evaluated to help inform the methodology in Chapter 3.

This chapter proceeds in four sections. In Section 2.1, we look into information-seeking behaviour, from how the sources people consult have evolved to what role trust plays in evaluating them. We discuss how robots are classified by physical form, how those forms shape user perceptions, and what specific features act as social signals in Section 2.2. We review evaluation frameworks and methodologies commonly used in HRI research to measure user perceptions and interaction outcomes in Section 2.3. In Section 2.4 we give a description of how trust can be measured and what it means in HCI generally before looking at how it is used in HRI. Finally, we summarise the findings from the literature in Section 2.5 on what trust means in this area, drawing on definitions and measurement approaches from automation and HRI to identify what dimensions of trust are most relevant to our investigation. Altogether, the sections help build a picture of the morphological features that exist, the ways they impact perceptions and trust, and what that means, to answer RQ1. This chapter also informs the methodology used for the studies that follow.

2.1 Information-Seeking

Information-seeking is a fundamental human behaviour, driven by the need to reduce uncertainty and solve problems [5]. At its core, it encompasses the behaviours and actions users undertake in acquiring information. Information-seeking is not a single act; this behaviour exists on a spectrum. It can be active, goal-oriented searches for specific answers, passive, push-based receiving of content from subscribed channels or social groups, or somewhere in between [27, 36].

The sources and environments through which people seek information have changed substantially over time. Traditionally, it involved sources such as libraries, human experts, and printed media [37]. Now, it includes websites [38], search engines [12], and more recently, automation such as AI agents (e.g., ChatGPT, Claude, and Gemini) [8], and is beginning to include interactive and embodied technologies like robots [39]. While the purpose of information-seeking is the same, the environment and process have changed because of these digital platforms. This domain expansion raises questions about how people evaluate and place trust in the sources they turn to.

Regardless of context, information-seeking is not a single action, but a sequence of behaviours [37]. Behaviours include locating and navigating sources, and evaluating and filtering them based on perceived reliability. In digital contexts, this process is mediated by the different technologies available to the user. How they perceive those tools plays an important role in the strategies used, the challenges encountered, and the outcomes achieved. As more types of sources become available, the evaluation dimension becomes more significant, with users having to consider how much they believe one type of source over another.

2.1.1 Question-Asking Behaviour

Taylor [40] distinguished the need for information a person has, from the question they end up asking. A person goes through multiple stages, starting with an unexpressed awareness that something is missing, and ending with a formed question that is put to a librarian or a system. Belkin [41] argued along similar lines that a user who could state precisely what they were missing would have less need to search for it. A question has therefore evolved and clarified before a source receives it. Part of what shapes a question is the perceived answer capabilities of the source. The same underlying query may be phrased one way for a colleague, and another way for a search engine, for example. In an embodied interaction, these expectations may be informed by morphology: a robot that appears knowledgeable, communicative, or socially capable may invite different questions from one that appears functional, toy-like, or limited.

Users also revise their questions as they go. Bates [42] described searching as

“berrypicking”, where each partial result helps the user iterate on what they are looking for, and the query is adjusted accordingly, so that information is collected over a series of attempts rather than returned at once. The same pattern was observed over longer research tasks [43], where uncertainty is highest before the searcher settles on a focus. Query log studies report the behaviour directly, with web queries averaging only a few terms and being reformulated within most sessions [44, 45]. Users differ in how well they do this, and the lack of proficiency discussed by Begany *et al.* [46] is partly a matter of not revising a query that did not receive a satisfactory answer.

The clearest evidence that the form of a question is not fixed by the underlying need comes from studies of children, where the gap between how a need is expressed and how a system expects to receive it is most pronounced. Children typically express an information need as a natural language question or phrase rather than as keywords. Bilal and Gwizdka [47] found phrase-like and question-like queries to be more than twice as common as keyword queries among 11- to 13-year-olds, with around half of all queries being reformulations. Vanderschantz and Hinze [48] reported that natural language accounted for close to two-thirds of the queries children eventually succeeded with. When a task was presented as a question, question-based queries were used, while tasks posed as instructions produced sentences or keywords. How a question is asked is largely dictated by surface-level features of the situation.

Typical strategies used to seek information are of limited help to children, as query suggestion is tuned to adult search behaviour and can hinder younger users rather than assist them [49, 50]. Speaking to a robot removes the need to translate a question into keywords, and a multi-turn exchange helps get a more accurate understanding of what is wanted than a single spoken query [51], but what is asked still depends on what the robot is perceived to be capable of answering.

Conversational systems change how this query formulation evolves. The effect predates large language models, as users of earlier chatbots and voice assistants formed expectations of a system’s intelligence and scope from how it presented itself, rather than from what it could do [52], and addressed it in shorter messages drawn from a narrower vocabulary than they used with a person [53]. Because the exchange is a two-way dialogue, rather than a single query and a list of results, the system can respond to a vague question by asking for clarification instead of answering it as posed [54]. However, this conversation can have drawbacks, since the questions a robot asks can misread intent, or add social overhead that distracts from the task [55, 56]. Changing from written search queries to spoken word can also increase the burden on the user. Questions asked aloud tend to be longer and closer to ordinary conversation than typed queries, and users who are misunderstood often respond by shortening what they say or repeating it more

clearly [57–59].

It may not be necessarily clear to the user that the framing of their question has affected the quality of the answer received. Conversational systems often assume that words and phrases mean the same thing to all users, even though subjective terms may be interpreted differently, increasing the likelihood that a user’s intended meaning is misunderstood [60]. A skilled human librarian can resolve uncertainty through a reference interview by establishing what a reader actually wants before answering [40]. A system that responds to whatever it is given does not do this, especially given that large-language models aim to give a generally good sounding answer rather than one that is more correct [61, 62]. A robot may consequently be negatively judged on its answer to a question, even if that question was not well posed, and its appearance is one of the things that will have influenced how the user posed it based on perceived capabilities. A user’s choice of words, reformulation or repetition of a question, and decision to verify an answer can therefore provide observable evidence of how successfully the interaction proceeds and how much reliance the user places on the robot.

2.1.2 Digital Information Sources

Understanding how users conceptualise digital tools is key to improving the information-seeking experience. The mental models that users construct influence the search strategies, satisfaction, and overall effectiveness. A 1999 study of academics by Bruce [63] found that they view the Internet either as a means of connectivity or as a repository of information. The Internet can be seen as useful for linking resources, but also as a disorganised library, which greatly affects their experience. Those who view the Internet as a “freeway system” or a “web of nodes” are often satisfied as long as connectivity is achieved. Whereas those who view the Internet as a large library are often dissatisfied with the lack of proper curation and indexing, making it difficult to find the information they want.

Expanding on the limitations of this older conceptualisation, Lynch [64] argued that the distributed nature of the Web fundamentally changed the expectation of having reliable, well-behaved databases and metadata. The digital age enables systematic deception and manipulative indexing to drive greater usage, making it very different to a curated library. Newer research reinforces the challenges associated with the disorganised library model, indicating that users care about ease of use, organisation, and categorisation [46]. The increasing use of AI agents and other digital sources introduces new complexities regarding source provenance and perceived trustworthiness, which increases scepticism about content reliability and authority now that information can be retrieved rapidly [11].

According to Dalglish and Hall [38], some of the main challenges with information-seeking on the Internet are speed, quality, and proficiency. Frustration with the

time required for information retrieval, with the slow speed of access, and the time-consuming nature of searching [26] has been addressed through improvements in infrastructure and the tools available.

The two other challenges remain and are growing. Given the unstructured nature of the web, users often struggle with the variability in information quality and reliability, leading them to have a “healthy scepticism” of online sources [38]. Scepticism is even harder to calibrate appropriately amid the rise of misinformation and the excessive prioritisation of AI-generated content. Similarly, many users lack a strong understanding of how search tools operate, leading to poor use of their capabilities. A lack of proficiency in formulating effective queries is a major barrier to successful information retrieval, preventing users from finding the information they need [46] and from making accurate assessments of that information and the source.

A user’s digital literacy is a critical factor that mediates these challenges. As defined by Yamamoto *et al.* [65], this refers to “the ability to examine web information critically and collect correct information using information access systems, such as search engines”. Literacy includes a user’s verification strategies, their skill in using search engines, and their resistance to cognitive biases that can arise when assessing information.

Users with strong verification attitudes exhibit more careful search behaviour, such as by investigating lower-ranked search results rather than just using the top hits, or adding the keyword “evidence” to their queries to scrutinise the sources [66]. A study of computing students seeking help for programming tasks found that materials that offer immediate, visual, and repeatable instruction (such as YouTube videos) were utilised much more than text-based resources [36]. Sources that provide answers quickly and easily are seen as more useful and thus used more.

2.1.3 Trust in Information-Seeking

Trust is central to how people evaluate information sources. Trust affects how users perceive and assess the qualities of an information source. Whether the source is a person, document, or virtual agent, perceived trustworthiness influences how users judge its quality, accessibility, and usefulness [6]. A systematic review by Sbaffi and Rowley [9] broke down the factors that influence judgements of trust and credibility in web-based health information into design-related and content-related cues. Design-related cues include the layout and design of the page, the author’s details, personalisation, and an absence of advertising. Content-related cues include the author’s authority, objectivity, currency, references, accuracy, and understandability. Factors such as readability and personalisation can also cross both areas.

People do not passively receive information; instead, they seek out “trust signals”

to reduce the uncertainty and perceived risks associated with relying on a source [67]. These signals and cues help users form a judgement on a source’s trustworthiness, giving them a degree of confidence. The need for assurance is even more complex now, as information sources have evolved from static documents to interactive and conversational agents. However, if incorrect information from the agent is discovered, user trust and satisfaction decline significantly [11].

Trust in this area refers to the level of confidence an individual has in a specific information source to provide fair and accurate information [68]. Confidence is assessed through components such as a source’s perceived credibility and the reliability of its content. The Trust in Online Health Information (TOHI) scale by Rowley *et al.* [69] breaks it down into credibility, content, style, usefulness, brand, verification, ease of use, and recommendation. Greater trust also leads to greater motivation to seek out information, along with more positive affect towards the source, like hope, optimism, and enthusiasm.

2.1.4 Robots as an Information Source

As conversational agents are used more for accessing information, users are no longer retrieving information but instead having the synthesis done for them. Services like ChatGPT provide answers much faster than a traditional search process [12], but may risk accuracy due to factual errors or AI hallucinations. Synthesised answers can reduce the opportunities users have to evaluate individual sources during the search process, which undermines the trust-building process. Search engines like Google are slower for finding information, but provide more reliable and verifiable information, as there is transparency of the source [8]. Although users can verify answers, if they do, this places a higher cognitive load on them, which may not be desirable. When a source requires greater effort to use or verify, users may prefer an easier alternative.

Trust in AI is affected by more than accuracy; perceptions are also formed by interaction design, the interface used, and psychological factors. Trust is also influenced by the presentation style of the information. Higher trust is reported in the structured, confident, and human-like answers provided by ChatGPT (compared to a list of links in Google) [7]. Furthermore, making the agent a more approachable information source may lower the “perceived social risk” of the interaction [70], contributing to higher trust and usage. Text-based interfaces have been shown to elicit higher levels of trust than speech-based or embodied interfaces [10], indicating that the medium of the interaction is important, not just the content itself.

Actively involving the user in the information-seeking process can be used as a mechanism to build trust. For example, asking about a user’s preferences was shown to significantly increase trust in a robot-based recommendation system [71]. This involvement mechanism supports deeper psychological needs for agency and

transparency in the interaction. By giving the user a sense of control, it mimics how other people are used as a source of information [28].

Providing insights into a system’s search process can address scepticism about a source’s transparency and credibility, including distrust in the organisation behind the technology [72]. Social norms, user preferences, and context can all affect the choice of source. Users tend to favour AI when their perceived information-gathering capability is low, and prefer people when they feel like they know more [28].

A study of five- to nine-year-old children found that they attributed cognitive abilities like thinking or learning to robots, but also correctly identified them as non-biological, human-made artefacts [39]. Attributing cognitive abilities fosters a willingness to trust and befriend the robot, viewing it as a capable and approachable social entity. The researchers also hypothesised that children might perceive the robot as non-judgemental, making them more willing to express their curiosity and admit to a lack of knowledge. Children were observed to ask more explanation-seeking questions than they did with a human adult.

A robot’s conversational style also impacts its effectiveness as an information-seeking partner. Sharma *et al.* [55] compared two distinct interaction styles for a robot supporting search tasks and found a clear trade-off between guidance and user agency. A robot that asked closed-ended, yes or no questions had the benefit of being seen as more helpful by being specific and on-topic, but also misunderstood user intent and jumped between topics. A user-directed conversational style was seen to better understand a user’s intent and feel more connected, but led to vague or off-topic questions.

The model by Shin *et al.* [73] classified questions a robot could ask across three axes: K-type, the knowledge being sought; C-type, the cognitive process required to answer; and Q-type, the pragmatic expression of the question. Together, these axes aim to enable robots to generate more insightful, context-aware questions that can resolve the robot’s own uncertainty while prompting users to provide more helpful and relevant responses.

However, the model points to a tension worth acknowledging. Even well-designed questioning strategies carry a risk of being ineffective if the social or conversational features introduced to elicit better responses become distracting in their own right. This questioning strategy echoes the tutoring scenario in the study by Kennedy *et al.* [56], where the robot’s social behaviours drew children’s attention away from learning. As a result, the conversation style employed by the “asocial” robot produced better outcomes. The design of robot questioning, therefore, requires not just consideration of what to ask, but how much social elaboration around that asking is appropriate for the context.

2.2 Robot Morphology

Robot morphology refers to the physical form and structure of a robot. There is more to morphology than just appearance; it includes the body plan and spatial assembly of features, such as sensors and motors [74]. Early work in the field of HRI has long identified morphology as a determinant of how people react to robot behaviour. Specifically, Paul [75] claimed that morphology is an active part of computation, determining the function and capability of a robot, rather than a cosmetic by-product. Physical characteristics affect perceptions of robots, from feelings of comfort, trust, or even discomfort, and they can influence a user’s willingness to cooperate with a robot.

2.2.1 Physical Characteristics

Various approaches to classifying robot morphologies have been proposed in literature. Dunstan and Hoffman [24] provide historical reasoning for robot design, grouping them into three categories: “the human replica, the futuristic machine, and the cute companion”. The human replica intends to recreate and control nature, drawing parallels with slavery, where unpleasant jobs are given to those less than human but requiring human capabilities and thus resemblance. The futuristic machine is a product of advances in manufacturing, the industrial era, and the space age. Function is hidden by a sleek form, even when that hidden appearance is not necessary, to evoke “a better-than-nature efficiency” [24].

The third category proposed by the authors is the cute companion, emphasising child-like or pet-like features to build an emotional connection and commodify the use of robots for the general public. Here, the robot’s cuteness is intended to increase acceptance and help set appropriate expectations for the robot’s capabilities, inviting it to be a social participant in the home. The authors claim that all robot designs have some basis in these three tropes. However, robot designs often do not fit neatly into one of these categories, such as industrial machines and autonomous vehicles. Even within social robots, it does not include intentionally creepy robots, smart appliances, and hospitality delivery robots, for example. Instead, it is more common to refer to robot forms by what they look like.

Much of the HRI research classifies robots into general embodiment types, which are assigned by researchers or participants in experiments. Form classifications are based on what robots resemble, such as humans, animals, or various kinds of machines. The review of social robot designs by Mahdi *et al.* [3, p. 27] resulted in the following types of embodiment: humanoid, mechanoid, hybrid, android, biomimetic, animal-like, and abstract.

The first four embodiments are often referred to as anthropomorphic, as seen in [18, 76, 77], because these robot morphologies have some human characteristics.

Humanoid robots resemble the general human form, with a head, torso, and limbs. Humanoids are not to be confused with androids, which are designed to look more like an actual person, rather than humanoid robots that are still clearly mechanical. Mechanoid robots are more functional, being designed for a specific role, and are often referred to as “droids”. A greater degree of anthropomorphism means that users will interact in a more natural (i.e., human-like) way [2, 78]. Although highly realistic androids can go too far, provoking discomfort if their near-human likeness becomes unsettling rather than assuring [79].

Biomimetic robots are inspired by living things, such as animals and plants, but are nondescript enough not to resemble anything in particular. Similar to anthropomorphism, zoomorphism is when robots take on animal-like qualities. Zoomorphic designs draw on the appearance or movements of non-human living creatures to elicit a positive affective response. These positive responses are often intended to support social engagement but without the same expectations that human-like appearances can promote. Finally, abstract robots carry no resemblance to any known living form and are often more artistic. While they may be less immediately relatable, they avoid the social and aesthetic expectations that anthropomorphic or zoomorphic designs can impose on an interaction.

The ABOT (anthropomorphic robot) database by Phillips *et al.* [18] contains a curated list of robot appearance characteristics. This list was refined by study participants, who judged whether images of robots from the IEEE Robots Guide¹ exhibited each characteristic. The resulting database can be used to determine which features a robot is perceived to possess, helping inform the design of anthropomorphic robots based on their human-likeness scores. The identified features and the database itself are frequently used as measures of robot anthropomorphism, for example in [1, 19, 80, 81].

Ringe *et al.* [80] take a more analytical approach to classifying the physical components that make up the morphology of a robot with the “METAMORPH” framework by using meta-modelling. The IEEE Robots Guide was also used here to form the study dataset. Focusing on “key physical parts (such as grippers, tools, eyes, or wheels) that are visually distinguishable by humans” provides a balance between abstract and highly specific definitions like the Unified Robot Description Format as part of ROS, allowing the framework to compare robots in a structured way.

2.2.2 Impact on Perceptions

According to Hoffmann *et al.* [82], morphological structures and features can influence aspects of social interactions and emotional responses. Jones *et al.* [83]

¹See <https://robotsguide.com>.

used morphology to make robots “needy” like children or pets to help reduce loneliness. Functional robots are perceived as less warm than anthropomorphic robots, and zoomorphic robots are rated as being significantly more likeable than machine-like robots [19]. Likeability can be attributed to the emotional response elicited by each robot’s design, due to the connotations associated with animal or human resemblance.

Even between two robots with the same general embodiment, the composition and styling of morphological features can result in different likeability ratings. For example, the NAO robot was perceived as significantly more likeable than the Pepper robot in one study, despite both robots being humanoids [84]. Pepper’s height may be seen as more imposing [20], but NAO is also seen as cuter and compared to a small toy [85], making it seem less threatening than Pepper. In contrast, Akalin *et al.* [86] found that Pepper’s high perceived safety was influenced by its “curvy design that is friendly looking and engaging”.

A high degree of anthropomorphism does not automatically mean a robot is preferable to one with a lower degree of anthropomorphism. There is a risk that when the appearance is nearly human-like but not quite, it can fall into the “uncanny valley” [79], leading to the robot being perceived as creepy or causing discomfort. Specifically, the uncanny valley hypothesis states that imperfect human likeness may elicit aversive emotional responses and an “eerie” feeling in users. Whether a robot makes people uneasy is an assessment of both the robot’s morphology and its behaviours [87]. Users may perceive a robot as creepy, but this does not necessarily result in avoidance.

Research by Kunold *et al.* [19] explored the relationship between morphology and user response systematically. Using the ABOT database [18], 46 images of robots were shown to participants to gauge user perceptions and evaluations of the robots’ capabilities. Capability perceptions were measured using the EmCorp-Scale [82], which assesses users’ embodiment-specific perceptions of a robot across four dimensions: corporeality (how physically real and present the robot appears to be), non-verbal expressiveness, tactile interaction and mobility, and shared perception. Perceived capabilities were then linked to evaluations of warmth, competence, and discomfort.

The results revealed that discomfort was most strongly predicted by high perceived expressiveness and corporeality. The more anthropomorphic robots provoked the greatest unease with their realism and unrestricted expressions. Anthropomorphic heads without bodies elicited the greatest discomfort, while purely functional robots elicited the least. Building on the uncanny valley effect, discomfort also encompasses the specific combination of expressive capability and perceived physical presence. Together, these effects give the robots a sense of social agency that users may find difficult to reconcile with the robot’s form.

Appearance-function fit is another signal that is somewhat intertwined with legibility. Where role legibility is the degree to which an observer can infer from form alone what a robot is intended to do, fit is the degree to which the role that is read matches what the robot can perform. Legibility concerns whether a signal can be read; fit concerns whether the appearance seems appropriate for the answer quality. A form can therefore be legible but misleading, projecting a role the robot cannot fulfil, or capable but illegible, offering no readable role at all.

Naya-Varela *et al.* [88] tested a positivity hypothesis, in which more appealing and more human-like robots are always better received, against a matching hypothesis, in which appearance should correspond to the task at hand. Participants preferred robots whose human-likeness matched the sociability the job required, and complied more with a robot whose manner matched the seriousness of the task, supporting the matching account. Both signals have been studied mainly for physical tasks, where a mismatch between implied and actual capability is exposed by what the robot does. This comparison has been less examined where a robot’s function is to be an information source, and how the two signals operate requires further investigation.

2.2.3 Morphological Signals

Before an interaction even takes place, perceptions of a robot begin to form [20] due to previous experiences and mental heuristics [89], exclusively from what can be seen. Embodiment type, size, and shape can each impact perceived safety. Haring *et al.* [25] found that an android-style robot was regarded as being much less safe than a more loosely anthropomorphic (i.e., humanoid) robot, as well as a non-biomimetic robot. As previously mentioned, the appearance of NAO also signalled better safety in addition to likeability compared to Pepper [84].

Making judgements based on appearance extends beyond safety to other aspects of trust. Schaefer *et al.* [20] identified a strong link between a robot’s physical form and perceived trustworthiness, with morphology acting as a proxy for capability in the absence of direct experience with the robot. Perceived intelligence is a key determinant of trust. The a robot’s morphology is what changes perceptions and attributions, as the underlying capabilities are often mostly similar. Differences in how intelligent and, therefore, how trustworthy a robot appears are driven by visible form rather than actual capabilities. A mismatch in these performance expectations can negatively affect the success and effectiveness of the interaction before it even starts.

In addition to affective responses, capability attributions are also affected by physical form. A robot’s constituent morphological features limit or enable different interactions, due to how it can manipulate the environment [19, 82]. Manipulators (e.g., hands and grabbers) imply, but do not necessarily mean, that a robot is

capable of handling objects, and eyes imply awareness of the world around the robot. The presence of hands contributes to an anthropomorphic appearance, and with that, expectations that the robot will be able to accomplish tasks a human can, and in a similar manner. Robots without these capabilities were perceived to be less competent by Kunold *et al.* [19].

The physical form of a robot helps set expectations for the interaction, encompassing both capability and intention, much like non-verbal social cues people use [90]. Heuristics from the form help people understand how a robot works and what it is trying to do, making it more predictable and approachable. Competence at completing a task is only part of the picture. Whether a user will put their faith in a robot is not solely an evaluation of objective task performance; it also comes down to perceptions [32].

According to Aly and Tapus [91], morphological features are the platform that enables robot expression, giving robots a personality that speech alone cannot achieve [92]. Eyes and other facial features are part of the affective communication channel [18] that help signal social expectations, hinting at what to expect. Lighting is one visible feature independent of embodiment type that can be used as a social cue [93]. Sound can have a similar purpose in explaining what a robot is “thinking”, but is not a part of morphology [94]. The inclusion of different features needs to be carefully considered, however, as unmet expectations can alter trust in the robot and in the interaction itself [88], and thus needs to be managed appropriately.

How an appearance is received varies by culture. In an interview by Li *et al.* [77], participants believed that a robot’s appearance influenced the task it was appropriate for. However, when these participants rated the robot’s performance after observing its interaction, their ratings did not differ significantly by appearance. The researchers did find that perceptions can vary across cultures. For example, they state that “masculine cultures prefer large and quick robots to small and slow ones, while people from feminine cultures hold the opposite opinion”.

Age is another factor shaping robot perceptions, with initial work suggesting that children respond differently to anthropomorphic and zoomorphic robots than adults do [95]. Context also matters when robots are integrated into human teams. Morphology affects how the robot as an individual and its identity as a group member are perceived, with anthropomorphic designs facilitating team cohesion [34].

A consistent pattern can be seen in people’s attributions of agency to robots. Both anthropomorphic and zoomorphic robots are more readily perceived as intelligent and intentional than more functional designs, often likened to living things capable of acting independently [81, 93]. Agency attribution shapes how users engage with and relate to a robot, making it more likely to be treated as a social actor rather than a tool. Changing these expectations of the type

of interaction changes the conversations had and how the robot’s behaviour is interpreted. These expectations then progress to their willingness to trust it. For designers, this means morphological choices carry social and relational implications that extend beyond appeal. A robot’s form signals what kinds of interactions can be had with it.

Addressing the gap between intended and perceived purpose has become an active area of HRI research recently. The HuGo robot [96], for example, enables tailoring its appearance through modular face components, allowing its form to more accurately represent its purpose in a given context. Allowing users to customise appearance can also help interaction success [97] and robot trust [35].

Especially for social robotics, facial expressions and movement animations help convey social cues, such as gaze, emotions, intentions, and gestures. Social cues can help support more successful and appropriate interactions. Applying animation principles to robot movements can make these cues more effective by making them more legible and assisting with perceived intent [98]. That use of legibility concerns the readability of motion and intent. Role legibility, as used in this thesis, extends the same idea to static form, with the readability of a robot’s intended role from its appearance alone. Custom appearances and animations can help align a robot’s form with a user’s own expectations, rather than relying on a designer to anticipate them. Together, these approaches help bridge the gap between a robot’s intended role and what expectations the design actually promotes.

2.3 Evaluation Frameworks & Methodologies

User acceptance is one of the primary determinants of whether a technology succeeds or fails when deployed [99]. Understanding how and why users interact with technology the way they do is therefore an important consideration when designing it. For many digital systems, acceptance is largely predicted by usefulness and ease of use, or by how efficient a user expects it to be. Robots, however, are more complex as they occupy physical space and can replicate human roles autonomously. Social, emotional, and psychological factors are also relevant, which is why researchers have developed a range of approaches to evaluate interactions.

2.3.1 Standard Measures

Davis [99] proposed the Technology Acceptance Model (TAM) in 1989 to help understand user behaviour when encountering new technology. The TAM and derivatives have been used to explain and predict the adoption of new technology, based on perceived usefulness and perceived ease of use. Perceived usefulness is

defined as “the degree to which a person believes that using a particular system would enhance his or her job performance” [99].

Usefulness stems from the belief that in corporate situations, improved job performance is often linked to promotions and raises. Usefulness then means that if using a technology does not contribute to this goal, it is not perceived as useful. Conversely, the technology is likely to be adopted if the user can see a clear and positive relationship between its use and professional success. Ease of use refers to the perceived degree of effort in using the system, where even if the technology is useful, if it is too hard to use, then the “performance benefits of usage are outweighed by the effort of using the application” [99], hindering adoption. The concepts of usefulness and ease of use form the basis of technology evaluations.

Ultimately, the success of introducing a new technology comes from the cost-benefit calculation the user makes when it is proposed. Empirical validation of the TAM revealed significant positive correlations between perceived usefulness and perceived ease of use, both of which were strongly linked to self-reported current and predicted future usage. Of the two, perceived usefulness proved the stronger and more direct predictor, with ease of use appearing to influence usage indirectly by shaping perceptions of usefulness. While the TAM provides a baseline framework for general information systems, the unique social and physical aspects of robotics have led to more specialised tools for measuring user perceptions, such as the Godspeed questionnaire and the Robot Social Attributes Scale.

The Godspeed questionnaire series, developed by Bartneck *et al.* [100], was a foundational effort to address the need for a standardised tool for comparing human-robot interactions across studies. The Godspeed questionnaire is frequently used in HRI research to quantitatively measure the subjective perceptions of robots. Five core concepts were proposed in the questionnaire for evaluating a robot: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety, each of which is measured with multiple semantic differential scales. Despite its widespread adoption, Godspeed has been criticised for lacking psychometric validity, with its dimensions not strongly tied to the concepts it was intended to cover [101].

The Robot Social Attributes Scale (RoSAS) by Carpinella *et al.* [102] aimed to simplify Godspeed down to just perceptions of warmth, competence, and disturbance. RoSAS also faced similar limitations, with ambiguous results when the robots shown were systematically varied, and the explained variance being below the threshold for strong construct validity [81].

Learning from this, Spatola *et al.* [81] developed the Human-Robot Interaction Evaluation Scale (HRIES), integrating psychosocial theory to enable more nuanced measurement of user perceptions. Through a series of validation studies, the scale demonstrated strong reliability across four dimensions: sociability, agency, animacy,

and disturbance. With this scale, higher reliability and sensitivity were demonstrated in real-world experiments than when static images of robots were shown, possibly indicating that perceptions are strongly affected by actually interacting with a robot. In one study, the HRIES explained 71.82% of the total variance in participant responses, exceeding the recommended 60% threshold that previous scales had failed to meet.

The context of the interaction plays an important role in measuring interactions. The situation in which a robot is used requires different approaches. In task-based interactions, success is readily apparent and often reflects efficiency, whereas for social robots, success may instead be measured by user engagement. Paetzel-Prüsmann *et al.* [103] found that for social robots, lower efficiency is preferred as a consequence of the robot being engaging and enjoyable, having longer sustained and more frequent interactions.

As robots become capable of increasingly complex and adaptive behaviours, the line between a functional tool and a social actor becomes harder to draw. When evaluating success, user satisfaction and task performance are not necessarily aligned. A robot can perform well by objective measures while still being poorly received, or vice versa. Evaluation methods must be chosen that are appropriate to the robot’s intended role and context. Because many factors can mediate how an interaction unfolds, the context of any study should also be carefully considered, reported, and accounted for when interpreting and comparing findings.

Taxonomies such as [1] and [80] provide holistic frameworks for analysing robot interactions, facilitating comparisons between studies. Studies may appear to test the same variable, but their structures can vary, so they may not actually be comparable. The taxonomy by Onnasch and Roesler [1] categorises interaction elements into three main clusters: context, robot classification, and team classification. Context includes environmental factors like obstacles and lighting. Context also includes the task, such as information exchange or cognitive stimulation, where robots engage humans through verbal and non-verbal communication [1].

Morphology falls under the robot component of the taxonomy. Morphology is relevant to the robot because visual associations with other objects drive initial expectations upon first contact. Closely related are a robot’s available communication channels, which are “how information from the human is ‘perceived’ by the robot” and how the output is “perceived by the human senses”. Together, morphology and communication channels make it clear what a robot supports, changing how an interaction will go.

Research often focuses on the controllable characteristics of robots, but the H in HRI is just as important. For example, the ratio of robots to humans is not inherently positive or negative, but it can significantly alter the dynamics. Together, the taxonomy addresses complexities that simpler tools, such as the Godspeed

questionnaires, cannot capture, highlighting the importance of tracking nuanced factors. A useful illustration is made that suggests R2-D2 is more likeable than C-3PO because of its non-verbal cues, emotional beeps and tones, and not being presented as intelligent. This example aligns with the effect of human emotional needs [83]. Using a comprehensive taxonomy to classify interactions can be used to intentionally structure studies and isolate the factors of interest.

Physical form is a primary part of user perceptions, shaping psychological outcomes, notably trust. Trust is defined in many different ways, as we discuss next in Section 2.4. One popular definition is “the reliance by an agent that actions prejudicial to their well-being will not be undertaken by influential others” [21]. Trust is often identified as a critical component of an HRI relationship, so accurately measuring how trust is calibrated is crucial.

A lack of trust can lead to neglect or disuse of a robot, but too much trust can result in disappointment or complacency [104], which is especially dangerous where robots are used in high-risk situations. High-risk situations make accurate trust measurement important, and a robot’s morphological dimensions and a user’s perspectives of them are key inputs.

A robot’s appearance directly shapes user expectations and trust calibration; the “cute companion” [24] is an example of leveraging specific aesthetic attributes to set user expectations and evoke a particular kind of trust. The sociability and agency attributes outlined in [81] serve as direct inputs to trust assessments. Agency covers perceived intelligence and intentionality, both of which predict trust in robots [20, 105], and sociability covers the affective side of trust built through perceived warmth rather than performance [17], so the scale measures the perceptions that feed a trust judgement rather than the judgement itself. Morphology is an input, not exclusively cosmetic; it is a channel for communicating the nature and intentions of a robot and, consequently, a user’s willingness to trust it, all of which can be evaluated.

Morphology is not a design choice made in a vacuum. Interpretations of how a robot looks can stem from cultural traditions and a user’s unique perception of its features, leading to different meanings for each person, as detailed in Section 2.2.3, and are therefore also relevant to evaluations. Robots that fall into similar categories, such as those labelled as “humanoid”, can also be drastically different. Because of this, more systematic breakdowns of robot appearance, like the ABOT database [18] and the METAMORPH framework [80], are essential to make morphological assessment consistent and comparable.

To build the ABOT database, the researchers conducted a principal component analysis from user ratings of robot images to identify four distinct appearance dimensions of anthropomorphic robots: surface look, body-manipulators, facial features, and mechanical locomotion. METAMORPH takes this further to classify

robots of all types, not just anthropomorphic ones. The framework looks at morphological subdivisions, descriptions, and silhouettes. To maximise relevance, Ringe *et al.* [80] state that morphology analysis depends on how it integrates with the broader context of the interaction. Specifically, the authors propose integrating it with the popular taxonomy by Onnasch and Roesler [1] to help evaluate interactions.

2.3.2 Experiment Designs

The design of a study affects what the measures can actually do. An experiment in HRI is defined by how participants are allocated to conditions, what is manipulated and what is held constant, the task and setting the interaction takes place in, and the kind of evidence of trust that is collected. We review how existing work has made these choices to set out some decisions that need to be made in planning a study of morphology and trust in information-seeking.

Adamik *et al.* [35] allocated participants between subjects. One group personalised the colour and name of a virtual robot before playing a dice bluffing game against it, while a control group only saw an introductory video. The robot behaved identically for both groups and the game was fixed, so every participant faced the same declarations and dice results, leaving personalisation as the only difference. Participants were not told that other conditions existed. Sun *et al.* [7, 10] took the opposite approach, having each participant complete health searches through a text interface, a smart speaker, and a small embodied robot, which were all connected to the same language model, in counterbalanced order. Allocating within subjects is more efficient and shows how trust differs within a person, though every participant sees the manipulation which may not be desirable in some circumstances.

Akalin *et al.* [86] used a mixed-methods design, varying whether faulty robot behaviour occurred early or late between subjects and using five scripted behaviour conditions within subjects. Their sessions ran up to an hour because a questionnaire battery followed every condition. A study such as this needs to explicitly recruit participants, as repeated administration of validated instruments assumes a captive participant and cannot be expected of a brief or voluntary encounter. Which method of participant allocation will suit a study depends on whether seeing more than one condition would compromise the manipulation. Interfaces that differ substantially, as in [10], are meant to be directly compared. Robots that differ only in appearance while behaving the same way are not, since a participant who meets several of them can infer the purpose of the study and overstate differences in ratings. Between-subjects allocation protects the manipulation at the price of more participants and comparisons that can only be made between group means.

Reported and enacted trust can diverge within one interaction, so the type of evidence a design collects can matter as much as the quantity. Adamik *et al.*

[35] collected both, administering the Trust Perception Scale-HRI [106] before and after the game, alongside behavioural measures taken from the game logs. Questionnaire scores did not differ significantly between the groups, but the median trust limit, the rank of the declared value at which a participant chose to challenge the robot, was significantly higher for those who had personalised it. A non-significant questionnaire result does not prove the groups felt the same, though a design relying on self-report alone would have reported no effect where behaviour showed one.

Qualitative data can help contextualise responses alongside other measures. Ratings in Akalin *et al.* [86] showed trust falling when the robot’s speech recognition failed, and the interviews then explained the drop, as participants could not tell whether the robot had malfunctioned, or whether their speech was simply not heard. Where trust forms is a further consideration, since the HRIES proved more sensitive during live interaction than when static images were shown [81], so perceptions formed by observation need not match those formed through use. A design that separates the two can show where trust is established. In all of these studies the evidence was gathered within a single sitting, describing how trust forms and is revised during a first encounter, rather than how it develops over repeated use.

The task and setting fix both the behavioural proxies available and how far findings generalise. For example, structured tasks can offer precision. The bluffing games in [35] and [107] produce a countable decision point, the quiz in [86] allowed faults to be introduced at controlled moments, and the tutoring task in [56] was able to measure learning outcomes. A negative of designs like these is that participants follow instructions rather than pursue anything of their own, so the findings do not necessarily reflect what users would do in the real world.

Akalin *et al.* [86] note this as a limitation of their study, reporting that participants felt comfortable because the experimenter was available, but call for data collected in uncontrolled environments. Their objective measures depended on that setting as well, since wristband sensors and video recording require fitting, consent, and supervision that an unattended interaction cannot provide. Information-seeking has been studied under similar constraints. Sun *et al.* [7, 10] assigned health questions to be answered through each interface and measured trust in the information separately from trust in the interface, finding that information from the embodied interface was trusted significantly less than the same information presented as text. Hannibal *et al.* [16] used an everyday shopping situation to observe vulnerability as it arises ordinarily. Only in a naturalistic setting does the choice to approach a robot become observable, and that choice is a trusting behaviour in its own right, which cannot occur where participation has already been agreed.

Studies also vary in what they manipulate. Haring *et al.* [25] manipulated morphology, comparing an android, a humanoid, and a non-biomimetic robot.

Kennedy *et al.* [56] manipulated behaviour and found a socially expressive tutor less effective than an asocial one. Hei *et al.* [70] manipulated the participant instead, briefing one group about a robot’s capabilities and finding that the briefed group sought more help and performed better. The result in [56] carries the most weight for design, since behavioural differences produced an effect at least as large as those elsewhere attributed to form. Attributing an effect to morphology therefore requires behaviour to be identical across robots, as achieved in [35] by scripting the robot and fixing the game, and [10] by using one language model behind every interface. Behaviour is not the only thing that has to be held constant, as setting, task, proximity, and the participant’s role also vary between studies and each can alter the result. The taxonomy by Onnasch and Roesler [1] provides a structured way to record these dimensions and confirm which of them differ between conditions.

Participant characteristics moderate all of these effects and are better planned for than discovered afterwards. Akalin *et al.* [86] found that male participants reported feeling safer and more in control, and that higher neuroticism was associated with lower perceived safety. Culture can shift ratings as well [25, 77]. Prior experience can matter too; Sun *et al.* [10] identified familiarity with an interface as the strongest determinant of trust in the information it delivered. As participants trusted the text interface because it resembled the tools they already used, a familiar consumer product does not enter a study neutrally. Such characteristics should be recorded where the length of the instrument allows it, and acknowledged as a limitation where it does not.

Several requirements for studying morphology and trust in information-seeking follow from this review. Participants should encounter a single morphology, so that robots programmed to behave identically are never compared directly by the same person, and both reported and enacted evidence should be collected because the two can diverge. The task should be one that participants choose to undertake, which makes engagement itself observable, and every dimension of the interaction apart from morphology should be recorded and held constant against an established taxonomy. Participant characteristics should be recorded where the design permits and acknowledged where it does not. A design of this kind cannot compare individuals against themselves or show how trust changes over repeated use, and its findings are tied to the setting studied. Chapter 3 contains the methodology built to meet these requirements.

2.4 Definitions of Trust

Trust is a key psychological factor that influences a person’s acceptance of, reliance on, and effectiveness of collaboration with a robot. Trust is not a simple variable, as it encompasses prior experience, context, a system’s perceived capabilities, and

the consequences of misplacing it. Complexity in the different aspects means that both too much and too little trust can lead to poor outcomes. In manufacturing and industrial automation, a lack of trust can lead to the disuse or underutilisation of systems that may improve productivity and safety. In social care and other high-stakes domains, the opposite problem is also important. An over-assignment of trust can result in misuse or complacency, which comes with safety risks.

Efforts must be made to support a user’s critical thinking. Support is achieved by helping a person calibrate trust so that their reliance on a robot accurately reflects its capabilities and limitations, and understanding how important relying on the information is. Achieving this requires an understanding of how trust forms, how it is maintained or broken, and how it can be measured. This section reviews how trust has been conceptualised from broader automation to the specific dynamics of trust in HRI.

2.4.1 Trust Measurements

Researchers across psychology, HCI, automation, and robotics have proposed various models and definitions of trust. In 2004, Lee and See [13] defined trust as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability”, which is frequently cited in computing research. This definition is supported by earlier cross-disciplinary work by Rousseau *et al.* [108], using the widely accepted conditions for trust: the presence of both risk and interdependence [15].

Subjective measures, specifically self-report questionnaires, are a simple and common approach for assessing a user’s thought processes and attitudes when dealing with a robot. A methodology using subjective measures aims to turn feelings of trust into quantitative data on beliefs, feelings, and intentions. For subjective measures to have valuable contributions, the questionnaires should be standardised and validated. Bartneck *et al.* [100] state that the use of unvalidated scales does not help research as they produce unreliable data and make it hard to compare results across studies, hence the creation and widespread use of the Godspeed scale. Rigorous psychometric scales are necessary to ensure that they consistently and accurately measure the intended construct.

As the context changes perceptions of an interaction, using an approach specific to it can be helpful. The human-robot collaboration scale by Charalambous *et al.* [29] was developed for industrial settings. Rather than adapting existing measures, they first explored using qualitative methods, such as semi-structured interviews, to identify recurring themes and concepts in robot trust that are relevant to the workers collaborating with the robots. Then, they generated a pool of questions and used principal component analysis to identify underlying factors, refine the items, and assess the scale for reliability and validity. Grounding the scale in

workplace experience distinguishes it from general measures that may not fully capture the relevant concerns.

There are many distinct facets of trust, as identified in the meta-review by Law and Scheutz [109] and confirmed by Campagna and Rehm [110], but all of these can be broadly framed as performance trust (task execution) and moral or relational trust (social behaviour). Malle and Ullman [111] presented this framework as a basis for understanding the performance and moral aspects of trust. Performance trust is particularly relevant in industrial and military settings, where robots must complete tasks safely and effectively. Moral and relational trust focuses more on how well a robot adheres to social and ethical norms, acting more like a person.

Performance trust covers reliability, where consistent and predictable operation is expected, capability in the skill and knowledge to perform a task, and competence demonstrated to achieve goals [111, 112]. Moral and relational trust includes sincerity, truthfulness and honesty in communication, integrity in adhering to principles, and benevolence, the belief that a robot’s intentions are aligned with the user’s well-being [111, 113]. While trust of information provided by a robot is implicitly a part of performance trust, as it is how well the robot communicates the information, in social robotics, the relationship impacts whether the robot is trusted as a source of information. All of the trust factors need to be considered together to determine trust in robots, particularly in information-seeking contexts where these factors are closely linked.

The Godspeed questionnaire series [100] measures HRI concepts related to trust, such as perceived intelligence and safety, but does not address trust as a construct itself. The Multidimensional Measure of Trust by Malle and Ullman [111] specifically aims to assess both performance trust and moral trust as mentioned. There are also instruments tailored to specific areas, like the industrial-focused Human-Robot Collaboration scale [29], and the HRI trust scale by Yagoda and Gillan [30] for robot integration into teams.

Subjective measures are a straightforward approach, but also come with some limitations. Questionnaires are typically administered after the interaction, meaning participants have to rely on recollection, so they may not accurately represent their true feelings in the moment [100]. Participants may also either consciously or unintentionally tailor their responses to fit what they believe to be the acceptable or expected answer, rather than expressing how they actually feel. Trust is also a dynamic cognitive value that can change over the course of an interaction, and a single questionnaire provides only a static measurement rather than capturing the variations that occur [110].

Objective measures often aim to address some of the limitations of subjective scales by focusing on quantifiable, observable aspects of an interaction rather than relying on human perceptions. Objective methods examine users’ behavioural

choices and physiological responses, which are assumed to be manifestations of trust, or at least correlated with it. An approach like this should be more impartial, but it is typically limited to measuring proxies of trust rather than trust itself. One of the most commonly measured proxies is reliance, which refers to a user’s willingness to delegate tasks to an automated system. Trust is then inferred from how much responsibility is assigned and the frequency of that, or how often the user steps in. More delegation and less intervention imply higher levels of trust [13].

Trust can also be inferred from a user’s choice to follow a robot’s advice [23]. Non-verbal cues from the person, such as posture and movement, can be monitored for signs of apprehension, as observed by Hald *et al.* [114]. Apprehension is indicated by cautious or hesitant movements, which suggest lower trust in the robot. Kahn Jr. *et al.* [115] systematically coded videos of interactions to identify and categorise specific social behaviours like dialogue, touch, and gestures. Contextualised behavioural coding captures trust-relevant responses as they naturally emerge during interaction, rather than relying on users to accurately self-report their feelings after the fact.

These two types of measures differ in the kind of evidence they provide. Lee and See [13] define trust as an attitude and distinguish it from reliance, the behaviour that follows from it. In this thesis, this distinction is referred to as reported trust and enacted trust. Reported trust is what a person states about the trust they place in a robot, whether through validated scales, ratings, or verbal answers.

Enacted trust is what a person is observed to do, such as choosing to approach a robot, rely on its answers, and not checking them. It is inferred from behaviour rather than stated by the person, and requires the context to be interpreted. The same action could indicate trust in one setting, but it could also be due to habit, compliance, or a lack of alternatives in another. The distinction is one of measurement rather than of interaction, since both kinds of evidence can be collected from the same encounter.

Reported trust can also cover enacted trust, such as how likely a person is to rely on a robot. Reported trust and reporting on enacted trust can differ based on what is asked. The distinction matters because the two do not always align. A person may rate a robot highly and then avoid it, or use it readily while describing it as unreliable. Neither form of evidence is more valid than the other as they relate to different types of trust being used. Reported trust describes the attitude a person holds and can articulate, while enacted trust describes the behaviour that attitude did or did not produce. Studies that measure only one are therefore making a narrower claim than their framing sometimes suggests, and comparisons across studies are complicated when the same term is used for both.

Adamik *et al.* [35] found that questionnaires administered after a genuine interaction showed similar trust across groups whose behaviour within that same

interaction had demonstrated different trust, showing that what people believe about their trust and what they are seen to do can diverge. Stating which kind of evidence a study captures is therefore necessary for its findings to be interpreted correctly.

The signals a person’s nervous system sends can be captured to identify subconscious reactions correlated with trust, stress, or comfort. Measuring physiological responses provides continuous, real-time data, but it cannot determine valence (whether the response is positive or negative) or intensity. A response does not say what the meaning behind the responses is without additional information, only that they are occurring, which can be matched to what happened in the interaction.

A validated subjective measure can serve as “ground truth” to train machine learning models, using rich sensor data as classifiers to determine what a user is feeling. Shayesteh *et al.* [31] used EEG signals to output corresponding trust levels, as either high or low. Even though the data only shows when there is a strong response, the model can learn the patterns to then output a measure of trust. However, this approach requires more training and validation to be a reliable way to measure trust exclusive from other emotions. Further, it may be regarded as invasive.

Similarly, Khalid *et al.* [116] used facial expressions and vocal features to detect the user’s trust state. A synthesis of other examples of inferring trust from other events can be seen in [110]. Capturing non-verbal communication can be used non-invasively by a robot to tailor its response, optimising team performance.

The Trust Perception Scale-HRI (TPS-HRI) [106] is a widely used measure in HRI research. Participants rate statements about the robot as the percentage of time each holds true, covering both its functional performance and how it conducts itself towards the user. The scale builds on the three-factor model of trust by Hancock *et al.* [117], developed through a meta-analysis of trust in HRI, which identified human-related, robot-related, and environmental factors as the primary determinants of trust; the TPS-HRI adapted these into its questions.

The HRI trust scale [30] takes a broader view by distinguishing between trust that exists before an interaction and trust that develops during an interaction. Dispositional trust reflects a user’s general attitude towards robots based on prior experience and expectation. The interaction trust is assessed across five factors: team configuration, team process, context, task, and system. By separating these, the scale acknowledges the existing beliefs that users bring, which changes how they interpret and respond to a robot’s behaviour. However, this scale has a strong focus on collaborative industrial settings, making it less appropriate for general tasks.

Revisiting and building on work from 2011 [117], Hancock *et al.* [21] conducted a meta-analysis on the factors that affect trust in HRI, drawing on quantitative

findings from psychology, computing, and other fields. The meta-analysis is distinct from work in automation, as it focuses on identifying factors specifically relevant to robots and recognises that not all robots possess automation capabilities. With trust factors, the “interactive dependencies illustrate how difficult it is to promulgate and verify any simple model of trust” [21]. The researchers recoded their findings and identified new ones from the last decade (up to 2021), resulting in 45 articles remaining after exclusion.

The new factors fit into the three dimensions previously identified, which were human (abilities and demographics), robot (performance and attributes), and context (environment, team dynamics, and task). The ultimate definition of trust settled on was “the reliance by an agent that actions prejudicial to their well-being will not be undertaken by influential others” [21]. The authors determine that trust is about loss avoidance by managing over- and under-reliance, disuse, or abuse. In other words, trust is not about good or bad, but about being appropriately aware of the situation.

The analysis found that robot-related factors were among the strongest predictors of trust, with an overall effect size of .21. Within this, morphology specifically had an effect size of .31, while robot personality, encompassing facial expression, empathy, likeability, and sociability, was the strongest single factor at .49. The findings also highlighted anthropomorphism as a human tendency rather than a design feature. The inclination to attribute human characteristics to non-human objects was itself a significant predictor of trust, and likely amplifies the effects of morphology in studies where participants tend to anthropomorphise robots. Further, the analysis found that perceiving a robot as more human-like leads to greater trust, reinforcing the importance of appearance as a trust-relevant variable.

Communication method and human personality were especially influential, as were shared factors (belonging to the same team or group). These human factors inform us “how this partnership should evolve, via design, training, and combinational work process development” [21]. The intention is to increase trust without negatively affecting other aspects of the interaction, as happened with the more sociable tutor in the work by Kennedy *et al.* [56]. The authors conclude with a prediction that time and dynamic dimensions will emerge as critical factors. This meta-analysis is limited to what has been investigated in existing literature, so it cannot speak to factors that have not yet been investigated or whose findings have not been published. The trust model can be seen in [21, p. 1212], which explains the different areas of trust and how they work together.

2.4.2 Trust in Automation

Whether people choose to use a new technology or not comes down to whether they can place trust in it. In the 1989 work on technology acceptance by Davis [99],

perceived usefulness and perceived ease of use were identified as the fundamental determinants of users' acceptance or rejection of technology, as detailed in Section 2.3.1. Ease of use drives how useful something is perceived to be, which in turn affects usage. Users value systems that are easy to operate, but ultimately prefer a tangible performance benefit.

Subsequent work by Muir and Moray [112] demonstrated that perceived competence is the dominant predictor of trust in automation. In their simulation, an operator's perception of the system's competence accounted for 81% of the variance in trust ratings. This work also highlighted the fragility of this trust, as any sign of incompetence or unreliability (through performance variations) eroded trust quickly and required considerable time and effort to rebuild, suggesting that a user's trust in a system comes from its worst rather than its average performance.

The review by Wagner and Arkin [15] establishes that trust is important to an interaction only in certain circumstances. The researchers found that the two preconditions for trust are risk and interdependence. Risk refers to the potential loss that may occur, like when relying on the system for something critical. Interdependence is whether mitigating that risk depends on that system, or the amount of user control in preventing risk. When these conditions exist, the user must decide whether to trust the system. Effectively, risk implies interdependence, which together can imply trust. In this case, trust is the belief that the system will mitigate the user's risk when they are vulnerable and relying on its performance. The need for trust and the circumstances around it then determine what the user will do with the technology.

Interdependence means the user plays an integral part in the trust equation, not just the robot. The system being judged is the basis for that equation, but it comes down to how the user makes the appraisal, based on their own capabilities, prior experience, emotional state, and perceived needs. Research by Liao *et al.* [28] revealed that a user's perceptions of their information-gathering ability were a strong moderator of trust. In particular, the choice between relying on AI and relying on a person. If a person is unsure of their ability to gather information, they will choose an AI over a person, but if they are confident in their abilities, they are more likely to choose a person for the quality of the information.

An artificial agent may be seen as a more capable and efficient partner, and therefore more trustworthy, or a person may be seen as more reliable in helping alleviate their information deficit. Building on [99], usefulness is now about the user's abilities relative to the AI's. Firmino de Souza *et al.* [118] reviewed human-related factors that facilitate or block trust in HRI, and found psychological safety, familiarity with technology, and tailoring the experience all facilitated trust, whereas emotions (e.g., anxiety, fear, and stress), complacency, and privacy concerns are barriers to trust. Gender, emotional state, a need for control, and the user's

willingness to take risks also played a part, confirming that trust is very dependent on the person and context, and has some factors that are static and some that are dynamic.

The design, behaviour, communication medium, and style of a technology are the inputs a user evaluates when deciding whether to rely on it. One behaviour is transparency; whether a user can understand a system’s purpose, inner workings, and ownership helps facilitate trust [6, 118, 119]. Ownership comes with responsibility and accountability, so knowing who the organisation behind the system is and how to contact them helps build trust in the system [120].

Third-party validation, recommendations, and testing provide a reason to trust a real-world system. A robot should communicate in a helpful way without being intrusive, as an annoying or pestering robot can be seen as manipulative, rather than working for the user’s best interest. Intent also plays a core part; clearly communicating the system’s purpose and capabilities helps create more accurate and stronger expectations of its behaviour.

Hertzum *et al.* [6] found that first impressions matter. Surface attributes, such as visual appearance, influenced a user’s initial trust. In focus groups, interactions with shopping websites revealed that users reacted more positively to naturalistic, photo-like images of people than to CGI talking heads, which were perceived as artificial and silly.

Appearance matching the domain meant the agent was more relevant, knowledgeable, and trustworthy. While seemingly superficial, these attributes provide cues for users when functional performance information is not present. Aside from what the system does, how it approaches it can affect trust [112, 118]. A system that performs worse overall but with little variation can have more trust placed in it than one that is, on average, better but inconsistently. Hesitancy signals a lack of competence, which undermines trust. Systems used for social interactions may also violate social norms that affect perceptions of competence and reliability.

Public sentiment and consensus shape an individual’s attitudes towards technology. Research on informational subjective norms [28] (social pressure with seeking information) found that they can polarise user preferences between humans and AI. In the American-based samples, the norms encouraged seeking information from humans but actively discouraged seeking information from AI agents. When there is a strong social preference, it can create an incentive to conform that overrides other factors.

The source of information about AI (not AI as a source) has a large impact on trust in that automation. In a survey of people in Malaysia, Indonesia, Singapore, and India by Neyazi *et al.* [121], perceptions of AI varied between those who obtained their information from traditional news sources and those who primarily got their news from social media. Social media users in these countries were

more likely to trust AI compared to traditional media consumers, possibly due to government narratives or their experience with technology.

Being more likely to trust AI was moderated by a concern about misinformation; traditional media consumers in Singapore, who were concerned about misinformation, were less likely to trust AI than social media consumers who were also concerned about misinformation. In fact, social media users in Singapore had strengthened relationships with AI despite concerns about misinformation. The research attributed the strong relationship to the perceived efforts by social media platforms to combat misinformation and to the wider variety of sources, compared with the professional, focused narrative from traditional media in other countries.

Sun *et al.* [7, 10] compared trust in Google and ChatGPT for health information-seeking, and found that users typically trust ChatGPT more because of its conversational style that provided structured, coherent, and professional responses. The ease of getting an answer enhanced its perceived reliability. The user’s familiarity with the system reinforces confidence in the interface; however, those with more technology experience tend to trust information from ChatGPT more. Text-based interfaces were preferred for serious topics, as they are less prone to commercial filtering or bias. Those with limited information on the topics preferred the greater control in searching offered by Google, allowing users to make their own decisions like a human source would.

2.4.3 Trust in Human-Robot Interaction

As a definition for what trust means in HRI, the meta-analysis by Hancock *et al.* [21] pulled together all of the definitions of trust to come up with “the reliance by an agent that actions prejudicial to their well-being will not be undertaken by influential others”. Being compiled from multiple papers, this definition goes further than definitions of trust in automation to highlight the elements of risk, vulnerability, and the social aspects associated with physical presence.

Several frameworks have decomposed trust in robots into distinct dimensions. Malle and Ullman [111] distinguish between performance trust, concerned with functional reliability and capabilities, and moral trust, involving perceived sincerity, integrity, and ethical standing. Moral aspects grow more relevant as robots take on more socially consequential roles.

Drawing on interpersonal trust models, Manor *et al.* [17] distinguish between the rational and emotional components of trust. Particularly for social robots, affective trust is developed by emotional bonds, feelings of closeness, empathy, and the perception that the robot is a caring and supportive partner in the interaction. Affective trust can be likened to the personal connection humans have, such as what a person feels when a barista remembers their order, and the faith they will get it right. The cognitive aspect stems from a logical assessment of a robot’s

performance, reliability, and predictability, rather than feelings.

A further distinct aspect of trust concerns its origins rather than its nature. Dispositional trust is a user’s propensity to trust others, including robots as social agents, whereas history-based trust is dynamic and evolves through interactions, along with the user’s experiences of the robot’s behaviour and performance. Disposition and history are the difference between a user’s baseline tendencies and the trust earned over time, respectively [30, 122]. The various components of trust in robots come from the social aspects, a robot’s design, and performance-based factors. HRI research has identified many aspects across various studies. Looking at the different parts together helps clarify what trust means in HRI.

The trust people place in robots can be measured through self-report measures as well as through observations of a user’s actions and decisions. Studies using the latter approach often use tasks that require participants to place high reliance on a robot. An example of reliance is the choice to follow a robot’s advice in a problem-solving game [107, 123]. Conversely, low reliance may be seen in a decision to seek a second opinion from a human when receiving a medical diagnosis [124]. Reliance could also be demonstrated by the choices made in games that involve financial risk and cooperation [33]. Each of these approaches measures trust in robots in different ways, across different aspects of trust and risk levels. Because of the various ways trust in a robot, or a lack of it, can be demonstrated, different approaches to measuring it need to be used.

As described by Hancock *et al.* [21], the factors relating to the robot itself are the strongest and most significant predictors of trust. Robot factors include design attributes and operational performance, which the user uses to form their initial and ongoing assessments, respectively. The meta-analysis found a significant positive correlation between robot reliability and human trust. Better reliability leads to more trust, and worse reliability leads to less trust.

In contrast, robot errors, malfunctions, and system failures damage trust [86, 125]. Performance problems violate established expectations and increase the perceived risk of collaborating with a robot. First impressions matter; Xu and Howard [123] demonstrated that a robot’s performance in the first interaction had a significant and lasting impact on perceptions. Users who saw a correct, high-performing robot on their first encounter were much more willing to trust its advice in subsequent interactions, even if it later made mistakes, compared to users whose first encounter was with a “faulty” robot.

Avoiding this kind of trust damage depends on the users being able to make sense of a robot’s behaviour before problems arise. Having clear intentions that align with expectations helps with trust. Akalin *et al.* [86] found that predictable and understandable behaviours and actions (current and future) led to higher perceptions of safety, contributing in turn to trust. When a robot’s actions are

predictable, its intentions are legible in the moment, and if those intentions are anticipated, the robot is seen as safe. Trust requires a robot to perform correctly, but also have clear reasoning and a transparent intent, which is a design concern as much as a functionality one.

Both the communication received from a robot and the approach taken influence trust. Especially when trust has been violated, clear communication can help rebuild trust and mitigate some of the negative effects of that violation. Ciocirlan *et al.* [125] found that a robot communicating about a task helped mitigate trust lost after an error, compared to communication that is informal, non-task related, or not present. Nesses *et al.* [124] showed that high transparency, explaining why things were happening, helped users understand the decisions made by a robot. Even if a robot is transparent, scepticism still increases if it makes an error.

When errors occur, trust repair strategies, such as apologies and explanations, have been shown to be effective in rebuilding trust after a violation [23]. Trust is not necessarily permanently lost, but the effectiveness of repair strategies depends on the nature of the violation and the strategy used. Subsequent failures will further erode the credibility of previous recovery attempts. Given that errors are inevitable, the ability to recover socially is a meaningful design consideration.

A robot’s appearance and embodiment influence trust, as evidenced by the significant positive correlation between trust and anthropomorphism [21]; however, uncanny valley effects can make a robot seem less safe than other forms [86]. A robot can be made more real in other ways. Giving a robot a name and voice helps establish greater trust in it [34], as it contributes to its identity. A personality provides a platform for assigning the reliability and intent attributes to, rather than an ambiguous identity.

Non-verbal behaviours, such as movement, light, and gaze, serve as a communication channel that helps form different perceptions. Manor *et al.* [17] found that a non-humanoid that was attentive by nodding and following a user’s gaze helped with affective and cognitive trust as it communicated attention. How fast a robot moves and the path it takes can indicate a robot’s affective state [126], which in turn affects the judgements users make. Shirol *et al.* [93] demonstrated that perceived trust and safety can be increased by the movement and light patterns used on an autonomous robot as they indicate where it is likely to go, making it seem more predictable.

Users are more likely to trust a robot they believe to be intelligent. Research by Schaefer *et al.* [20] and Kim *et al.* [105] have both found that users are more willing to let their guard down when a robot is perceived as smart, with a direct positive correlation between perceived intelligence and perceived trustworthiness ratings. However, since perceived intelligence is shaped substantially by appearance and behaviour, users may overextend trust based on how intelligent it appears, rather

than how capable it actually is.

Findings around dispositional trust are more tentative. While a user’s personality (e.g., agreeableness and neuroticism) has some impact on trust, a propensity to trust was not a significant predictor of trust in robots [21]. Decontextualisation may explain why personality was not a significant predictor, as the context of a human-robot interaction matters greatly. In a situation designed to manipulate comfort with a robot, Akalin *et al.* [86] found male participants reported feeling significantly safer during the interaction than female participants; however, Hancock *et al.* [21] did not find gender as a significant factor.

Firmino de Souza *et al.* [118] identified familiarity with technology as an important factor affecting trust, but again, it was not one of the strongest correlations found by Hancock *et al.* [21]. Simple experience and actual usage are both referred to as familiarity. The result in [118] may actually mean that the quality of prior interactions matters more than awareness as an important familiarity factor. Counterexamples to the meta-analysis suggest that certain factors may be relevant in specific contexts and controlled studies; however, their effects may be diluted across the diverse contexts, robot types, and tasks included in the meta-analysis, so they cannot be generalised.

Kim *et al.* [105] created a framework called the perfect automation schema (PAS) to explain a user’s beliefs regarding the performance of automated systems. Their research found that a user’s PAS was significantly correlated with their assessment of a robot’s trustworthiness. This belief in automation, in general, can be applied to specific robots. The agency a user feels during an interaction is strongly positively correlated with their comfort and trust, according to [86]. The level of control a user has over the robot, interaction, and situation affects how safe they feel.

As previously discussed, context also plays a significant role. According to Wagner and Arkin [15], risk is necessary for trust to be psychologically relevant. Without some vulnerability, there is no reason for a user to care about relying on another person or robot.

Vulnerability does not necessarily require a high-stakes scenario [16]. Even everyday risks, such as spending money on a bad recommendation, can affect trust. If the information received is even slightly relevant to a user, then trustworthiness is also relevant. Situations involving financial decisions are common times when trust is important, making them useful for assessing trust.

The social structure of the context matters too. When a robot is integrated into a team as a contributor to shared goals rather than operated as a tool, then people are more likely to trust it [127, 128]. In this sense, trust is shaped not only by what is at stake but by the relational role the robot occupies within a given situation.

Trust is a complex topic, and we look at multiple kinds of trust in this thesis. The definitions concern trust in the robot, in whether it will operate reliably and act in the user’s interest. Where a robot’s function is to answer questions, the user must also judge the information itself, and the exposure at stake is being misinformed rather than being harmed or let down. In this thesis, we treat epistemic trust not as an entirely separate form of trust, but as a context-specific judgement drawing on performance dimensions such as competence and reliability, relational dimensions such as sincerity, and a behavioural willingness to rely on information without verification. The first two are the dimensions set out in [111], applied to a robot’s answers rather than to its actions, and correspond to the separation Hovland and Weiss [129] draw between a source’s expertise and its intention to report honestly. The third follows the treatment of reliance as the behavioural expression of trust in automation [13].

2.5 Chapter Summary

In this chapter, we reviewed literature on information-seeking, robot morphology, and trust, which together suggest a connected process: a robot’s appearance shapes expectations, those expectations contribute to trust judgements, and those judgements influence whether and how the robot is used as an information source. Four findings from this review set up the work in the remainder of this thesis.

First, we will use the trust definition by Hancock *et al.* [21] in this thesis, which includes reliance, risk, and interdependence. Without some vulnerability managed by the robot, trust is not meaningfully at stake [15, 67]. Trust is multidimensional, spanning performance trust in a robot’s reliability and competence and moral or relational trust in its intentions [111]. These facets are not expected to move together. Studies that have measured them apart have shown they can diverge, as trust in an interface and trust in the information delivered through it were affected differently by the same manipulation [10], so a composite score can conceal opposing movement in the dimensions that compose it. The trust measures we use in this thesis are therefore not a composite. It is also dynamic and fragile, updated as behaviour confirms or violates expectations and sharply reduced by errors [112]. The way that errors can break trust is a property that makes reliability especially consequential for a robot whose primary role is to provide information.

Second, morphology functions as a signalling system. It is the primary information a user has about a robot before any interaction, acting as a proxy for capability in the absence of direct experience. Morphology was among the strongest robot-related predictors of trust in the meta-analysis by Hancock *et al.* [21]. Both overall embodiment type and individual features such as faces, gaze, movement, and lighting communicate what a robot can do and what kind of interaction it

supports [17, 81, 90, 93]. When appearance and actual function are misaligned, expectations may not be met, which damages trust [88].

Third, information-seeking is a trust-relevant context because evaluating the source and evaluating the robot are the same act. It satisfies both preconditions for trust: risk, since acting on incorrect information has consequences even in everyday scenarios [16], and interdependence, since the robot controls access to the information [15]. The trust signals users seek in digital sources, such as design cues, authority, and transparency [9], map directly onto what morphology communicates, making a robot’s form an active variable in informational trust rather than mere presentation. Trust in this context must also be appropriately calibrated. A user who overextends trust to a robot may act on inaccurate information without sufficient scrutiny, while one who does not trust it enough forgoes a genuinely helpful resource. Further, receiving synthesised answers removes the user from evaluating sources themselves, which makes that calibration harder and affects their likelihood of verifying what they are told.

Fourth, the research areas examined in this review have not been brought together. Trust in information-seeking has been studied with disembodied digital interfaces, while morphology and trust in HRI have been studied in industrial and task-oriented contexts. No identified study has directly examined how robot morphology influences trust in information-seeking interactions. This gap in the research motivates the empirical work of this thesis in the subsequent chapters.

In answer to RQ1, the literature indicates that morphological features initiate and shape how users calibrate trust by signalling capability, intelligence, sociability, intention, predictability, and role. These effects are moderated by context, user characteristics such as culture, age, and gender [77, 86, 95], prior experience [70], and the degree to which appearance matches actual function [88]. Whether these expectations hold in a real information-seeking interaction is the subject of RQ2; Chapter 3 sets out the methodology used to test it.

Chapter 3

Methodology

In Chapter 2, we concluded that although morphology has been established as a determinant of trust, and information-seeking is a trust-relevant context, no study has investigated what this means together. In this chapter, we justify the approach taken to address that gap by aiming to answer RQ2: How does robot morphology affect epistemic trust during information-seeking scenarios? RQ2 follows from the findings of the literature review, in which we answered RQ1 by identifying how morphology affects trust. RQ2 is investigated empirically through an online survey and a naturalistic field study. The aim is to inform the design of robots that allow users to make accurate assessments of the robot’s purpose and the information it provides.

The reasoning behind the study designs is explained in this chapter, whereas the methods sections for each study (Sections 4.1 and 5.1) are in the subsequent chapters. In Section 3.1, the choice to conduct two similar but distinct studies is explained. In Section 3.2, the choice of the robot morphologies shared by both studies is described. In Section 3.3, the reasoning behind the information-seeking context used in the studies is presented. Finally, in Section 3.4, there is an explanation of why different measures were chosen.

3.1 Research Design

As seen in [35], survey ratings do not necessarily align with real-world behaviour. As such, two studies will be used to investigate RQ2 and determine what users believe they trust versus what they are observed to trust. A survey can identify whether morphology changes reported perceptions, but not whether those perceptions carry over into how the robot is trusted and relied upon in actual use. Because the design for both studies is between-subjects across the four morphologies, obtaining sufficiently stable group estimates requires substantially more participants than a

within-subjects design. Recruiting and scheduling enough in-person participants to complete the full validated scales would have been impractical, and would make the interaction less natural, with fewer people willing to complete it in full. As such, an online survey will be used to reach the numbers for quantitative comparison between groups, and the user study provides contextualisation through qualitative responses [130].

Because of the difficulties in recruiting participants described, the survey will measure perceptions formed through passive observation of a robot. A naturalistic field study will complement the survey by allowing participants to choose whether to approach and use the robot. Unlike a lab-based study, this does not simulate an information-seeking use case; it is a real tool people can use and will need to trust to have value. The user study is self-directed, and therefore provides evidence of how participants behave during genuine use, compared to general perceptions. As it is naturalistic and to minimise the burden on participants, only a small amount of qualitative data is collected via an optional and short questionnaire synthesised from the literature review of relevant trust dimensions. Treating these as two studies instead of collapsing them into one also allows for the difference between reported and enacted trust to be examined.

Both studies adopt a mixed-methods approach, using quantitative ratings to measure the effect of morphology on trust and qualitative responses to contextualise those ratings. As demonstrated by Akalin *et al.* [86], qualitative data can reveal changes in trust that quantitative measures alone cannot explain. This approach addresses the gap between reported and enacted trust, supporting the use of two complementary studies.

Each participant in both studies will encounter only one robot. A between-subjects design avoids carryover between robots and prevents participants from noticing that the robots behave identically. Exposure to multiple robots could reveal the purpose of the study and influence the differences participants report between conditions. A between-subjects design also keeps the survey short enough to limit fatigue or straight-lining in the responses and reduce drop-out. In the user study, a similar concern arises, and it would be less realistic to have multiple devices for the same purpose. The cost is that individual responses cannot be compared across morphologies, so the analysis compares the mean ratings of each robot between groups. Using group means follows established practice to identify how design choices affect a population [100]. Naturally, individuals will vary within a group, but with enough participants, the group means should stabilise [86].

Because the studies are only informative when they measure trust under the same conditions, the taxonomy proposed by Onnasch and Roesler [1] has been used to characterise each interaction condition and to identify any contextual differences that may alter the effect morphology has. The taxonomy is a classi-

fication framework rather than a measurement scale. It describes an interaction across three clusters. The interaction context covers the field of application, the robot’s task, whether the robot is embodied or depicted, whether the setting is the field or a laboratory, and the temporal and physical proximity between the human and the robot. The robot classification covers the robot’s morphology, its degree of autonomy across four stages from information acquisition to action implementation, and its input and output communication channels. The team classification covers the role humans play in the interaction and the ratio of humans to robots. An interaction is classified across all dimensions using the graphical template provided by the authors, resulting in a one-page profile. Two conditions can then be compared dimension by dimension, revealing structural differences that could otherwise be overlooked.

The taxonomy was applied when the studies were designed, rather than retrospectively, serving as a checklist to keep all dimensions other than the robot constant. A profile was completed for each robot in each study format, resulting in eight profiles in total, which are provided in Appendix A. Within each study, the only entries that change between profiles are those in the robot classification that vary with the device itself, primarily its morphology. All contextual and team dimensions are identical within each study, which strengthens the attribution of any within-study differences in trust to morphology rather than to other aspects of the interaction.

Comparing the profiles across the two study formats shows that four dimensions differ. The user study exposure is classified as embodied because the robot is physically present with the user, but in the online video, the exposure changes to depicted. The user study setting is in the real world, classified as field, and the taxonomy specifically states that online studies are a laboratory setting. For physical proximity, a participant communicates nearby to the robot but does not touch it, classified as approaching, and an online participant shares no environment at all, so the proximity is none. Finally, the human role also changes, as when the participant actually uses the robot, they are pursuing an answer to a question together, so they are a collaborator, whereas in passive observation, the participant is a bystander. All other parts of the taxonomy are consistent. These context dimensions have previously been identified as factors that can influence human-robot interactions, meaning that any differences between the studies may be attributed to changes in context rather than to the robots themselves.

Recording the conditions in this way also supports reproducibility and helps researchers judge how far the findings generalise to other HRI scenarios. This taxonomy was selected because it focuses on structuring interactions rather than measuring perceptions. Additionally, Ringe *et al.* [80] suggest integrating META-MORPH with it, ensuring that the framework used to choose the robots in Section

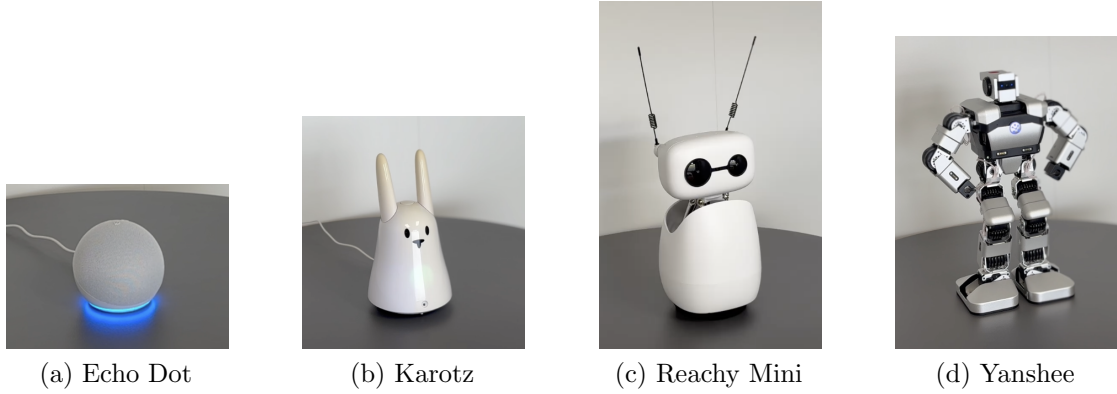


Figure 3.1: Devices Selected for Studies

3.2 complements the framework for characterising interactions.

3.2 Robot Platforms

Using METAMORPH and the stereotypical categories (covered in Section 2.2.1), four devices were selected for use in the studies, each representing a different embodiment type. As a baseline, an Amazon Echo Dot smart speaker will be used to isolate the effect of embodiment on perceived information-seeking performance. The Mindscape Karotz is a small zoomorphic robot that adds embodiment and lights to make the speech functionality feel more alive. The Karotz device was chosen because it adds only a minimal set of features beyond a body, compared to the Echo Dot. The next device is a Pollen Robotics Reachy Mini, a small droid that can manipulate its features to perform different behaviours and expressions, designed to feel more engaging to interact with and to signal intent [126]. Finally, the UBTECH Yanshee was chosen as it is a humanoid robot that can walk, track faces, and use its limbs to perform behaviours. These four devices can be seen in Figure 3.1. Each of these devices adds a morphological capability to determine the effect of features on perceptions and represents a distinct embodiment level. Table 3.1 summarises these main feature distinctions.

In Table 3.1, behaviours refer to the expressive physical actions a device can perform. Expressive movement was treated as a dynamic aspect of morphology in this study because the platforms’ physical structures enabled different visible movements, and these behaviours serve as visual cues that help signal intent. It is distinct from embodiment, even though the body’s features can limit or enable these behaviours, as they may not be utilised, functioning as a dynamic aspect of the form. Behaviours matter because they can determine how alive, capable, and

Table 3.1: Device Morphology Matrix

Robot	Embodied	Behaviours	Appearance Type
Echo Dot	No	None	Speaker
Karotz	Yes	Basic	Zoomorphic
Reachy Mini	Yes	Advanced	Mechanomorphic
Yanshee	Yes	Advanced	Humanoid

intentional a robot seems, which are factors that affect how much trust is extended to it. The Echo has no behaviours as it does not move, and the Karotz can only spin its ears, while the Reachy and the Yanshee have advanced articulation that allows them to express intent and respond to the user.

The Echo serves as a control because it performs the same information-seeking function without a body or distinct features, which isolates the effect of embodiment. The Echo is not, however, a fully neutral baseline as it is a familiar consumer product, so participants may arrive with expectations of how a smart speaker behaves. A lack of neutrality is acknowledged here and noted as a limitation in the discussion of each study. Each robot communicates intent visually, and the Echo does too by using a ring of lights to show when it is speaking, thinking, and answering.

The METAMORPH framework [80] was used to quantify how distinct the four morphologies are, representing each device as a labelled feature graph (Figure 3.2), and the graph edit distance between every pair was computed (Table 3.2). Graph edit distance counts the feature additions, deletions, and substitutions needed to turn one robot’s graph into another’s, so a larger value means the two robots share fewer morphological features and structural relationships.

The distances confirm that the four devices are morphologically distinct and span a meaningful range rather than clustering together. Distance from the disembodied Echo Dot increases steadily with embodiment, from the zoomorphic Karotz (11) to the mechanomorphic Reachy Mini (15) to the humanoid Yanshee (25), mirroring the order in which each device adds morphological features. The Yanshee sits furthest from the other devices, while the Karotz and Reachy Mini are closest to one another (10), as both have small bodies and expressive features without a full humanoid form. The largest distance separates the two extremes, the disembodied Echo Dot and the humanoid Yanshee, and is comparable to the distances Ringe *et al.* [80] report between unrelated robots in the IEEE Robots Guide.

These values use the standard unit-cost graph edit distance, in which every edit is weighted equally, so adding a whole limb costs the same as changing a single descriptor on a shared part. Because some devices share subdivisions that differ only in descriptor, a scheme that penalises descriptor changes less than structural

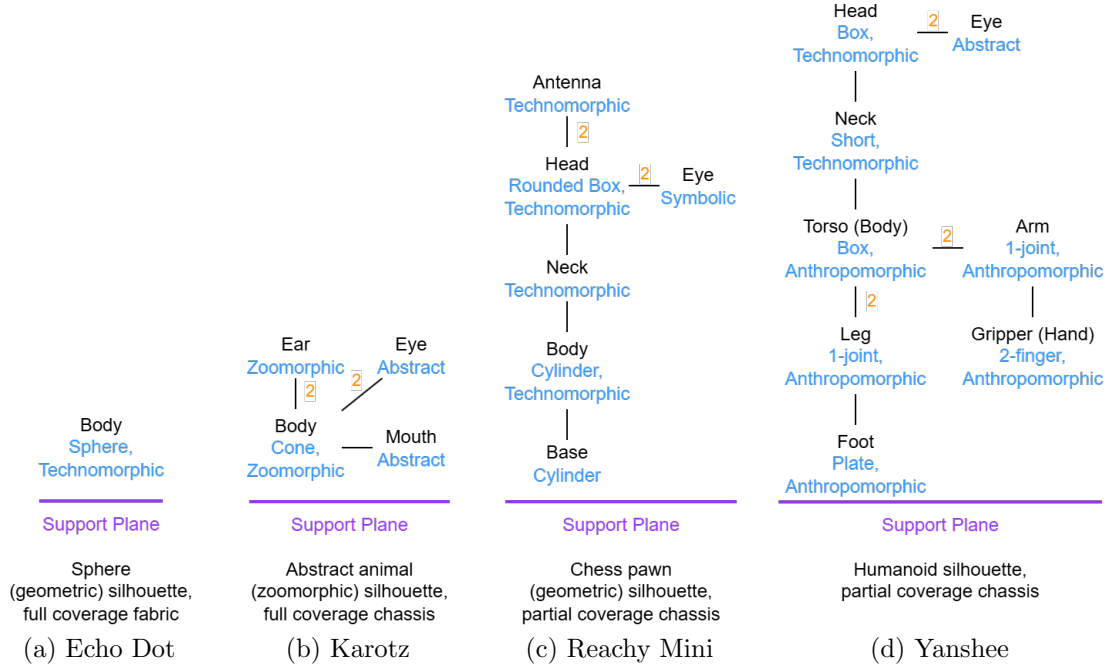


Figure 3.2: METAMORPH Representation of Morphologies

Table 3.2: Robot Graph Edit Distance

Robot	Echo Dot	Karotz	Reachy Mini	Yanshee
Echo Dot	–	11	15	25
Karotz	11	–	10	20
Reachy Mini	15	10	–	18
Yanshee	25	20	18	–

ones would shrink the distances between the more closely related robots, most noticeably among the small embodied devices such as the Karotz and Reachy Mini, while leaving the distances of structurally dissimilar pairs such as the Echo Dot and Yanshee essentially unchanged. The relative ordering of the morphologies is preserved either way. Ringe *et al.* [80] note that every editing operation is treated equally in their implementation and identify the choice of optimal edit weights as future work, so this was not reported here.

3.3 Study Context

A low-stakes everyday information-seeking scenario will be used for both studies. The setting is the University of Waikato’s Tauranga campus during Orientation Week, which features a range of events to welcome students and engage them with university life. Participants will interact with the robot to seek information about events happening in the first week, such as when and where they will be held and where to get support. Information-seeking is a fundamental and constant human behaviour [5], and its everyday, low-stakes form is the kind encountered most often. Information-seeking is also among the most plausible emerging uses of a social service robot, where it answers simple factual questions in a public setting as an initial point of help and guidance.

Studying trust in this setting produces findings that apply where an information-seeking robot would actually be deployed. Choosing a realistic situation in which a person might authentically choose to ask a robot a question, rather than a contrived task, also supports observing real self-directed behaviour over someone following instructions. The study arose from a research goal, but the Waikato Students’ Union (WSU) are genuinely interested in this use case, giving the investigation practical value.

Although the stakes are low, trust remains relevant because it only matters where there is risk and interdependence [15]. Acting on incorrect information still has consequences, although minor, like disappointment or a wrong turn that makes the person take slightly longer to get somewhere. The robot controls the information users seek, even if only out of curiosity or to confirm what they already know. Vulnerability, therefore, does not require a high-stakes scenario; everyday risks are sufficient to engage trust in another entity and the information from it, making even small trust violations meaningful [16]. The low-stakes context minimises the consequences of errors without removing the prerequisites for trust, allowing robots to be compared on this basis.

An information-seeking context also addresses a gap in the literature, as trust in information-seeking has not yet been extended to embodied agents, despite the application of work on morphology and trust to robots. Existing HRI work typically focuses on higher-stakes, industrial, or collaborative tasks or settings. Everyday information-seeking with a robot has received limited attention, especially outside these domains, despite it being one of the most probable uses. A university provides a natural setting because information-seeking is central to learning, and there is always new information present. Experiments are also more acceptable at a university, with less scepticism towards them. Orientation Week further provides authentic questions as students are unsure in their first week, but also naturally curious. This opportunity offers valuable real-world usage that researcher-directed lab-based studies cannot replicate.

The low-stakes nature of the setting also makes the naturalistic design possible. Because participants could not be meaningfully harmed or even too concerned by acting on inaccurate information, the user study could be conducted without a researcher present, making it more realistic again. Removing the researcher also prevents them from influencing participants with expected usage and results. A higher-stakes study could not be run responsibly in the same manner. Allowing individuals to decide whether to approach the robot also reflects real-world adoption, and the decision to engage is itself a trusting behaviour influenced solely by appearance.

Using the same context across both studies enables direct comparison because the robots, task, and topic remain constant while only the mode of participation changes. The trade-off is that the findings are restricted to this context and may not be as general as replicating the study in other contexts and levels of vulnerability. When the consequences of incorrect information are small, participants may feel little need to verify what they are told. Trust may then operate differently and more strongly in higher-stakes settings such as healthcare [105]. This limitation of the stakes is discussed further in Chapter 6.

3.4 Measurement Approach

The survey will use two validated HRI scales. The TPS-HRI [106] is well-suited to examining how trust varies with morphology as the independent variable. The performance trust component is shaped by the accuracy and reliability of the information provided, and it also captures the relational and dispositional dimensions of trust that morphology influences. The HRIES [81] complements it by capturing the sociable, intentional, and emotional perceptions that surround a trust judgement. A comparison of alternative scales is covered in Section 2.3. These perception areas are affected by morphology and adjust people’s expectations of how a robot should be interpreted before interacting. Using both scales together shows how perceptions of morphology evolve into trust, offering richer insight than assessing the trust outcome alone. How the scales are administered is described in Section 4.1.

A brief and voluntary interaction cannot support a long questionnaire; hence the split of studies. Instead, the user study will measure trust with three short questions asked verbally afterwards. With no researcher present, a verbal questionnaire also needs to be short and simple, as any difficulty cannot be resolved by the robot. Demographics will also not be collected for similar reasons. The three questions target trust at different points, because information-seeking involves a source, a message, and the credibility of each.

Helpfulness concerns how the robot performed its function, corresponding to

performance trust in system reliability, capability, and competence [111], separate from answer quality. Robot-related factors are the strongest predictors of trust in HRI research [21], and perceived competence is a dominant factor in automation [112].

Answer quality is a judgement about the information. Whether a robot seems fit-for-purpose or useful is less about the robot, and more a judgement about the content delivered, and the perceived credibility and capability to provide accurate information [129]. Answer quality draws on content-related cues that help with trust in digital information more generally [9, 69]. A robot can be considered a capable agent even if its answers are poor, and perceived quality can modify trust independently of accuracy [10]. Assessing this separately from helpfulness shows the distinction between trust in the source and trust in the message the source provides.

The likelihood of verifying the information provided relates to human factors. The third question, on the likelihood of verifying what the robot said, is the closest of the three to enacted trust, as choosing not to check an answer is a form of reliance and a standard behavioural proxy for trust [13]. A higher likelihood to verify the information means that the robot is less believable. It is, however, a self-report measure, which is why behavioural measures are collected alongside it. Whether someone verifies depends on the risk they perceive in relying on the answer and on the dispositional trust they bring, from their prior knowledge, their ability to find the information elsewhere, and how much an accurate answer matters to them [21, 28]. Because the chosen context is low-stakes, and there are various other sources readily available, placing high epistemic trust in the robot in this situation shows how credible the robot seems to be. As these three questions assess distinct aspects of trust, they do not assess the same underlying judgement. The specifics of these measures are explained in more detail in Section 5.1.

The trustworthiness reported by users may not necessarily align with the different types of trust they exhibit. To investigate this, a new survey will be administered after the user study has concluded. Individuals on campus will be shown pictures of each robot and asked which one they would trust, without any additional information to influence their choices. Analysing whether user perceptions of trust align with their actual trust levels provides valuable context, even though this approach is less formal and rigorous. It is worth highlighting that, across the survey, the user study, and this questionnaire, responses are mostly first impressions, but trust also has a temporal factor [122]. In this study, we do not aim to investigate how morphology affects trust that evolves over an extended period.

Trust is multidimensional, so multiple approaches are needed to capture the different aspects. The rows below follow the three components of epistemic trust set out in Section 2.4: performance dimensions such as competence and reliability,

relational dimensions such as sincerity, and a behavioural willingness to rely on information without verifying it. Table 3.3 sets out how each measure relates to the trust dimensions reviewed in Chapter 2, and in which study it is collected.

Table 3.3: Map of Trust Dimensions to Measures

Trust dimension	Measure	Study
Performance trust (competence and reliability)	TPS-HRI and helpful item	Survey and user study
Social perceptions that feed into trust	HRIES	Survey
Source expertise and answer credibility	Useful item	User study
Reliance on the answer	Believable item	User study
Enacted reliance	Questions asked and interaction duration	User study
Moral and relational trust	Qualitative explanations	User study
Trust attributed to appearance	Comparative choice of most trustworthy robot	Follow-up survey
Dispositional trust and prior familiarity	Familiarity covariate	Survey

Performance trust in reliability, capability, and competence will be measured by the TPS-HRI and the rating items, which further separate the robot as a source from the information it provides. The HRIES captures the social perceptions morphology establishes before any judgement forms. Because reported and enacted trust can diverge within one interaction, the behavioural measures will record what participants did rather than what they say, as not checking an answer is a proxy for reliance. The qualitative explanations cover the moral and relational aspects that the brief Likert questions cannot support. Appearance alone is isolated in the follow-up survey. Familiarity is a covariate in the survey so that dispositional trust is separated from the effect of morphology.

3.5 Chapter Summary

In this chapter, we set out the approach used to answer RQ2: How does robot morphology affect epistemic trust during information-seeking scenarios? Building on the literature review, which answered RQ1 by establishing morphology as a determinant of trust, a two-study approach was chosen to gather real-world data, and in large quantities. Using two separate study designs also demonstrates the impact of in-person interaction, as survey ratings do not always reflect real-world

behaviour [35]. An online survey will measure perceptions formed through passive observation of a robot, while a naturalistic field study will measure the trust people enact when they choose to approach a robot and ask their own questions. Both combine quantitative ratings with short qualitative responses, so that participant responses can be contextualised by the reasons they give.

Across both studies, morphology will be the independent variable. Robot platform is the experimental condition, with the four platforms selected to represent substantially different morphology categories. Conversational capability, task, topic, and interaction procedure are to be held as constant as practicable. Because the platforms also differ in movement, familiarity, and other device-specific characteristics, differences between conditions are interpreted as effects associated with morphology rather than effects attributable to any single morphological feature.

The task and topic will be held constant, using the taxonomy by Onnasch and Roesler [1] to characterise and confirm any contextual differences between the interaction conditions. Four devices were selected to span a range of embodiments: the Echo Dot, a disembodied voice as a control; a zoomorphic Karotz; a mechanomorphic Reachy Mini; and a humanoid Yanshee. The setting is an everyday, low-stakes information-seeking scenario at the University of Waikato's Tauranga campus during Orientation Week, chosen because it reflects a realistic, likely use case for a social robot. The context does however mean the findings may not generalise to higher-stakes scenarios.

The two studies measure trust differently to suit how each is run. The survey will use two validated scales, the TPS-HRI and the HRIES, while the brief and voluntary field interactions will be assessed with three short verbal questions covering how helpful the robot is, how useful the information it provides is, and how likely participants are to verify the information, supported by behavioural measures and a separate trust survey comparing appearances. The following chapters report these studies in turn. In Chapter 4, we present the survey on perceptions formed through passive observation. In Chapter 5, we present the user study and the trust enacted in real use. In these two chapters together, we answer RQ2 by distinguishing between what users report trusting and what they are observed to trust.

Chapter 4

Survey

In this chapter, we report the approach and results of the first study justified in Chapter 3. This study was an online survey examining how morphology affects perceptions and trust. As described in the methodology, the survey investigated the perceptions formed during passive observation of an information-seeking task with a robot. Each participant watched a short video of a single robot being used for this task, and then rated that robot using validated trust and interaction scales. The survey complements the naturalistic field-based user study reported in Chapter 5, which captures how those perceptions translated into enacted trust during genuine usage. Together, the two studies answer RQ2 by distinguishing trust assessments from passive observation to direct, self-initiated interaction.

This chapter is organised into two sections. In Section 4.1, we describe how the survey was designed and administered, covering the experiment structure, the survey content including the video of the interaction, the participants and their recruitment, and the analysis plan. In Section 4.2, we report the findings, beginning with the sample achieved, then proceeding through the descriptive statistics, the effect of morphology on each scale, and the influence of participant characteristics.

4.1 Methods

The survey was created using Qualtrics and used a between-subjects design for the robot morphologies. A written copy of the survey is provided in Appendix C. The survey was approved by the University of Waikato’s STEM Human Research Ethics Committee with number STEM_HREC(2025)#65. The approval letter and participant information sheet are included in Appendix B.1.

4.1.1 Survey Design

To gather trust perceptions of different morphologies, an online survey was developed where participants would observe a short video of a robot, and then rate that robot against trust and interaction perception scales. Four distinct robot morphologies were used to provide a range of appearances (as covered in Section 3.2) to determine whether the different features present affect the trust rating given.

To maximise responses, the survey duration was kept under 10 minutes to reduce the burden on participants. As a result, each participant viewed only one robot, with groups distributed equally using the Qualtrics even-presentation randomiser to ensure equal numbers of ratings for each of the four morphologies. As justified in Section 3.1, this between-subjects design trades the ability to compare individuals across morphologies for lower fatigue and abandonment; group mean scores were compared per survey item. No part of the survey required an answer, besides the consent and eligibility questions, to support the participant’s right to participate as much or as little as they desired.

After receiving a brief on the survey content and purpose, the participants chose whether to provide consent. Basic demographic information was then collected to support the analysis of differences in perceptions by age, gender, and robot familiarity. Age and gender were collected as categorical variables, and familiarity was measured on a 5-point Likert scale and analysed as an ordinal variable. At this point, those ineligible to participate in the study (those under 16 years of age) were excluded for ethical reasons.

Participants were shown a short information-seeking interaction video of the robot, which demonstrated its morphological features and capabilities. As the HRIES assesses interactions with the robot, the video needed to include someone using the robot, given the nature of online surveys and the inability to provide each participant with the robot to interact with themselves. The video used the information-seeking context justified in Section 3.3, allowing comparability with the field study. The context was a short snippet of a typical interaction, described in the observational study section. A description and full transcript of the video are provided in Appendix D.

Participants could view the video as much as they wanted to help them answer the rating questions. The measures used were the Human-Robot Interaction Evaluation Scale by Spatola *et al.* [81], and the Trust Perception Scale-HRI by Schaefer [106], as previously detailed in Sections 2.3.1 and 2.4. As per author recommendation, the HRIES scale item order was randomised to reduce semantic bias. Participants could provide open-text responses to explain their scores if they wished, but this was not a priority to minimise their time investment.

As detailed in Section 2.3.1, the HRIES uses four traits per dimension (sociability, agency, animacy, disturbance) on a 7-point scale, scored per dimension rather than

as a single index; the TPS-HRI uses 0-100% ratings averaged into an overall trust percentage. The short, 11-item version of the TPS-HRI was used to make it easier on participants¹.

4.1.2 Participants & Recruitment

An initial pool of participants was recruited through channels on the University of Waikato’s Tauranga campus for both students and staff, including Tauranga Moodle pages, Tauranga Staff Teams groups, and WSU social media pages. A wide range of participant backgrounds was desired to avoid bias in the data and ensure compatibility with other research. To gather a range of prior robot familiarity, no groups (e.g., computer science students) were targeted or excluded. The survey was available for six weeks throughout A trimester of 2026. As the information-seeking context in the second study focused on the University, the sample was initially composed of staff and students for comparability, intended to reflect the diversity of ages, genders, and cultural backgrounds of the campus. To increase the statistical power of the results, a second pool of participants was recruited via the Prolific² crowdsourcing platform, and they were paid for their responses, with as many participants as budget allowed. This different group of participants reduces comparability with the user study but increases the power of the survey’s findings.

When the survey was distributed, participants were informed that the survey would take up to 10 minutes to complete. Participants were given a unique ID that they could use to contact the researchers to withdraw their entry. They could optionally allow Qualtrics to email them a copy of their response, where they could retroactively withdraw their consent. Participation (or non-participation) was entirely voluntary and had no impact on the employment or academic work of staff, or the grades and other progress of students. The researchers did not know who had participated. Due to ethical considerations, people under 16 cannot give consent to participate. The survey recorded each participant’s age, gender, and familiarity with robots, because these are known to moderate trust in robots: age and gender are common moderators [16, 21, 86, 131, 132], and familiarity reflects the prior experience a participant brings to the interaction.

4.1.3 Analytical Approach

Analyses were conducted in Python using pandas and NumPy for data handling. SciPy and Pingouin were used for univariate tests, post-hoc comparisons, and effect sizes, and statsmodels for the multivariate analysis. The two recruitment pools

¹N.B., the authors refer to the short version as having 14 items, but only 11 are marked in their table [106, p. 213–214].

²<https://www.prolific.com/>

were exported as a CSV from Qualtrics and combined into a single dataset. We excluded incomplete responses and those flagged as low-quality from the rating calculations.

The two scales were analysed independently as they assess different concepts. The HRIES gives four subscale scores (sociability, agency, animacy, and disturbance), so these were examined together using a one-way multivariate analysis of variance (MANOVA) to determine whether the morphology results differ across the four subscales together (the approach in [81]), and reported as Pillai’s Trace. Because the morphologies were tested on two scales, there is an increased chance of a false positive, so a Bonferroni-adjusted significance threshold of $\alpha = .025$ was applied to the HRIES MANOVA and the TPS-HRI ANOVA omnibus tests.

Where the MANOVA indicated an overall difference, each subscale was then tested with a univariate ANOVA to determine which groups differed on that dimension. As the MANOVA was already significant here, no adjustment was made on the rationale that it already guards the error rate on the set of subscales. The TPS-HRI trust score was analysed in the same way with its own one-way ANOVA. If an ANOVA showed that groups differed, post-hoc pairwise comparisons identified which pairs of morphologies differed significantly. A Tukey’s range test was used for these comparisons due to the comparable spread of scores.

Two findings are reported for each test. First, whether the difference is statistically significant and the difference in scores are reported. Second, the magnitude of that difference, expressed as the effect size (the amount of variance attributable to morphology), is reported. For clarity, η_p^2 measures the share of the variation of scores associated with morphology for MANOVA and ANOVA, and Hedges’ g is the size of the gap between two groups in terms of standard deviations for pairwise comparisons.

To check that any effects were not due to the characteristics of participants assigned to each group, the analyses were repeated as sensitivity checks while controlling for age, gender, and prior familiarity. The analysis involved using a multivariate analysis of covariance (MANCOVA) for the HRIES subscales and an analysis of covariance (ANCOVA) for the TPS-HRI score. By controlling for these background characteristics, the covariate analyses estimated the effect of morphology independent of differences in participant demographics and prior experience with robots. The covariate analyses also show whether the covariates themselves are predictors of the scores.

4.2 Results

In this section, we report the survey findings, following the analysis plan outlined in the methods (Section 4.1). We begin the results with an overview of the achieved

sample and its demographic composition, followed by descriptive statistics for the two scales. The effect of morphology is tested against the HRIES subscales and the TPS-HRI trust score. The influence of participant characteristics is examined through covariate analysis.

4.2.1 Responses

A total of 263 responses were received, with 223 participants recruited from the Prolific platform and 40 from within the University. There was a total of 252 responses after incomplete submissions were removed. The groups of robots observed were intended to be distributed evenly by the platform, with 66 participants observing the Reachy, 60 the Yanshee, 64 the Echo, and 62 the Karotz. The unequal count is due to some participants not completing the survey and how the Qualtrics platform randomises; non-submissions were excluded from analysis.

Demographics were not selected for; instead, they came from the pool of available participants on Prolific and from the Tauranga campus. Almost half of the participants were in the 25-34 age bracket, as seen in Table 4.1. Only 13% of the participants were over the age of 44. Slightly more participants were male than female (Table 4.2).

Table 4.1: Survey Participant Ages (N = 252)

Age bracket	n	Percentage
16-24	62	25%
25-34	116	46%
35-44	40	16%
45-54	20	8%
55-64	9	4%
65-74	5	2%
75+	0	0%

Table 4.2: Survey Participant Genders (N = 252)

Gender	n	Percentage
Male	139	55%
Female	112	44%
Non-binary	1	<1%

Most participants reported a middling level of familiarity with robots; 90% of respondents were slightly, moderately, or very familiar with robots, as seen in Table 4.3.

Table 4.3: Survey Participants Robot Familiarity (N = 252)

Familiarity	n	Percentage
Not familiar at all	15	6%
Slightly familiar	84	33%
Moderately familiar	84	33%
Very familiar	59	23%
Extremely familiar	10	4%

4.2.2 Descriptive Statistics

Across the different morphologies, participants rated the robots above the HRIES 7-point Likert midpoint of 4 for the sociability and agency dimensions, showing some level of anthropomorphic attribution in all the robots. Animacy was lower than the midpoint, while disturbance was well below, meaning participants did not perceive the robot as particularly dynamic, but not unsettling either. Means and standard deviations for the HRIES subscales are reported in Table 4.4. The order of the robots in the tables is to be consistent with the presentation order in Chapter 5.

Table 4.4: Survey HRIES Descriptive Statistics

Morphology	Sociability		Agency		Animacy		Disturbance	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Reachy	5.10	1.13	4.93	1.20	3.32	1.41	2.23	1.16
Yanshee	4.51	1.46	4.75	1.29	3.05	1.45	2.48	1.31
Echo	5.10	1.28	5.02	.97	3.86	1.45	2.00	1.07
Karotz	4.96	1.19	4.70	1.08	3.46	1.34	2.29	1.19
Overall	4.92	1.28	4.85	1.14	3.43	1.43	2.25	1.19

For the TPS-HRI, trust across all morphologies was high, with an average score of 76.12%. Means and standard deviations are presented in Table 4.5. As can be seen for the HRIES in Figure 4.1 and the TPS-HRI in Figure 4.2, the mean scores are similar with overlapping confidence intervals.

4.2.3 Effects of Morphology

Following the approach taken in the original HRIES paper, a one-way MANOVA first tested whether morphology affected the four HRIES dimensions jointly. The multivariate effect was significant (Pillai's Trace = .100, $F(12, 741) = 2.136$, $p = .013$), so follow-up univariate ANOVAs were conducted to examine the individual

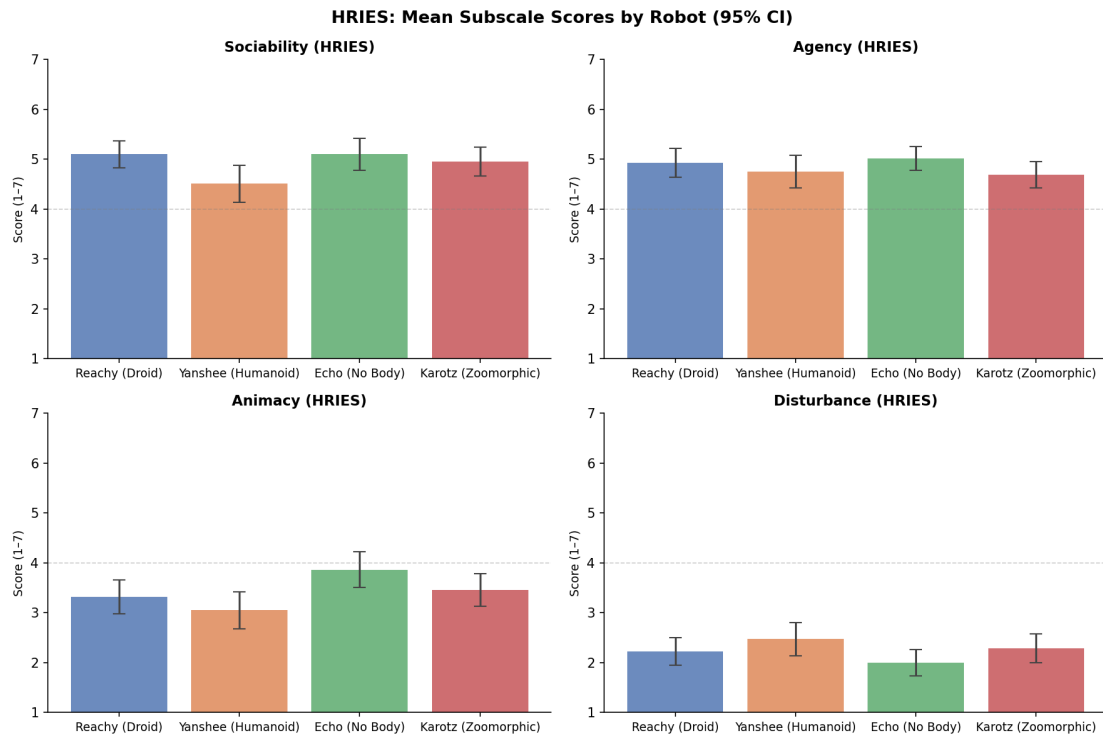


Figure 4.1: HRIES Dimensions Morphology Bar Charts

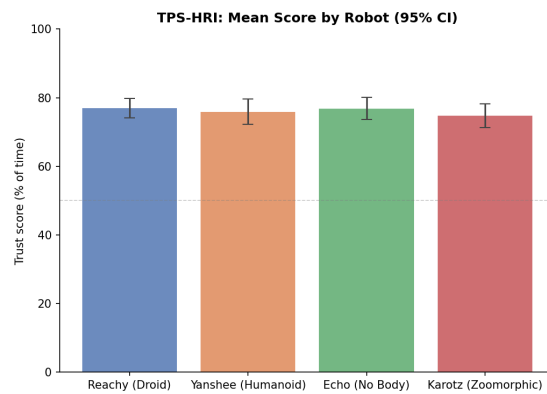


Figure 4.2: TPS-HRI Morphology Bar Chart

Table 4.5: Survey TPS-HRI Descriptive Statistics

Morphology	M	SD
Reachy	76.99%	11.88%
Yanshee	75.92%	14.41%
Echo	76.87%	13.14%
Karotz	74.71%	14.13%
Overall	76.12%	13.35%

dimensions. The multivariate effect was small and was carried mainly by sociability and animacy; agency and disturbance did not differ across morphologies (Table 4.6). A stricter correction across the four subscales would treat both as non-significant, so they are interpreted here as indicative rather than definitive.

Table 4.6: Univariate ANOVA for HRIES Dimensions

Dimension	$F(3, 248)$	p	η_p^2
Sociability	2.99	.032	.035
Agency	1.10	.351	.013
Animacy	3.52	.016	.041
Disturbance	1.68	.172	.020

Pairwise Tukey comparisons within these two dimensions were consistent with that pattern, finding two significant differences. The Reachy received slightly higher ratings than the Yanshee for sociability ($M_{\text{diff}} = .59, p = .046, g = .45$). For animacy, the Echo was higher than the Yanshee ($M_{\text{diff}} = .81, p = .009, g = .56$).

A one-way ANOVA found no effect of morphology on TPS-HRI scores, with $F(3, 248) = .40, p = .753, \eta_p^2 = .005$. The Reachy had the highest trust score, and the Karotz the lowest, but this was not statistically significant, as the effect sizes were negligible. A Tukey’s range test further confirmed there were no significant pairwise differences.

4.2.4 Effects of Participant Characteristics

Participant backgrounds were not collected, so we could not assess the representativeness of the sample. As many HRI studies recruit participants from predominantly WEIRD (Western, Educated, Industrialised, Rich, and Democratic) populations [133], and considering the Prolific platform’s likely user base, it is possible that the observed level of robot familiarity reflects sampling bias; however, this cannot be confirmed from the available data.

A MANCOVA controlled for age, gender, and robot familiarity; morphology retained a significant effect on HRIES scores (Pillai's Trace = .098, $F(12, 732) = 2.057, p = .018$). Participant robot familiarity was the strongest predictor of HRIES perceptions ($F(4, 242) = 7.486, p < .001$), with more familiar participants attributing more anthropomorphic characteristics regardless of morphology. Age group and gender did not significantly predict outcomes.

Similarly, for the TPS-HRI, an ANCOVA was conducted with morphology as the independent variable, controlling for the age, gender, and familiarity covariates. Morphology was still not a significant predictor of trust ($F(3, 245) = .39, p = .763, \eta_p^2 = .005$). For the TPS-HRI, robot familiarity was not a predictor of trust here, nor were age and gender.

4.3 Chapter Summary

We presented the first of the two studies addressing RQ2 in this chapter. An online between-subjects survey was used to capture the trust and social interaction perceptions that form when a robot is observed in a video, using the HRIES [81] and the TPS-HRI [106], with morphology as the independent variable to examine what users report trusting.

Overall, participants evaluated all four robots positively, with high sociability, agency, and trust scores and low disturbance scores. Morphology had a small but statistically significant effect on the HRIES, limited to the sociability and animacy dimensions, but did not significantly influence TPS-HRI trust scores, which were uniformly high across the four robots. Prior familiarity with robots was instead the strongest predictor of HRIES perceptions, though not of trust. More familiar participants attributed more humanising traits to every robot, but were no more trusting than those less familiar. Together, these results suggest that, without genuine interaction, the morphology shown had, at most, a small effect on social perceptions and no detectable effect on TPS-HRI trust. Prior familiarity was a stronger predictor of HRIES perceptions than morphology, although the small multivariate morphology effect remained after controlling for participant characteristics. These reported perceptions are compared with the trust enacted during real use in Chapter 5.

Chapter 5

User Study

We present the user study in this chapter, the second of the two investigations that answer RQ2. While the survey (Chapter 4) captured perceptions formed through passive observation, the user study examined how those perceptions translate into trust enacted during real-world use. Following the reasoning in the methodology (Chapter 3), the study was run as a naturalistic field deployment of an information-seeking robot on the University of Waikato’s Tauranga campus during Orientation Week. Participants could choose if they wanted to approach the robot and ask questions they actually wanted to be answered. The four robots were the same as those used in the survey, and each was deployed on a separate day so that morphology was the intended manipulated variable. In Section 5.1, we describe the methods and in Section 5.2, we report the results. The findings of this study are interpreted alongside the survey results in the combined discussion in Chapter 6.

5.1 Methods

This section describes how the user study was carried out, beginning with the setting and overall approach, followed by the participants involved. The interaction procedure is described, along with how the system works technically. The measures used and the approach for analysing them are detailed; the justification is provided in Chapter 3. The user study was approved by the University of Waikato’s STEM Human Research Ethics Committee with number STEM_HREC(2025)#66. The approval letter and participant information sheet are included in Appendix B.2.

5.1.1 Setting & Approach

Data collection took place in person near the student hub on the Tauranga campus during the first week of A trimester 2026, known as “Orientation Week”. The

robot was intended to support the WSU in welcoming new students with the events planned for the week. The same robots with the distinct morphologies as in the first study were used, each in the same location.

Each robot was deployed on a separate day of the same week, as presenting the morphologies concurrently or alternating them within a day would have allowed direct comparison between robots and revealed the manipulation, undermining the between-subjects design. Orientation Week was selected because it is the busiest week of the semester, as all first-year students, alongside the wider campus population, are reliably on campus, and because the WSU ran the same schedule of events at the same times each day. This kept both the population present and the reasons people had for approaching the robot as consistent as possible across the four deployment days. Assigning one morphology per day nonetheless means effects due to the day cannot be fully controlled, which is acknowledged in the limitations in Section 6.2.6.

Written information about the study was available next to the robot, and the robot provided a brief summary at the beginning of the interaction. Participants were told that there would be different robots each day to encourage them to come back and check what was happening that day. All of the robots were configured to provide identical informational capabilities, ensuring that morphology was the primary variable assessed.

Each robot operated using the same conversational system implemented with the Amazon Nova Sonic 2 speech-to-speech platform. The platform supported speech recognition, natural language processing, and dialogue generation from a provided knowledge base in real-time with minimal delay. Using a speech AI system allowed information to be sought about the provided topic in a natural, conversational way. Audio input was processed by the platform, and recordings and conversation transcriptions were not retained. Instead, just the ratings provided in answer to questions and transcriptions of only the qualitative answers at the end were entered into a database by the conversational agent. The system fully resets after each interaction to prevent data from persisting between participants.

5.1.2 Participants

Students and staff on campus were made aware of the robot’s availability, intended purpose, and location through the same University channels as the survey. Recruitment relied on participants choosing to approach and interact with the robot, regardless of whether they were initially aware of the research. Not recruiting participants meant there was no way to know who participated, and required participants to voluntarily and passively recruit themselves. No direct prompting or requests were made to specific groups, and demographics were not recorded, so the sample diversity is unknown. However, it is intended to reflect the campus demographics,

given its university context. An initial target of 20 completed interactions (five for each robot) was set to support preliminary quantitative analysis between the robot morphologies.

5.1.3 Interaction Flow

A user study session was initiated when a participant approached the robot of their own accord and began the interaction using the wake command, as described in the brief next to the robot. Before any data was collected, participants received a short verbal overview of the study and were directed to the written participant information sheets, which were also positioned adjacent to the robot. The robot then requested verbal consent to proceed. If consent was not granted, the interaction terminated immediately, the participant was thanked, and no data was recorded.

Upon providing consent, participants engaged with the robot in an open-ended, user-driven information-seeking scenario of their choosing. Typical enquiries were explained as relating to events happening that week, such as the free breakfasts and lunches, or the social gatherings. If the question could not be answered (e.g., an unrelated topic or out of scope), the robot informed the user and directed them to where the information might be found. Interaction length varied according to participant engagement and generally ranged from one to three minutes.

Following the interaction, the robot administered a short post-interaction questionnaire verbally, to which participants responded with numerical answers. At the conclusion of the session, the robot provided the participant with a unique interaction identification number that could be used to request the withdrawal of their data at a later time. Participants were also informed that they could alternatively provide the researchers with the estimated date and time of the interaction in case they could not remember their identifier. No researcher was present during the interactions. The robot autonomously managed the consent process, interaction flow, and questionnaire administration to minimise researcher influence, preserve the naturalistic nature of the encounter, and ensure feasibility for the researchers. Having the robot entirely administer the study also required minimal additional time from the participant to ensure successful completion of the interaction. A diagram of the interaction flow is shown in Figure 5.1.

5.1.4 Technical Implementation

The conversational system was developed in Python to support real-time spoken interaction between the participants and the robot, using the Nova Sonic 2 speech-to-speech AI model API hosted by AWS. The design is modular to allow easy

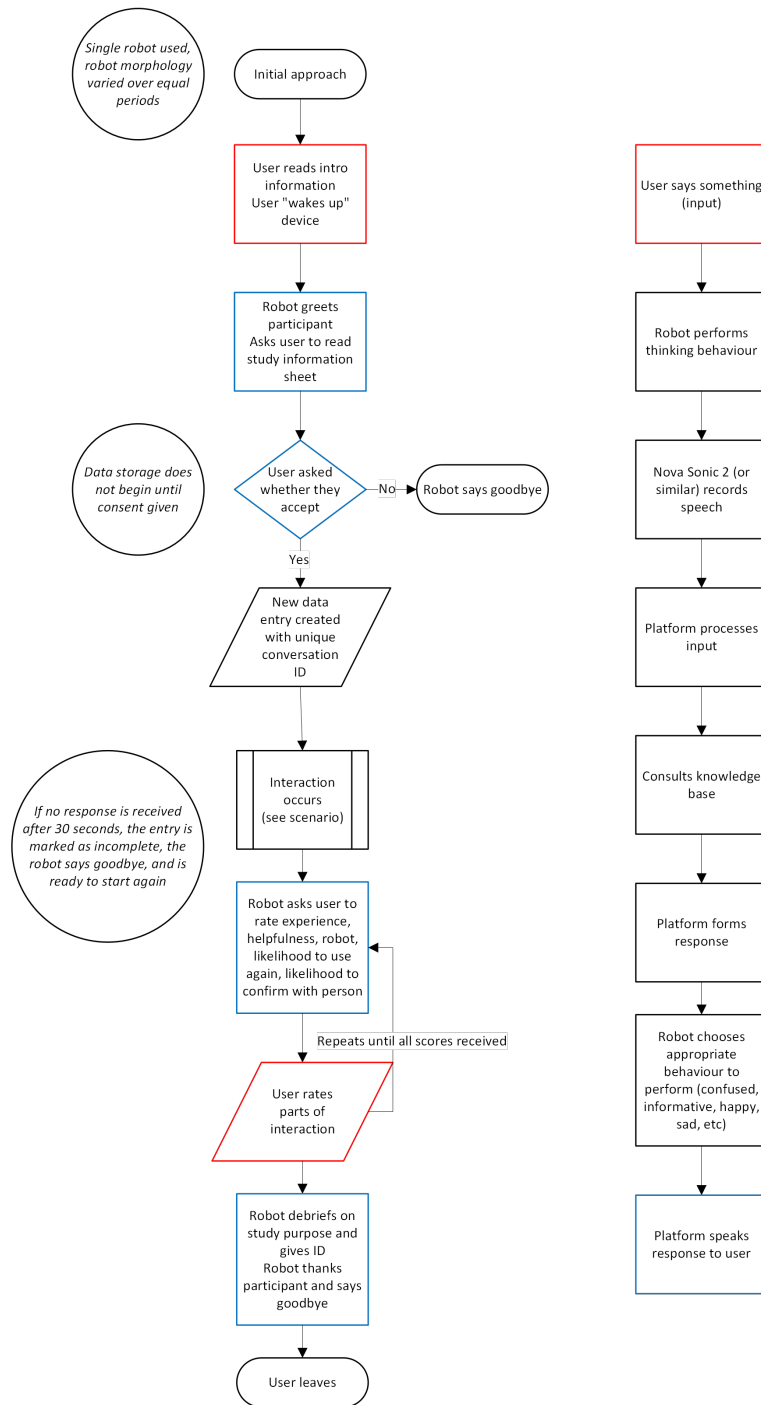


Figure 5.1: Robot Conversation Process

integration with different robot platforms. Example code from GitHub¹ was broken down into audio I/O management, API connection, and tool usage components to allow robot emotion management to be integrated. The code was run on a Raspberry Pi, with a USB microphone and Bluetooth speaker. The Amazon Echo Dot was used as the speaker for consistency, but it was hidden when paired with a robot. This kept the audio consistent and reliable as the robots handled audio inconsistently and with significant delay. The complete system is available on GitHub² for other researchers to adapt for similar projects that require conversation and behaviours, use of a knowledge base, or robot-run studies where a researcher is not present.

Three AWS services were used: Bedrock running Nova Sonic 2 for model inference, Bedrock Knowledge Base for retrieving specific information, and S3 to store documents containing the WSU information. These AWS services are connected to enable the robot to respond and query the knowledge base’s stored information.

Audio

Real-time speech interaction is managed by the AudioStreamer class, which handles bidirectional audio through PyAudio. PipeWire is used to apply WebRTC echo cancellation to the microphone input to prevent the model from hearing itself. Microphone input is captured using a callback in an asyncio event loop, which is sent to the API for processing. Audio chunks from the API are written to the PipeWire playback device as the robot’s spoken response. The API supports a barge-in mechanism that lets the user interrupt the response and ask something else, enabling more natural conversation.

As the API times out after 30 seconds of no events being sent or received, custom wake-word detection (“Hey Robot”) is implemented using Picovoice Porcupine to start an interaction. The timeout and clear new session start also allow for responses for individual interactions and to show an animation to utilise the morphology. Upon detection, the emotion system is activated, and the Amazon Bedrock stream is started, which then waits for a prompt from the user.

AI Connection

The BedrockStreamManager module orchestrates the persistent bidirectional stream to AWS Bedrock, using the ap-northeast-1 region (Tokyo, which is the closest server

¹https://github.com/aws-samples/amazon-nova-samples/blob/2517fd8b58ad6d1083397c7977bafbff87fb0885/speech-to-speech/amazon-nova-2-sonic/sample-codes/console-python/nova_sonic.py

²<https://github.com/cybnz/RobotCommunication>

that can run the AI to minimise conversation latency). Event templates, passed via the Bedrock protocols, handle communication with the API through the session lifecycle, content, and tool usage. Text and audio prompts are sent as these events, and then the system can speak while listening or decide to invoke one of the tools asynchronously. This code also sets the model's temperature, which was lowered to make the responses as consistent as possible. This retrieval-augmented generation (RAG) approach bases AI responses on what is known rather than hallucinating facts.

Access to AWS services is configured through a dedicated IAM user for both the model and the knowledge base to authenticate the code, as the Bedrock API does not allow authentication with the root account. The IAM user was granted AmazonBedrockFullAccess and an S3 policy, the credentials for which can be loaded in as an environment variable to allow the code to work.

Through the AWS Bedrock console UI, the knowledge base is loaded with a text file containing the WSU Orientation Week calendar and the results from a web crawler of the WSU website and the University of Waikato's student support pages. The knowledge base is the information the model can query by invoking a tool when the user asks about events or the WSU. This information is stored in an Amazon S3 bucket and connected to the knowledge base as a data source. To identify data, embeddings were generated using the Amazon Titan Text Embeddings model and indexed into an Amazon OpenSearch Serverless vector store database.

Tool Usage

The system supports dynamic tool invocation by the model at inference time. The ToolProcessor module routes calls to the knowledge base query model and the user rating recorder module as appropriate. Knowledge base retrieval is implemented using the AWS Bedrock Agent Runtime, which searches for matching terms and returns metadata to support information-seeking capabilities. The rating recorder adds an entry to a MariaDB table with the participant's ID, interaction duration, date and time, the robot identifier, the number of questions asked, responses to numerical questions, and transcriptions of optional qualitative responses. Automatically administering the questions allows the researcher to be absent and makes the collection easier for the participant.

Emotion Controller

The robots' physical behaviours are managed by the EmotionManager module, which runs asynchronously to handle robot-specific tasks and control their movements. The manager implements a singleton pattern and exposes a non-blocking emotion method that adds movement commands to a queue, preventing hardware

and networking latency from affecting the conversation. On instantiation, the manager checks the environment variables for the robot to connect to, which then branches into the relevant networking code.

Each robot platform implements an abstract `BaseEmotionController`, defining a common interface for lifecycle management and emotion playback. The main conversation loop then gets the manager instance and adds an emotion to be played; the manager then handles the connection and translates the simple command into the motor controls or animations for the appropriate robot. For robots without specific emotions to be displayed (i.e., the Echo), no methods are overridden, allowing the main conversation loop to function without changes. Emotions included `start_convo`, which the robot performs as an embodiment-appropriate “waking up” animation, `end_convo`, a falling asleep animation, and `thinking`, for the small delay in querying the knowledge base. These three animations utilise the specific morphological features of the robot to ensure the conversation feels dynamic and to make each robot unique, thereby invoking different perceptions. Because movement is a visual feature, it is important to include it as part of assessing morphology.

5.1.5 Measurement and Data Analysis

Both quantitative and qualitative data was collected, with strict controls in place to maintain participant anonymity. For each interaction, the data was retained as a new database entry, containing the numerical ratings, interaction duration, number of questions asked, date and time, robot used, and the unique participant interaction ID. If qualitative explanations for the ratings were provided, transcriptions of only those explanations were recorded and not the rest of the conversation.

Audio input was processed in real time by the Amazon Nova Sonic 2 platform to enable interaction, but was not retained by Amazon or the researchers. Conversational content or transcriptions relating to the task were also not stored or analysed. No video recordings, personal identifying information, or task-specific information were collected. This study focused on participants’ perceptions of the robot rather than the actual accuracy or substance of the information provided (i.e., the information-seeking and delivery capabilities were not being tested).

The post-interaction questionnaire consisted of three verbally administered 5-point rating-scale items:

1. Out of five stars, how helpful was the robot at providing information?
2. Out of five stars, how well did the robot answer your questions?
3. On a scale from one to five, where one is “not at all” and five is “absolutely”, how likely are you to verify the information provided?

For each of these, participants were asked if they would like to explain what about the appearance made them give that answer.

These items assess the robot as an information source, regarding perceptions of usefulness, competence, and trust that it could provide correct information. Statistical tests were performed on this quantitative data to determine differences in ratings for each question across the robot morphologies. Each questionnaire item was analysed independently to determine whether robot morphology had a statistically significant effect on perceived trust and interaction quality, and whether the morphologies differed significantly from one another. Qualitative findings are used to contextualise quantitative results and to identify whether any trust considerations emerged that were not captured by numerical measures. The qualitative data also reveals what participants believe affects their perceptions, even if it is not reflected in their quantitative responses.

As set out in Section 3.4, the verification item runs in the opposite direction to the other two, as not checking an answer signals greater trust, so it was reverse-coded prior to analysis, giving all three items a common direction, where higher means more trust. After recoding, a score of five on any item denotes the most helpful, most fit-for-purpose, and most trusted response. For conciseness, these three dependent variables are referred to throughout as helpful, useful, and believable.

In addition to the self-report items, two behavioural measures were derived automatically from each interaction: the number of questions a participant asked and the total interaction duration. Unlike the rating items, these record what participants did rather than what they reported, bringing the analysis closer to enacted trust. The relationship between these behaviours and trust is not strictly directional; more questions could reflect either reliance or scepticism. Gathering behavioural data was exploratory and analysed identically to the rating items; interactions were brief, so these measures were not expected to explain much variance. Further, all interactions were brief because the use case was not complex, so these measures are likely to offer little explanation of enacted trust.

Qualitative responses were analysed using an inductive thematic approach [134]. First, the primary researcher read all of the transcribed responses to become familiar with the data. The reviewer then used open coding, assigning descriptive codes to each response without using a predefined coding framework. Similar codes were grouped into themes, which were reviewed and refined by comparing them with the full dataset. Where responses within a theme showed meaningful differences, subthemes were created. The coding was then presented to the other researchers, discussed, and refined. To show how the themes were developed from the data, illustrative quotations are included alongside each theme.

As source, message, and disposition are distinct assessments of trust as described in Section 3.4, the three items were analysed separately rather than as a composite;

Cronbach’s alpha was computed to test whether that separation is empirically justified. As the rating items are five-point and ordinal, and group sizes are likely to be small and unequal, the data was assumed not to be normally distributed, so non-parametric methods were used for analysis. For each dependent variable, a Kruskal-Wallis test was used to assess whether the distribution of ratings differed across the four morphologies.

Because three tests examining overall ratings were conducted, the Benjamini-Hochberg procedure was used to reduce the risk of false-positive results due to multiple testing. When one of these omnibus tests remained significant after this correction, Dunn’s test was used to compare the morphologies for this rating question. Pairwise, this means each robot was compared with the other three for a total of six comparisons, with a Bonferroni adjustment applied to the p-values. The two behavioural measures were also analysed with the same Kruskal-Wallis procedure.

Because recruitment was uncontrolled and the sample size was expected to be modest, a sensitivity analysis was conducted to establish the smallest effect the design could reliably detect at 80% power ($\alpha = .05$). Unmanaged recruitment bounds the interpretation of any non-significant result.

The unit assessed was individual participants, even though some were in groups, as they gave answers individually. All analyses were conducted in Python, using SciPy for the Kruskal-Wallis and chi-squared tests, scikit for Dunn’s test, statsmodels for the Benjamini-Hochberg correction, and Pingouin for Cronbach’s alpha and the sensitivity analysis.

The user study measured trust as it was enacted during interaction, avoiding asking participants directly about trust or morphology so as to not influence their answers. To examine whether the actualised most trusted robot matches what people believe looks trustworthy, a short follow-up survey was conducted on the day following the final robot deployment. Here, the robots could be directly compared in a picture of them together, and passers-by on campus were asked which of the four looked most trustworthy for providing information, with no other context. Responses, therefore, reflect first impressions rather than an experience of its functionality. Participants could optionally offer a brief verbal reason for their choice.

The follow-up survey was a separate, self-recruited sample, and as a single forced-choice judgement, it is less rigorous than the interaction-based measures; it is included as a complementary point of comparison rather than as a primary result. Responses were analysed with a chi-squared goodness-of-fit test, comparing the observed votes against the uniform distribution expected if morphology had no bearing on which robot was judged most trustworthy. The results are reported at the end of this section.

5.2 Results

This section reports the findings. Participation numbers and descriptive statistics are summarised. Following this are the tests used to infer what the findings say about morphology from the self-report and behavioural measures. Pairwise comparisons identify those effects with the differences in morphology. The quantitative results are pulled together before reporting themes from participants' explanations of their ratings. Finally, the results of a follow-up survey are reported, and the findings are summarised ahead of interpretation in Chapter 6.

5.2.1 Participants & Overview

Over the course of the week the study was conducted, there were 54 participants recorded over 33 interactions (because some people interacted in groups). Due to the naturalistic observational design, whether people chose to approach and interact was self-directed and could not be controlled; thus, group sizes were unequal each day. Monday with the Reachy Mini robot had 18 participants; Tuesday with the UBTECH Yanshee had 9; Wednesday with the Amazon Echo Dot had 10; and Thursday with the Karotz had 17 participants. The presentation order was chosen by the WSU to maximise engagement. Some participants interacted with more than one robot, but this was not recorded because identities were not known. As the repeat participants were unknown, these have been treated as unique participants.

27 of the 54 participants provided qualitative explanations for the quantitative ratings. Participants responded to three dependent variables based on their interaction: the perceived helpfulness of the robot, the perceived answer quality, and the likelihood to verify, referred to as helpful, useful, and believable respectively throughout this section. The believability rating was scored from one to five, where one is “not at all” and five is “absolutely”, in terms of how likely they were to verify the information provided, but reverse coded for analysis. For clarity, in the analysis, a score of five for each variable means the information was the most helpful, the most useful, and the most believable.

5.2.2 Descriptive Statistics

The three dependent variables representing different aspects of trust from the literature review were evaluated for their psychometric properties and internal consistency. The Cronbach's alpha for the three variables was calculated to be .36 or “unacceptable” reliability, well below the .7 threshold considered acceptable, meaning the three items do not form an internally consistent scale; consistent with the literature review findings that these aspects of trust are unique and come together to form the greater picture of trust. Because of this low internal reliability,

each item was analysed as a separate dependent variable throughout the results. Note that five-point Likert scores are ordinal data, so are not expected to follow a normal distribution, so non-parametric tests were used. Table 5.1 shows a summary of the means and standard deviations of each question for each robot morphology.

Table 5.1: User Study Descriptive Statistics

Robot	Helpful		Useful		Believable	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Reachy	4.2	.88	3.8	1.26	3.6	1.34
Yanshee	4.4	1.01	4.7	.50	2.0	.71
Echo	3.8	1.23	3.8	1.14	2.5	1.43
Karotz	3.3	1.11	3.2	.90	2.2	1.03

For the helpfulness question, the Yanshee returned the highest mean helpful rating at 4.4, and the Karotz the lowest at 3.3. The Reachy and the Echo were closer to the middle, with overlapping distributions. The standard deviations were moderate and broadly similar across groups, with the Echo showing the most variance in responses, with a spread of 1.2. The Yanshee also had the highest mean usefulness score, at 4.7, with a small standard deviation of .50, indicating high agreement among participants. The Karotz again received the lowest score of 3.2, while the Reachy and the Echo had similar scores in the middle. When looking at the believable rating, however, the Reachy had the highest mean at 3.6, indicating participants were least inclined to verify its information. Contrasting with the other two questions, the Yanshee had the lowest score here with 2.0, followed closely by the Karotz’s 2.2, indicating participants were most inclined to verify the information provided by these robots. All robots saw a drop in scores, but the Yanshee’s inversion from being the highest score on the first two questions to the lowest score on the last question is notable.

5.2.3 Inferential Statistics

Kruskal-Wallis tests were conducted to formally determine whether morphology had a statistically significant effect on each dependent variable, shown in Table 5.2. Each question had an independent test, with Benjamini-Hochberg false discovery rate correction applied across them to control false positives (Type I error inflation).

All effect sizes are large, and the effect is significant after correction. For each question, a large effect size is notable given the relatively small sample size. A sensitivity analysis confirmed the design had 80% power to detect effects of Cohen’s $f \geq .47$ ($\epsilon^2 = .18$), corresponding to a large effect. The believable effect ($\epsilon^2 = .19$) exceeds this threshold; the helpful (.14) and useful (.17) effects fall marginally

Table 5.2: User Study Kruskal-Wallis Tests

Question	H (3 df)	p	p -adjusted	ϵ^2
Helpful	10.11	.018	.018	.14
Useful	11.55	.009	.014	.17
Believable	12.34	.006	.014	.19

below it, so those two should be treated as detected at slightly lower-than-nominal power. The study was adequately powered to detect large effects, though it would be underpowered to detect small-to-medium effects. Given the results, morphology significantly affected perceived helpfulness, perceived answer quality, and perceived believability. The believable question had the smallest adjusted p-value and largest effect size, indicating that morphology had the strongest impact on perceived believability of the robot’s information. All three effects survived correction, so the probability of any one of the results being a false positive is controlled.

Beyond the self-report ratings, two behavioural measures were also examined: the number of questions asked and the interaction duration. Kruskal-Wallis tests found no significant effect of morphology on questions asked ($H(3) = 2.18, p = .535, \epsilon^2 = .00$) or on interaction duration ($H(3) = 4.80, p = .187, \epsilon^2 = .04$), with negligible effect sizes in both cases. Given that the design was powered only to detect large effects, these results indicate the absence of a detectable large effect rather than positive evidence of no difference. Further, given the simplicity of the interaction, there is likely to be no effect regardless.

5.2.4 Post-Hoc Pairwise Comparisons

Dunn’s tests were conducted to determine whether morphological differences drive the effect on each question, given that all were significant. Each morphology was compared against every other to find whether there are significant differences between any morphology pair (of which there are six unique pairs). Bonferroni-adjusted p-values for each question are presented in Tables 5.3, 5.4, and 5.5.

As seen for the helpful rating in Table 5.3, only the Yanshee-Karotz contrast survived correction with a p-value of .03, meaning the Yanshee’s morphology was perceived as significantly more helpful than the Karotz, whereas all other pairwise comparisons were non-significant after correction. That is, no effect was significant between the rest of the morphologies and how helpful they seemed. This pattern means the overall Kruskal-Wallis effect on the helpfulness question must be localised to the Yanshee-Karotz contrast.

Again, for the useful rating, only the Yanshee-Karotz comparison was seen to be significant in Table 5.4, and this p-value is the smallest in any of the questions,

Table 5.3: Helpful Rating Morphology Pairwise Comparison

Comparison	<i>p-adjusted</i>
Reachy vs. Yanshee	1.000
Reachy vs. Echo	1.000
Reachy vs. Karotz	.071
Yanshee vs. Echo	.908
Yanshee vs. Karotz	.030
Echo vs. Karotz	1.000

being distinctly below the .05 threshold at .004. The pattern of only the Yanshee-Karotz comparison being relevant is consistent with the helpful question but with a stronger contrast, which is consistent with the smaller standard deviation observed for the Yanshee robot in the useful question (Table 5.1). Despite the Reachy having a lower mean usefulness score than the Yanshee, this is not statistically significant, so it cannot be said that the Reachy is less fit-for-purpose than the Yanshee.

Table 5.4: Useful Rating Morphology Pairwise Comparison

Comparison	<i>p-adjusted</i>
Reachy vs. Yanshee	.316
Reachy vs. Echo	1.000
Reachy vs. Karotz	.455
Yanshee vs. Echo	.387
Yanshee vs. Karotz	.004
Echo vs. Karotz	1.000

For the believability rating in Table 5.5, two contrasts were significant for the believable question, with the Reachy-Yanshee and the Reachy-Karotz comparisons surviving correction, whereas the Yanshee-Karotz comparison was not significant for this question, in contrast to the findings for the other two questions. In terms of whether the information provided by the robot was believable, the Reachy is the morphology that was differentiated from the others.

5.2.5 Summary of Quantitative Findings

In the self-report ratings, morphology was found to significantly affect all three perception dimensions, helpfulness, usefulness, and believability, with large effect sizes throughout. However, the pattern of the effects varied by dimension, in that helpfulness and usefulness were differentiated between the Yanshee and the Karotz, while believability was differentiated with the Reachy outperforming both

Table 5.5: Believable Rating Morphology Pairwise Comparison

Comparison	<i>p-adjusted</i>
Reachy vs. Yanshee	.024
Reachy vs. Echo	.212
Reachy vs. Karotz	.021
Yanshee vs. Echo	1.000
Yanshee vs. Karotz	1.000
Echo vs. Karotz	1.000

the Yanshee and the Karotz. The Yanshee robot was found to have higher mean helpfulness and usefulness scores than the other robots, but for believability, the Reachy had the highest scores, meaning these outcomes are distinct. A robot being given higher trust in one area does not mean it is trustworthy in another. This divergence in scores is consistent with the scale’s low internal consistency and justifies treating the variables as distinct rather than as a single combined score in the analysis, as they are not related.

5.2.6 Qualitative Findings

After scoring the robot, 27 participants provided context and reasoning for their answers. These qualitative responses frequently addressed the underlying AI model rather than the morphology, but the assessment of a robot’s capabilities still varied despite the model staying consistent. Themes were identified from the 27 sets of explanations using the inductive procedure described in the method (Section 5.1).

Theme 1: Appearance as a Signal of Function

Many participants noted that how a robot looks changed the implied purpose, shaping how the engagement went and the interpretation of the intent. For the Reachy, the “face is believable”, and the “combination of face and movement helped make it seem responsive”, allowing users to infer the information-seeking role. The Reachy’s positive use of appearance signals is emphasised by a participant who noted after interacting with the Yanshee that “Reachy looks like it provides information, but this one [Yanshee] doesn’t make that communication capability obvious” despite its humanoid form. With the more recognisable purpose of the Echo, a participant said it is “like a kiosk, the focus is on the information not the interface or mode of delivery”, meaning preconceptions do help shape purpose. With the Karotz, multiple participants commented that its appearance did not suggest it could hold a conversation, saying it was “cute but I don’t have anything to ask it. Why would I talk to a robot?” and another asking “what do I ask her”,

indicating they were not sure what it could do and therefore did not end up asking that robot questions after beginning the interaction.

Theme 2: Movement and Expressivity

Intent was signalled by the movement and morphological features, with the Reachy seeming more dynamic in its expressivity, with a positive note: “combination of face and movement helped make it seem responsive”, with the antenna, eyes, and “jiggly” head being cute and appealing to participants. The Yanshee’s form generated comments such as “looks cool”, “arms and legs give different vibes”, but were not necessarily linked to the function it performed. The Yanshee’s humanoid appearance does imply that it can move around; however, it did not walk during the study, which may have diminished the anthropomorphism trust effect seen in literature. No participants commented on the Karotz’s animations, likely due to its limited movement capabilities.

Theme 3: Cuteness and Approachability

Cuteness was repeatedly cited as a characteristic that positively influenced participants’ decisions to approach and engage. Participants frequently described cute robots as more approachable, expressing greater interest in initiating interaction and asking questions. Further, some participants noted that it even helped them trust the robot, with participants mentioning that a robot is “more trustworthy if cute”, that the Karotz was “adorable so clearly trustworthy”, and that the Reachy was “cute, if it was ugly I would be less inclined to talk to it”. Cuteness appeared to improve participants’ confidence to engage. Across responses, cuteness was associated not only with positive affect but also with increased willingness to approach the robot, engage in conversation, and rely on it for information.

Theme 4: Type of Information and Trust

Several participants explicitly described trust as conditional on how high the stakes of relying on the information would be, as it “depends how important the information is”. Participants were not overly concerned given the context of this experiment, as it involved “not sensitive information” and they were “not worried about checking, pretty low stakes”. The lack of concern suggests that participants interpreted the believability question differently depending on the topic they discussed, thereby introducing some within-group variance.

Theme 5: Prior Expectations and Familiarity

Regarding the Echo, participants tended to recognise it as a smart speaker, whether from its general aesthetic or personal experience, so they brought pre-existing expectations about how it should perform. “If you know what it does you expect it to work to the best of its abilities”, and the looks of the Echo specifically “come with expectations of a service”. Meaning, the Echo is not a neutral baseline; rather, it is a recognisable consumer product with an established reputation. Participants also believed that the “higher-quality” a robot looks, the more likely it is to perform well because of the effort put into it by the creators.

Theme 6: Confounding AI Response Quality

One robot was described as having “answered well, but because I knew it was wrong [it received a low score]”. By users familiar with multiple of the robots, the Yanshee “gave more wrong answers”, and the Karotz “couldn’t answer specifics”. These comments suggest the AI response quality varied across robots, which may have affected the scores given independent of morphology. The Karotz was described negatively in one case, allegedly “fast-paced like an influencer” and “annoying”. However, with the low model temperature and context controls, there was very little change between responses. Perceived differences in performance indicate that the appearance of the robot changed how participants received the information from the robot, as the underlying speech-to-speech model was the same.

5.2.7 Direct Perceptions of Trust in Appearance

On Friday of the same week, the day following the last robot deployment, a picture of the four robots together was shown around campus, and participants were directly asked which robot looked the most trustworthy, to evaluate whether the robot considered to be the most trustworthy is the same as what was determined through the user study. This survey directly asked about morphology and trust, whereas the user study did not mention these concepts directly.

51 people participated in this survey, with 24 (47%) reporting that the Reachy robot looked the most trustworthy. In contrast to the user study, the Yanshee robot was not the clear winner here, with 12 (24%) of the participants voting it as having the most trustworthy appearance. Again, the Echo and the Karotz robots were low performers, with seven and eight votes, respectively. A chi-squared goodness-of-fit test compared these counts against the 12.75 votes per robot expected if all morphologies were equally perceived to be trustworthy. The distribution departed significantly from that expectation, $\chi^2(3) = 14.33, p = .002$, so the votes were not spread evenly across the four robots. The Reachy accounted for almost half of all

votes, twice as many as the Yanshee and more than the Echo and Karotz combined.

Participants who voted the Echo as most trustworthy reported that it was because “Alexa is established and trusted by a big company, so you expect it to work well”. Themes similar to the user study qualitative responses emerged. Those who voted for the Reachy did so because “Reachy is cute and expressive, which makes it much more friendly”. The Reachy was likened to popular droids from movies such as Baymax, WALL-E, R2-D2, and BB-8, and was the only robot so compared. Some voted for the Yanshee because “[it] is closest to a person and I would trust a person much more”; however, others did not vote for it and mentioned “Yanshee is a bit scary but not too bad because it is small” and “it would be threatening if bigger”. Again, the Karotz was not trusted, with a participant saying “why would I want to get information from an animal?”.

5.3 Chapter Summary

In this chapter, we covered the second of two studies answering RQ2, with a user study measuring trust enacted when information-seeking with a robot, complementing the perceived trust captured by the survey (covered in Chapter 4). Four robots with different morphologies but a consistent conversational system were deployed so that students could ask questions about Orientation Week. Distinct aspects were measured using three 5-point rating questions (helpful, useful, and believable), and two behavioural measures (questions asked and interaction duration).

Morphology significantly affected all three rating questions, with large effect sizes, but the pattern varied by dimension. The Yanshee scored highest on helpfulness and fit-for-purpose, driven by its contrast with the Karotz, while believability was instead differentiated by the Reachy, where participants were least likely to double-check its information. The three items showed low internal consistency, as seen by the diverging patterns. The facets of trust operate independently; a robot trusted in one respect was not necessarily trusted in another. Neither behavioural measure differed significantly by morphology. Qualitative responses highlighted cuteness as encouraging people to approach. Explanations also showed the need for a clear link between appearance and function, and that device familiarity and perceived answer quality also contributed to the ratings given. A follow-up survey directly comparing the robots’ appearances placed the Reachy as the most trustworthy-looking source of information. These findings are discussed in Section 6.2 and compared with the first study in Section 6.3.

Chapter 6

Discussion

Following the results of the survey in Chapter 4 and the user study in Chapter 5, in this chapter, we discuss what those results mean for how morphology affects trust in robots when information-seeking. The two studies give distinct and somewhat opposing findings. During passive observation, when participants watched a robot being used, morphology had little effect. With direct interaction, when participants chose to approach and question the same robots, morphology had a much stronger effect, shaping trust across every aspect measured. Because the robots, task, and topic were held constant, this divergence cannot be fully attributed to the robots, and instead needs interpreting in terms of what changed between the two settings. We argue that it reveals how trust is bounded by the type of interaction and its context, as well as embodiment, and the nature of the interaction. These change whether a robot’s appearance is merely observed or is actively relied upon as a channel for judgement.

In this chapter we first interpret the studies individually then bring them together to account for their differing findings and to answer RQ2. The discussion will cover the implications of the survey results in Section 6.1, the user study in Section 6.2, and the differing results between the two studies in Section 6.3. What the studies answer about RQ2 in light of the findings of how robot morphology can affect trust from RQ1’s literature review is discussed in Section 6.4.

6.1 Survey

The survey investigated whether robot form affects people’s perceptions using the HRIES and TPS-HRI scales, by observing an information-seeking interaction. The results showed that morphology did not significantly affect the ratings given across the different robots for the TPS-HRI, but it did for the HRIES. Participant familiarity with robots in general, however, was revealed to be the strongest

predictor of HRIES results but not trust scores. This section will discuss the small effect of morphology, patterns observed in the scales, the effect of familiarity, limitations, and future work.

6.1.1 Small Effect of Morphology

Morphology was found to have a significant but small effect on parts of the HRIES. With over 60 results per group, there is some confidence that any effects that exist would have been found. With these small effect sizes, fewer participants would make subtle differences hard to detect. Power analysis indicated that the HRIES MANOVA was adequately powered, with 91% power to detect these small effects. With the TPS-HRI, there was insufficient statistical power to reliably detect small effects, but it was powered to detect large ones, so this null result means that large effects would have been detected. While the results do not support the conception that there would be large differences across parts of interaction (i.e., sociability, agency, animacy, disturbance, and trust), the presence of statistical significance indicates that morphology still had small effects.

Although morphology had a significant effect on sociability and animacy, the effect sizes were small, with morphology accounting for only 3.5% and 4.1%, respectively. The remaining variance must be attributed to other factors, such as prior familiarity. The two significant pairwise effect sizes were around medium, so larger differences in scores may be found if participants saw two robots. Further, comparing multiple robots may make the relative difference more apparent, leading to more distinct scores. However, if participants were to watch the same video with different robots, they would realise they function identically, which would also normalise scores.

The larger difference in mean scores between the Yanshee (3.05) and the Echo (3.86) on animacy suggests that the static smart speaker, with no body, feels more dynamic and alive than the humanoid robot, with its crafted design and animations. The Reachy, in particular, is capable of the most movement and behaviour, so it would be expected to have the highest animacy rating, although most of the communication was fairly static and did not make good use of these capabilities. With a more focused study of the animation features, these patterns could be investigated in greater detail to confirm them.

The information-seeking context was intentionally low-stakes in order to facilitate the second study. However, high-vulnerability scenarios can have a strong effect on trust [105]. Emotional engagement with a robot is influenced by personal interaction [126, 127], with a robot's visible features serving as social cues. The participants in this study did not personally interact with the robot, which likely means they have less attachment than if they had. The video-based study in the HRIES paper had less strong effects than real-world interaction, but they were still

present.

As the task was communicative, participants may also have focused more on task performance and on what the robot was saying, rather than on how the appearance made them feel. In the meta-analysis by Hancock *et al.* [21], task and performance are key components of trust, and if more attention were placed on those, then participants might not have thought much about appearance. The equalisation of results is therefore more due to the task and behaviours being identical. If participants were told to explicitly focus on appearance, the effects may have seemed larger and more important than they actually were. Particularly in information-seeking, the robot’s behaviour changes anthropomorphic attributions [56].

The four robots are commercially available and designed so that people will want to use them. As such, they do not represent extreme morphologies. If one were designed to be intentionally creepy, for example, then the interaction ratings would have been greatly affected, like in the work by Turner *et al.* [87]. The morphologies shown were distinct, but given the short, non-interactive video, they may not have elicited a strong enough reaction to be measurable, especially without a direct comparison to another robot.

6.1.2 Scale Dimensions

The Yanshee scoring the lowest for sociability is a counterintuitive finding, as the HRIES framework predicts that human-like behaviours and morphologies should increase the attributed warmth and friendliness. Instead, the opposite was observed, with the Yanshee’s mean rating being significantly lower than the Reachy’s. How sociable a robot seems is more complex than just its form; its animations and cuteness are also considerations, as discovered in Chapter 5.

Having human-like properties may also raise expectations about how the robot should behave, rather than just enhancing perceived social performance. A humanoid form comes with anthropomorphic standards of interpersonal behaviour, and if the robot was not seen to express the expected warmth of a person, then this form may not have worked in its favour. The lack of facial expression, small scale, and voice did not help meet these expectations, whereas these were likely not comparisons made of the other robots. As the Yanshee is small and was placed on a table, some participants interpreted it as more like a toy, sending the wrong signals of expected task and performance. A larger humanoid may produce a different result, so this finding cannot be generalised to all humanoid robots.

Without a body, the Echo was rated as significantly more alive and human-like than the Yanshee on the animacy dimension. Simple physical presence did not mean more animacy, and this may be due to the equal speech capabilities. Speech is a channel through which people can gauge consciousness and intentionality. The

fluent and responsive voice applied to the Echo meant there was no contradiction of this, whereas the Yanshee’s plastic appearance and stiff movements may have reminded participants that it was a machine.

Many participants recognised the Echo, which may also have helped its ratings, as it is more physically present and recognisable in their homes. A less familiar smart speaker may still have been likened to the Echo, and it is difficult to establish a baseline in a different way when focusing on morphology, so these results are unlikely to change in a similar study under the same conditions.

Agency ratings were uniformly high across all morphologies. Since the robots had the same functionality as an information-seeking tool, there was little difference in the capabilities each robot was seen to have. The perceived rationality and goal-directedness aspects are seemingly attributed to what the robot does, not what it looks like.

Disturbance was uniformly low, with a mean of 2.25 out of 7. Despite the Yanshee having the highest score (i.e., more disturbing), this was not a significant result and was unlikely to be an uncanny valley effect. Low disturbance is possibly due to the helpful nature of the robots, the generally disarming, toy-like appearance of their bodies, or, in the Echo’s case, its ability to blend with appliances and furniture.

Trust scores were flat, indicating that interaction perceptions and functional trust are not strictly intertwined. The HRIES captures what participants think the robot seems like, and the TPS-HRI captures what they think it will do. Because the robots did the same task with identical performance, trust estimates are based on factors other than morphology. Given the participants’ limited experience, the robot’s reliability was the same, and morphology was not found to be a consideration. Having such similar trust scores also suggests that physicality does not give a trust advantage when faced with the same behaviour.

Taken together, these different parts suggest that perceptions are only partially sensitive to morphology. Appearance shapes how warm and lively a robot seems, but not how rational or intentional it seems, which is instead inferred from observed behaviour, regardless of morphology. Low disturbance provides context to sociability, showing that there is no discomfort due to appearance, but rather an unmet social expectation. Trust was entirely unaffected by morphology in this study, separating social perceptions from functionality, with the robot’s task performance trusted equivalently regardless of how sociable or alive it seemed.

6.1.3 Prior Familiarity

Familiarity with robots emerged as a consistent predictor assessed in this study, more so than morphology. For the HRIES, familiarity was a significant covariate with a p-value of $< .001$, but not for the TPS-HRI with a p-value of .115. Unlike

morphology, familiarity predicted higher scores on every HRIES dimension and was independent of that morphology.

Anthropomorphic perceptions are appraisals of interactions and are shaped by prior attitudes, mental models, and expectations, all of which depend on prior experience with robots. Those who are more familiar with robots bring more complete preconceptions of what they think is intended, which seems to make their perceptions more personal, thus assigning more human qualities. Morphology can help set those initial expectations of a robot’s purpose, but with minimal interaction, the user is required to base their judgements on what they know, rather than what they have seen.

In contrast, trust and the questions asked about it related to task performance. The ANCOVA found no significant predictor of trust in this study, meaning trust is entirely based on other factors, such as what they saw the robot do, rather than what they brought as individuals to the encounter.

Greater familiarity likely reduced uncertainty and negative preconceptions towards the robot, which is consistent with findings from Hei *et al.* [70]. Because of this familiarity, participants may have been more charitable with their attitudes. Breaking down familiarity into personal experience and awareness from media or other sources would help determine where participant attitudes come from. The absence of a morphology effect helps confirm that familiarity is more based on a person’s general disposition than on the chosen independent variable.

Familiarity influenced perceptions equally regardless of morphology, as more familiar participants rated all robots more positively by a similar amount. Those who were familiar with robots did not rate the Yanshee as more sociable relative to the Reachy, for example. Conversely, less familiar participants are not dramatically influenced by form; how a robot looks does not make up for the rest of the interaction, no matter how familiar the person is with robots.

In contexts where users are likely to be less familiar with robots or are first-time users, such as support in aged care or developing regions, they will have less positive attitudes towards any robot encountered [66, 77]. Helping to expose users to a robot’s purpose may improve attitudes and reception more than an appealing design alone.

6.1.4 Limitations

Participants watched a one-minute video rather than using the robot themselves. This was necessary to reach a large enough sample, but it means the ratings describe observations of robots, rather than the users themselves relying upon the robot, and the small effects reported here should be read on those terms. Each participant also saw only one robot. This was a deliberate choice, as showing several would

have made it obvious that performance was identical across conditions, but it left participants judging against prior experience rather than a direct comparison.

One robot represented each embodiment type, so the findings describe these four robots rather than their categories as a whole; the Yanshee is small and toy-like, and a larger, more realistic humanoid may well have been rated differently. The sample was mostly aged 16 to 34, and no data was collected on background or country of residence, so how well it represents a wider population is unknown. The two recruitment groups were also pooled for statistical power despite likely differing in background. Finally, the task made little use of the robots' bodies, so the flat trust result may not extend to tasks such as caregiving or collaboration, where the body is more visible and part of the interaction.

6.1.5 Future Work

Using a wider range of robots within each embodiment type would show how much of the effect comes from body type and how much from the design of a particular robot. Studies in which participants see and directly compare several robots would also make differences more visible, although the task would need to be designed so that morphology remains the main thing that changes, and pairwise comparisons would require a large sample.

Because speech was the main channel in this task, further work could separate the contribution of voice from that of form, and test whether the flat trust result holds in contexts where the body is more relevant. Measuring familiarity in more detail would also help, distinguishing personal experience with robots from awareness through media and other sources, and recording more demographic information would show which participant characteristics account for the variance that morphology did not explain. Repeated or longer exposure would allow familiarity with a specific robot to develop, and would support varying reliability over time to see how trust responds, as in Schaefer [106].

6.1.6 Survey Summary

From this study, morphology has been shown to have no effect on many different aspects of perceptions, and only a small effect when significant. This small or absent effect pattern challenges the general sentiment that a human-like form means more positive perceptions and more trust. Morphology was found to matter very little in this context with this style of interaction, with familiarity being a more consistent predictor of perceptions than morphology.

Morphological design choices produced modest and somewhat unexpected differences in perceptions when people are not involved in information-seeking with a robot. When the users themselves do not need to place trust in the robot, their

trust is not dependent on perceived performance. Common conceptions about design choices may not be as important as previously thought.

The HRIES and TPS-HRI have different amounts of significance, showing that perceptions and trust are complex and work at different sensitivities. It is key in research to use validated scales carefully to determine where effects come from. Morphology can be a major factor, as will be discussed with the second study, but interacting in-person with a robot changes perceptions compared with watching a short video.

6.2 User Study

The user study set out to give qualitative answers to RQ2, how epistemic trust during information-seeking scenarios is affected by robot morphology. As the results show, morphology significantly affects trust in a robot as an information source, but the effect is not uniform. Morphology affects the dimensions of trust in different ways; this discussion will explore the qualitative findings to determine how. This section will include an interpretation of the findings, the connection to literature, limitations, and future directions.

6.2.1 Interpretation of the Findings

The user study's main contribution is not that morphology impacted perceptions, but rather how it mattered in different ways across the trust dimensions. The humanoid robot (Yanshee) was rated highest on both perceived helpfulness and perceived answer quality, while the mechanomorphic robot (Reachy) was rated highest on believability.

While the first two dimensions are important for people to place their faith in a robot as an information source before interacting, the last dimension represents the actual trust after the interaction has taken place, whether they actually trust what the robot says, represented by the low likelihood of verifying the information provided. The mechanomorphic and the humanoid robots differed significantly, but not in the direction expected from the literature. The robot rated most capable at providing good answers was simultaneously among the least trusted to provide accurate ones. The Yanshee's superiority in usefulness ratings did not extend to believability, and the Reachy being more believable did not mean more useful.

Dissociation between the trust aspects is reinforced by the scale's low internal consistency. A Cronbach's alpha of .36 falls well below the conventional threshold of .70 considered acceptable for treating items as a unified scale, and the diverging patterns in the pairwise analysis explain why; the three items do not assess the same underlying construct. Had these items been combined into a composite score,

as seen in some studies [135], the opposing effects would have cancelled each other out, thereby losing this finding. Treating the items separately is not a concession to poor reliability but rather a reflection of the complex aspects of trust.

In the pairwise comparisons, the Echo was never significant, and it did not serve as a baseline in the expected way. The Echo was intended to be the control for this experiment as it functioned in the same way as the robots, but did not have distinct morphological features. Instead, the Echo’s neutral ratings may be due to the preconceptions and experiences some participants reported they had around smart speakers. Participants said good performance is expected from the Echo, so the scores are middling and not significant when compared to any of the robots.

Only the Karotz robot had a participant project a gender onto it, perceiving its morphology as distinctly female. Participants likely saw the other robots as objects, albeit animate, rather than living beings. This objectification may also align with how only the Reachy robot was compared to existing fictional robots from movies, suggesting how popular culture affects perceptions and preconceptions, as in [136]. The Yanshee’s morphology gave off “scary vibes” but also performed best quantitatively on helpfulness and answer quality. The humanoid being scary but performing well suggests that anthropomorphism does have an effect, but it is not as strong as seen in literature, such as [76], where anthropomorphism was a key predictor of trust.

The distinction seen between how effective a robot is perceived to be and how trustworthy it is maps onto existing frameworks in information-seeking. Hovland and Weiss [129] identified source credibility as comprising two parts: its expertise, the perceived ability of a source to provide accurate information; and trustworthiness, the perceived intention and reliability of the source independent of capability. Applied to human-robot interaction, the meta-analysis in [117] showed both components influence user behaviour and perceptions in separate ways. The findings here are consistent with this, in that perceived answer quality and helpfulness corresponded to whether the robot is a good information source. Also, the likelihood of verifying aligns with the trust dimensions, as well as the scrutiny required in the information provided itself.

The three measures map onto this two-part split rather than competing with it. Helpfulness and answer quality separate trust in the robot from trust in the information, but both speak to the source’s expertise, whereas the likelihood to verify reflects its trustworthiness and the user’s own decision about whether to rely on what it said. The fact that these dimensions of trust were affected differently by the same morphology suggests that robot form changes the attributions of whether a robot is trustworthy and fit-for-purpose through different mechanisms, which requires further empirical exploration.

6.2.2 Appearance-Function Fit

How well each robot’s function matched what its appearance led people to expect was the main factor behind the patterns in the qualitative data. A robot’s form signals the kind of interaction it is built for, and users read that signal before they know anything about its capabilities. The first signal is role legibility, meaning how clearly the form communicates an intended role to someone who has not seen the robot before. The second is appearance-function fit, meaning whether the robot then performs in line with the role that was read. Where the signal is missing or misleading, users can make incorrect assumptions, and are more likely to disengage or be disappointed when performance does not match what they expected.

Intentional design was most successfully used by the Reachy robot. Its mechanomorphic form, smooth and plain aesthetic, and expressive movement together elicited an appearance that participants understood as communicative and informational in purpose. Qualitative responses were consistent on this point, with participants highlighting the “trustworthy face” and that the “combination of face and movement helped make it seem responsive”, and characteristic of a “movie character”. References like this reflect the cultural familiarity with stereotypical non-humanoid droid designs in science fiction, which typically portray them as capable autonomous agents rather than toys or companions reliant on a human. The Reachy’s voice was described as not sounding “weird like Disney”, emphasising its informational utility rather than as a character for entertainment.

Across these responses, the Reachy’s appearance was legible as that of a device capable of conversing and sharing information, which appears to have aided participants’ trust assessments. Importantly, this trust advantage was not dependent on it providing the best answers, as evidenced by slightly lower fit-for-purpose scores, suggesting that trustworthiness was driven by morphology rather than information-seeking performance. In other words, it provided the right capability signals even if it was less useful.

The Yanshee is the inverse case, and shows that embodiment alone does not decide perceptions. As one participant put it, “Reachy looks like it provides information, but [the Yanshee] doesn’t make that communication capability obvious”. A humanoid form carries expectations of mobility, physical work, and social companionship rather than standing in place and answering questions. Its most visible capability, locomotion, went unused, leaving a torso, head, and limbs doing a job that none of them explained, and there is no well-known equivalent of a humanoid information point in popular media, against which participants could calibrate, unlike the Reachy. Appearance and perceived function were therefore misaligned, and the resulting trust deficit reflects the difficulty participants had in mapping what they saw onto what they were using the robot for.

The Karotz robot represents another failure in fitting appearance to function.

Here there was no clear but mismatched signal like the Yanshee, instead there was no signal at all for what the intended purpose was. Responses reflected the uncertainty in interacting, with the zoomorphic form not being recognisable as having an existing use case. The Fisher-Price toy comparison a participant made suggests that the only signal the appearance gives is that it is for children. Being toy-like attracted people’s eyes and drew them to approach, but the cuteness factor did not convert into engagement or credibility. Adding embodiment does not improve perceptions if that body does not communicate the role effectively.

The Echo, as the disembodied voice, appears to oppose the theory proposed so far. The lack of a body would be presumed not to communicate an informational role, but awareness of it as an existing product, and the segment in general, likely helped it receive the moderate and stable trust ratings. Here, the role legibility also stems from cultural familiarity, with preconceptions and mental models of how a smart speaker would behave in an interaction, based on prior experiences and awareness of publicly available devices. Appearance-function fit followed from that as the device behaved as the user’s expectations predicted.

Qualitative responses indicated that the participants had decided what the Echo was for before beginning the interaction. Work on technology acceptance and trust calibration [99] shows that prior experience with a device category is a strong predictor of trust, regardless of actual system performance [13, 22]. However, this complicates whether a neutral baseline is really possible with this study design.

Together, the appearance-function pattern contradicts a simple embodiment gradient equalling more trust, and supports a more nuanced definition of role legibility from the purpose a robot projects, as it can be achieved through both morphology and familiarity. In information-seeking scenarios, trust in robots appears to depend less on using a sophisticated form or a humanoid appearance, and more on whether the morphology can reliably communicate what it does to an audience of experienced users and naïve observers. Effective signalling has implications for the fields of information-seeking and social robotics, and how trust factors into the equation.

There is no truly neutral morphology, so any design choice should be carefully considered to ensure it has the intended effect, particularly if users are to trust the information a robot provides. Morphology is a channel of communication, so using cuteness or anthropomorphic sophistication for impressiveness or engagement can come at the expense of role legibility, undermining the trust it may be thought to generate subject to the constraints discussed in the limitations.

6.2.3 Cute Factor

A secondary yet frequently recurring theme in the qualitative data was the role of cuteness in shaping participants’ preferences, perceptions, and responses. References

to cuteness were made across multiple robots and in relation to all three trust dimensions, making it one of the most consistent signals. However, cuteness and the perceived outcomes were not directly correlated. Robots that were seen as the cutest did not necessarily receive the highest ratings; in particular, the Karotz received the lowest ratings on helpfulness and usefulness. The Karotz’s low ratings contrast with the Reachy’s, which was the most trustworthy in the subsequent questionnaire and performed well across the dimensions. Fully understanding why a cute morphology operated differently across robots and dimensions is a complex question in itself and would require results before and after an interaction, as well as information on what affected the decision to approach and engage, which was not measured in this study.

Cuteness functioned as a consistent signal of approachability in the user study, with participants describing the Reachy as something they would be “less inclined to talk to” if it were not cute, and characterising the Karotz as “adorable so clearly trustworthy”. The benefit of cuteness aligns with findings from other first encounters with a specific robot [137]. In that study, participants reported three related effects: the robot’s childlike appearance reduced hesitance to engage, participants frequently adopted gendered pronouns when referring to the robot, and the robot’s cuteness appeared to buffer against negative preconceptions. In both this work and ours, physical appeal appeared to lower the threshold at which people are willing to initiate interaction with an unfamiliar robot.

However, the Karotz results help solidify that this effect applies to the initial approach rather than to the rest of the mental evaluation that follows. The Karotz attracted more cuteness references than any other robot, yet scored lowest on multiple aspects of trust. After participants moved beyond initial engagement and began assessing it as an information source, the appeal did not translate into credibility. Cuteness drew participants in, but did not sustain confidence.

The Reachy is the complementary case. It was described as cute, but also as facially expressive and communicatively responsive. The cute features likely gave participants a stronger base of confidence besides just warmth. Its high trust scores suggest that when cuteness is utilised as part of the communication channel, rather than merely as an aesthetic, it reinforces perceived competence rather than serving as a separate aspect of the robot.

In a practical implementation of robots in public information roles, cuteness could be a strategy for maximising voluntary engagement, but it is not a substitute for other design features, morphology or otherwise, that communicate credibility. The Reachy shows that a robot can be both appealing and trustworthy; the two are not mutually exclusive, and one need not diminish the other. The warmth that an appearance generates must be supported by other signals that convey role and competence to someone with no prior experience of the robot’s capabilities.

6.2.4 Trust Calibration

Trust in the information provided by a robot appears to be primarily a response to robot appearance. The quantitative results support this framing, as the believability question results were significantly affected by morphology, with large effect sizes and the post-hoc pattern pointing towards the Reachy’s design as a main driver of that effect.

Qualitative data complicates this, however. Across the robots, participants described their likelihood of verifying as dependent on the nature of the information rather than on the qualities of the robot’s morphology. The believability measure appears to capture more than morphology; instead, it reflects the underlying AI model’s response quality and, more strongly, the importance of the information being requested. Trust is thus calibrated to an appropriate level based on whether the information is important or needs to be relied upon. The relationship between morphology and epistemic trust is more context-dependent than the statistical analysis can show, with only a single low-stakes context investigated.

Participant explanations for the believable question were consistent across robots, with believability weighed as a cost against the stakes of relying on the information, rather than as a judgement about the robot itself. Whether they would check was described as something that “depends how important the information is”, with minimal worry in the low-stakes context. Questions were reportedly constrained to topics participants were already aware of, in this case, having already heard some information about the events. Because this pattern was not concentrated in any morphology, it reflects a general evaluation strategy that people bring to interactions, irrespective of the robots used.

The variation in the likelihood of verification affects the interpretation of the results, but also reflects a deliberate methodological choice. The study context was selected because it involved everyday, low-stakes information-seeking, in which voluntarily using a robot was reasonable and plausible, and it integrated well into a setting where it would receive sufficient use. Participants had no sense of urgency or concern about the information they were asking for and receiving, as alternatives were available.

Trust in an information source involves both the affective response to the appearance or behaviours, and a (conscious or involuntary) assessment of the competence given the stakes of decisions made based on the information [21, 138]. A participant who is not concerned about verifying because the information is not relatively important, as most were here, is using a different psychological process than someone who genuinely trusts the robot. However, both types produce a similar score on the believability rating. This measure is therefore a combination of these two processes, rather than a direct evaluation of trust in the robot. That said, since the context was restricted to low-stakes, the comparison between robots can

be based mostly on trust rather than on a lack of concern. The believability item is also a reported measure of enacted trust, capturing a stated intention to verify rather than verification itself; this could not be observed due to the naturalistic design.

Morphology can be seen as having a greater impact on affective trust, the emotional connection when looking at the robot, than the deliberate trust placed in terms of whether the information provided is believable. Robot form consistently shaped initial perceptions, but topic stakes and prior knowledge regulated what they would do with the information, which appearance influenced less. Design interventions that aim to increase reliance on or adoption of both robots and robot-provided information, therefore, need to address contextual factors after evaluating morphological ones. Appeal needs to be balanced with features that allow users to place an appropriate level of reliance on the robot.

6.2.5 Perceived Answer Quality

All four devices used the same underlying model with the same settings, yet participants who used more than one reported differences in the quality of the answers they received. The Yanshee was described as giving a number of less accurate answers, the Karotz as struggling with some questions, and the Reachy as being too verbose. Some variation between individual responses is expected, since the model does not produce identical output every time and participants asked unique questions. Even so, the differences were reported confidently and were attributed to the robot in front of the participant rather than to the system behind it.

This suggests that morphology affects not only how a robot is regarded as a source, but also how its answers are received once given. The two components of source credibility described by Hovland and Weiss [129] are useful here. A form that signals expertise appears to raise the perceived quality of the content delivered through it, and a form that signals the wrong role appears to lower it. Judgement of the message is coloured by judgement of the messenger, which is consistent with the wider finding that the same information was not trusted equally across the robots.

This matters more in information-seeking than it would in a physical task, where performance can be observed directly. A user has no independent way of checking an answer at the point it is given, so the available cues, including appearance, carry more weight in deciding whether it was a good answer. There is also a methodological consequence, as a study that assumes content is identical because one system produces it may still be measuring morphology, so ratings of answer quality should be understood as ratings of the robot and its answers together.

6.2.6 Limitations

The naturalistic setting was chosen so that participants would approach voluntarily and ask their own questions, which produced data on real use that a laboratory study could not, at the cost of experimental control. Participants were not randomly assigned, no demographic data was collected, interactions were not standardised, and each morphology was deployed on a different day, so day-level factors such as weather and foot traffic cannot be separated from the effect of morphology. The repeating daily schedule of Orientation Week was chosen to keep these as similar as possible. With 54 participants in group sizes of 9 to 18, the corrected pairwise comparisons have limited support and the null post-hoc results should not be read as evidence of no effect, though the Kruskal-Wallis tests give reasonable confidence in the main findings. Some participants used more than one robot and are treated as independent observations, as repeat participants could not be determined in this setting.

One robot represented each embodiment type, so the ratings reflect each robot as a whole rather than its body type alone. For example, the Reachy’s believability cannot be separated from its face, its animations, or its size, and the same applies to the other three. The study establishes that differences at the level of embodiment matter, but not which features within a morphology are responsible. For the same reason, the answers participants received cannot be fully separated from the robot delivering them.

Finally, the believability question was understood as referring to different aspects by the participants. It measured an affective response to the robot, as well as a judgement about the information and how much it mattered. Because the setting was low-stakes, many participants verbally stated they were unlikely to verify anything regardless of the robot, which narrowed the range of that measure and means the effect of morphology may be larger where the stakes are higher, as in healthcare or safety. The Echo was included to represent the absence of a body, but participants brought established expectations of smart speakers to it, so its ratings reflect familiarity as much as form.

6.2.7 Future Work

Following this, a more structured and controlled experiment should be designed to decouple morphology from AI response quality. A Wizard of Oz system with identical scripted questions and responses would ensure information content is consistent, allowing the only change to be the morphology. A scripted setup would clearly show how morphology affects trust, though it would give less useful results for an information-seeking context. If the fit-for-purpose and believability ratings still diverge between robots, then form can be a more confident predictor of trust.

A Wizard of Oz design would also uncover whether participants still perceive a difference between robots.

More delineated measures to assess trust are also needed. A validated short scale would help confirm the results for studies like this one, where there is little opportunity to ask many questions. More work on trust scales that use different dimensions, particularly for social robots, is a clear research gap. Future studies should develop and use validated multi-item instruments that accurately assess the constructs and components that make up trust, which would be valuable for future HRI research, rather than relying on common but suboptimal scales. A questionnaire to determine what people think about different morphologies would also give more explanation to other findings.

Using only one robot per embodiment category limits the analysis of which features affect perceptions. Systematic studies that vary embodiment while keeping features identical, or use the same embodiment, but change features such as facial expressivity, size, or movement, and their combinations, would help clarify which design choices affect which aspects of perceptions. Systematically varying morphological features may reveal that certain combinations affect perceptions the most, as seen in this study, instead of just individual features having an isolated effect. Parametric variation across a modifiable platform, as in the work started by Gabor and Palinko [96], is one approach that could work for this.

Investigating different situations and contexts to determine the impact of information stakes on the verification question would address the restricted range of raw verify ratings observed in this study. Further, comparing participant responses across different stakes while keeping everything else the same would reveal whether the morphological effects on trust behaviours are stronger when the consequences of acting on possibly incorrect information are meaningful to participants, perhaps by being told they need to use the information somehow. Having consequences of relying on information is directly relevant to common social robot deployment contexts, such as healthcare, where trust is essential.

Finally, a more complex study that integrates the data collected in both studies could show how trust varies across a diverse range of participants. Collecting information on a participant’s prior robot experience, technology familiarity, and trust dispositions would systematically reveal these differences but would require recruiting a very large number of participants. A longitudinal study would also show how repeated interactions affect trust, rather than just first impressions, and whether perceptions of information-retrieval performance fluctuate. More time spent would reveal whether a kind of personal familiarity exists with anthropomorphic robots, as seen in other work. The Echo seemingly benefited from general awareness and cultural familiarity. If a robot is to be used more than once, rather than as a novelty, then familiarity is clearly an important factor.

6.2.8 User Study Summary

This study asked whether robot morphology affects how a robot is perceived as an information source, to which the answer is yes and in many different ways, providing avenues for future work to focus on specific parts. Across helpfulness, fit-for-purpose, and believability, changing morphology produced significant effects with large effect sizes. However, the pattern was not uniform. Being fit-for-purpose as an information source and having that information be believed were not the same; different forms performed better in one aspect than the other. The most useful robot was not the most trusted, and the most sophisticated robot did not always produce better results, contrary to initial assumptions.

Robot form being able to communicate purpose legibly was found to have the biggest impact rather than embodiment, as this was reported to have helped the Reachy and the Echo, and hindered the Yanshee and the Karotz. Having unclear or mismatched expectations affects the trust calculation, regardless of capability. The legibility, or degree to which a robot's purpose is readable from its appearance, appears to be a more important driver of trust than anthropomorphic features or visual appeal on their own. The combination of features and subtle stylistic choices is what needs to be considered more than embodiment.

The AI response, context, and questions asked all affect how much can be read into the results, as discussed. This study is best placed as foundational work on robot morphology and trust, as well as robot trust and information-seeking. It achieved findings that show that morphology does affect how information from a robot is perceived and signals are detected in a real-world situation, providing many opportunities for future hypotheses to be explored. As social robots are adopted more to provide information, robot design needs to be carefully considered. This study provides a starting point for reasons why robot form matters, and how.

6.3 Combined Study Findings

The two studies were directed by the same research question, but produced different answers. The survey found no effect of morphology on trust and only small effects on some perception dimensions, while the user study found significant effects across all three trust measures with large effect sizes. The robots, the task, and the topic were the same in both, so the difference lies in what participants were asked to do. In the survey they watched a short video of a robot being used. In the user study they chose to approach a robot in a public setting and asked it their own questions.

The survey's null result is informative rather than inconclusive. With 60 or more participants per group, medium and large effects would have been detected, so morphology produces at most small perceptual differences, and no trust differences,

when participants are an external observer to an information-seeking interaction. Set against the large effects in the user study, the null result indicates that morphology matters when a person is involved in the interaction and has something, however small, invested in the outcome. It is consistent with the position that trust requires some vulnerability [15, 21], and with the distinction between interaction types drawn from the taxonomy of Onnasch and Roesler [1] in Section 3.1.

Involvement began before the interaction did. Participants in the user study were not recruited; it was their decision to walk up to the robot. This decision to engage with the robot was based on the only information available at that point, appearance. The decision to approach is a trusting behaviour in itself, and it remained part of the interaction that followed, with participants describing their choice to engage in terms of whether a robot looked worth talking to. Observers in the survey had no such choice, so the expectations that appearance sets, and that are then met or disappointed, were not relevant. This is the part of the process an online video-based study cannot reach. The appearance-function fit signalling only comes into play when someone is deciding whether to rely on the robot in front of them.

Our interpretation rests mainly on reported trust, as both studies used self-report measures. Willingness to approach differed between morphologies but cannot be separated from the day on which each robot was deployed. The other behavioural measures, questions asked and interaction duration, showed no differences, as would be expected over such short interactions. Directly measuring enacted trust, such as whether users verify what they are told or act on it, remains an important step for future work. The practical consequence is that evaluating robots through video or passive observation will tend to underestimate how much morphology matters. An effect that looks negligible under observation can be substantial in use. Answering RQ2, robot morphology does affect trust in information-seeking, but the size of the effect depends on whether participants are personally engaged.

6.4 Relation of RQ2 to RQ1

Different approaches were taken to answer the two research questions. RQ1 used a literature review to identify which morphological features are understood to affect trust, and the mechanisms through which they act. We found that morphology acts as a pre-interaction signal about a robot’s capabilities, and that trust is multidimensional and only meaningful where the user takes on some risk. The review also identified a gap in that trust during information-seeking has been studied mainly with digital interfaces and disembodied conversational agents, such as in [10], where embodiment is not a variable. Work on morphology and trust in HRI has concentrated on industrial and collaborative tasks, such as in [29, 30, 127].

The only research found on conversational information-seeking with a robot did not vary morphology and was concerned with conversation style rather than trust in the information [55]. RQ2 tested the expectations from RQ1 in that gap. In this section, we set out where the studies agree with the literature, where they qualify it, and what follows from that.

The studies support the literature on several points. Morphological features act as social cues, with a face and expressive movement supporting trust independently of human-likeness [17, 93], which is evident in the qualitative responses to the Reachy. The requirement that trust involves vulnerability [21] is consistent with the survey’s null result, as observers had nothing at stake. The multidimensional nature of trust [111] is evident in the dissociation between perceived usefulness and believability, and in the low internal consistency of the three user study items. Prior experience as a moderator [70] appears in the survey, where familiarity predicted perceptions more consistently than morphology did.

Our studies also narrow what the literature claims. The meta-analysis by Hancock *et al.* [21] reports morphology as a strong predictor of trust, but taking an average effect size cannot show whether that strength depends on how involved the user is. The divergence seen between the survey and the user study indicates that engagement type and risk change how relevant morphology is. The same meta-analysis treats an appearance-function mismatch as one trust factor among many, but in our user study, role legibility and the fit judged against it were much more important to trust, more than anthropomorphism.

The studies also sit against the assumption that a human-like form is inherently more trustworthy [25]. The humanoid robot was not the most trusted across the dimensions, and was the least believable. However, disregarding anthropomorphism has to be stated narrowly, as only one humanoid was used and its size and styling cannot be separated from its human-likeness, so the finding does not extend to humanoid robots in general. What it does show is that a human-like form can lose its advantage when the capabilities it implies go unused. Anthropomorphism remains relevant, but it works through the expectations it creates rather than producing trust directly.

Taken together, the two questions describe morphology as a conditional and dynamic signal rather than a fixed property of a robot. It sets expectations about a role before any interaction begins, and those expectations are then confirmed or contradicted by what the robot does. How much those expectations matter depends on how involved the user is, and how important the information is. For design, this means the signals an appearance sends should be matched to the role the robot is being asked to perform, and that appeal and clarity are not the same objective. For research, it means form and task context are difficult to study individually.

6.5 Chapter Summary

In this chapter, we discussed what the two studies mean for how robot morphology affects trust in information-seeking, following on from what was learned from the literature review. The chapter discussed how the minimal effect of morphology on trust under passive observation contrasted with the strong effect during active involvement. The dimensions of trust did not move together, attributed to the appearance-function fit, the role of cuteness, and how context changes how people calibrate their trust to the stakes of the information. How the two studies contrasted was attributed to interaction engagement style, with morphology shown to be a pre-interaction signal that matters mostly when the user is directly involved.

These findings were related back to RQ1, confirming parts of the reviewed literature, such as how morphology sets expectations and how trust requires some level of vulnerability or risk. However, these findings qualified other parts of the literature, such as the assumption that human-like form is inherently better for trust; what was found is that trust is much more complex than that, as is morphology. The implications from this discussion should be interpreted within the scope of the contexts examined in this thesis. Nevertheless, the combination of findings from the survey and field study suggests that robot morphology conveys meaningful social signals that users draw upon when forming expectations and trust judgements, thereby furthering the existing understanding in the research area. Understanding how those signals operate across different tasks, environments, and levels of risk remains an important direction for future work, but the present findings indicate that morphology is not simply a matter of appearance; it can shape how users understand a robot's role and how they engage with it. In the following chapter, we return to the research questions set out in the introduction and present the conclusions of the thesis.

Chapter 7

Conclusion

The increasing presence of robots in public and domestic spaces and their growing use as information sources raises the practical question this thesis set out to answer. Does the form of a robot shape how much people trust what it tells them? As established, an inappropriate amount of trust can be consequential in both directions. Too little trust can mean a useful tool goes unused, but too much trust can mean a person relies on information that should have been checked. Information-seeking is an area where trust calibration matters directly, because the robot is both the interface and the source. A user's judgement of the robot and their judgement of the information are therefore intertwined. With concern around misinformation and AI hallucinations, allowing people to place appropriate trust is more important than ever.

Morphology is a foundational part of that judgement, as it communicates expectations about the robot before any words are exchanged. Morphology is therefore a plausible influence on whether a robot's answers are believed, but this relationship has received little direct attention. This thesis asserted that a robot's morphology affects its perceived trustworthiness as an information source, and that the relationship morphology has with trust must be understood to design robots people can rely on appropriately.

That assertion was investigated in two stages. First, RQ1 asked which morphological features affect perceived trust, and was addressed through the literature review. Then, RQ2 assessed in what way epistemic trust during information-seeking is affected by morphology, addressed empirically through an online survey and a naturalistic field study. The two studies were designed to separate what people report as trusting from what trust they are observed to enact. In this chapter, we draw the investigations together, stating the answers to the research questions, and setting out the contributions that follow from them.

7.1 Key Findings

Altogether, this research gives a clear account of morphology’s impact in a low-stakes information-seeking setting. The effect of morphology is that, before any information is exchanged, it sets expectations about what a robot is for and what it can do. Trust rises when the robot’s purpose confirms that expectation, and falls when the form cannot be interpreted (poor role legibility) or implies capabilities the robot does not have (mismatch in appearance-function fit). At the beginning of this thesis, we asserted that a robot’s morphology affects perceived trustworthiness in the information it provides. Our findings support this thesis statement, but qualify it in two important ways.

The effect role legibility plays is conditional on the user’s personal engagement, registering only when a person is actually seeking information, rather than watching someone else do so. Morphology does not have a single effect; it modifies trust in complex ways. Morphology can move the dimensions of trust in different and sometimes opposing directions. For the low-stakes, real-world information-seeking scenario considered in this thesis, the implication is that robots should be designed with appearances that clearly signal their role and match their function if they are expected to be trusted. A form that can honestly and accurately signal what the robot is for will support an appropriately calibrated amount of reliance on it. The subsections that follow set out how each research question led to this conclusion.

7.1.1 Answer to RQ1

RQ1 asked which morphological features influence perceived trust in robots, and by what mechanisms. The answer established through the literature review is that trust is not determined by embodiment type but by resemblance and human-likeness, gaze and the presence of eyes or cameras, and expressive movement, all of which help signal what a robot is for and what it can do. These features act as social heuristics, evaluated unconsciously before an interaction begins, and each can independently support or undermine trust, regardless of how anthropomorphic the robot is overall.

The literature review established morphology as a strong and consistent predictor of trust, and clarified the mechanism through which it is relevant. A robot’s form is the initial evidence a person has of how it might perform before interacting with it. The appearance causes an implicit evaluation using social heuristics, with the resemblance, gaze, movement, and light patterns hinting at what the robot is for and what it can do. These signals are not reducible to an embodiment type. Individual features can support or undermine trust independently of anthropomorphism, and emotional expressions remain interpretable even when the body is not human-like.

Trust itself was found to be multidimensional. It encompasses performance,

relational, and moral judgements, with the person, the robot, and the context all having a role. Trust is dynamic throughout an interaction and fragile, easily lost when errors occur. Trust is also only meaningful when the user is exposed to some risk and has to rely on the robot in some way, both of which are conditions that information-seeking satisfies even in low-stakes contexts.

The literature review also identified the gap this thesis helped fill. Work on trust in information-seeking has concentrated on digital interfaces and disembodied agents. Embodiment is a variable more complex socially than those mediums. Work on trust in HRI has focused on industrial and collaborative tasks, but the intersection of information-seeking and robot trust has not been investigated, particularly when morphology serves as the interface. RQ1 produced a set of expectations and a question that was tested in RQ2, where they had not previously been applied.

7.1.2 Answer to RQ2

RQ2 asked how robot trust during information-seeking scenarios is impacted by morphology. The answer is that morphology acts before any information is exchanged, setting expectations about what the robot is for, and trust then follows the fit between those expectations and the role the robot actually performs. This effect is conditional on the type of engagement. Under passive observation, morphology produced no significant differences in trust. But when people genuinely sought information, its effects were large. Those effects were also not uniform. Morphology shifted the dimensions of trust in different, sometimes opposing, directions, with the humanoid rated most capable yet most often double-checked, and the mechanomorphic robot most relied upon because its form legibly signalled its communicative purpose. Role legibility and appearance-function fit together appeared more important than human-likeness in shaping trust calibration with the morphologies and low-stakes context studied.

The survey, in which participants only observed a robot, found that morphology had no significant effect on trust and only small effects on social perceptions. The humanoid did not benefit from anthropomorphic attributions, being rated among the least sociable and animate, while the disembodied smart speaker was perceived as more alive than the humanoid. With a large number of participants per robot and adequate statistical power, the null results are informative. Under passive observation, morphology produces at most small differences in perceptions and no difference in trust. Instead, prior familiarity with robots did predict changes in perceptions, suggesting that when not personally at risk, judgements are based on what people already believe about robots rather than on what they are shown.

The user study, in which participants made their own decisions about whether to approach a robot and seek information they genuinely wanted, produced large

and significant effects across each trust dimension. Interestingly, those dimensions did not universally increase or decrease with the robot. The humanoid robot was perceived as most helpful and fit-for-purpose, yet participants were more likely to double-check what it said. Participants placed more trust in what the mechanomorphic robot said, as they were least likely to verify the information provided. In this study, the robot form that was rated as most capable was simultaneously least trusted to be accurate.

The trust items in the user study were distinct dimensions rather than forming a single scale. A robot that was trusted in one respect was not necessarily trusted in another. This dissociation aligns with the distinction in information-seeking literature between a source's expertise and its trustworthiness. Morphology helped raise the perception that a robot was competent at answering questions, but lowered the perception that those answers could be relied upon. Treating performance trust and social trust as a combined trust score would conceal how different robot forms can pull those dimensions in opposite directions.

What morphology determined about the direction of trust was the match between what the appearance implied and the role it actually performed. Appearance-function fit emerged as the clearest explanation for the observed pattern, more than embodiment type. The morphology of the mechanomorphic robot was read as communicative and informative about its purpose, so it was trusted accordingly. Its face and expressive movements were legible as a device capable of conversing and sharing information. The humanoid's form implied physical and social capabilities like mobility and companionship, but these features were not relevant to its static informational purpose. The expectations set and what it actually did produced an informational trust deficit despite being rated as most capable. The zoomorphic robot was appealing, but signalled nothing to users about its purpose, leaving them unsure what to ask of it. The smart speaker, with no body whatsoever, did give clear signals to users, as it could utilise established expectations of what a smart speaker is for.

Taken together, these show that a simple embodiment type is not what is important, but rather what the complete morphology and features communicate about the device. The fact that a recognised product was readily trusted despite less sophistication, along with the human-like robot not always being rated the highest, is the strongest evidence against the common assumption that anthropomorphism grants more trust. What did matter was legibility of the robot's purpose. There proved to be no neutral morphology; instead, just appearances that signal things more or less clearly. A short follow-up survey directly asking which robot looked most trustworthy to provide information placed the mechanomorphic robot well ahead of the humanoid, reinforcing the same point.

Two other patterns qualify the effect morphology has. Cuteness recurred in

participants' explanations and lowered the threshold for approaching and engaging with a robot. By itself, cuteness does not signal credibility, as the zoomorphic robot was cute but not well-received as an information source. Cuteness instead helps support other signals of communication competence. It is therefore best thought of as a catalyst for engagement, one that still requires legibility. Trust in the information was further calibrated based on the risks of acting on it. In the low-stakes context studied, willingness to rely on an answer was reportedly due to how much that answer mattered, as well as any robot-related factors. The impact of the stakes changes the affective response to a robot more than the deliberate assessment and decision to rely on what it says. Neither pattern displaces legibility as the central driver, as cuteness invites the interaction and stakes calibrate reliance, but it is the clarity of the robot's purpose that directs where trust is placed.

7.2 Contributions

We provide an early empirical account of how morphology shapes user trust in robots as information sources. While previous work has examined trust in digital information sources or in robots performing physical tasks, in this thesis, we establish what happens at their intersection. Within the context studied, appearance-function fit was more informative than human-likeness in explaining trust judgements. The evidence is immediately useful to those wanting to use robots for information-seeking because it comes from real-world usage. People chose to use the robots in a real setting rather than being recruited for a laboratory task, reflecting how such robots are likely to be received in practice.

Our research also contributes some concepts that will help develop future work. First, the fit between a robot's appearance and its actual role was more important than human-likeness. How well the form signalled things about the robot changed trust. Anthropomorphism can negatively impact trust if those personal qualities are implied but not used. Second, trust is complex, and its aspects do not necessarily move together. Morphology can make a robot seem more capable but also less believable; capability and credibility are not the same thing. Third, cuteness draws people in, but the rest of what the robot signals is what will sustain confidence. Fourth, the last main contribution is that morphology only affected trust once people were engaged. Engagement means that morphology cannot be considered relevant for trust independent of context. Morphology had its clearest and strongest effects when participants were directly engaged in seeking information.

The two-study methodology follows on from the fourth contribution. The intention was to combine a large quantitative sample using validated scales with qualitative evidence from real-world interaction. Instead, the two-study design was more useful as a way to compare what users say against what they actually

do. It showed that research with observation only, like those that use videos, may understate more emotional effects associated with direct engagement, in this case, how morphology functions. If only the survey had been conducted, there would be a less substantial contribution, but the combination is what makes this work more useful to the field.

A methodological contribution of the user study was that the data collection was handled entirely by the robot instead of by a human researcher. Removing the researcher in this way may allow for data to be collected that might not be possible due to observer effects, caused by human researchers being present during data collection. Other studies may wish to detach the researcher in this way to keep people's behaviours more natural. To support this, the full implementation is released as open-source software (Section 5.1), giving researchers a practical starting point for similar autonomous field deployments. A limitation of this approach is that lengthy questionnaires are impractical in brief, voluntary field interactions. This points to a need for short, validated trust scales suitable for real-world studies and information-seeking.

The implications for robot development are as follows. Informational robots should be designed to make their purpose clear as a first priority, followed by preferences to make them appealing or of a certain type. For other use cases, priorities may differ, but when trust is important, robots should be designed with honesty in mind to not set the wrong expectations. No form is neutral, since every choice of appearance communicates something different to the user. Instead, the question that should direct design is what the form says to the user. Simply making a robot appealing is not the same as making it trustworthy; the two can work in opposition, so they are separate design considerations that need to be tested as such.

These contributions should be read within the overall scope defined in Chapters 1 and 3, and alongside the limitations discussed in Chapter 6. Together, the two studies establish a foundation for understanding how robot morphology shapes trust in information-seeking interactions. Continuing work can build directly on this foundation, and the most productive next step is the systematic variation of individual morphological features, so that the appearance-function fit, observed here at the level of an entire morphology, can be traced to the specific design choices that produce that fit.

7.3 Concluding Remarks

So, can you trust what a robot says? The answer is, of course, it depends. A better question would be, should you trust what a robot says? The answer this thesis supports is that users should be helped to make that decision. Judgement of

the information provided begins before the robot even says anything. Appearance should not be limited to serving a cosmetic purpose, because it does not. It is communicative; it tells a prospective user what the robot is for. People will use that signal in deciding whether to engage and how far to rely on what they are told. If social robots are to be placed in situations where they will answer questions, it is important for designers and implementers to make as much as possible about their appearance legible to support the assessments users will make. Trust should neither be withheld nor overextended; instead, it should be earned appropriately.

Bibliography

- [1] L. Onnasch and E. Roesler, “A taxonomy to structure and analyze human–robot interaction,” *International Journal of Social Robotics*, vol. 13, no. 4, pp. 833–849, Jul. 2021, ISSN: 1875-4805. DOI: 10.1007/s12369-020-00666-5. [Online]. Available: <https://link.springer.com/article/10.1007/s12369-020-00666-5>.
- [2] C. Breazeal, *Designing sociable robots*. The MIT Press, May 2002, ISBN: 978-0-262-25583-7. DOI: 10.7551/mitpress/2376.001.0001. [Online]. Available: <https://direct.mit.edu/books/monograph/2309/Designing-Sociable-Robots>.
- [3] H. Mahdi, S. A. Akgun, S. Saleh, and K. Dautenhahn, “A survey on the design and evolution of social robots — Past, present and future,” *Robotics and Autonomous Systems*, vol. 156, p. 104193, Oct. 2022, ISSN: 0921-8890. DOI: 10.1016/j.robot.2022.104193. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889022001117>.
- [4] P. Agarwal, S. Swami, and S. K. Malhotra, “Artificial intelligence adoption in the post COVID-19 new-normal and role of smart technologies in transforming business: A review,” *Journal of Science and Technology Policy Management*, vol. 15, no. 3, pp. 506–529, Feb. 2022, ISSN: 2053-4620. DOI: 10.1108/JSTPM-08-2021-0122. [Online]. Available: <https://doi.org/10.1108/JSTPM-08-2021-0122>.
- [5] T. Wilson, “Models in information behaviour research,” *Journal of Documentation*, vol. 55, no. 3, pp. 249–270, Aug. 1999, ISSN: 0022-0418. DOI: 10.1108/EUM0000000007145. [Online]. Available: <https://doi.org/10.1108/EUM0000000007145>.
- [6] M. Hertzum, H. H. Andersen, V. Andersen, and C. B. Hansen, “Trust in information sources: Seeking information from people, documents, and virtual agents,” *Interacting with Computers*, vol. 14, no. 5, pp. 575–599, Oct. 2002, ISSN: 0953-5438. DOI: 10.1016/S0953-5438(02)00023-1. [Online]. Available: [https://doi.org/10.1016/S0953-5438\(02\)00023-1](https://doi.org/10.1016/S0953-5438(02)00023-1).

- [7] X. Sun *et al.*, *Trusting the search: Unraveling human trust in health information from Google and ChatGPT*, Mar. 2024. DOI: 10.48550/arXiv.2403.09987. [Online]. Available: <http://arxiv.org/abs/2403.09987>.
- [8] O. F. Karpinski, L. N. Mcfetridge, M. C. Ta, F. N. Ashriem, and A. Singh, “Evaluating AI-assisted search vs. traditional keyword search: A comparative study of ChatGPT and Google,” in *2025 IEEE World AI IoT Congress (AIIoT)*, May 2025, pp. 0982–0987. DOI: 10.1109/AIIoT65859.2025.11105219. [Online]. Available: <https://ieeexplore.ieee.org/document/11105219>.
- [9] L. Sbaffi and J. Rowley, “Trust and credibility in web-based health information: A review and agenda for future research,” *Journal of Medical Internet Research*, vol. 19, no. 6, e7579, Jun. 2017. DOI: 10.2196/jmir.7579. [Online]. Available: <https://www.jmir.org/2017/6/e218>.
- [10] X. Sun, Y. Liu, J. De Wit, J. A. Bosch, and Z. Li, “Trust by interface: How different user interfaces shape human trust in health information from large language models,” in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’24, New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–7, ISBN: 979-8-4007-0331-7. DOI: 10.1145/3613905.3650837. [Online]. Available: <https://dl.acm.org/doi/10.1145/3613905.3650837>.
- [11] I. Amaro, P. Barra, A. D. Greca, R. Francese, and C. Tucci, “Believe in artificial intelligence? A user study on the ChatGPT’s fake information impact,” *IEEE Transactions on Computational Social Systems*, vol. 11, no. 4, pp. 5168–5177, Aug. 2024, ISSN: 2329-924X. DOI: 10.1109/TCSS.2023.3291539. [Online]. Available: <https://ieeexplore.ieee.org/document/10183726>.
- [12] C. Shah and E. M. Bender, “Envisioning information access systems: What makes for good tools and a healthy web?” *ACM Trans. Web*, vol. 18, no. 3, 33:1–33:24, Apr. 2024, ISSN: 1559-1131. DOI: 10.1145/3649468. [Online]. Available: <https://dl.acm.org/doi/10.1145/3649468>.
- [13] J. D. Lee and K. A. See, “Trust in automation: Designing for appropriate reliance,” *Human Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004, ISSN: 0018-7208. DOI: 10.1518/hfes.46.1.50_30392. [Online]. Available: <https://journals.sagepub.com/action/showAbstract>.
- [14] B. C. Kok and H. Soh, “Trust in robots: Challenges and opportunities,” *Current Robotics Reports*, vol. 1, no. 4, pp. 297–309, Dec. 2020, ISSN: 2662-4087. DOI: 10.1007/s43154-020-00029-y. [Online]. Available: <https://link.springer.com/article/10.1007/s43154-020-00029-y>.

- [15] A. R. Wagner and R. C. Arkin, “Recognizing situations that demand trust,” in *2011 RO-MAN*, Jul. 2011, pp. 7–14. DOI: 10.1109/ROMAN.2011.6005228. [Online]. Available: <https://ieeexplore.ieee.org/document/6005228>.
- [16] G. Hannibal, A. Weiss, and V. Charisi, “”The robot may not notice my discomfort” – Examining the experience of vulnerability for trust in human-robot interaction,” in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Aug. 2021, pp. 704–711. DOI: 10.1109/RO-MAN50785.2021.9515513. [Online]. Available: <https://ieeexplore.ieee.org/document/9515513>.
- [17] A. Manor, A. Parush, and H. Erel, “Attentiveness: A key factor in fostering affective and cognitive trust with non-humanoid robots,” in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, Aug. 2024, pp. 469–476. DOI: 10.1109/RO-MAN60168.2024.10731320. [Online]. Available: <https://ieeexplore.ieee.org/document/10731320>.
- [18] E. Phillips, X. Zhao, D. Ullman, and B. F. Malle, “What is Human-like? Decomposing robots’ human-like appearance using the Anthropomorphic roBOT (ABOT) database,” in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’18, New York, NY, USA: Association for Computing Machinery, Feb. 2018, pp. 105–113, ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171268. [Online]. Available: <https://dl.acm.org/doi/10.1145/3171221.3171268>.
- [19] L. Kunold, N. Bock, and A. Rosenthal-von der Pütten, “Not all robots are evaluated equally: The impact of morphological features on robots’ assessment through capability attributions,” *J. Hum.-Robot Interact.*, vol. 12, no. 1, 3:1–3:31, Feb. 2023. DOI: 10.1145/3549532. [Online]. Available: <https://dl.acm.org/doi/10.1145/3549532>.
- [20] K. E. Schaefer, T. L. Sanders, R. E. Yordon, D. R. Billings, and P. Hancock, “Classification of robot form: Factors predicting perceived trustworthiness,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1, pp. 1548–1552, Sep. 2012, ISSN: 1071-1813. DOI: 10.1177/1071181312561308. [Online]. Available: <https://doi.org/10.1177/1071181312561308>.
- [21] P. A. Hancock, T. T. Kessler, A. D. Kaplan, J. C. Brill, and J. L. Szalma, “Evolving trust in robots: Specification through sequential and comparative meta-analyses,” *Human Factors*, vol. 63, no. 7, pp. 1196–1229, Nov. 2021, ISSN: 0018-7208. DOI: 10.1177/0018720820922080. [Online]. Available: <https://doi.org/10.1177/0018720820922080>.

- [22] D. Y. Geiskkovitch and J. E. Young, “Trust calibration through intentional errors: Designing robot errors to decrease children’s trust towards robots,” in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2023, pp. 1402–1406. DOI: 10.1109/RO-MAN57019.2023.10309659. [Online]. Available: <https://ieeexplore.ieee.org/document/10309659>.
- [23] C. Esterwood and L. P. Robert, “A literature review of trust repair in HRI,” in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2022, pp. 1641–1646. DOI: 10.1109/RO-MAN53752.2022.9900667. [Online]. Available: <https://ieeexplore.ieee.org/document/9900667>.
- [24] B. J. Dunstan and G. Hoffman, “Social robot morphology: Cultural histories of robot design,” in *Cultural Robotics: Social Robots and Their Emergent Cultural Ecologies*, B. J. Dunstan, J. T. K. V. Koh, D. Turnbull Tillman, and S. A. Brown, Eds., Cham: Springer International Publishing, 2023, pp. 13–34, ISBN: 978-3-031-28138-9. DOI: 10.1007/978-3-031-28138-9_2. [Online]. Available: https://doi.org/10.1007/978-3-031-28138-9_2.
- [25] K. S. Haring, D. Silvera-Tawil, T. Takahashi, K. Watanabe, and M. Velonaki, “How people perceive different robot types: A direct comparison of an android, humanoid, and non-biomimetic robot,” in *2016 8th International Conference on Knowledge and Smart Technology (KST)*, Feb. 2016, pp. 265–270. DOI: 10.1109/KST.2016.7440504. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7440504>.
- [26] M. Kobayashi and K. Takeda, “Information retrieval on the web,” *ACM Comput. Surv.*, vol. 32, no. 2, pp. 144–173, Jun. 2000, ISSN: 0360-0300. DOI: 10.1145/358923.358934. [Online]. Available: <https://dl.acm.org/doi/10.1145/358923.358934>.
- [27] K. Srinivasan, M. L.-K. Tan, D. C.-K. Peh, D. H.-L. Goh, and C. S. Lee, “Information seeking in Pokemon Go: A preliminary study,” in *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Jun. 2019, pp. 376–377. DOI: 10.1109/JCDL.2019.00078. [Online]. Available: <https://ieeexplore.ieee.org/document/8791171>.
- [28] W. Liao, W. Weisman, and A. Thakur, “On the motivations to seek information from artificial intelligence agents versus humans: A risk information seeking and processing perspective,” *Science Communication*, vol. 46, no. 4, pp. 458–486, Aug. 2024, ISSN: 1075-5470. DOI: 10.1177/10755470241232993. [Online]. Available: <https://doi.org/10.1177/10755470241232993>.

- [29] G. Charalambous, S. Fletcher, and P. Webb, “The development of a scale to evaluate trust in industrial human-robot collaboration,” *International Journal of Social Robotics*, vol. 8, no. 2, pp. 193–209, Apr. 2016, ISSN: 1875-4805. DOI: 10.1007/s12369-015-0333-8. [Online]. Available: <https://link.springer.com/article/10.1007/s12369-015-0333-8>.
- [30] R. E. Yagoda and D. J. Gillan, “You want me to trust a ROBOT? The development of a human–robot interaction trust scale,” *International Journal of Social Robotics*, vol. 4, no. 3, pp. 235–248, Aug. 2012, ISSN: 1875-4805. DOI: 10.1007/s12369-012-0144-0. [Online]. Available: <https://doi.org/10.1007/s12369-012-0144-0>.
- [31] S. Shayesteh, A. Ojha, and H. Jebelli, “Workers’ trust in collaborative construction robots: EEG-based trust recognition in an immersive environment,” in *Automation and Robotics in the Architecture, Engineering, and Construction Industry*, H. Jebelli, M. Habibnezhad, S. Shayesteh, S. Asadi, and S. Lee, Eds., Cham: Springer International Publishing, 2022, pp. 201–215, ISBN: 978-3-030-77163-8. DOI: 10.1007/978-3-030-77163-8_10. [Online]. Available: https://doi.org/10.1007/978-3-030-77163-8_10.
- [32] F. Correia, P. Alves-Oliveira, T. Ribeiro, F. Melo, and A. Paiva, “A social robot as a card game player,” *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 13, no. 1, pp. 23–29, 2017, ISSN: 2334-0924. DOI: 10.1609/aiide.v13i1.12936. [Online]. Available: <https://ojs.aaai.org/index.php/AIIDE/article/view/12936>.
- [33] R. C. R. Mota, D. J. Rea, A. Le Tran, J. E. Young, E. Sharlin, and M. C. Sousa, “Playing the ‘trust game’ with robots: Social strategies and experiences,” in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2016, pp. 519–524. DOI: 10.1109/ROMAN.2016.7745167. [Online]. Available: <https://ieeexplore.ieee.org/document/7745167>.
- [34] A. Bejarano and T. Williams, “No name, no voice, less trust: Robot group identity performance, entitativity, and trust distribution,” in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2023, pp. 1339–1346. DOI: 10.1109/RO-MAN57019.2023.10309416. [Online]. Available: <https://ieeexplore.ieee.org/document/10309416>.
- [35] M. Adamik, K. Dudzinska, A. J. Herskind, and M. Rehm, “The difference between trust measurement and behavior: Investigating the effect of personalizing a robot’s appearance on trust in HRI,” in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Aug.

- 2021, pp. 880–885. DOI: 10.1109/RO-MAN50785.2021.9515487. [Online]. Available: <https://ieeexplore.ieee.org/document/9515487>.
- [36] S. Lausa, R. Bringula, A. E. Catacutan-Bangit, and C. Santiago, “Information-seeking behavior of computing students while programming: Educational learning materials usage, satisfaction of use, and inconveniences,” in *2021 IEEE Global Engineering Education Conference (EDUCON)*, Apr. 2021, pp. 761–765. DOI: 10.1109/EDUCON46332.2021.9453905. [Online]. Available: <https://ieeexplore.ieee.org/document/9453905>.
- [37] D. Ellis, “A behavioural model for information retrieval system design,” *Journal of Information Science*, vol. 15, no. 4-5, pp. 237–247, Aug. 1989, ISSN: 0165-5515. DOI: 10.1177/016555158901500406. [Online]. Available: <https://doi.org/10.1177/016555158901500406>.
- [38] A. Dalglish and R. Hall, “Uses and perceptions of the World Wide Web in an information-seeking environment,” *Journal of Librarianship and Information Science*, vol. 32, no. 3, pp. 104–116, Sep. 2000, ISSN: 0961-0006. DOI: 10.1177/096100060003200302. [Online]. Available: <https://doi.org/10.1177/096100060003200302>.
- [39] B. Ünlütürk and D. E. Barkana, “Children’s information search with social robots: A focus on Children’s question-asking behavior,” *International Journal of Child-Computer Interaction*, vol. 45, p. 100751, Sep. 2025, ISSN: 2212-8689. DOI: 10.1016/j.ijcci.2025.100751. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2212868925000315>.
- [40] R. S. Taylor, “Question-negotiation and information seeking in libraries,” *College & Research Libraries*, vol. 29, no. 3, pp. 178–194, Apr. 1968, ISSN: 2150-6701. DOI: 10.5860/crl_29_03_178. [Online]. Available: <https://crl.acrl.org/index.php/crl/article/view/12027>.
- [41] N. J. Belkin, “Anomalous states of knowledge as a basis for information retrieval,” *Canadian journal of information science*, vol. 5, no. 1, pp. 133–143, 1980. [Online]. Available: <https://tefkos.cominfo.rutgers.edu/Courses/612/Articles/BelkinAnomolous.pdf>.
- [42] M. J. Bates, “The design of browsing and berrypicking techniques for the online search interface,” *Online Review*, vol. 13, no. 5, pp. 407–424, May 1989, ISSN: 0309-314X. DOI: 10.1108/eb024320. [Online]. Available: <https://doi.org/10.1108/eb024320>.
- [43] C. C. Kuhlthau, “Inside the search process: Information seeking from the user’s perspective,” *Journal of the American Society for Information Science*, vol. 42, no. 5, pp. 361–371, 1991, ISSN: 1097-4571. DOI: 10.1002/

- (SICI) 1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199106%2942%3A5%3C361%3A%3AAID-ASI6%3E3.0.CO%3B2-%23>.
- [44] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, “Analysis of a very large web search engine query log,” *ACM SIGIR Forum*, vol. 33, no. 1, pp. 6–12, Sep. 1999, ISSN: 0163-5840. DOI: 10.1145/331403.331405. [Online]. Available: <https://dl.acm.org/doi/10.1145/331403.331405>.
 - [45] B. J. Jansen and A. Spink, “How are we searching the World Wide Web? A comparison of nine search engine transaction logs,” *Information Processing & Management*, Formal Methods for Information Retrieval, vol. 42, no. 1, pp. 248–263, Jan. 2006, ISSN: 0306-4573. DOI: 10.1016/j.ipm.2004.10.007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457304001396>.
 - [46] G. M. Begany, N. Sa, and X. Yuan, “Factors affecting user perception of a spoken language vs. textual search interface: A content analysis,” *Interacting with Computers*, vol. 28, no. 2, pp. 170–180, Mar. 2016, ISSN: 0953-5438. DOI: 10.1093/iwc/iwv029. [Online]. Available: <https://doi.org/10.1093/iwc/iwv029>.
 - [47] D. Bilal and J. Gwizdka, “Children’s query types and reformulations in Google search,” *Information Processing & Management*, vol. 54, no. 6, pp. 1022–1041, Nov. 2018, ISSN: 0306-4573. DOI: 10.1016/j.ipm.2018.06.008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457317308889>.
 - [48] N. Vanderschantz and A. Hinze, “Children’s query formulation and search result exploration,” *International Journal on Digital Libraries*, vol. 22, no. 4, pp. 385–410, Dec. 2021, ISSN: 1432-1300. DOI: 10.1007/s00799-021-00316-9. [Online]. Available: <https://doi.org/10.1007/s00799-021-00316-9>.
 - [49] J. A. Fails, M. S. Pera, O. Anuyah, C. Kennington, K. L. Wright, and W. Bigirimana, “Query formulation assistance for kids: What is available, when to help & what kids want,” in *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, ser. IDC ’19, New York, NY, USA: Association for Computing Machinery, Jun. 2019, pp. 109–120, ISBN: 978-1-4503-6690-8. DOI: 10.1145/3311927.3323131. [Online]. Available: <https://dl.acm.org/doi/10.1145/3311927.3323131>.
 - [50] O. Anuyah, J. A. Fails, and M. S. Pera, “Investigating query formulation assistance for children,” in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, ser. IDC ’18, New York, NY, USA:

- Association for Computing Machinery, Jun. 2018, pp. 581–586, ISBN: 978-1-4503-5152-2. DOI: 10.1145/3202185.3210779. [Online]. Available: <https://dl.acm.org/doi/10.1145/3202185.3210779>.
- [51] T. Beelen, R. Ordelman, K. P. Truong, V. Evers, and T. Huibers, “Does conversation lead to better searches? Investigating single-shot and multi-turn spoken searches with children,” *International Journal of Child-Computer Interaction*, vol. 41, p. 100668, Sep. 2024, ISSN: 2212-8689. DOI: 10.1016/j.ijcci.2024.100668. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2212868924000369>.
 - [52] E. Luger and A. Sellen, ““Like having a really bad PA”: The gulf between user expectation and experience of conversational agents,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’16, New York, NY, USA: Association for Computing Machinery, May 2016, pp. 5286–5297, ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858288. [Online]. Available: <https://dl.acm.org/doi/10.1145/2858036.2858288>.
 - [53] J. Hill, W. Randolph Ford, and I. G. Farreras, “Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations,” *Computers in Human Behavior*, vol. 49, pp. 245–250, Aug. 2015, ISSN: 0747-5632. DOI: 10.1016/j.chb.2015.02.026. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0747563215001247>.
 - [54] F. Radlinski and N. Craswell, “A theoretical framework for conversational search,” in *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, ser. CHIIR ’17, New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 117–126, ISBN: 978-1-4503-4677-1. DOI: 10.1145/3020165.3020183. [Online]. Available: <https://dl.acm.org/doi/10.1145/3020165.3020183>.
 - [55] S. Sharma, T. Beelen, and K. P. Truong, “Robot-supported information search: Which conversational interaction style do children prefer?” In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’23, New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 466–470, ISBN: 978-1-4503-9970-8. DOI: 10.1145/3568294.3580128. [Online]. Available: <https://dl.acm.org/doi/10.1145/3568294.3580128>.
 - [56] J. Kennedy, P. Baxter, and T. Belpaeme, “The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning,” in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’15, New York, NY, USA: Association

- for Computing Machinery, Mar. 2015, pp. 67–74, ISBN: 978-1-4503-2883-8. DOI: 10.1145/2696454.2696457. [Online]. Available: <https://dl.acm.org/doi/10.1145/2696454.2696457>.
- [57] Z. Xing, X. Yuan, D. Wu, Y. Huang, and J. Mostafa, “Understanding voice search behavior: Review and synthesis of research,” in *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, C. Stephanidis, M. Kurosu, H. Degen, and L. Reinerman-Jones, Eds., Cham: Springer International Publishing, 2020, pp. 305–320, ISBN: 978-3-030-60117-1. DOI: 10.1007/978-3-030-60117-1_23.
 - [58] K. Keyvan and J. X. Huang, “How to approach ambiguous queries in conversational search: A survey of techniques, approaches, tools, and challenges,” *ACM Computing Surveys*, vol. 55, no. 6, pp. 129:1–129:40, Dec. 2022, ISSN: 0360-0300. DOI: 10.1145/3534965. [Online]. Available: <https://dl.acm.org/doi/10.1145/3534965>.
 - [59] S. Melumad, “Vocalizing search: How voice technologies alter consumer search processes and satisfaction,” *Journal of Consumer Research*, vol. 50, no. 3, pp. 533–553, Oct. 2023, ISSN: 0093-5301. DOI: 10.1093/jcr/ucad009. [Online]. Available: <https://doi.org/10.1093/jcr/ucad009>.
 - [60] F. Radlinski, C. Boutilier, D. Ramachandran, and I. Vendrov, “Subjective attributes in conversational recommendation systems: Challenges and opportunities,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 12 287–12 293, Jun. 2022, ISSN: 2374-3468. DOI: 10.1609/aaai.v36i11.21492. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/21492>.
 - [61] C. Agarwal, S. H. Tanneru, and H. Lakkaraju, *Faithfulness vs. Plausibility: On the (un)reliability of explanations from large language models*, Mar. 2024. DOI: 10.48550/arXiv.2402.04614. [Online]. Available: <http://arxiv.org/abs/2402.04614>.
 - [62] L. Munn, L. Magee, and V. Arora, “Truth machines: Synthesizing veracity in AI language models,” *AI & SOCIETY*, vol. 39, no. 6, pp. 2759–2773, Dec. 2024, ISSN: 1435-5655. DOI: 10.1007/s00146-023-01756-4. [Online]. Available: <https://doi.org/10.1007/s00146-023-01756-4>.
 - [63] H. Bruce, “Perceptions of the Internet: What people think when they search the Internet for information,” *Internet Research*, vol. 9, no. 3, pp. 187–199, Aug. 1999, ISSN: 1066-2243. DOI: 10.1108/10662249910274575. [Online]. Available: <https://doi.org/10.1108/10662249910274575>.

- [64] C. A. Lynch, “When documents deceive: Trust and provenance as new factors for information retrieval in a tangled web,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 52, no. 1, pp. 12–17, Jan. 2001, ISSN: 1532-2882.
- [65] Y. Yamamoto, T. Yamamoto, H. Ohshima, and H. Kawakami, “Web access literacy scale to evaluate how critically users can browse and search for web information,” in *Proceedings of the 10th ACM Conference on Web Science*, ser. WebSci ’18, New York, NY, USA: Association for Computing Machinery, May 2018, pp. 97–106, ISBN: 978-1-4503-5563-6. DOI: 10.1145/3201064.3201072. [Online]. Available: <https://dl.acm.org/doi/10.1145/3201064.3201072>.
- [66] T. Yamamoto, Y. Yamamoto, and S. Fujita, “Exploring people’s attitudes and behaviors toward careful information seeking in web search,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, ser. CIKM ’18, New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 963–972, ISBN: 978-1-4503-6014-2. DOI: 10.1145/3269206.3271799. [Online]. Available: <https://dl.acm.org/doi/10.1145/3269206.3271799>.
- [67] A. Morton and M. A. Sasse, “Desperately seeking assurances: Segmenting users by their information-seeking preferences,” in *2014 Twelfth Annual International Conference on Privacy, Security and Trust*, Jul. 2014, pp. 102–111. DOI: 10.1109/PST.2014.6890929. [Online]. Available: <https://ieeexplore.ieee.org/document/6890929>.
- [68] J. R. Allen Catellier and Z. J. Yang, “Trust and affect: How do they impact risk information seeking in a health context?” *Journal of Risk Research*, vol. 15, no. 8, pp. 897–911, Sep. 2012, ISSN: 1366-9877. DOI: 10.1080/13669877.2012.686048. [Online]. Available: <https://doi.org/10.1080/13669877.2012.686048>.
- [69] J. Rowley, F. Johnson, and L. Sbaffi, “Students’ trust judgements in online health information seeking,” *Health Informatics Journal*, vol. 21, no. 4, pp. 316–327, Dec. 2015, ISSN: 1460-4582. DOI: 10.1177/1460458214546772. [Online]. Available: <https://doi.org/10.1177/1460458214546772>.
- [70] X. Hei, H. Zhang, and A. Tapus, “Exploring help-seeking behavior, performance, and cognitive load in individual tutoring: A comparative study between human tutors and social robots,” in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, Aug. 2024, pp. 1035–1041. DOI: 10.1109/RO-MAN60168.2024.10731328. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10731328>.

- [71] S. Herse *et al.*, “Do you trust me, blindly? Factors influencing trust towards a robot recommender system,” in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2018, pp. 7–14. DOI: 10.1109/ROMAN.2018.8525581. [Online]. Available: <https://ieeexplore.ieee.org/document/8525581>.
- [72] C. N. Harrington and L. Egede, “Trust, comfort and relatability: Understanding black older adults’ perceptions of chatbot design for health information seeking,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’23, New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–18, ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3580719. [Online]. Available: <https://dl.acm.org/doi/10.1145/3544548.3580719>.
- [73] M. Shin, M. Jang, M. Cho, and J.-K. Ryu, “Uncertainty-resolving questions for social robots,” in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’23, New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 226–230, ISBN: 978-1-4503-9970-8. DOI: 10.1145/3568294.3580077. [Online]. Available: <https://dl.acm.org/doi/10.1145/3568294.3580077>.
- [74] H. Lund, J. Hallam, and W.-P. Lee, “Evolving robot morphology,” in *Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC ’97)*, ser. ICEC, Indianapolis, United States of America: IEEE, Apr. 1997, pp. 197–202. DOI: 10.1109/ICEC.1997.592295. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/592295>.
- [75] C. Paul, “Morphological computation: A basis for the analysis of morphology and control requirements,” *Robotics and Autonomous Systems*, Morphology, Control and Passive Dynamics, vol. 54, no. 8, pp. 619–630, Aug. 2006, ISSN: 0921-8890. DOI: 10.1016/j.robot.2006.03.003. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889006000613>.
- [76] K. S. Haring, D. Silvera-Tawil, T. Takahashi, M. Velonaki, and K. Watanabe, “Perception of a humanoid robot: A cross-cultural comparison,” in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2015, pp. 821–826. DOI: 10.1109/ROMAN.2015.7333613. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7333613>.
- [77] D. Li, P. L. P. Rau, and Y. Li, “A cross-cultural study: Effect of robot appearance and task,” *International Journal of Social Robotics*, vol. 2, no. 2, pp. 175–186, Jun. 2010, ISSN: 1875-4805. DOI: 10.1007/s12369-010-0056-

9. [Online]. Available: <https://link.springer.com/article/10.1007/s12369-010-0056-9>.
- [78] N. Mavridis, “A review of verbal and non-verbal human–robot interactive communication,” *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, Jan. 2015, ISSN: 0921-8890. DOI: 10.1016/j.robot.2014.09.031. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889014002164>.
- [79] M. Mori, K. F. MacDorman, and N. Kageki, “The uncanny valley [from the field],” *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98–100, Jun. 2012, ISSN: 1558-223X. DOI: 10.1109/MRA.2012.2192811. [Online]. Available: <https://ieeexplore.ieee.org/document/6213238>.
- [80] R. Ringe, R. Nolte, N. Zargham, R. Porzel, and R. Malaka, “METAMORPH - A metamodeling approach for robot morphology,” in *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2025, pp. 627–636. DOI: 10.1109/HRI61500.2025.10973806. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10973806>.
- [81] N. Spatola, B. Kühnlenz, and G. Cheng, “Perception and evaluation in human–robot interaction: The Human–Robot Interaction Evaluation Scale (HRIES)—A multicomponent approach of anthropomorphism,” *International Journal of Social Robotics*, vol. 13, no. 7, pp. 1517–1539, Nov. 2021, ISSN: 1875-4805. DOI: 10.1007/s12369-020-00667-4. [Online]. Available: <https://link.springer.com/article/10.1007/s12369-020-00667-4>.
- [82] L. Hoffmann, N. Bock, and A. M. Rosenthal v.d. Pütten, “The peculiarities of robot embodiment (EmCorp-Scale): Development, validation and initial test of the Embodiment and Corporeality of Artificial Agents Scale,” in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’18, New York, NY, USA: Association for Computing Machinery, Feb. 2018, pp. 370–378, ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171242. [Online]. Available: <https://dl.acm.org/doi/10.1145/3171221.3171242>.
- [83] N. Jones, F. Loizides, and K. Jones, “A theoretical framework for the development of “needy” socially assistive robots,” in *Human-Computer Interaction – INTERACT 2023*, Springer, Cham, 2023, pp. 396–401, ISBN: 978-3-031-42293-5. DOI: 10.1007/978-3-031-42293-5_42. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-42293-5_42.

- [84] S. Thunberg, S. Thellman, and T. Ziemke, “Don’t judge a book by its cover: A study of the social acceptance of NAO vs. Pepper,” in *Proceedings of the 5th International Conference on Human Agent Interaction*, ser. HAI ’17, New York, NY, USA: Association for Computing Machinery, Oct. 2017, pp. 443–446, ISBN: 978-1-4503-5113-3. DOI: 10.1145/3125739.3132583. [Online]. Available: <https://dl.acm.org/doi/10.1145/3125739.3132583>.
- [85] J. Turner, J. Bowen, J. König, K. Stawarz, and N. Vanderschantz, “Applying the perceived creepiness of technology scale to social robots,” in *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’25, Melbourne, Australia: IEEE Press, Mar. 2025, pp. 1695–1699.
- [86] N. Akalin, A. Kristoffersson, and A. Loutfi, “Do you feel safe with your robot? Factors influencing perceived safety in human-robot interaction based on subjective and objective measures,” *International Journal of Human-Computer Studies*, vol. 158, p. 102744, Feb. 2022, ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2021.102744. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1071581921001622>.
- [87] J. Turner, N. Vanderschantz, J. L. König, and R. Siddika, ““it almost wanted to hurt someone”: The impact of intentional creepiness on user perceptions,” in *Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction*, Edinburgh Scotland UK: ACM, Mar. 2026, pp. 907–912, ISBN: 979-8-4007-2321-6. DOI: 10.1145/3776734.3794526. [Online]. Available: <https://dl.acm.org/doi/10.1145/3776734.3794526>.
- [88] M. Naya-Varela, A. Faíña, and R. J. Duro, “Morphological development in robotic learning: A survey,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 4, pp. 750–768, Dec. 2021, ISSN: 2379-8939. DOI: 10.1109/TCDS.2021.3052548. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9328449>.
- [89] K. S. Haring, D. Silvera-Tawil, K. Watanabe, and M. Velonaki, “The influence of robot appearance and interactive ability in HRI: A cross-cultural study,” in *Social Robotics*, A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He, Eds., Cham: Springer International Publishing, 2016, pp. 392–401, ISBN: 978-3-319-47437-3. DOI: 10.1007/978-3-319-47437-3_38.
- [90] J. Urakami and K. Seaborn, “Nonverbal cues in human–robot interaction: A communication studies perspective,” *J. Hum.-Robot Interact.*, vol. 12, no. 2, 22:1–22:21, Mar. 2023. DOI: 10.1145/3570169. [Online]. Available: <https://dl.acm.org/doi/10.1145/3570169>.

- [91] A. Aly and A. Tapus, “A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction,” in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2013, pp. 325–332. DOI: 10.1109/HRI.2013.6483606. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/6483606>.
- [92] J. S. Back, P. Cheong, and H. S. Ahn, “Is the Earth flat? Mythbusting with HRI,” in *Proceedings of the 2025 ACM/IEEE international conference on human-robot interaction*, ser. HRI ’25, Melbourne, Australia: IEEE Press, 2025, pp. 1810–1813.
- [93] S. Shirol, J. La Delfa, I. Leite, and E. Yadollahi, “Designing social behaviours for autonomous mobile robots: The role of movement and light in communicating intent,” in *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’25, Melbourne, Australia: IEEE Press, Mar. 2025, pp. 1638–1643.
- [94] B. J. Zhang and N. T. Fitter, “Nonverbal sound in human-robot interaction: A systematic review,” *J. Hum.-Robot Interact.*, vol. 12, no. 4, 46:1–46:46, Dec. 2023. DOI: 10.1145/3583743. [Online]. Available: <https://dl.acm.org/doi/10.1145/3583743>.
- [95] A. Barco, C. de Jong, J. Peter, R. Kühne, and C. L. van Straten, “Robot morphology and children’s perception of social robots: An exploratory study,” in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’20, New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 125–127, ISBN: 978-1-4503-7057-8. DOI: 10.1145/3371382.3378348. [Online]. Available: <https://dl.acm.org/doi/10.1145/3371382.3378348>.
- [96] T. Gabor and O. Palinko, “The HuGo humanoid project: A modular social robot,” in *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’25, Melbourne, Australia: IEEE Press, Mar. 2025, pp. 1776–1778.
- [97] G. Vaiani and F. Paternò, “End-user development for human-robot interaction: Results and trends in an emerging field,” *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. EICS, 252:1–252:40, Jun. 2024. DOI: 10.1145/3661146. [Online]. Available: <https://dl.acm.org/doi/10.1145/3661146>.
- [98] L. Takayama, D. Dooley, and W. Ju, “Expressing thought: Improving robot readability with animation principles,” in *Proceedings of the 6th international conference on Human-robot interaction*, ser. HRI ’11, New York, NY, USA: Association for Computing Machinery, Mar. 2011, pp. 69–76, ISBN: 978-1-

- 4503-0561-7. DOI: 10.1145/1957656.1957674. [Online]. Available: <https://dl.acm.org/doi/10.1145/1957656.1957674>.
- [99] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989, ISSN: 0276-7783. DOI: 10.2307/249008. [Online]. Available: <https://www.jstor.org/stable/249008>.
 - [100] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, Jan. 2009, ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-008-0001-3. [Online]. Available: <http://link.springer.com/10.1007/s12369-008-0001-3>.
 - [101] C.-C. Ho and K. F. MacDorman, “Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices,” *Computers in Human Behavior*, Online Interactivity: Role of Technology in Behavior Change, vol. 26, no. 6, pp. 1508–1518, Nov. 2010, ISSN: 0747-5632. DOI: 10.1016/j.chb.2010.05.015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0747563210001536>.
 - [102] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, “The Robotic Social Attributes Scale (RoSAS): Development and validation,” in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’17, New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 254–262, ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020208. [Online]. Available: <https://dl.acm.org/doi/10.1145/2909824.3020208>.
 - [103] M. Paetzel-Prüsmann, J. F. Lehman, C. J. Gomez, and J. Kennedy, “An automatic evaluation framework for social conversations with robots,” in *Proceedings of the 2024 International Symposium on Technological Advances in Human-Robot Interaction*, ser. TAHRI ’24, New York, NY, USA: Association for Computing Machinery, May 2024, pp. 56–64, ISBN: 979-8-4007-1661-4. DOI: 10.1145/3648536.3648543. [Online]. Available: <https://dl.acm.org/doi/10.1145/3648536.3648543>.
 - [104] R. Parasuraman and V. Riley, “Humans and automation: Use, misuse, disuse, abuse,” *Human Factors*, vol. 39, no. 2, pp. 230–253, Jun. 1997, ISSN: 0018-7208. DOI: 10.1518/001872097778543886. [Online]. Available: <https://doi.org/10.1518/001872097778543886>.

- [105] W. Kim, N. Kim, J. B. Lyons, and C. S. Nam, "Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach," *Applied Ergonomics*, vol. 85, p. 103056, May 2020, ISSN: 0003-6870. DOI: 10.1016/j.apergo.2020.103056. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0003687020300132>.
- [106] K. E. Schaefer, "Measuring trust in human robot interactions: Development of the "Trust Perception Scale-HRI"," in *Robust Intelligence and Trust in Autonomous Systems*, R. Mittu, D. Sofge, A. Wagner, and W. Lawless, Eds., Boston, MA: Springer US, 2016, pp. 191–218, ISBN: 978-1-4899-7668-0. DOI: 10.1007/978-1-4899-7668-0_10. [Online]. Available: https://doi.org/10.1007/978-1-4899-7668-0_10.
- [107] A. Alzahrani, J. Nasir, A. J. Tayeb, E. André, and M. I. Ahmad, "What do the face and voice reveal? Investigating trust dynamics during human-robot interaction," in *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '25, Melbourne, Australia: IEEE Press, Mar. 2025, pp. 400–409.
- [108] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, "Not so different after all: A cross-discipline view of trust," *Academy of Management Review*, vol. 23, no. 3, pp. 393–404, Jul. 1998, ISSN: 0363-7425. DOI: 10.5465/amr.1998.926617. [Online]. Available: <https://journals.aom.org/doi/10.5465/amr.1998.926617>.
- [109] T. Law and M. Scheutz, "Chapter 2 - Trust: Recent concepts and evaluations in human-robot interaction," in *Trust in Human-Robot Interaction*, C. S. Nam and J. B. Lyons, Eds., Academic Press, Jan. 2021, pp. 27–57, ISBN: 978-0-12-819472-0. DOI: 10.1016/B978-0-12-819472-0.00002-2. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128194720000022>.
- [110] G. Campagna and M. Rehm, "A systematic review of trust assessments in human-robot interaction," *J. Hum.-Robot Interact.*, vol. 14, no. 2, 30:1–30:35, Jan. 2025. DOI: 10.1145/3706123. [Online]. Available: <https://dl.acm.org/doi/10.1145/3706123>.
- [111] B. F. Malle and D. Ullman, "Chapter 1 - A multidimensional conception and measure of human-robot trust," in *Trust in Human-Robot Interaction*, C. S. Nam and J. B. Lyons, Eds., Academic Press, Jan. 2021, pp. 3–25, ISBN: 978-0-12-819472-0. DOI: 10.1016/B978-0-12-819472-0.00001-0. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128194720000010>.

- [112] B. M. Muir and N. Moray, “Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation,” *Ergonomics*, vol. 39, no. 3, pp. 429–460, Mar. 1996, ISSN: 0014-0139. DOI: 10.1080/00140139608964474. [Online]. Available: <https://doi.org/10.1080/00140139608964474>.
- [113] R. C. Mayer, J. H. Davis, and F. D. Schoorman, “An integrative model of organizational trust,” *The Academy of Management Review*, vol. 20, no. 3, pp. 709–734, 1995, ISSN: 0363-7425. DOI: 10.2307/258792. [Online]. Available: <https://www.jstor.org/stable/258792>.
- [114] K. Hald, M. Rehm, and T. B. Moeslund, “Human-robot trust assessment using top-down visual tracking after robot task execution mistakes,” in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Aug. 2021, pp. 892–898. DOI: 10.1109/RO-MAN50785.2021.9515501. [Online]. Available: <https://ieeexplore.ieee.org/document/9515501>.
- [115] P. H. Kahn Jr. *et al.*, ““Robovie, you’ll have to go into the closet now”: Children’s social and moral relationships with a humanoid robot,” *Developmental Psychology*, vol. 48, no. 2, pp. 303–314, 2012, ISSN: 1939-0599. DOI: 10.1037/a0027033.
- [116] H. Khalid, W. S. Liew, B. S. Voong, and M. Helander, “Creativity in measuring trust in human-robot interaction using interactive dialogs,” in *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, and Y. Fujita, Eds., Cham: Springer International Publishing, 2019, pp. 1175–1190, ISBN: 978-3-319-96071-5. DOI: 10.1007/978-3-319-96071-5_119.
- [117] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, “A meta-analysis of factors affecting trust in human-robot interaction,” *Human Factors*, vol. 53, no. 5, pp. 517–527, Oct. 2011, ISSN: 0018-7208. DOI: 10.1177/0018720811417254. [Online]. Available: <https://doi.org/10.1177/0018720811417254>.
- [118] D. Firmino de Souza *et al.*, “Trust and trustworthiness from human-centered perspective in human–robot interaction (HRI)—A systematic literature review,” *Electronics*, vol. 14, no. 8, p. 1557, Jan. 2025, ISSN: 2079-9292. DOI: 10.3390/electronics14081557. [Online]. Available: <https://www.mdpi.com/2079-9292/14/8/1557>.
- [119] Y. Li, B. Wu, Y. Huang, and S. Luan, “Developing trustworthy artificial intelligence: Insights from research on interpersonal, human-automation, and human-AI trust,” *Frontiers in Psychology*, vol. 15, Apr. 2024, ISSN:

- 1664-1078. DOI: 10.3389/fpsyg.2024.1382693. [Online]. Available: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1382693/full>.
- [120] P. Schroepfer and C. Pradalier, “Why there is no definition of trust: A systems approach with a metamodel representation,” in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2023, pp. 1245–1251. DOI: 10.1109/RO-MAN57019.2023.10309503. [Online]. Available: <https://ieeexplore.ieee.org/document/10309503>.
 - [121] T. A. Neyazi, T. Khai Ee, A. Nadaf, and R. Schroeder, “The effect of information seeking behaviour on trust in AI in Asia: The moderating role of misinformation concern,” *New Media & Society*, vol. 27, no. 4, pp. 2414–2433, Apr. 2025, ISSN: 1461-4448. DOI: 10.1177/14614448231212804. [Online]. Available: <https://doi.org/10.1177/14614448231212804>.
 - [122] K. Lazányi and B. Hajdu, “Trust in human-robot interactions,” in *2017 IEEE 14th International Scientific Conference on Informatics*, Nov. 2017, pp. 216–220. DOI: 10.1109/INFORMATICS.2017.8327249. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8327249>.
 - [123] J. Xu and A. Howard, “The impact of first impressions on human-robot trust during problem-solving scenarios,” in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2018, pp. 435–441. DOI: 10.1109/ROMAN.2018.8525669. [Online]. Available: <https://ieeexplore.ieee.org/document/8525669>.
 - [124] B. Nasset, D. A. Robb, J. Lopes, and H. Hastie, “Transparency in HRI: Trust and decision making in the face of robot errors,” in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’21 Companion, New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 313–317, ISBN: 978-1-4503-8290-8. DOI: 10.1145/3434074.3447183. [Online]. Available: <https://dl.acm.org/doi/10.1145/3434074.3447183>.
 - [125] S.-D. Ciocirlan, R. Agrigoroaie, and A. Tapus, “Human-robot team: Effects of communication in analyzing trust,” in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Oct. 2019, pp. 1–7. DOI: 10.1109/RO-MAN46459.2019.8956345. [Online]. Available: <https://ieeexplore.ieee.org/document/8956345>.
 - [126] M. Saerbeck and C. Bartneck, “Perception of affect elicited by robot motion,” in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2010, pp. 53–60. DOI: 10.1109/HRI.2010.5453269. [Online]. Available: <https://ieeexplore.ieee.org/document/5453269>.

- [127] F. Correia, S. Mascarenhas, R. Prada, F. S. Melo, and A. Paiva, “Group-based emotions in teams of humans and robots,” in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’18, New York, NY, USA: Association for Computing Machinery, Feb. 2018, pp. 261–269, ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171252. [Online]. Available: <https://dl.acm.org/doi/10.1145/3171221.3171252>.
- [128] F. Correia, F. S. Melo, and A. Paiva, “When a robot is your teammate,” *Topics in Cognitive Science*, vol. 16, no. 3, pp. 527–553, 2024, ISSN: 1756-8765. DOI: 10.1111/tops.12634. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/tops.12634>.
- [129] C. I. Hovland and W. Weiss, “The influence of source credibility on communication effectiveness,” *The Public Opinion Quarterly*, vol. 15, no. 4, pp. 635–650, 1951, ISSN: 0033-362X. [Online]. Available: <https://www.jstor.org/stable/2745952>.
- [130] V. Braun, V. Clarke, E. Boulton, L. Davey, and C. McEvoy, “The online survey as a qualitative research tool,” *International Journal of Social Research Methodology*, vol. 24, no. 6, pp. 641–654, Nov. 2021, ISSN: 1364-5579. DOI: 10.1080/13645579.2020.1805550. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/13645579.2020.1805550>.
- [131] V. Lim, M. Rooksby, and E. S. Cross, “Social robots on a global stage: Establishing a role for culture during human–robot interaction,” *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1307–1333, Sep. 2021, ISSN: 1875-4805. DOI: 10.1007/s12369-020-00710-4. [Online]. Available: <https://link.springer.com/article/10.1007/s12369-020-00710-4>.
- [132] A. Pinto, S. Sousa, A. Simões, and J. Santos, “A trust scale for human-robot interaction: Translation, adaptation, and validation of a human computer trust scale,” *Human Behavior and Emerging Technologies*, vol. 2022, no. 1, p. 6437441, 2022, ISSN: 2578-1863. DOI: 10.1155/2022/6437441. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2022/6437441>.
- [133] K. Seaborn, G. Barbareschi, and S. Chandra, “Not only WEIRD but “uncanny”? A systematic review of diversity in human–robot interaction research,” *International Journal of Social Robotics*, vol. 15, no. 11, pp. 1841–1870, Nov. 2023, ISSN: 1875-4805. DOI: 10.1007/s12369-023-00968-4. [Online]. Available: <https://link.springer.com/article/10.1007/s12369-023-00968-4>.

- [134] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006, ISSN: 1478-0887. DOI: 10.1191/1478088706qp063oa. [Online]. Available: <https://doi.org/10.1191/1478088706qp063oa>.
- [135] L. Saad, Eileen Roesler, E. K. Phillips, and G. J. Trafton, *Choosing the “perfect” scale: A primer to evaluate existing scales in HRI*. [Online]. Available: <http://hriscaledatabase.psychology.gmu.edu/>.
- [136] I. Bode, H. Huelss, A. Nadibaidze, and T. F. Watts, “Cross-cultural narratives of weaponised artificial intelligence: Comparing France, India, Japan and the United States,” *Big Data & Society*, Dec. 2024. DOI: 10.1177/20539517241303151. [Online]. Available: <https://journals.sagepub.com/doi/full/10.1177/20539517241303151>.
- [137] E. Sobolewska, N. Vanderschantz, and F. Bitcon, ““He’s adorable and I want to take him home”. Trust perceptions before and after first-time encounters with social robots,” London, UK: BCS Learning & Development, Jul. 2021, pp. 324–328. DOI: 10.14236/ewic/HCI2021.35. [Online]. Available: <https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2021.35>.
- [138] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, “Assessing acceptance of assistive social agent technology by older adults: The Almere model,” *International Journal of Social Robotics*, vol. 2, no. 4, pp. 361–375, Dec. 2010, ISSN: 1875-4805. DOI: 10.1007/s12369-010-0068-5. [Online]. Available: <https://doi.org/10.1007/s12369-010-0068-5>.

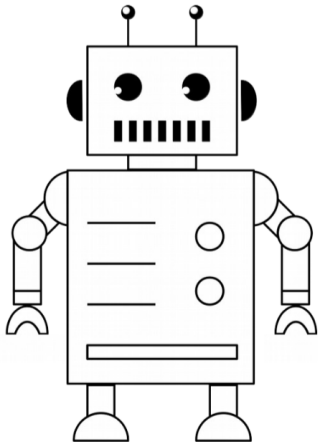
Appendix A

Interaction Taxonomy

In this appendix is the completed interaction taxonomy by Onnasch and Roesler [1], applied to each device for both the online survey and the user study, for a total of eight sheets completed (four robots for the two study settings). It classifies every device by the different dimensions that can affect an interaction, and we used it to investigate differences between the robots and the studies and help determine the possible causes of those differences. Justification and written explanations of the differences can be seen in Section 3.1 of the methodology.

Robot Description & Illustration

Amazon Echo Dot is a compact smart speaker designed for voice-based assistance, information access, and smart home control. It is able to respond to spoken commands, answer questions, play music, manage calendars and reminders, and control compatible smart devices.



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot

embodied
 depicted

settings

field
laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☐	☐	☒
communication	☒	☐	☐
movement	☐	☐	☒
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
 collaborator
 cooperator
bystander

Communication Channel

input

electronic
 mechanical
acoustic
 optic

output

tactile
acoustic
 visual

Proximity

temporal

synchronous
 asynchronous

physical

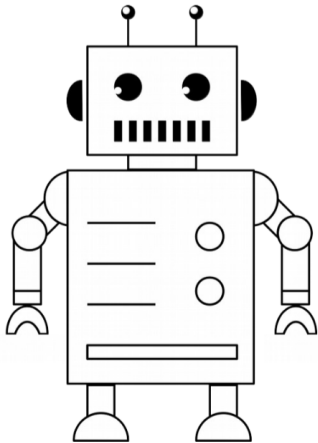
following
 touching
 approaching
 passing
 avoidance
none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Robot Description & Illustration

Amazon Echo Dot is a compact smart speaker designed for voice-based assistance, information access, and smart home control. It is able to respond to spoken commands, answer questions, play music, manage calendars and reminders, and control compatible smart devices.



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot
embodied
 depicted

settings
field
 laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☐	☐	☒
communication	☒	☐	☐
movement	☐	☐	☒
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
collaborator
 cooperator
 bystander

Communication Channel

input

electronic
 mechanical
acoustic
 optic

output

tactile
acoustic
 visual

Proximity

temporal

synchronous
 asynchronous

physical

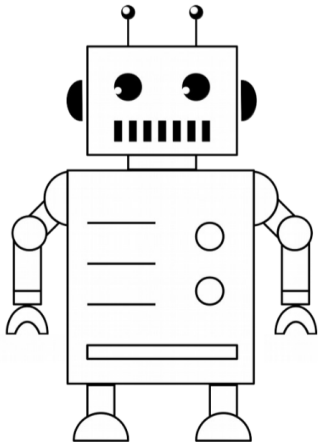
following
 touching
approaching
 passing
 avoidance
 none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Robot Description & Illustration

Mindscape Karotz is a compact internet-connected house companion designed for communication, entertainment, and smart-home interaction. It is able to read social media updates, notify about emails, and describe the weather, play music and radio, and respond to voice commands. It interacts with the user through speech, lighting, movable ears, and RFID tags.



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot

embodied
 depicted

settings

field
laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☐	☒	☐
communication	☒	☐	☐
movement	☐	☒	☐
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
 collaborator
 cooperator
bystander

Communication Channel

input

electronic
 mechanical
acoustic
 optic

output

tactile
acoustic
 visual

Proximity

temporal

synchronous
 asynchronous

physical

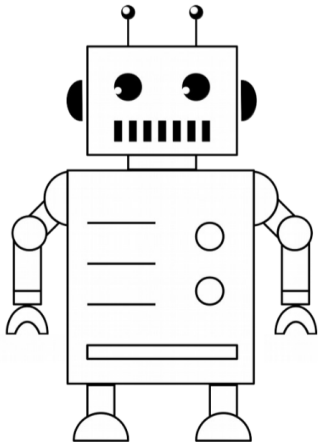
following
 touching
 approaching
 passing
 avoidance
none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Robot Description & Illustration

Mindscape Karotz is a compact internet-connected house companion designed for communication, entertainment, and smart-home interaction. It is able to read social media updates, notify about emails, and describe the weather, play music and radio, and respond to voice commands. It interacts with the user through speech, lighting, movable ears, and RFID tags.



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot

embodied
 depicted

settings

field
 laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☐	☒	☐
communication	☒	☐	☐
movement	☐	☒	☐
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
collaborator
 cooperator
 bystander

Communication Channel

input

electronic
 mechanical
acoustic
 optic

output

tactile
 acoustic
 visual

Proximity

temporal

synchronous
asynchronous

physical

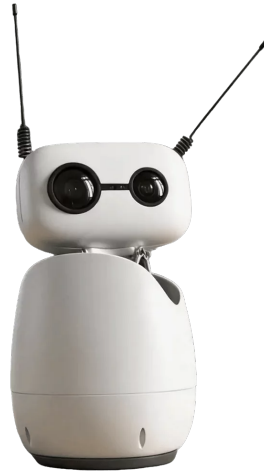
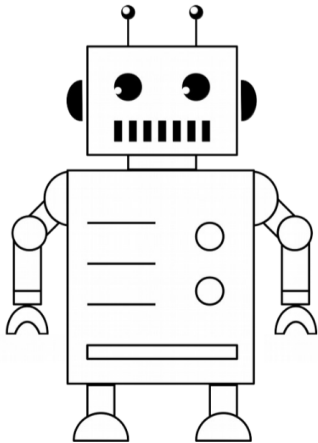
following
 touching
approaching
 passing
 avoidance
 none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Robot Description & Illustration

Reachy Mini is a compact desktop companion designed for human-robot interaction, education, and AI experimentation. It is able to perceive its environment through a camera, engage in natural interactions using speech and movement animations, and serves as a platform for developing and testing applications.



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot

embodied
 depicted

settings

field
laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☐	☐	☒
communication	☒	☐	☐
movement	☒	☐	☐
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
 collaborator
 cooperator
bystander

Communication Channel

input

electronic
 mechanical
 acoustic
 optic

output

tactile
 acoustic
 visual

Proximity

temporal

synchronous
 asynchronous

physical

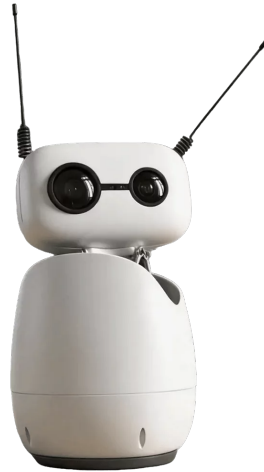
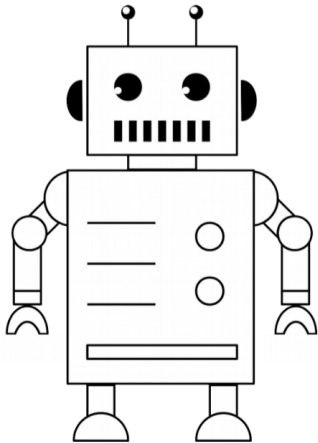
following
 touching
 approaching
 passing
 avoidance
none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Robot Description & Illustration

Reachy Mini is a compact desktop companion designed for human-robot interaction, education, and AI experimentation. It is able to perceive its environment through a camera, engage in natural interactions using speech and movement animations, and serves as a platform for developing and testing applications.



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot

embodied
 depicted

settings

field
 laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☐	☐	☒
communication	☒	☐	☐
movement	☒	☐	☐
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
collaborator
 cooperator
 bystander

Communication Channel

input

electronic
 mechanical
acoustic
 optic

output

tactile
acoustic
 visual

Proximity

temporal

synchronous
 asynchronous

physical

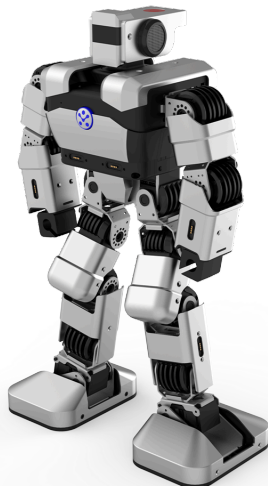
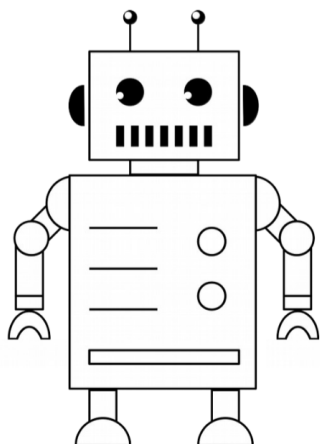
following
 touching
approaching
 passing
 avoidance
 none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Robot Description & Illustration

UBTECH Yanshee is a small humanoid robot designed for education, research, and human-robot interaction. It is able to perform expressive movements, recognise speech and visual information, execute programmable behaviours for learning and experimentation.



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot
embodied
 depicted

settings
 field
laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☒	☐	☐
communication	☒	☐	☐
movement	☒	☐	☐
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
 collaborator
 cooperator
bystander

Communication Channel

input

electronic
 mechanical
acoustic
 optic

output

tactile
acoustic
 visual

Proximity

temporal

synchronous
 asynchronous

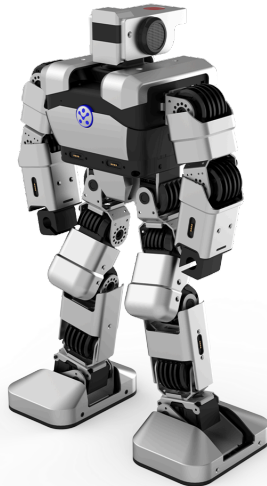
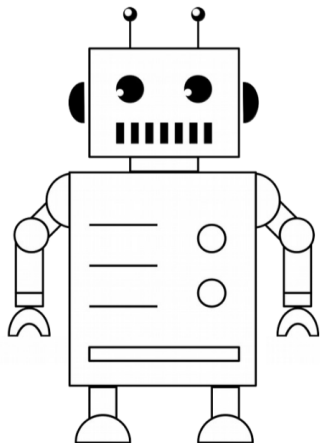
physical

following
 touching
 approaching
 passing
 avoidance
none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Robot Description & Illustration



Field of Application

industry
service
 military & police
 space expedition
 therapy
education
entertainment
 none

Exposure to

robot
embodied
 depicted

settings
field
 laboratory

Robot Task Specification

information exchange
 precision
 physical load reduction
 transport
 manipulation
 cognitive stimulation
 emotional stimulation
 physical stimulation

Robot Morphology

	a	z	t
appearance	☒	☐	☐
communication	☒	☐	☐
movement	☒	☐	☐
context	☒	☐	☐

a (anthropomorphic),
 z (zoomorphic),
 t (technical)

Degree of Robot Autonomy

	-	+
information acquisition	☐	☒
information analyses	☐	☒
decision-making	☐	☒
action implementation	☐	☒

Human Role

supervisor
 operator
collaborator
 cooperator
 bystander

Communication Channel

input

electronic
 mechanical
acoustic
 optic

output

tactile
acoustic
 visual

Proximity

temporal

synchronous
 asynchronous

physical

following
 touching
approaching
 passing
 avoidance
 none

Team Composition

☒ $N_H = N_R$
☐ $N_H > N_R$
☐ $N_H < N_R$

Appendix B

Ethical Approval

This appendix contains the formal ethical approval for both the survey and the user study, as granted by the University of Waikato's STEM Human Research Ethics Committee; evidence that the research was conducted ethically. Following each approval, the Participant Information Sheet is provided, explaining the study to participants and their rights to withdraw. Instead of physical sheets, participants selected an option in the survey, and in the user study, they could either sign or the Committee allowed participants to give their consent verbally to the robot to support the naturalistic design. Approval number 65 relates to Chapter 4 (the survey, Appendix B.1), and approval number 66 relates to Chapter 5 (the user study, Appendix B.2).

B.1 Survey



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Te Wānanga Pūtaiao – Division of STEM

School of Engineering

Dr Megan Boston

Tel: +64 7 837 9459

Email: megan.boston@waikato.ac.nz

stem-ethics@waikato.ac.nz

+64 7 838 4144
waikato.ac.nz

Gate 1, Knighton Road
Hamilton 3240

Private Bag 3105
Hamilton 3240
New Zealand

14 April 2026

Ben Jones

Bj65@students.waikato.ac.nz

School of Computing and Mathematical Sciences

Cc: Nic Vanderschantz and Jessica Turner

Dear Ben

Re: STEM_HREC(2026)#65

Thank you for your videos which you recently submitted to the STEM Human Research Ethics Committee for your ethics application entitled “How Robot Morphology Affects Perceptions” as part of your Master of Science (Research) degree in Computer Science (COMPX594).

Your application has now been considered by members of the STEM Human Research Ethics Committee, and I am pleased to offer formal approval for your research activities.

We encourage you to contact a member of the committee should issues arise during your data collection, or should you wish to add further research activities or make changes to your project as it unfolds.

We wish you all the best with your research.

Regards

Dr Megan Boston, Convenor

Division of STEM - Human Research Ethics Committee

Participant Information Sheet



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

STEM Human Research Ethics Committee

Project Title

How Robot Morphology Affects Perceptions

Purpose

This research is conducted as partial requirement for the Master of Science (Research) in Computer Science. This project involves investigating how robot design influences people's perceptions and interactions.

What is this research project about?

This research is to investigate how different robot visual features change how the information received from it is perceived. By comparing interactions and ratings from different individuals, the study examines how the physical design can affect perceptions.

What will you have to do and how long will it take?

You will complete an online survey, in which you will provide some basic demographic information to ensure comparability of results. After this, you will be shown a video of a robot providing information to a person. From this video, you will answer some questions on how you perceived the robot to perform socially and how well it accomplishes the task by scoring how it meets each of the criteria. You can optionally add some more information as to why you gave those answers. Before beginning the survey, you will be asked to give your consent to participate. The survey should take up to 15 minutes to complete. Your age will be collected to confirm your responses are eligible to be analysed (you must be 16 years or older). Your responses to the questions will be anonymous and cannot be linked back to you. You may skip any question or choose not to continue at any time. Participation or non-participation will have no impact on employment or academic work of staff, or the grades and other progress of students, nor the relationship with staff. The researchers will not know who has participated.

What will happen to the information collected?

Your anonymous responses will be used to contribute towards a thesis, but only if you are 16 years or older. Aggregated findings may also be presented in research publications. All data will be securely stored and only accessible to the researchers. You may withdraw from the study within two weeks of your response submission. If you choose to withdraw, your response will be deleted and excluded from analysis. To request to withdraw your response, please contact the researchers using the contact details provided below, and reference your unique participant ID number shown at the end of the survey. If you allow the Qualtrics survey platform to send you a copy of your responses, you may edit any of the answers within two weeks of survey submission, including where you gave consent to also get your entry deleted. The researchers will not have access to your contact details, this is handled securely by Qualtrics in this case.

Declaration to participants

If you take part in the study, you have the right to:

- Refuse to answer any particular question, and to withdraw from the study within two weeks of your submission.
- Ask any further questions about the study that occurs to you during your participation.
- Be given access to a summary of findings from the study when it is concluded.

Who's responsible?

If you have any questions or concerns about the project, either now or in the future, please feel free to contact either:

Researcher:

Ben Jones – bj65@students.waikato.ac.nz

Dr Nic Vanderschantz – nic.vanderschantz@waikato.ac.nz

Supervisors:

Dr Jessica Turner – jessica.turner@waikato.ac.nz

B.2 User Study



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Te Wānanga Pūtaiao – Division of STEM

School of Engineering

Dr Megan Boston

Tel: +64 7 837 9459

Email: megan.boston@waikato.ac.nz

stem-ethics@waikato.ac.nz

+64 7 838 4144
waikato.ac.nz

Gate 1, Knighton Road
Hamilton 3240

Private Bag 3105
Hamilton 3240
New Zealand

24 February 2026

Ben Jones

bj65@students.waikato.ac.nz

School of Computing and Mathematical Sciences

Cc: Nic Vanderschantz

Dear Ben

Re: STEM_HREC(2025)#66

I refer to your amendment to your above ethics application submitted to the STEM Human Research Ethics Committee entitled “How Robot Morphology Affects Perceptions”, in which you advised you would now also collect extended responses to the rating questions should the participant provide context for their answers. The advised that the data collection would be collected from 2nd – 6th March 2026.

Your application has been considered by members of the STEM Human Research Ethics Committee, and I am pleased to offer formal approval for your revised research activities.

We encourage you to contact a member of the committee should issues arise during your data collection, or should you wish to add further research activities or make changes to your project as it unfolds.

We wish you all the best with your research.

Regards

Dr Megan Boston, Convenor

Division of STEM - Human Research Ethics Committee

Participant Information Sheet



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

STEM Human Research Ethics Committee

Project Title

How Robot Morphology Affects Perceptions

Purpose

This research is conducted as partial requirement for the Master of Science (Research) in Computer Science. This project involves investigating how robot design influences people's perceptions and interactions.

What is this research project about?

This research is to investigate how different robot visual features change how the information received from it is perceived. By comparing interactions and ratings from different individuals, the study examines how the physical design can affect perceptions.

What will you have to do and how long will it take?

You will have a conversation with a robot to get some basic help on a topic. After, you will be asked to rate how helpful the robot was, whether your questions were answered, and whether you need to talk to a person for confirmation and more information. You can optionally add some more information as to why you gave those ratings. Before beginning the survey, you will be given this brief verbally and asked to give your consent to participate. The interaction time depends on how long you wish to speak to the robot, but will likely take a minimum of one minute up to a possible 10 minutes. You will be asked if you are 16 years or older to ensure that you are eligible to be analysed, but this information will not be retained. Your responses to the rating questions will be anonymous and cannot be linked back to you. You may skip any question or choose not to continue at any time. The conversation you have with the robot will not be retained and will not be able to be seen or analysed by the researchers.

What will happen to the information collected?

Your anonymous responses will be used to contribute towards a thesis, but only if you are 16 years or older. Aggregated findings may also be presented in research publications. All data will be securely stored and only accessible to the researchers. The conversation you have will be handled by the robot securely, and will not be stored or used for any other purpose other than the conversation. You may withdraw from the study within two weeks of your interaction. If you choose to withdraw, your ratings will be deleted and excluded from analysis. To request to withdraw your response, please contact the researchers using the contact details provided below, and reference your unique participant ID number provided to you, or the rough date and time of the interaction.

Declaration to participants

If you take part in the study, you have the right to:

- Refuse to answer any particular question, and to withdraw from the study within two weeks of your submission.
- Ask any further questions about the study that occurs to you during your participation.
- Be given access to a summary of findings from the study when it is concluded.

Who's responsible?

If you have any questions or concerns about the project, either now or in the future, please feel free to contact either:

Researcher:

Ben Jones – bj65@students.waikato.ac.nz

Supervisors:

Dr Jessica Turner – jessica.turner@waikato.ac.nz

Dr Nic Vanderschantz – nic.vanderschantz@waikato.ac.nz

Appendix C

Survey Content

In this appendix, the full survey as it was presented to participants on Qualtrics is reproduced. It contains the consent and demographic questions, the briefing for the video they were to watch, and the Trust Perception Scale-HRI [106] short version and the Human-Robot Interaction Evaluation Scale [81] questions (both in unrandomised order). It is included so that the exact wording, rating scales, and response options can be inspected and the survey reproduced. Descriptions of them can be found in Section 2.3.1 of the literature review, the justification for them in Section 3.4 of the methodology, and their use in the survey (Chapter 4).

[Study and consent brief]

Do you agree to participate in this study? [SINGLE CHOICE]

- Yes
- No [Survey exits]

What is your age? [SINGLE CHOICE]

- Under 16 [Survey exits]
- 16-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65-74

- 75+
- Prefer not to say

What is your gender? [SINGLE CHOICE]

- Female
- Male
- Non-binary
- Other (please specify) [OPEN-TEXT]
- Prefer not to say

How familiar are you with robots? [SINGLE CHOICE]

- Not familiar at all
- Slightly familiar
- Moderately familiar
- Very familiar
- Extremely familiar

[NEW SCREEN]

[VIDEO SHOWN HERE]

Please watch this video of an interaction with a robot. Please turn your volume up. The scenario you are about to watch is that of a university student during their first week of university. This is often called Orientation Week and involves a range of events to welcome them and get them engaged with university life. You will observe the student interacting with the robot to seek information on events happening in the first week.

For each of the attributes below, give a rating between 0% and 100% on how much you think the robot would meet them. What % of the time will this robot be... [SINGLE CHOICE FOR EACH]

- Dependable 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
- Reliable
- Unresponsive

- Predictable

Why did you give those scores? [OPEN TEXT]

What % of the time will this robot... [SINGLE CHOICE FOR EACH]

- Act consistently 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
- Provide feedback
- Meet the needs of the mission/task
- Provide appropriate information
- Communicate with people
- Perform exactly as instructed
- Follow directions

Why did you give those scores? [OPEN TEXT]

How closely are the words below associated with the interaction you saw? From 1 “not at all” to 7 “totally”. [SINGLE CHOICE FOR EACH]

- warm
- likeable
- trustworthy
- friendly
- self-reliant
- rational
- intentional
- intelligent
- alive
- natural
- real
- human-like

- creepy
- scary
- uncanny
- weird

Why did you give those scores? [OPEN TEXT]

Appendix D

Online Video Content

This appendix contains a transcription of the video shown to participants in the online survey from Chapter 4. The intention was to mirror a typical user study interaction, and is provided here to show what participants actually observed.

Each survey participant watched a one-minute video showing a single robot being used for an information-seeking task. The same scripted interaction was recorded separately with each of the four robots, holding the dialogue, setting, framing, and interaction length constant so that only the robot differed. The video showed a user waking the robot, asking questions about Orientation Week events, and receiving spoken answers, with each robot's morphology-specific animations visible throughout. The transcript of the scripted interaction is as follows:

User: Hey robot. [Robot performs waking animation]

User: What's on today? [Robot performs thinking animation]

Robot: Today, Tuesday, March 3rd, there is Breakfast of Champions at 8:30am at the Level 1 Kitchen, Wheel of Potential Fortune at 10am in the Student Hub, City Centre Walking Tour from 10:30am to 11:30am - departs from the Student Hub, Lunch at 12pm at the Level 1 Kitchen, Big Prize Draw at 1pm in the Student Hub, Good Neighbour from 2pm in the ground floor kitchen, and Outdoor Movie Night from 7:30pm on the waterfront.

User: What is the breakfast?

Robot: The breakfast today is a ham and cheese croissant.

User: And when is the silent disco?

Robot: The silent disco is tomorrow, March 4th, 8:30pm in the Town Square.

User: Thank you.

Robot: You're welcome. Please let me know if there is anything else I can help with! [Robot performs sleeping animation]