Degree project

Making Robot Motions Understandable: A Qualitative Study on Human Interpretation of Gripper Motion intents

Master Degree Project in Virtual Product Realisation (Intelligent Automation)

Second Cycle 30 credits

Spring semester 2025

Student: Rency Raju

Supervisor: Peter Thorvald

Examiner: Dan Högberg

Abstract

This project explores how people understand a robot's actions when it communicates only through its movements. In many human-robot collaborative settings, robots use lights, sounds, or displays to show their intentions. While these signals can be useful, they often distract from the natural way humans read meaning from motion. This study focuses on *motion intent*, i.e. the idea that the robot's movements alone can express what it is about to do. Earlier research has concentrated on visual or audio signals, leaving *motion intent* less explored. This project addresses that gap by studying how people interpret different gripper motions, such as opening, closing, lifting, or rotating, when no other cues are given. A qualitative approach was used so that participants could describe what they thought each motion meant and how they felt about the robot's behaviour. The goal is to understand how clear and natural robot motions can make collaboration safer, smoother, and easier to trust. By focusing on motion as the main form of communication, this work highlights how thoughtful movement design can help robots interact more intuitively with people in shared workplaces.

Keywords: Human–Robot Collaboration, Motion Legibility, Gripper Communication, Qualitative Analysis, Thematic Analysis, Intent Recognition, Industry 4.0

Table of Contents

List of Figures

List of Tables

Acknowledgement

I would like to express my heartfelt gratitude to my thesis supervisor, Peter Thorvald, for his continuous guidance, support, and insightful feedback throughout this thesis project.

His expertise and mentorship have been supportive in shaping the direction and enhancing the quality of this thesis. His patience and encouragement have been invaluable, providing me with the confidence to navigate the complexities of my research.

Furthermore, I sincerely thank the University of Skövde for the enriching experience and knowledge acquired while working on the thesis on campus. The exposure to real-world challenges and solutions has been profoundly beneficial.

Skövde,2025

Rency Raju

Certification and Declaration on the Use of Artificial Intelligence (AI) Tools during the Degree Project

This report has been submitted by Rency Raju to the University of Skövde as a requirement for the degree of Master of Science in Intelligent Automation, Virtual Product Realisation

The undersigned certifies that all the material in this report that is not my own has been properly acknowledged using accepted referencing practices and, further, that the report includes no material for which I have previously received academic credit.

As part of my commitment to academic integrity and transparency, I hereby declare the extent to which I have used Artificial Intelligence (AI) tools during the preparation and writing of this report. This declaration does not apply to the use of basic tools for checking grammar, spelling, references etc.

Please check all that apply:

☐ **No use of AI tools**

I have not used any AI-based tools or services in any part of the research, writing, analysis, or presentation of this report.

☑ **Language editing assistance only**

I have used AI tools exclusively for language correction, grammar checking, or stylistic improvements. No substantive content was generated or modified by AI.

☐ **Idea generation / brainstorming**

I used AI tools to help generate initial ideas, brainstorm possible research questions, or explore conceptual frameworks. The final work and conclusions are my own.

☐ **Methodology design**

I used AI tools to assist in designing the research methodology (e.g., structuring experiments, surveys, or models). The methodology has been critically reviewed and adapted by me.

☑ **Programming or code assistance**

I used AI-based coding assistants to help write or debug code. All code has been reviewed, tested, and modified by me.

☐ **Data analysis / processing supports**

I used AI tools to assist with data analysis, visualization, or statistical processing. Interpretations and conclusions were drawn independently.

☐ **Content generation (text, figures, tables)**

AI tools were used to generate portions of text, diagrams, or tables that appear in this report. Such content has been critically reviewed, fact-checked, and appropriately cited where applicable.

☐ **Other (please specify):**

Responsibility Statement

I confirm that all AI-assisted contributions have been used ethically and in accordance with institutional guidelines. I take full responsibility for the accuracy, originality, and academic integrity of the content presented in this report.

I understand that any misuse or misrepresentation of AI tool usage may be considered a breach of academic conduct and could result in disciplinary actions, including revocation of the report and/or the academic degree awarded.

Signature: 

Name: Rency Raju

Date: 19/12/2025

1 Introduction

Over the past several decades, industrial robotics has evolved from isolated automation toward closer collaboration with human operators. Early industrial robots were primarily designed for repetitive and high-precision tasks and were typically enclosed behind safety barriers, reflecting a paradigm in which humans and robots operated in strictly separated workspaces (Villani et al., 2018). These systems offered strength, endurance, and consistency but provided limited flexibility and little support for interaction or communication, leading robots to be perceived mainly as programmable tools rather than collaborative partners (Matheson et al., 2019). The emergence of Industry 4.0 has contributed to a shift away from this separation-based model, emphasising intelligent, connected, and adaptive production systems that increasingly require safe and intuitive human–robot collaboration (Villani et al., 2018). Modern production systems are increasingly characterised by intelligence, connectivity, and adaptability (Pascher et al., 2023). Mass customisation, reduced batch sizes, and the need for rapid changeovers have replaced the static automation logic of the past. To meet these demands, collaborative robots or cobots have been developed to operate safely alongside humans, merging the precision and repeatability of machines with human creativity, adaptability, and contextual reasoning (Dragan et al., 2013; Lemasurier et al., 2021). In principle, cobots enable a hybrid workspace where each partner contributes complementary strengths. In practice, however, this integration introduces new cognitive requirements on both partners. Human operators must be able to interpret and anticipate the robot's actions, while robots increasingly incorporate mechanisms to sense and respond to human behaviour. Human–robot collaboration therefore depends on a reciprocal process of mutual understanding rather than unidirectional adaptation. For effective collaboration, it is therefore crucial that human operators can correctly interpret robot motion, as movement often serves as the primary cue through which a robot's intended next action is communicated.

A fundamental question in Human–Robot Collaboration (HRC) remains deceptively simple: "What is the robot going to do next?" (Szafir et al., 2015). In human teamwork, subtle cues such as gaze direction, hand gestures, or body posture allow individuals to coordinate seamlessly without explicit dialogue. Similarly, when people work in close proximity to robots, their ability to predict robotic intentions directly affects safety, efficiency, and trust (Kirschner et al., 2022). Misinterpretation of motion can result in hesitation, inefficient coordination, or even accidents. Thus, intent communication is not a secondary feature but a prerequisite for effective collaboration (Pascher et al., 2023).

Historically, researchers have attempted to address this issue through explicit signalling mechanisms such as lights, audio cues, and displays (Lemasurier et al., 2021). While these can improve clarity, they often compete with sensory overload in industrial environments like blinking LEDs or auditory alarms may be missed or misread. This limitation has shifted attention toward implicit

communication specifically, the use of robot motion itself as an expressive channel.

The present study concerns non-humanoid articulated robots equipped with a simple parallel gripper rather than anthropomorphic hands. These robots, commonly used in industrial and research settings, communicate primarily through arm trajectories and discrete gripper actions rather than finger articulations or expressive body motion. Because the end-effector serves as the robot's functional interface with the environment, observers often focus their attention on the gripper during manipulation. This context differs from humanoid Human Robot Interaction (HRI) studies, where anthropomorphic hands provide richer communicative cues, and therefore requires examining how intent is interpreted from simplified, non-anthropomorphic motion signals.

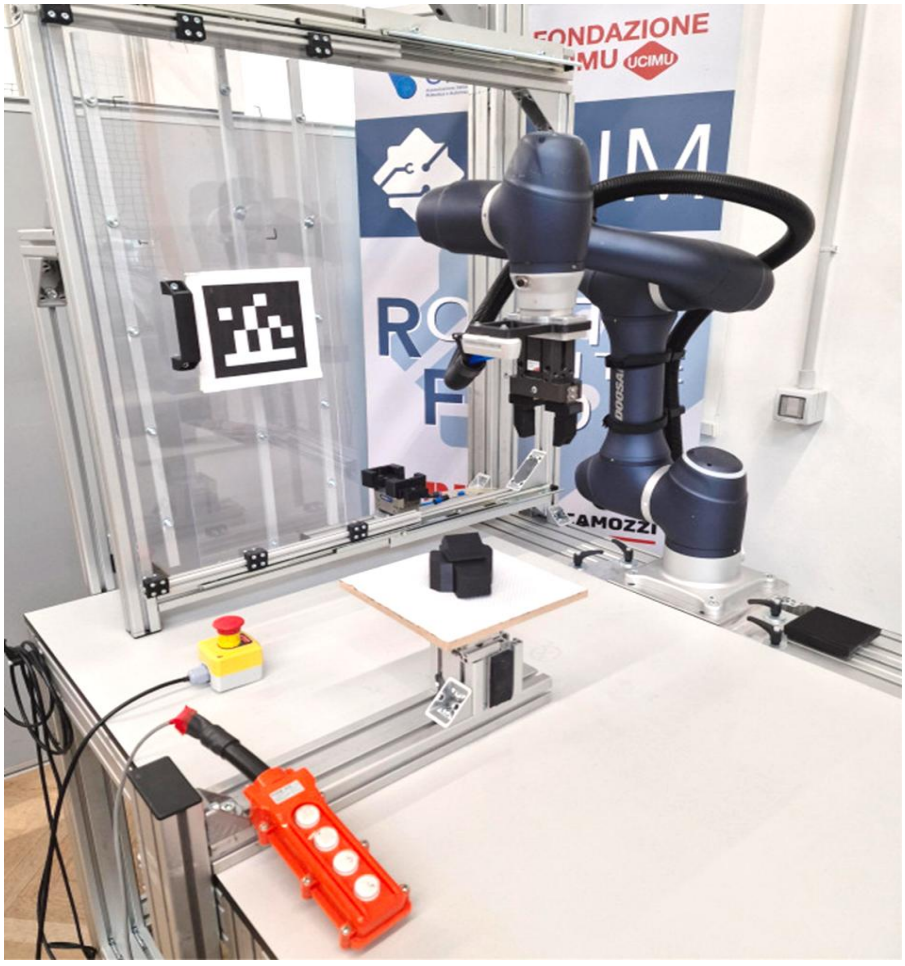

Figure 1 Industrial Robot with an end effector (Zappa et al., 2026).

Motion is inherently meaningful. Every trajectory, velocity profile, or pause carries information that human observers intuitively interpret (Dragan et al., 2013). In their seminal distinction, Dragan and colleagues differentiated between predictable and legible motion. Predictable motion aligns with what an observer expects once the goal is known, whereas legible motion allows the

observer to infer the goal merely from watching the motion unfold. For example, a robot that subtly arcs its reach toward a target may appear inefficient but reveals its intention earlier to a human partner while enhancing legibility at the cost of minimal path optimality. Here, 'path optimality' refers to the geometrically shortest or most efficient path between start and goal. Motion intent has been systematised as the explicit communication of future actions through movement properties, location, and timing (Pascher et al., 2023). Collectively, these studies establish "motion as a non-verbal communication medium central to HRC design".

Yet, despite this growing recognition, most research concentrates on arm trajectories or multimodal cues combining motion with lights and sounds. The gripper, which serves as the robot's primary manipulation interface and functional analogue to the human hand, has received limited dedicated research attention, despite often attracting the human partner's visual focus during manipulation (Villani et al., 2018; Lemasurier et al., 2021). Gripper movements inherently carry communicative potential: an opening gesture suggests readiness to grasp or release; a pause indicates hesitation or completion; a rotation implies alignment. Understanding how humans interpret these micro-motions is critical for designing robots whose actions are not merely safe, but intuitively understandable.

A second limitation in existing literature lies in methodological bias. The majority of prior studies adopt quantitative evaluation which include task completion times, recognition accuracy, or usability scores (Guest et al., 2006). Such metrics reveal what participants did but not why. They fail to capture the reasoning processes, contextual cues, and perceptual heuristics by which humans make sense of robot behaviour. Consequently, design recommendations often remain superficial, optimising performance without uncovering the cognitive underpinnings of understanding.

Addressing this limitation demands a qualitative perspective grounded in interpretation. Approaches such as think-aloud protocols, semi-structured interviews, and video elicitation enable participants to articulate their thought processes while observing robot motion (Braun & Clarke, 2006). Thematic analysis of such data can uncover latent patterns, how observers construct meaning from trajectory shapes, timing, or context thereby identifying both convergent and divergent interpretations. As Braun and Clarke emphasize, thematic analysis provides a structured yet flexible method for revealing how people experience and attribute meaning to events, making it particularly suitable for studying perception and intent in HRC.

From an industrial viewpoint, legible motion is not only about comfort; it directly affects operational efficiency and safety. Here, comfort refers to the human operator's psychological and perceptual ease when interpreting nearby robot motion, i.e. specifically the reduction of stress, surprise, and uncertainty during collaborative tasks. Robots that act predictably reduce hesitation and downtime, while robots that act legibly foster trust and smoother workflow synchronization (Kirschner et al., 2022; Abe et al., 2024). Psychological safety

refers to the assurance that a robot's movements will not surprise or endanger a human partner. It is now regarded as an important dimension of cognitive ergonomics (Kirschner et al., 2022). It has been shown that positive perceptions of autonomous robots depend strongly on whether their intentions are correctly inferred, while robots whose purposes are not understood are perceived as uncooperative, highlighting how comprehension directly governs acceptance (Abe et al., 2024). Here, 'autonomous' refers to robots that generate or adapt their movements without direct human teleoperation, yet still operate in close proximity to people.

Therefore, an integrated understanding of motion, perception, and communication is necessary for the design of collaborative systems. In order for HRC to be effective, robots must move both within perceptual expectation bounds, i.e. motions that correspond with how humans perceive intentional action and within biomechanical safety limits, which are established limitations on speed, force, and impact energy to prevent physical harm. By using these principles, robots become socially interpretable motion design rather than just kinematic optimisation.

This project responds to the identified research gap by focusing specifically on gripper-based motion communication and on qualitative interpretation rather than quantitative performance. It investigates how humans perceive, describe, and reason about various gripper motions in collaborative scenarios, using interpretive methods to uncover the cognitive logic behind their judgments. Through systematic thematic analysis, the study aims to map the semantic space of gripper motion, what each movement "means" to a human observer and to derive practical design principles for legible and trustworthy robot behaviour.

2 Background

2.1 Human–Robot Collaboration in Modern Workspaces

For many decades, industrial robots operated separately from humans because the available technology and safety regulations required physical separation. Early robots were designed for high-speed, high-power, repetitive tasks and lacked the sensing, force-limiting, or perception capabilities needed for close human interaction. As a result, they were installed inside cages or behind protective barriers (De Santis et al., 2008; Villani et al., 2018), and human access was restricted to programming and maintenance activities. This enforced separation shaped how robots were perceived in factories, i.e. not as partners, but as specialised automation units performing narrowly defined tasks. This situation has begun to change as Industry 4.0 increases the demand for flexible, adaptive production systems. Instead of rigid automation designed around a single fixed task, modern factories require systems that can support frequent changeovers, small batch sizes, and dynamic workflows. Collaborative robots (cobots) emerged in response to these technological and organisational needs.

Unlike traditional industrial robots, cobots are built to operate safely alongside humans, combining the precision and repeatability of automation with human strengths in problem-solving, dexterity, and adaptability (Kirschner et al., 2022).

The transition to HRC presents both new challenges and exciting opportunities. On the plus side, when properly designed, it promises safer working conditions, less physical strain on employees, and more effective use of resources. This improvement in safety arises because collaborative robots are equipped with force and speed limiting mechanisms, which reduce the risk of high-impact collisions, and because they can take over physically demanding or ergonomically harmful tasks that typically lead to worker injuries (Villani et al., 2018). For instance, humans can concentrate on tasks requiring judgment or fine control while robots handle repetitive motions or heavy lifting. However, working so close to the robot can cause hesitation or even accidents if there is any uncertainty about its next move. People may stop working, overreact, or put themselves in danger if they find it difficult to comprehend what the robot is about to do. People must be able to read the robot's behaviour and predict its intentions, just like they would with a human coworker, for collaboration to be truly successful. Small cues, such as a glance, a hand gesture, or a slight movement, help partners in human teams stay in sync. Similar to this, collaborative robots must clearly convey their next moves so that people can react instinctively and without anxiety (Dragan et al., 2013).

2.2 Importance of Communicating Intent

The question, "What is the robot going to do next?" is one of the most straightforward but challenging problems in HRC. In order to predict each other's next move, coworkers constantly read each other's tone of voice, body language, and even tiny hand gestures. Tasks can proceed without needless pauses or confusion thanks to this seamless anticipation. In a similar vein, humans must be able to comprehend and predict the actions of robots in order for them to be trusted partners. Collaboration is made much simpler when a robot expresses its intentions clearly. When the robot is about to enter their area, humans can move aside to make room for it, they can get ready for the next task without having to wait, and they can feel secure knowing that they won't be startled by sudden or unexpected movements.

The advantages are obvious:

Efficiency: There are fewer hiccups or delays in the work process.

Safety: By keeping people out of a robot's path, collisions are less likely to occur.

Trust: Employees become more assured that the robot is behaving in a trustworthy and comprehensible manner.

However, people may hesitate or even stop what they're doing if they can't read a robot's behaviour. For instance, consider a robot that rotates its gripper. Is it spinning aimlessly or is it aligning an object to place it? The intention is not immediately apparent to the human partner. In addition to slowing down the task, this kind of ambiguity can make people feel uneasy or unsafe. Uncertainty can occasionally even result in errors, like when a worker reaches for something at the same time as the robot. Researchers attempted to address this issue by including explicit intent signals. These include digital displays that inform the operator of the robot's actions, warning lights, beeps, and alarms. These approaches have disadvantages even though they can be useful. Such signals can be easily missed or even become annoying in noisy, flashing factory settings. They also require additional hardware, which raises the cost and complexity of the system. Furthermore, in contrast to how people typically communicate with one another, these cues frequently feel strange.

Using the robot's own motions as the signal is a more organic option. Humans can learn to decipher robot motions, just as people can decipher body language or gestures to determine another person's next move. For instance, a human observer can quickly determine what a robot is attempting to grab if the robot slightly exaggerates its reach toward an object. In real life, this could entail guiding a clearer trajectory toward a target instead of taking a shortcut, or slowing down before a grasp to indicate readiness. These motion adjustments do not require additional equipment, but they can significantly enhance human-robot collaboration. Robots can decrease hesitation, boost trust, and facilitate more seamless collaboration in shared workspaces by making their actions easier to read.

2.3 Motion as a Communication Medium

Recent studies have demonstrated that a robot's movement can be a potent indicator of its purpose. Not only does motion accomplish a task, like picking up or moving an object, but it also conveys information to onlookers. Here, (Dragan et al., 2013) made a crucial distinction. According to their description, predictable motion is a movement that is simple to follow if the observer is already aware of the robot's intended action. For instance, a straight-line reach toward the right side of a table will appear predictable if someone is aware that the robot's objective is to place an object there. Legible motion, on the other hand, is made so that, simply by observing the robot's movements, a person who is not yet aware of the objective can make an accurate guess. The human partner can better comprehend the intention in this situation if the path is shaped or exaggerated to reveal the target early. Expanding on this,(Szafir et al., 2015) demonstrated that even minor variations in a robot's motion, like its speed, whether it pauses momentarily before acting, or the angle at which it approaches, can alter how humans perceive what the robot is going to do. This indicates that motion is a type of nonverbal communication in addition to being a technical aspect of trajectory planning. While a slower, more methodical motion might imply caution, a faster motion might convey urgency. In a similar

way, a robot may emphasize a point by repeating a small motion, bringing the human's attention to it.

Not every motion is equally simple to read. Human expectations are naturally met by certain movements. Most people will understand the purpose of a robot arm, for instance, if it reaches straight for a cup on the table. Because they mimic how people would behave in the same circumstance, these motions are intuitive. Other kinds of movements, however, can be much less obvious. A lateral sweep across the table might be interpreted as a programming error, as clearing space, or as searching. Similarly, a gripper rotation may indicate alignment with an object, but it may seem random or pointless without context. As a result, effectively programming robots to perform tasks is not the only challenge in HRI. It also involves creating motions that are simple and reliable for people to understand. Put another way, communication is just as important to robot motion design as mechanics. Human partners' perceptions of various trajectories, speeds, and patterns must be taken into account by engineers and designers. From the robot's perspective, an efficient movement might not always be the one that makes the most sense to a coworker nearby.

3 Problem Formulation

3.1 The Role of Gripper-Based Motions

The end effector or gripper's movements are particularly crucial among the various ways a robot can communicate through motion. Since the gripper is the component that touches and manipulates objects directly, it is frequently referred to as the robot's hand. Human partners naturally focus on it because it is the most active component of the robot when performing tasks like grasping, lifting, or placing. People frequently look at the robot's gripper to understand its intentions, much like we watch someone else's hands to understand what they are about to do.

Humans typically use their personal hand movement experiences to interpret gripper motions. People typically interpret the gripper opening, for instance, as an indication that they are ready to grab or release something. It obviously implies holding onto something when it closes. In a similar way, when the gripper moves downward, it may indicate that the robot is placing something on a surface, whereas when it moves upward, it frequently appears as though the robot is lifting something. Because they closely resemble human behaviour and are connected to obvious interactions with the surroundings, these motions are simple to comprehend.

Larger arm trajectories or systems that combine movements with other cues like lights, sounds, or displays have garnered more research attention than gripper-based signals, despite the gripper's powerful ability to capture human attention. The majority of the research to date has focused on quantitative metrics, such as how well humans can recognize a robot's objective or how fast a task can be finished using specific motions. These figures are helpful, but they only provide

a portion of the picture. They do not describe the mental processes that led people to interpret a movement in a particular way or why. Beyond measuring task times or counting correct answers, a deeper understanding of gripper motions is required. A more thorough examination of the human viewpoint how participants explain what they observe, the rationale behind their conclusions, and the types of associations they form when interpreting robot behaviour is what is lacking in the literature. Designing motions that are not only useful but also understandable and readable by human partners requires this kind of insight.

3.2 Need for a Qualitative Approach

Measuring metrics like accuracy or error rates is insufficient when researching how humans depict robot intent signals. Although these quantitative metrics can indicate whether a participant made a correct guess, they are unable to provide an explanation for their particular interpretation of the motion. A qualitative approach that examines meaning, interpretation, and the mental processes underlying human judgments is necessary to fully understand how people interpret robot movements. For instance, one participant might interpret a robot's sweeping motion as "waving goodbye," while another might interpret it as "searching for something." This difference cannot be captured by a mere accuracy score. What made them think that way is a question we must ask. Perhaps it was the robot's lack of an object at the moment, the speed of the sweep, or the motion's repetition? In the same way, the task's larger context is crucial. When the robot is aligning an object, a rotation may seem intentional, but if the task context is unclear, the same movement may seem pointless.

Researchers investigate these more profound issues with the aid of qualitative methods. Methods like think-aloud protocols, in which participants describe their ideas while observing the robot, offer instant insight into their logic. While video elicitation, which involves playing recordings of the robot's movements, encourages participants to reevaluate and provide more thorough explanations for their interpretations, semi-structured interviews enable them to think more deeply about their experiences. These techniques produce detailed and illustrative data that surpasses mere counts of correct and incorrect responses.

Following collection, the data can be examined through the use of thematic analysis (Braun & Clarke, 2006), which entails methodically coding the responses, classifying them into themes, and spotting trends among participants. This procedure calls attention to areas of misunderstanding or disagreement in addition to areas where there is widespread agreement. Both are useful: divergence reveals which signals require redesign or extra support, while consensus indicates which motions are inherently readable.

When it comes to offering design recommendations, the importance of this qualitative viewpoint is particularly evident. Even though a strictly quantitative study might find that a particular motion type had the highest recognition rate, designers would not be able to use that information to enhance or improve the motion. A qualitative study, on the other hand, can explain why one motion was

obvious possibly because it mimicked the movements of the human hand and why another was unclear possibly because it was too quick or lacked a clear objective. Designers can modify robot behaviours to make them more meaningful, intuitive, and user-friendly by using these insights. Qualitative approaches enable us to do more than just rank motions according to performance. By demonstrating how humans see, decipher, and occasionally misinterpret robot signals, they provide a window into the human side of collaboration. This knowledge is essential for creating robots that can actually express their intentions in ways that people can understand and trust.

3.3 Research Gap

Although motion-based communication is widely recognised as important in HRI, several gaps remain. Most previous studies have focused on measurable outcomes such as task completion time, accuracy, and efficiency. While these metrics are useful, they do not explain how people understand robot motions or why they interpret a motion in a particular way. As a result, the human perspective behind motion interpretation is often missing.

Another gap is the limited focus on gripper motion. The gripper is the robot's main point of contact during manipulation and is often where people look first. However, few studies examine gripper motion on its own. Instead, research has mainly focused on arm movement or combined motion with lights and sounds. This is a missed opportunity, as gripper motion may provide a simple and natural way for robots to communicate intent.

There is a lack of qualitative research that explores how people explain and reason about robot motion. Many studies report whether a signal was noticed or preferred, but not how participants understood it. Motion-based and light-based signals were compared with differences in comfort and effectiveness reported, while participant interpretation of the motions was not the primary focus (Lemasurier et al., 2021). Building on this work, the present study uses a motion-only and qualitative approach to examine how people understand robot gripper motions, and how expectations and task context influence their interpretations.

4 Aim

4.1 Aim and Research Questions

The objective of this study is to better understand how humans perceive and comprehend various gripper-based robot movements in a cooperative environment. This study draws attention to the human perspective how people perceive, explain, and justify the robot's actions by emphasizing interpretation rather than performance.

The goal of this research is to generate insights that transcend statistics and accuracy rates by gathering and examining participant reflections using qualitative methods. Converting these insights into useful design suggestions is the ultimate objective. These suggestions can help engineers and designers create intuitive and functional robot behaviours so that people interacting with them can quickly and confidently understand what the robots are trying to do. In this sense, the research advances the creation of robot systems that are more human-centered, reliable, and feel more natural in collaborative work environments. The research question of this study is:

"***How do participants interpret and make sense of different gripper motions, and what contextual or motion features shape their understanding?***"

5 Literature Review

5.1 Robot Intent and Motion Communication

Understanding how robots communicate their motion intent

ions to humans lies at the core of effective human–robot collaboration. The evolution of this field reveals a gradual shift from examining isolated signalling modalities to establishing comprehensive frameworks that capture legibility, predictability, and contextual meaning. Early work by (Lemasurier et al., 2021) highlighted that collaboration in shared industrial spaces demands intuitive communication methods to prevent accidents and improve fluency. Through experiments comparing motion-based and light-based intent signals using a Baxter robot, they demonstrated that proximity and modality distinction play decisive roles in human perception. Light cues positioned near the end effector, particularly an LED bracelet, proved most effective because they were perceptually independent from task-related motion and therefore easier to interpret. Participants consistently rated this modality as clearer and more comfortable, revealing that human collaborators instinctively associate separate, non-task-related cues with communicative intent. The study advanced practical guidance for industrial robots by proving that visible, proximate, and distinctive signals enhance interpretability and safety, while also uncovering social projection tendencies participants often attributed avoidance behaviour or cooperation motives to the robot itself. This insight underscores that intent communication is not purely technical but embedded in human cognitive and social expectations.

Building on the challenge of making motion understandable, (Dragan et al., 2013) established a conceptual and mathematical distinction between legibility and predictability. They argued that predictable motion follows a goal-to-action inference what trajectory a human expects given a goal whereas legible motion follows an action-to-goal inference what goal best explains an observed action. The two properties, though related, often conflict: a predictable trajectory may be efficient but ambiguous, while a legible one deliberately deviates to clarify

intent. Their Bayesian framework modelled how observers infer goals from partial trajectories, demonstrating empirically that legible paths improve recognition speed and confidence, even if they appear less efficient. This reframed motion planning as a problem of human interpretability rather than mere optimization. For collaborative robotics, the implication is clear: motion must communicate why an action is happening, not only where it is heading. The distinction established by (Dragan et al., 2013) continues to inform studies seeking to balance transparency and efficiency in human-robot cooperation.

Parallel efforts explored communication beyond physical manipulation. (Szafir et al., 2015) investigated how small flying robots can visually express their flight intentions through explicit light-based cues. Addressing what they called the "gulf of evaluation" between robotic behaviour and human comprehension, they introduced four signalling designs are Blinker, Beacon, Thruster, and Gaze each inspired by familiar human or mechanical metaphors. Their experiments confirmed that such signals significantly improved users' ability to anticipate robot movement and enhanced perceived clarity and confidence. Importantly, the study demonstrated that visual cues grounded in everyday metaphors (such as turn indicators or gaze direction) can bridge the interpretive gap between humans and autonomous agents, translating motion into an accessible form of non-verbal language. The underlying message resonates with Dragan's framework: predictability may arise from conformity to expectation, but legibility emerges from communicative intent designed for human perception.

As the field matured, (Pascher et al., 2023) provided a unifying framework through a scoping review of seventy-seven studies on robot motion intent. Their Intent Communication Model categorised communication along three dimensions which constitute intent type, intent information, and intent location linking them to specific human roles such as collaborator, observer, coworker, or bystander. They defined four primary intent types: motion (future actions), attention (pre-movement cues), state (system condition), and instruction (requests for human assistance). Furthermore, Pascher et al. emphasised that multimodal combinations, rather than single-channel signals, yield the most robust understanding. Their synthesis framed motion intent communication as a structured design problem where spatial proximity, temporal timing, and user role jointly determine effectiveness. They concluded that consistent intent representation is a prerequisite for explainable and trustworthy robotics, calling for integrative approaches that blend motion, visual, and contextual cues.

It has been demonstrated that perceptually distinct motion cues enhance user comfort during interaction (Lemasurier et al., 2021), that the cognitive mechanisms underlying the interpretation of robot trajectories can be systematically explained (Dragan et al., 2013) and that motion expressivity can be extended to aerial robots through metaphor-based design (Szafir et al., 2015). This body of work has further been consolidated into a cross-modal taxonomy of robot intent communication (Pascher et al., 2023), establishing that transparency in robot behaviour depends on the alignment of physical movement, visual feedback, and human cognitive models.

5.2 Multimodal and Expressive Interaction

As research on robot intent evolved from single-modality signalling toward richer communicative behaviour, attention shifted to how robots integrate multiple sensory channels to enhance expressivity, engagement, and understanding. (Bonarini, 2020) conceptualized communication in HRI as a multimodal process spanning auditory, visual, and tactile dimensions, asserting that it extends "beyond language-based interaction" and encompasses all perceptual channels available to embodied agents. He emphasized that effective HRI depends on the coherence of these channels, how sound, motion, and light combine to form an integrated message. Within this framework, communication occurs not only through deliberate signals but also through involuntary cues arising from movement and spatial proximity, reflecting the impossibility of non-communication when a perceptual channel remains open between agents.

Auditory cues form one of the most established channels of robot communication. Bonarini discussed how robots use synthetic speech, non-verbal sounds, and prosody to convey emotional states and intentions. Sound patterns such as beeps or melodic tones popularized in cinematic robots like R2D2 serve as intuitive affective signals. Advances in text-to-speech synthesis and dialogue management allow robots to adapt tone, rhythm, and emotional expression to context. On the receptive side, prosodic analysis and speech recognition enable robots to infer human emotions and intentions, completing the communicative loop. Bonarini argued that emotional nuance conveyed through voice and sound fosters familiarity and reduces perceived mechanical detachment, strengthening social trust in robotic partners.

The visual modality contributes both expressivity and clarity. Light cues, colour changes, and movement patterns can all serve as intentional communication media. Bonarini noted that light intensity, rhythm, and hue can encode emotional or operational states, while movement of particularly head, arm, or body motion conveys attention and intent. Facial displays or screen-based expressions, though less embodied, offer a structured and easily interpretable channel for conveying affective information. Bioinspired motion, such as gestures modelled after human or animal behaviour, reinforces perceived animacy and aids comprehension. However, the author also recognized persistent limitations in visual interpretation: despite extensive research, emotion recognition from facial expressions and gesture detection under real-world variability remain technically challenging. Nonetheless, the visual channel remains central for creating natural, believable interactions between humans and robots.

Tactile communication completes the triad by linking physical contact to emotional and safety cues. Robots equipped with capacitive or resistive sensors can interpret affective gestures such as touches, caresses, or resistance, which Bonarini described as critical for fostering trust and comfort, especially in therapeutic or assistive contexts. The nature of contact such as gentle, firm, or

resistive implicitly communicates intent and emotional state. By bridging social and physical dimensions, tactile interaction underscores the embodied character of human–robot communication.

Machine learning further strengthens multimodal integration by enabling adaptive perception and response. Bonarini highlighted imitation and reinforcement learning as techniques through which robots can acquire interaction patterns from experience, dynamically refining how they express or interpret signals. Yet he cautioned that limited datasets restrict progress, reinforcing the need for alternative learning paradigms capable of generalizing across unstructured settings. Evaluation and benchmarking were identified as emerging priorities: the quality of communication must eventually be assessed under objective, reproducible conditions to ensure both performance and safety in real-world deployment. Bonarini's synthesis thus established a holistic foundation for understanding communication as an inherently multimodal and adaptive process.

Building on this foundation, (Gielniak & Thomaz, 2012) explored how exaggerated motion can enhance the social and cognitive effectiveness of robot behaviour. Drawing inspiration from classical animation, they proposed that deliberate exaggeration amplifies communicative clarity by emphasizing key motion elements, guiding human attention, and increasing memorability. Implemented on the humanoid robot Simon, their algorithm modified gesture trajectories in real time to create cartoon-like emphasis. Experiments comparing exaggerated and natural motions revealed that participants consistently perceived the exaggerated versions as more engaging, entertaining, and emotionally expressive. Crucially, observers retained more details from interactions involving exaggerated motion and directed their gaze longer toward exaggerated body parts, confirming that exaggerated kinematics not only improve subjective enjoyment but also support cognitive processing. The authors concluded that controlled exaggeration transforms motion from a purely functional act into a communicative performance bridging engineering precision with human perception. In the context of collaborative robotics, this implies that even utilitarian movements can be stylized to signal intent, reduce ambiguity, and sustain engagement without verbal explanation.

Expressivity is not limited to entertainment or attention; it also plays a pedagogical role in shaping human learning and compliance. (Ramachandran et al., 2018) demonstrated this by integrating social expressivity into a robot tutoring system designed to encourage children's metacognitive "think-aloud" strategies. Their experiment combined robot embodiment with cognitive prompting during mathematical problem solving. Results showed that children interacting with the humanoid robot displayed greater engagement, persistence, and learning gains than those in non-robot or silent conditions. The robot's physical presence and responsiveness created a social context that sustained attention and fostered verbal reasoning. Immediate learning improvements were attributed to embodiment and feedback, while longer-term retention stemmed from active verbalization. The study confirmed that social presence and expressivity can facilitate cognitive processes, positioning robots

not merely as instructors but as interactive partners that influence motivation and reflection.

Across these works, a consistent theme emerges: communication in HRI relies on the coordination of multiple expressive channels that align physical behaviour with human cognitive and emotional expectations. Bonarini established that true interaction demands multimodal coherence; Gielniak and Thomaz demonstrated how exaggerated motion enhances salience and memory; and Ramachandran et al. proved that embodiment and expressivity strengthen engagement and learning. Together, these insights reveal that motion, sound, and behaviour operate synergistically to create meaning in robot actions. This theoretical grounding informs the present project by highlighting that the legibility of a gripper's motion cannot be separated from its expressive qualities such as the timing, speed, and curvature of movement function as a visual language that humans instinctively interpret as intentional communication.

5.3 Human Perception, Trust, and Psychological Safety

As collaborative robots gain autonomy, a critical question arises: how do humans interpret and emotionally respond when robots act without explicit instruction? (Abe et al., 2024) tackled this challenge by studying how people perceive unanticipated robot actions during physical interaction. Their work captured an often-overlooked reality: in future workspaces, robots will occasionally initiate their own movements, and human partners must interpret those movements as either cooperative or disruptive. Through a game-based experiment in which a robot alternated between a compliant mode and an active mode that pursued higher-valued targets without warning, Abe et al. revealed the psychological processes behind this interpretation. Participants first experienced confusion, then investigation, and finally adaptation progressing from resistance to cooperation as they learned to anticipate the robot's intent.

The decisive factor determining whether participants perceived the robot as cooperative was not its physical motion alone, but whether they could correctly infer its intention. Those who understood that the robot sought to maximize rewards described it as "helpful" and performed better in the task; those who misunderstood it viewed the same behaviour as intrusive. Statistical analysis confirmed that traits such as Extraversion and Openness correlated with both understanding and positive perception, suggesting that personality mediates trust and interpretability. This dynamic interplay between cognition, emotion, and performance highlights that transparency is not achieved merely by displaying efficient motion it depends on how humans assign meaning to that motion. The study's conceptual framework of confusion–investigation–adaptation illustrates how trust is progressively constructed when humans recognize consistent patterns in autonomous behaviour.

Abe et al.'s findings redefine the boundary between autonomy and cooperation. Humans interpret agency through embodied experience, not through technical indicators. When robots act unpredictably but purposefully, the human mind

searches for intent and once that intent becomes legible, trust rapidly replaces uncertainty. This insight underscores that the perception of safety in HRI extends beyond physical collision avoidance; it includes cognitive and emotional predictability. For industrial or assistive robots, therefore, clarity of purpose is as essential as mechanical precision.

While Abe et al. addressed interpretive trust, (Kirschner et al., 2022) approached safety from complementary angle bridging biomechanics with cognitive expectation. Their development of the Expectable Motion Unit (EMU) introduced a new safety paradigm grounded in human psychology. Traditional frameworks, such as the Safe Motion Unit (SMU), focus on measurable physical limits such as speed, force, and proximity. Kirschner et al. argued that such constraints ignore the human perceptual state: even when a robot moves safely, unexpected acceleration or direction changes can startle humans and trigger involuntary motion (IM). These reflexive reactions, rather than the robot's trajectory itself, often generate the real hazard.

The EMU model therefore redefined safety as expectation fulfilment. By monitoring human–robot distance and velocity, it dynamically adjusts robot speed to maintain an "expectation curve" which is a probabilistic boundary within which human observers are unlikely to be startled. Experiments confirmed that the EMU avoided involuntary motion in five of six test cases while allowing higher operational speeds once psychological comfort was ensured. This dual-layered control is combining biomechanical and cognitive safety that creates what the authors described as a "holistic safety framework" that preserves both efficiency and trust. The results revealed that aligning robot motion with human expectation not only prevents reflexive responses but also enhances confidence, enabling smoother and faster cooperation.

Beyond the quantitative gains, Kirschner et al. highlighted a profound qualitative shift: by incorporating perceptual comfort into control logic, robots become trust-aware agents. The expectation curve is inherently human-centred and it adjusts not to the robot's physical limits but to the user's mental model of what "feels normal." This approach transforms the human from a passive safety constraint into an active reference for motion planning. By fulfilling expected behaviour patterns, the robot communicates reliability without words, pre-emptively reinforcing trust.

Taken together, (Abe et al., 2024) and (Kirschner et al., 2022) show that safety and trust emerge from cognitive alignment, not just compliance with standards. Abe et al. revealed that understanding intent transforms surprise into cooperation; Kirschner et al. demonstrated that maintaining expect ability transforms proximity into comfort. Both converge on a crucial insight: humans judge robots not only by what they do, but by how well their behaviour fits human mental models of purpose and predictability.

For the present project, this cognitive framework is used to guide how gripper motion is analysed and interpreted. It supports the view that robot movements are not perceived only as physical trajectories, but as actions that humans

actively interpret as meaningful signals. Based on this perspective, the study examines how gripper motions are understood as indicators of intent, such as goal direction, attention, or readiness to cooperate. The analytical focus of this work is informed by the framework proposed in which motion legibility is conceptualised as a cognitive process grounded in shared expectations and mental models rather than solely in motion dynamics (Lemaignan et al., 2012).

5.4 Grounded Cognition and Cognitive Modeling

Beyond physical movement and expressivity, effective human–robot communication depends on the robot's capacity to interpret and generate meaning grounded in shared context. (Lemaignan et al., 2012) advanced one of the most influential frameworks for this purpose, demonstrating how perception, reasoning, and language can be unified into a cognitive architecture that supports natural, situated dialogue between humans and robots. Their approach rests on a simple but powerful premise: for a robot to interact meaningfully, it must not only perceive the environment but also understand what those perceptions represent within a shared conceptual world.

The authors illustrated this principle through everyday scenarios—for instance, when a human says "give me that" while looking at an object, the robot must infer which object is referenced, anchoring the linguistic cue "that" to perceptual data. This process, termed symbol grounding, transforms sensory input into symbolic knowledge that both human and robot can refer to consistently. Lemaignan et al. argued that grounding is the missing link between low-level perception and high-level reasoning: without it, communication remains shallow and ambiguous.

Their cognitive architecture integrates three interdependent modules. The first, situation assessment, constructs and continuously updates a 3D model of the physical world, incorporating objects, agents, and their spatial relations. The second, a symbolic knowledge base implemented through the OpenRobots Ontology (ORO), allows the robot to store and reason about facts, beliefs, and agent perspectives using description logics. The third, dialogue processing, connects perception and reasoning to natural language through parsing, disambiguation, and clarification routines. This tri-layered system enables the robot to not only interpret human instructions but also question and refine them, mirroring the collaborative dynamics of human conversation.

One of the architecture's most innovative capabilities is perspective taking the ability to model not only its own belief state but also those of others. Using the SPARK reasoning module, the robot can simulate another agent's viewpoint to resolve ambiguities such as "the object to your left." This multi-agent reasoning marks a critical leap in HRI, as it allows robots to interpret deictic and context-dependent expressions that dominate human speech. By reasoning about how a scene appears from the human's perspective, the robot establishes a shared understanding, reducing communicative friction and errors.

Lemaignan et al. validated their framework through experiments such as the "Naming" and "Spy Game" tasks, in which robots learned object labels, asked clarifying questions, and differentiated objects based on relational context. The results demonstrated that integrating grounding and perspective awareness allows robots to engage in fluid, adaptive dialogue that reflects genuine understanding rather than scripted responses. Their findings underscored that communication is not simply an exchange of information but a process of aligning internal representations across agents.

This work redefined communication in HRI as a cognitive negotiation between perspectives. The robot becomes not a reactive tool but a reasoning partner that interprets, hypothesizes, and refines shared meaning through iterative feedback. Such grounding mechanisms lay the foundation for explainable and context-sensitive behaviour qualities essential for trust in autonomous systems. They also establish the theoretical bridge between motion and cognition: when a robot moves its arm or gripper, that movement is not just mechanical—it embodies an internal model of purpose, linked to symbolic reasoning about tasks, objects, and intentions.

For the present project, this cognitive framework provides the conceptual backbone for interpreting gripper motion as communicative behaviour. By relating physical actions to grounded representations, the project extends beyond trajectory design to the study of meaningful motion: how humans interpret actions as signals of goal, attention, or cooperation. Lemaignan et al.'s architecture thus anchors the theoretical premise that motion legibility depends not only on physical dynamics but also on shared cognitive models that enable humans to understand what a robot's movement stands for.

5.5 Qualitative and Methodological Foundations

Understanding how humans interpret robot motion requires more than technical observation; it demands a systematic approach to analysing meaning, perception, and behaviour. Qualitative research methods provide the depth and flexibility necessary for such inquiry, allowing exploration of how individuals construct understanding through interaction. (Braun & Clarke, 2006) seminal work on thematic analysis established a rigorous yet adaptable foundation for identifying and interpreting patterns within qualitative data. They defined thematic analysis as "a method for identifying, analysing, and reporting patterns (themes) within data," bridging descriptive richness with interpretative insight. Unlike many prescriptive methods, thematic analysis is not tied to a single theoretical position; it can operate within both realist and constructionist paradigms, making it particularly suitable for studying human interpretations of robotic behaviour where objective phenomena and subjective meaning coexist.

Braun and Clarke outlined a six-phase process that are familiarisation, coding, theme search, review, definition, and reporting emphasising that analysis is recursive rather than linear. This cyclical engagement allows themes to evolve as researchers move between data and interpretation, refining understanding

iteratively. Their insistence that themes do not "emerge" automatically but are actively constructed by the researcher underscores the interpretive nature of qualitative analysis. The researcher's role, therefore, is not passive observation but disciplined interpretation grounded in transparency and consistency. This principle directly informs the analytical approach of the present study, which interprets participants' perceptions of robot motion as reflections of deeper cognitive models rather than surface-level descriptions.

Complementing this methodological foundation, (Guest et al., 2006) provided empirical guidance on achieving data saturation with the point at which additional data yield no new themes. Analysing sixty in-depth interviews across cultural contexts, they demonstrated that core themes typically stabilise within twelve interviews when experiences are relatively homogeneous. Their work transformed the abstract notion of saturation into measurable evidence, confirming that qualitative validity arises from conceptual completeness, not sample size. For the current research, this insight justifies a focus on analytical depth and diversity of perspective rather than large-scale replication. It reinforces that understanding motion legibility depends on how participants articulate meaning, not on how many participants are observed.

(Adams et al., 2008) extended this methodological dialogue into the field of Human–Computer Interaction (HCI), arguing that complex, socially embedded phenomena require qualitative inquiry capable of capturing emotion, context, and human motivation. They highlighted grounded theory as a particularly effective approach for generating new theoretical insights from experiential data without relying on prior hypotheses. Grounded theory proceeds through iterative cycles of data collection, coding, and conceptual refinement until theoretical saturation is achieved. Through open, axial, and selective coding, researchers distil raw narratives into structured conceptual frameworks that explain observed behaviour. This process mirrors the dynamic between human interpretation and robot behaviour: just as a robot refines its understanding through feedback, the researcher refines theoretical understanding through iterative engagement with data. Adams et al. emphasized reflexivity is the conscious examination of how the researcher influences interpretation as a safeguard against bias and a cornerstone of credibility. This perspective strengthens the epistemological alignment of the current project, which recognises the researcher as an interpretive participant in the meaning-making process.

Expanding these methodological principles into the socio-technical domain, (Hoda, 2025) introduced the Socio-Technical Grounded Theory (STGT) framework is a modern evolution of grounded theory tailored to contexts where human and technological systems interact dynamically, such as Human–Robot Interaction. Hoda argued that studying technology in isolation from its social environment produces incomplete understanding; the technical and human components must be analysed as an integrated system. STGT provides structured tools such as open coding, constant comparison, and memoing that guide the researcher from descriptive data toward conceptual models explaining socio-technical behaviour. The method has been successfully applied in fields

ranging from software engineering to robotics, yielding theoretical models that connect human experience with technical design. For HRI research, STGT ensures that analysis captures both subjective perception and objective system properties, aligning with the dual cognitive–technical nature of robot intent communication.

Hoda also emphasised that qualitative rigor in HRI requires clear documentation, reflexivity, and transparency to transform observations into transferable theoretical insights. This aligns directly with the present study's objective of constructing a grounded understanding of how humans infer intent from gripper motion. The approach goes beyond description: it seeks to model the underlying processes by which humans translate kinetic cues into cognitive interpretations of cooperation, caution, or purpose. Through STGT's systematic coding and constant comparison, such interpretations can be formalised into categories and relationships that reveal the social logic embedded in mechanical motion.

Collectively, these methodological perspectives such as Braun and Clarke's thematic analysis, Guest et al.'s empirical saturation model, Adams et al.'s grounded theory methodology, and Hoda's socio-technical extension form a comprehensive epistemological backbone for this project. They converge on three essential principles: first, meaning in human–robot interaction must be interpreted rather than measured; second, rigorous qualitative analysis ensures this interpretation is systematic and defensible; and third, socio-technical awareness ensures that neither human perception nor robot behaviour is examined in isolation.

For the present study, this methodological synthesis legitimises the use of qualitative data to examine how humans derive understanding from robot motion. Each motion observed is treated as a communicative act, and each interpretation as data revealing cognitive and emotional structures underlying collaboration. By adhering to these established frameworks, the research achieves both analytical depth and methodological transparency ensuring that its conclusions on motion legibility, trust, and interpretability rest on solid scientific foundations.

6 Methodology

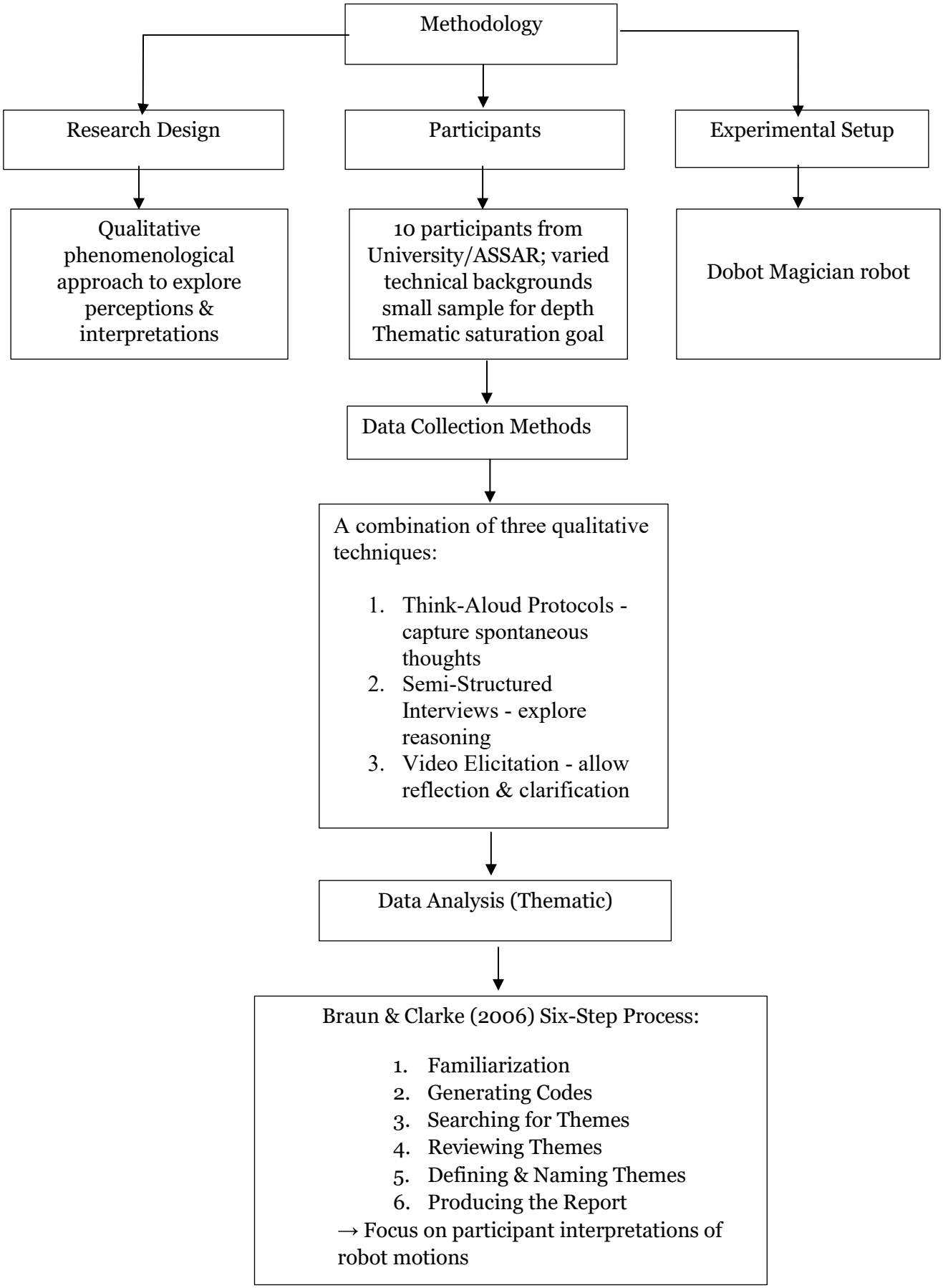

Figure 2 Methodological flow diagram illustrating the structure of the study from research design and data collection to thematic analysis.

Figure 2 provides an overview of the methodological structure of this study. It illustrates how the qualitative research design, participant selection, experimental setup, data collection methods, and thematic analysis are organised and connected. The figure shows the methodology chapter, guiding through the sequential steps of the study from research design to data analysis, which are described in detail in the following sections.

6.1 Research Design

This study adopts a qualitative, phenomenological research design to examine how humans perceive and interpret robot motion during collaborative interaction. Rather than evaluating motion effectiveness through quantitative performance metrics, the study focuses on participants lived experiences and meaning-making processes when observing robot behaviour.

A qualitative approach is therefore appropriate, as it enables exploration of how motion characteristics such as direction, speed, repetition, and task context shape human interpretation. The methodological rationale for prioritising interpretation over performance metrics is established in earlier sections of the report and is operationalised here through think-aloud protocols, semi-structured interviews, and video elicitation, followed by thematic analysis.

6.2 Participants

Ten participants were chosen from the university and ASSAR Innovation Arena for this study. Every participant was an adult with a range of professional and educational backgrounds. While some had minimal robotics experience, others possessed technical expertise in automation or engineering. Crucially, none of the participants were specialists in robotics. This combination was deliberate because it made it possible for the study to record a greater variety of viewpoints regarding how people interpret robot motions. In real workplaces, industrial robots are not only used by engineers but also by operators, technicians, and other staff. It is therefore more realistic to have participants with varying degrees of familiarity.

The number of participants in the study was deliberately kept small, which is common and appropriate in qualitative research. Qualitative research is more focused on the depth of understanding gained from each individual than quantitative research, which strives for large samples to yield statistically reliable results. A smaller sample size enables the researcher to collect rich, in-depth accounts of experiences and interpretations and to devote more time to thoroughly examining each participant's answers.

The sample size in this study was selected with the intention of achieving thematic saturation (Guest et al., 2006). This is the point in the data where similar themes start to repeat and new or distinct insights are no longer offered by new participants. When saturation is reached, it indicates that the researcher has gathered enough information to identify the key trends in how people

interpret the robot's movements. As a result, even though the sample size may seem small in comparison to quantitative research, it is adequate and suitable for this qualitative study.

6.3 Experimental Setup

A Dobot Magician robotic arm, a compact and adaptable tabletop robot frequently used for research and teaching, was used to conduct the experiment. Participants could easily watch the robot's actions from a comfortable distance because it was positioned on a level workspace. This configuration was selected because it offered a controlled setting free from equipment distractions so that the robot's movements could be seen clearly.

Figure 3 Dobot Magician DT-MG-4R005-02E.

6.3.1 Robot Control Logic and Motion intent Programming

The robotic experiments were executed using a DOBOT Magician (EAN 6150409056079), programmed through the manufacturer's Python-based API interface (dType library). The control program was developed to standardize the execution of all motion-intent trials while ensuring precise repeatability across participants.

At the beginning of each run, the robot was initialized by clearing all previous commands, defining a home position, and activating the gripper as the end effector. Two primary coordinate zones were defined: a pick-up position and a placement position, representing the operational area of the task. These coordinates were kept constant to maintain consistency between trials.

The program was structured around five predefined Motion intents, each representing a distinct communicative motion pattern:

Intent 1 – Gripper Open and Close: Simulated grasping and releasing actions to represent object interaction. This served as the most intuitive reference motion for participants.

Intent 2 – Gripper Rotation: Introduced angular changes about the tool axis to mimic alignment or reorientation behaviour, testing whether rotational movement alone conveyed purpose or caused confusion.

Intent 3 – Forward–Backward Movement: Represented approach and withdrawal behaviour, exploring how directional translation was perceived in terms of readiness or task completion.

Intent 4 – Up–Down Movement: Simulated lifting and placing actions, commonly associated with vertical manipulation or object handling.

Intent 5 – Left–Right Movement: Produced lateral sweeps across the workspace, used to investigate how ambiguous side motions were interpreted by observers.

Each motion sequence began and ended at the robot's base (neutral) position to eliminate carry-over effects between motions. A short delay was introduced between commands to allow clear visual distinction for participants. Every intent also included a brief pick-and-place cycle to preserve contextual relevance while observing motion cues.

The robot returned to its home position and stopped all motion after executing each complete intent sequence. This structure ensured that all participants observed identical motion trajectories, controlled timing, and a uniform presentation of robot behaviour throughout the study.

This experiment was intended to concentrate solely on motion cues, in contrast to some earlier research that integrated robot movements with light signals, sounds, or displays. This choice was made with the intention of separating movement's function as a means of communication from other signalling techniques. By doing this, the study sought to gain a better understanding of the amount of meaning that participants could extract from the robot's actions alone. Additionally, by using a tabletop, the experiment was made to resemble a shared workspace situation in which a human and a robot might need to coordinate tasks in close proximity. This arrangement is representative of real-world collaborative settings, like assembly lines or lab benches, where effective and safe teamwork depends on clear communication through motion.

6.3.2 Experimental Procedure

Each participant observed all five motion categories in a controlled laboratory setting. For each motion, the robot executed a pre-programmed trajectory designed to demonstrate a distinct action, such as lifting, rotating, or moving

laterally. This standardized programming ensured that all participants were exposed to identical motion sequences, allowing consistent observation conditions across trials.

After observing each motion, participants completed a structured response sheet that captured both quantitative and qualitative feedback. The first section included eight Likert-scale items assessing key perceptual dimensions clarity, confidence, possibility of misinterpretation, intuitiveness, ease of following, distinctness, and perceived exaggeration. These quantitative ratings provided an overview of how participants evaluated the interpretability and quality of each robot motion.

The second section of the response sheet contained two open-ended questions inviting participants to suggest possible improvements and share general reflections on the observed motions. In addition to written responses, participants were encouraged to provide verbal comments throughout the observation phase following a think-aloud approach. These spontaneous statements often revealed participants' initial impressions, expectations, or uncertainties regarding the robot's intent. After completing the response sheet, participants also provided written reflections in a review section to elaborate on their understanding or confusion regarding specific motions.

All sessions were video-recorded for contextual verification, allowing the researcher to revisit participants' behaviors, verbal remarks, and timing of reactions during analysis. This structured procedure ensured that every participant engaged with the same motion set while still allowing spontaneous qualitative responses. By combining quantitative ratings with open-ended reflections and live commentary, the study captured both immediate perceptual reactions and deeper cognitive interpretations of robot intent.

6.4 Data Collection Methods

This study combined three qualitative techniques to obtain significant insights into how participants understood the robot's motions. Each approach was selected because it records a distinct kind of response, and when combined, they offered a more comprehensive view of the experiences of the participants.

6.4.1 Think-Aloud Protocols

Participants were instructed to "think aloud" while they observed the robot executing each motion during the observation phase. They were therefore encouraged to speak aloud whatever was on their minds at the time, whether it was an explanation of what the robot appeared to be doing, an estimate of its purpose, or even a sense of perplexity. This approach was designed to record participants' initial thoughts and intuitive interpretations before they had a chance to think things through or overthink them. Because they offer direct access to participants' thought processes as they develop in real time, think-aloud protocols are helpful in this type of research.

6.4.2 Semi-Structured Interviews

Participants were asked to participate in a semi-structured interview after finishing the observation and think-aloud phase. Semi-structured interviews employ open-ended questions that encourage participants to explain their experiences in their own words, in contrast to rigid questionnaires. Detailed questionnaire is provided in Appendix 1**.** In addition to using a flexible guide, the interviewer asked follow-up questions based on the participant's responses. Participants were able to share personal viewpoints or analogies, think more deeply about what they had witnessed, and explain the logic behind their interpretations. For instance, some participants associated robot motions with machine-like behaviours, while others compared them to human gestures. These reflections provided context and depth that would not have been apparent during the think-aloud phase alone.

6.4.3 Video Elicitation

Participants watched brief video recordings of the robot repeating the same actions during the last phase. This provided them with another chance to express their opinions, frequently after they had had more time to consider their first impressions. Because people occasionally notice new details on a second viewing or may alter their interpretation when watching the motion in a different context, video elicitation is particularly useful. Additionally, it enabled participants to elaborate on their previous remarks, guaranteeing that their opinions were appropriately recorded.

These three techniques think-aloud, interviews, and video elicitation were combined in the study to gather data that represented both quick, reflexive responses and more deliberate, introspective ones. This multi-layered approach was crucial because comprehending robot intent signals involves not only what people see at the time but also how they interpret those signals later on. When combined, these techniques produced a rich dataset for analysis, enabling the researcher to record participant opinions as well as the reasons behind them. The full data collection results are shown in Appendix 3. As shown in Table 1, participant evaluations were collected using a structured framework combining Likert-scale ratings with binary judgments and open-ended responses for each motion intent.

Table 1 Model for Participant evaluation framework used to assess robot Motion intents.

Motion Type	Clarity (Q2)	Confidence (Q3)	Misinterpretation Risk (Q4)	Intuitiveness (Q5)	Ease of Following (Q6)	Distinctness (Q7)	Exaggeration (Q8)	Improvement Suggestions (Q9–10)	Comments
Forward–Backward	Binary (Yes/No)	5-point Likert scale	Binary (Yes/No)	Binary (Yes/No)	5-point Likert scale	5-point Likert scale	Binary (Yes/No)	Open-ended	Open-ended
Left–Right	Binary (Yes/No)	5-point Likert scale	Binary (Yes/No)	Binary (Yes/No)	5-point Likert scale	5-point Likert scale	Binary (Yes/No)	Open-ended	Open-ended
Up–Down	Binary (Yes/No)	5-point Likert scale	Binary (Yes/No)	Binary (Yes/No)	5-point Likert scale	5-point Likert scale	Binary (Yes/No)	Open-ended	Open-ended
Rotation	Binary (Yes/No)	5-point Likert scale	Binary (Yes/No)	Binary (Yes/No)	5-point Likert scale	5-point Likert scale	Binary (Yes/No)	Open-ended	Open-ended
Gripper (Open–Close)	Binary (Yes/No)	5-point Likert scale	Binary (Yes/No)	Binary (Yes/No)	5-point Likert scale	5-point Likert scale	Binary (Yes/No)	Open-ended	Open-ended

6.5 Data Analysis

All qualitative data, including think-aloud transcripts, interview transcripts, and video-elicitation notes, were analysed using Braun and Clarke's six-step thematic analysis framework.

6.5.1 NVivo for Qualitative Analysis

NVivo was used as a supporting tool to organise, manage, and visualise the qualitative dataset generated from think-aloud protocols, semi-structured interviews, and video-elicitation sessions. The software assisted the researcher during the initial familiarisation phase by enabling systematic importing of transcripts, attaching memos, and structuring data into cases based on motion type and method. NVivo's auto-coding and word-frequency functions were used only as exploratory aids to identify prominent concepts within the raw dataset. A word cloud was generated to visualise frequently occurring terms, offering an initial indication of which motion types triggered descriptive expressions such as "rotation", "confusing", "forward", "open", and "placing".

However, in line with Braun and Clarke's (2006) guidelines, all substantive analysis, theme development, and interpretation were carried out manually by the researcher. The NVivo visualisations did not determine the themes; they merely supported early data familiarisation and organisational clarity. The thematic structure was grounded in iterative coding, constant comparison, and analytical memoing rather than automated outputs.

6.5.2 Coding Process

The coding framework, categories, and code–theme relationships were developed through iterative engagement with the data following Braun and Clarke's six-phase thematic analysis process.

Step 1: Familiarization

Gaining a thorough understanding of the data was the first step. Every transcript was read and reread multiple times, and when needed, portions of the recordings were played back. In doing so, the researcher was able to gain a general understanding of the subject matter and start identifying intriguing details, such as moments of surprise, agreement, or confusion.

Step 2: Generating Codes

The data was then methodically coded. Coding is the process of underlining significant passages of text and assigning brief labels that encapsulate the main ideas. A participant's statement, "The sideways motion looked like it was waving at me," for instance, might be classified as "lateral motion interpreted as waving." In a similar way, another participant's statement, "When the gripper opened, I knew it wanted to grab something," could be classified as "clear grasping interpretation." At this point, capturing as many intriguing features as possible was more important than grouping the data.

Step 3: Searching for Themes

Following their creation, the original codes were grouped into more general categories known as themes. Themes are patterns found throughout the dataset that stand for significant concepts. For instance, a theme known as "Clarity of hand-like actions" was created by combining a number of codes regarding grasping and releasing motions being obvious to participants. Conversely, codes pertaining to confusing side-to-side sweeps and rotations were categorized under the theme "Ambiguity in non-standard motions."

Step 4: Reviewing Themes

The themes were thoroughly examined in this step to ensure they aligned with the data. If a theme covered too much ground, it was divided into smaller sub-themes, while other themes were combined if they overlapped. For instance, a sub-theme for "Rotation without context" was developed within the theme of "Ambiguity in non-standard motions" because various participants frequently expressed this kind of confusion.

Step 5: Defining and Naming Themes

After that, each theme was given a precise definition and a meaningful name. For instance:

- Clarity of hand-like actions: This refers to actions that participants could easily understand, like opening, closing, and lifting.
- Ambiguity in non-standard motions, such as rotations or lateral sweeps, which participants interpreted differently or found confusing.
- Contextual influence: describing how participants' interpretations of the same motion were altered by the presence or absence of objects.

This method of naming themes made the analysis easier to understand and more organized.

Step 6: Producing the Report

The themes were documented as part of the findings. Here, direct quotes from the participants were used to highlight the results and give the participants a voice. For example, the passage "It looked like the robot was just spinning for no reason" clarifies for the reader why rotation was frequently viewed as unclear. In addition to enhancing the analysis's credibility, using quotes makes the findings more relatable and clearer.

6.5.3 Codebook Extract

The Figure 4 below shows an excerpt of the final codebook used during the analysis. Each code includes a name, an operational definition, triggers, and representative quotes. This codebook guided the consistent application of codes across think-aloud transcripts, interview data, and video-elicitation responses. Detailed codebook is provided in Appendix 2.

Robot Motion Intent Study

Codes

Name	Description	Files	References
Theme 2 Ambiguity in Non-Standard Motions		1	21
Left - Right		1	11
Confusion or ambiguity	"This sweep is confusing." "It doesn't convey a clear goal even after rewatching."	1	2
Searching or Scanning	"Why is it moving side-to-side? Looks like it's searching." "It looks like scanning or exploring." "The side-to-side sweep was confusing — it looked like it was searching or maybe cleaning."	1	9
Rotation		1	10
Alignment	"Slow rotation feels like alignment." "When the gripper rotated slowly, I thought it was aligning with something." "Slow rotation clearly aligns with the object on replay."	1	8
Confusion or ambiguity	"The rotation just looked like spinning for no reason." "Rotation seems playful or random." "Intent remains unclear."	1	2
Theme 3 Influence of Motion Parameters		1	40

23 Nov 2025 Page 1 of 4

Figure 4 Final Codebook Used for Thematic Analysis from NVivo.

Below are several raw extracts illustrating how participant statements were coded. These examples illustrate the analytical development from participant language to initial codes and subsequently to higher-order themes. Including these extracts provides the traceability required for qualitative rigor.

Table 2 Example of Coded Data Segments.

Raw Data Extract (Participant Quote)	Assigned Code
"When the gripper opened, I immediately thought it wanted to grab something."	Clear grasping interpretation
"The rotation just looked like spinning for no reason."	Unclear rotation or purposeless motion
"The up-and-down motion was too quick, so I couldn't tell if it was lifting or dropping."	Confusion due to motion speed
"When it paused before moving forward, I could guess what it was about to do."	Pause-as-intention
"The sideways movement felt like it was gesturing to me, maybe waving."	Anthropomorphic gesture interpretation

The progression from codes to themes followed a consistent clustering of interpretive patterns across the dataset. Codes relating to grasping, holding, lifting, and placing naturally grouped together, forming Theme 1 (Clarity in Hand-Like Actions) because participants interpreted these motions using familiar human motor experiences. In contrast, codes describing confusion, purposeless spinning, or unclear lateral sweeps converged into Theme 2 (Ambiguity in Non-Standard Motions), reflecting how unfamiliar or context-free movements created uncertainty. Codes that referenced variations in speed, pauses, timing, or smoothness aligned under Theme 3 (Influence of Motion Parameters), demonstrating that participants' interpretations were strongly shaped by the robot's motion dynamics. Meanwhile, statements that indicated the importance of task context, object presence, or reset-like behaviour informed Theme 4 (Context-Based Interpretation), showing that identical motions were understood differently depending on situational cues. Finally, codes capturing perceptions of waving, signalling, or human-like expression formed Theme 5 (Anthropomorphic Interpretations), illustrating how participants often attributed social meaning to certain motion patterns.

7 Result and Findings

7.1 Overview of Coding Output

This section summarises the coding output generated during the analysis. The coding was conducted across all three data sources such as think-aloud transcripts, semi-structured interviews, and video-elicitation reflections resulting in a rich set of interpretive codes. The visualisation below displays the

distribution of coding references across all participants, illustrating how frequently each thematic category was triggered throughout the dataset. Presenting this coding pattern provides explicit evidence of the analytical process, addressing both dependability and traceability within the thematic analysis.

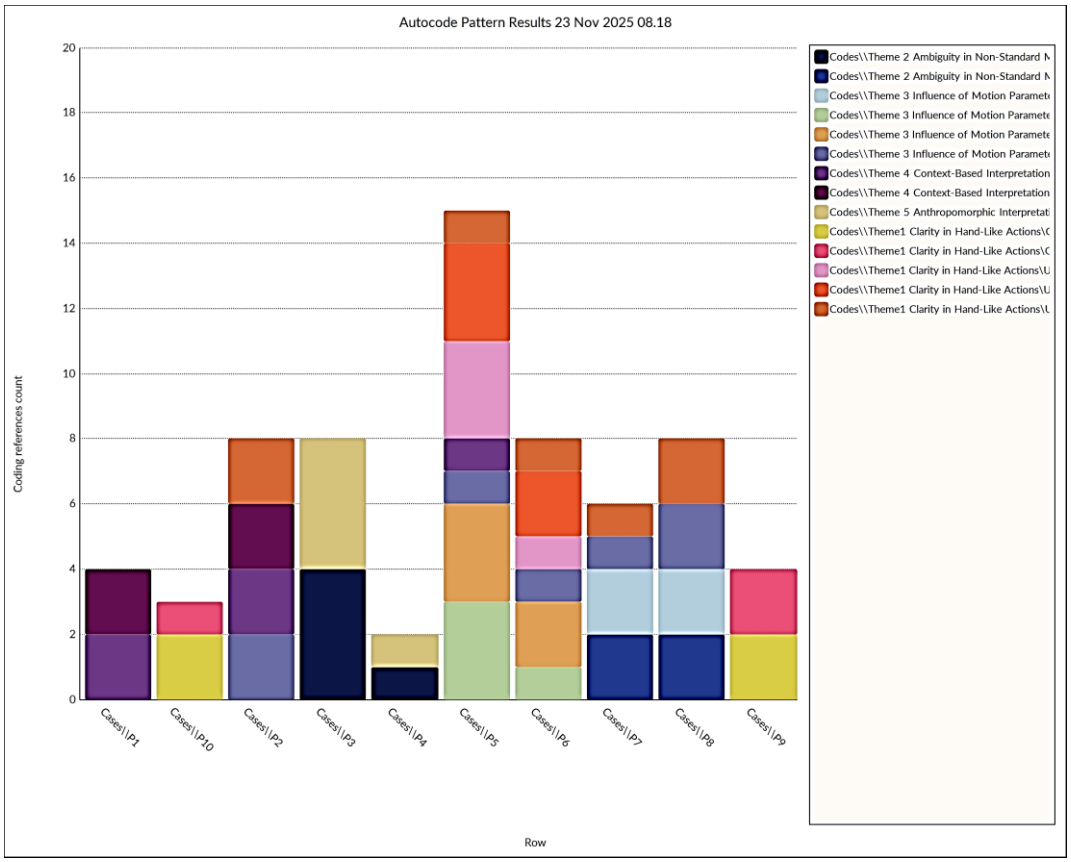

Figure 5 Distribution of coding references across participants generated during NVivo coding.

7.2 Qualitative Evaluation of Motion intents

To ensure that the analysis reflected the full breadth of participant interpretations, data were collected through three complementary qualitative techniques: think-aloud protocols, semi-structured interviews, and video-elicitation. Each method captured a different layer of participant reasoning spontaneous reactions, reflective explanations, and second-viewing reinterpretations. To demonstrate triangulation, the coding output was examined separately for each method. The method-wise coding figures presented below show how frequently each code appeared within each dataset, confirming that the final themes were not dependent on a single source. Instead, they emerged consistently across real-time commentary, follow-up discussions, and reflective video-based observations. This multi-method consistency strengthens the validity of the findings, demonstrating that the thematic structure is grounded in diverse forms of participant expression rather than isolated responses.

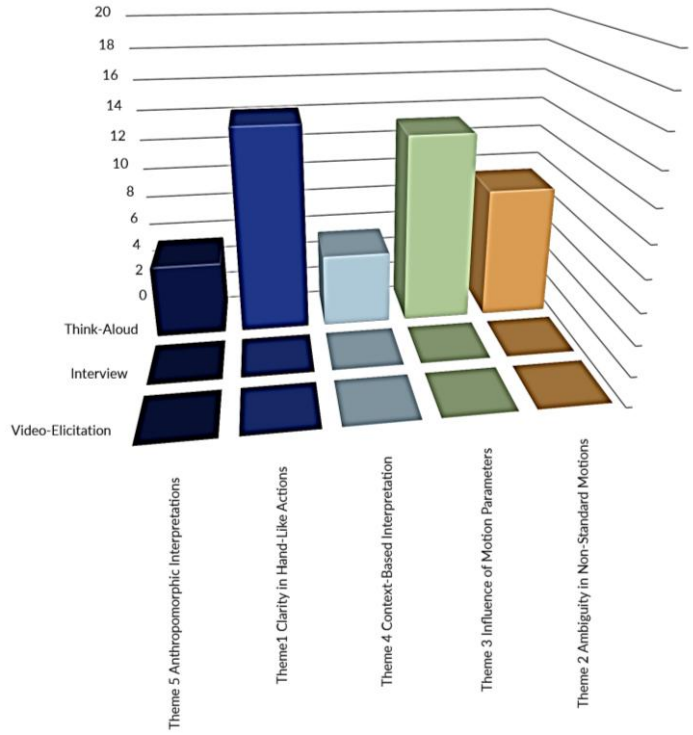

Figure 6 Coding frequency for Method 1 Think aloud from NVivo.

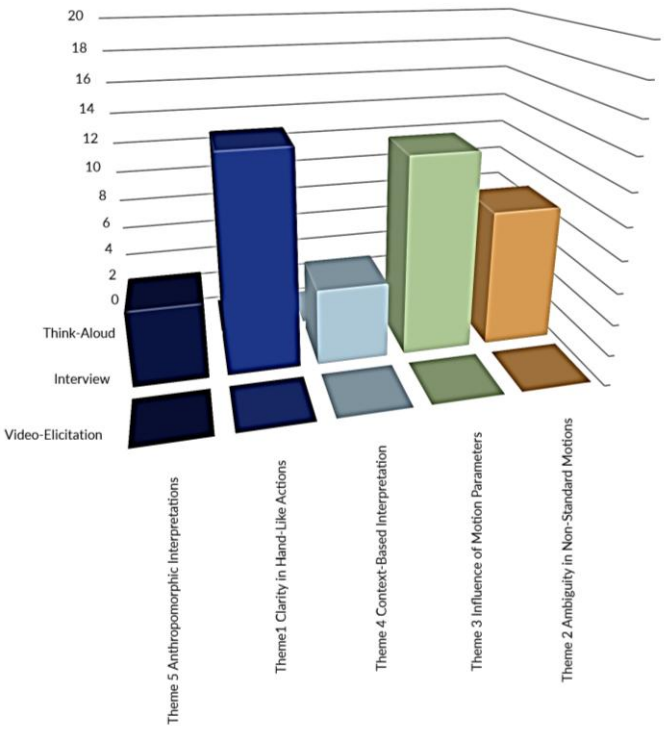

Figure 7 Coding frequency for Method 2 Semi-Structured Interview from NVivo.

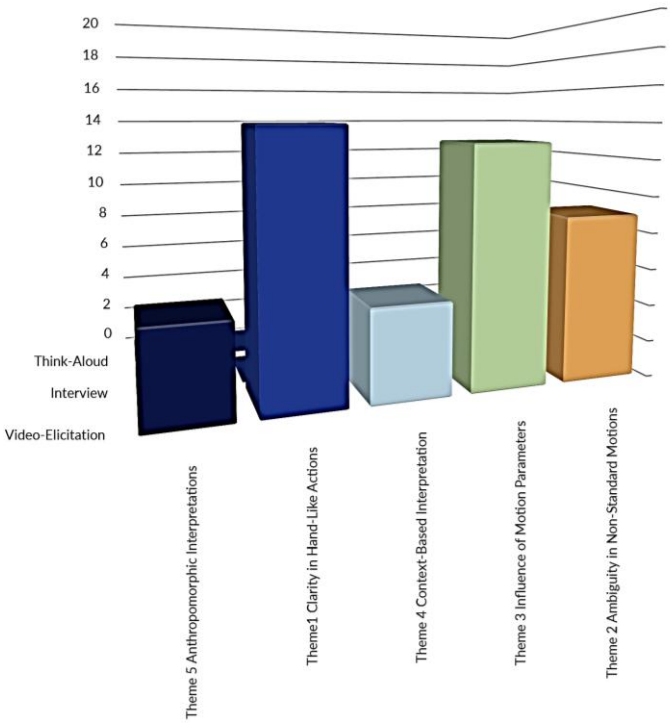

Figure 8 Coding frequency for Method 3 Video-Elicitation from NVivo.

To evaluate the clarity and interpretability of each robot motion, participant feedback was collected through handwritten questionnaires after each observation session. Detailed questionnaire is provided in Appendix 1**.** Each response sheet related to a certain motion type such as gripper open or close, up-down, rotation, forward-backward, and left-right. It comprised both scaled ratings and open-ended comments regarding clarity, confidence, potential misinterpretation, and recommendations for enhancement.

Across all motions, participants consistently identified the gripper open or close motion as the most intuitive, often rating their confidence between 4 and 5 on the five-point scale. The up–down motion was also widely understood, commonly interpreted as lifting or placing an object. In contrast, rotational and lateral left–right motions produced uncertainty and frequently required explanation or demonstration. Several participants described these motions as "confusing at first but understandable after demonstration," indicating a lower level of inherent legibility.

Participants also offered practical suggestions for improving clarity. Many recommended slower movements speed, brief pauses before and after each action, and more distinctive starting positions to make the robot's intention easier to infer. A few participants suggested supplementing motion cues with sound or visual signals, though most expressed a preference for maintaining motion as the primary communication channel.

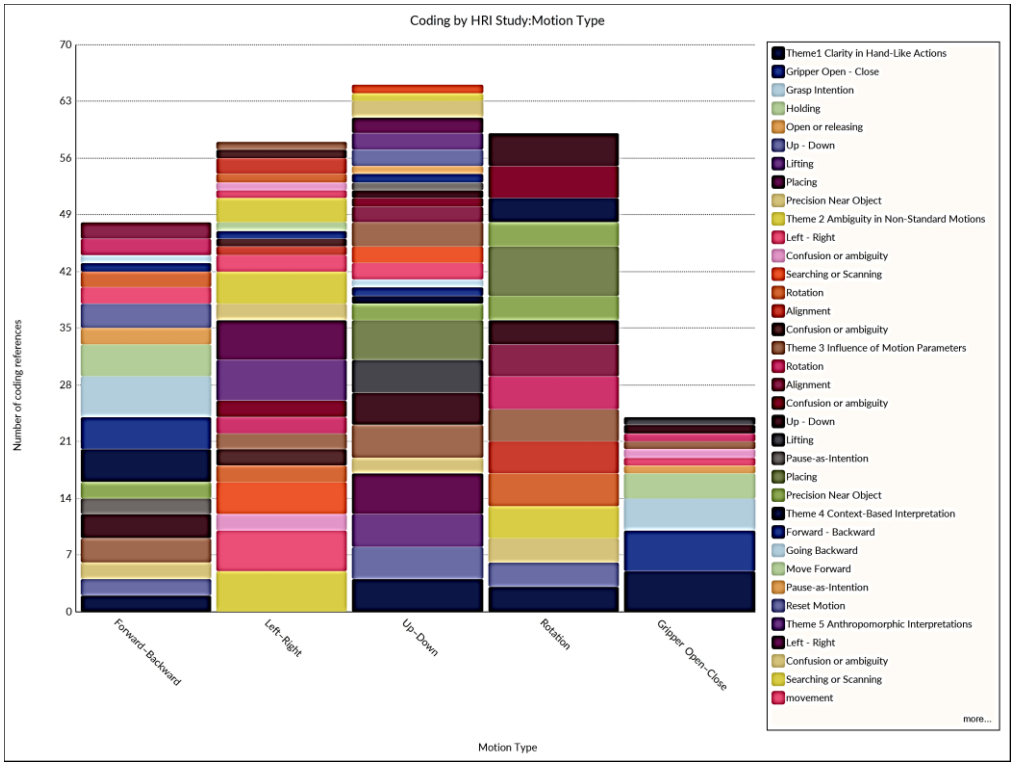

Figure 9 Consolidated coding output categorised by motion type.

The thematic analysis of the participant responses revealed several key patterns in how people interpreted the robot's motion-based intent signals. Five major themes emerged from the coded data, each of which reflected a distinct facet of how people interpret, evaluate, and comprehend robot motions. These themes demonstrate how participants' interpretations were impacted by the robot's actions as well as its movements, the action's setting, and their own human-like perceptions.

Table 3 Mean Participant Scores for Different Robot Motion Types Across Evaluation Criteria.

Motion	Clarity	Confidence	Intuitiveness	Ease of Following	Distinctness
Forward–Backward	3.6	3.8	3.4	4.2	3.7
Left–Right	3.2	3.1	3.0	3.8	3.4
Up–Down	4.6	4.5	4.7	4.8	4.6
Rotation	2.8	2.3	2.7	3.9	2.9
Gripper (Open/Close)	4.8	4.7	4.9	4.9	4.8

7.2.1 Theme Development from Codes

To ensure transparency in the qualitative analysis and to demonstrate how the final themes were constructed, this subsection provides an audit trail linking the raw data to their corresponding codes and thematic categories. The analysis followed Braun and Clarke's six-phase framework, beginning with open coding

of all think-aloud statements, interview transcripts, and video-elicitation reflections. Codes were then compared, clustered, and refined through iterative review until higher-level themes emerged.

Table 4 below presents an excerpt from the full code–theme table, showing how participant statements were coded and how these codes were grouped into sub-themes and final themes. This evidence illustrates the analytic progression: individual meaning units were first captured through descriptive codes, similar codes were merged or collapsed based on conceptual similarity, and these clusters formed the basis for each theme. For example, several statements referring to grasping, holding, lifting, or placing were consistently interpreted using human-hand analogies, and these were therefore grouped under the theme Clarity in Hand-Like Actions. Conversely, statements describing uncertainty, purposeless spinning, or unclear lateral sweeps consistently indicated confusion, forming the basis of Ambiguity in Non-Standard Motions. Codes referencing motion speed, pauses, and smoothness converged into Influence of Motion Parameters, whereas codes tied to task context or reset behaviour informed Context-Based Interpretation. Finally, a smaller set of codes reflecting wave-like gestures or social projections developed into Anthropomorphic Interpretations.

Table 4 Excerpt from Code–Theme Development Table.

Raw Participant Quote	Code	Sub-Theme	Theme
"When the gripper opened, I immediately thought it wanted to grab something."	Clear grasping interpretation	Opening/closing as natural signals	Theme 1: Clarity in Hand-Like Actions
"The rotation just looked like spinning for no reason."	Unclear purposeless rotation	Rotation without context	Theme 2: Ambiguity in Non-Standard Motions
"The up-and-down motion was too quick, so I couldn't tell if it was lifting or dropping."	Confusion due to motion speed	Speed affects clarity	Theme 3: Influence of Motion Parameters
"When it paused before moving forward, I could guess what it was about to do."	Pause as intention	Pauses increase predictability	Theme 4: Context-Based Interpretation
"The sideways movement felt like it was gesturing to me, maybe waving."	Gesture-like sideways motion	Side sweeps as social cues	Theme 5: Anthropomorphic Interpretations

7.3 Theme 1: Clarity in Hand-Like Actions

The easiest actions for participants to interpret were those that resembled human hand gestures, such as opening, closing, lifting, and placing. These were characterized by many as natural and instantly comprehensible. One participant stated, for example, "When the gripper opened, I immediately thought it wanted to grab something," while another clarified, "When it moved down, I thought it was placing something on the table."

These claims demonstrate that participants were able to relate the robot's motions to well-known human behaviours. Humans are able to predict the robot's intents with ease thanks to these hand-like motions, which offer a straightforward and natural form of communication. The robot appeared more understandable and comfortable to deal with because of the familiarity of these motions. Hand-like or naturalistic movements were the most effective for conveying intent because they align closely with human physical experience.

Sub-themes: Opening/closing as natural signals; Upward/downward motion linked to object handling.

7.4 Theme 2: Ambiguity in Non-Standard Motions

On the other hand, when non-standard motions like side-to-side sweeps and rotations were observed, participants frequently voiced perplexity. While some thought these actions were pointless or unnecessary, others recognized a purpose in them. For example, one participant noted, "The rotation just looked like spinning for no reason," while another said, "The side-to-side sweep was confusing—it looked like it was searching or maybe cleaning."

These results imply that individuals find it challenging to understand what a robot is doing or intends to achieve when it performs novel or context-free actions. Participants couldn't tell if the robot's motions were random or goal-directed in the absence of visual clues like an object to interact with. When the robot's next move is unclear, people may be reluctant to take action, which can reduce trust and coordination. The findings demonstrate that familiarity and the existence of contextual cues affect motion clarity.

Sub-themes: Side sweeps as searching or waving; Rotation without clear context.

7.5 Theme 3: Influence of Motion Parameters

Another significant discovery was that participants focused on the robot's movement, not just its direction or type of motion, but also its speed, pauses, and timing. Participants reported that their interpretation of intent was affected by variations in rhythm or tempo. For example, one participant reported, "I couldn't tell if it was lifting or dropping because the up-and-down motion was too quick," while another noted, "I could guess what it was about to do when it

paused before moving forward." As one participant observed, "When it slowed down near the object, it looked more careful and precise." Similarly, a slower motion close to an object was frequently perceived as cautious or precise.

The examples provided demonstrate how perceived meaning is shaped by motion factors. While slow movements can imply precision or compassion, a pause can build suspense. Conversely, rapid or jerky movements can appear unpredictable or mechanical. The study emphasizes that designers can make robot behaviour more understandable and human-friendly by utilizing time and speed as potent communicative cues.

Sub-themes: Speed affects clarity; Pauses increase predictability; Slower motion improves perceived precision.

7.6 Theme 4: Context-Based Interpretation

Participants also revealed that the task's context had a significant impact on their understanding of the robot's actions. Depending on the participant's perception of the robot's objective, the identical gesture was read differently. For example, forward motion could be interpreted as a social gesture aimed at the human or as the robot reaching for an object. "I felt like it was reaching toward me when it moved forward," claimed one participant, while "the backward motion seemed like it was done after completing a task, like pulling back," added another.

This demonstrates the situational nature of human understanding of robot mobility. To deduce meaning, people look to their surroundings and perceived purpose. Participants found it easier to guess the robot's goal when the context was obvious, such as when it was near an object. They frequently guessed based on past experiences or presumptions when context was lacking. This illustrates how motion interpretation is situational and not fixed.

Sub-themes: Forward motion as approach; Backward motion as completion or retreat.

7.7 Theme 5: Anthropomorphic Interpretations

A smaller but interesting pattern was the tendency for participants to attribute human-like qualities to certain robot motions. Side-to-side motions, for instance, were occasionally interpreted as friendly gestures or phrases. "The sideways motion felt like it was gesturing to me, maybe waving," said one participant.

This suggests that people frequently endow robot actions with social meaning, even when such meaning is not intended. If the gesture is not intended for communication, these anthropomorphic interpretations run the danger of causing misunderstandings, but they can also make conversations more relatable and interesting. The results demonstrate how easily humans may see

robots through a human perspective, particularly when the robot's movements seem rhythmic or expressive.

Sub-themes: Side sweeps as social cues; Perception of “friendly” or “expressive” robot behaviour.

The findings reveal a distinct pattern: non-standard or ambiguous actions (rotations, sweeps) were frequently confusing unless supported by context or timing, but hand-like, conventional motions (grasping, lifting, placing) were regularly well comprehended. In order to determine purpose, participants also mainly relied on motion dynamics including speed and pauses. Humans' perceptions of the robot's objectives and confidence were significantly impacted by these minute details of motion. The results show that human psychology and perception, in addition to mechanical design, influence robot intent communication. Robot designers should intentionally incorporate motion expressiveness such as speed control, pauses, and trajectory clarity into robot communication to facilitate collaboration. Comprehending these perceptual aspects can aid in the development of robots that move in ways that are understandable, predictable, and cozy for human collaborators.

The findings of this study align strongly with established theories on legible motion and human perception. According to (Dragan et al., 2013), legibility arises when motion allows observers to infer the robot’s goal as it unfolds, rather than only after completion. In this study, participants consistently identified hand-like actions such as grasping or lifting as legible because these motions closely matched human motor expectations. This supports Dragan’s legibility–predictability distinction, showing that small deviations toward expressive, human-like motion improve understanding even at the cost of path efficiency. Furthermore, (Kirschner et al., 2022) Expectable Motion Unit (EMU) concept, which defines safety as the fulfilment of human expectations, resonates with participants’ comfort when motions were smooth, predictable, and properly timed. Similarly, (Lemaignan et al., 2012) described grounded cognition as the process of connecting perception and meaning within shared context; the contextual interpretations observed here where identical motions were understood differently depending on surrounding objects extend that theory into motion-based communication. Together, these connections confirm that motion legibility is not only mechanical but also cognitive and social, bridging the gap between robotic control and human perception.

8 Discussion

This study investigated how motion design influences human comprehension, predictability, and comfort, and how participants interpret various robot motions during collaborative tasks. The results supported by the codebook audit trail Table 4 and the method-wise triangulation patterns Figure 6, Figure 7 and Figure 8 show that participants consistently relied on human motor analogies when interpreting the robot’s actions. Hand-like movements such as opening, closing, lifting, and placing produced the clearest and most confident

interpretations. These motions collectively contributed the highest number of clarity-related codes in the motion-type distribution Figure 9, demonstrating that participants immediately recognised these actions because they mirror familiar human behaviours.

By contrast, non-standard motions that are particularly side-to-side sweeps and rotations were generated as the greatest interpretive uncertainty. Theme 2 (Ambiguity in Non-Standard Motions), which accounted for a substantial proportion of ambiguity-related codes in the consolidated coding map Figure 9, reflects this pattern. Statements such as "The rotation just looked like spinning for no reason" and "The sideways sweep looked like searching or maybe cleaning" (see Table 4) illustrate how participants struggled to assign purpose when contextual cues were absent. These patterns were consistently present across the three qualitative methods, confirming that ambiguity was not a method-specific artefact but a stable interpretive issue.

Participants also highlighted the importance of motion parameters such as timing, pauses, and speed. Theme 3 (Influence of Motion Parameters) was strongly supported by think-aloud data, where participants frequently commented on the robot's motion rhythm. In the codebook Table 4, multiple extracts emphasise that slower approaches, momentary pauses, or deceleration near objects increased clarity, whereas fast or abrupt motions appeared unplanned. These observations align with the motion-type distribution, where speed-related codes were heavily associated with up–down and forward–backward behaviours. Figure 9, indicating that temporal dynamics shaped perceived intention as much as motion direction.

The findings also confirm the major role of context in shaping interpretation. Theme 4 (Context-Based Interpretation) emerged from numerous statements coded as "approach," "retreat," or "reset," particularly within the forward–backward motion category. As shown in Figure 9, these motions produced dense clusters of context-dependent codes. Participants repeatedly indicated that identical motions were easier to understand when an object was present, consistent with grounded cognition theories. The codebook excerpts demonstrate this link clearly, with participants expressing that they needed something in the environment to "anchor" the meaning of the motion.

A smaller but meaningful pattern involved anthropomorphic interpretations. Theme 5 (Anthropomorphic Interpretations) received fewer but consistent codes across datasets Figure 6, Figure 7 and Figure 8. Participants interpreted certain lateral motions as "waving" or "gesturing," highlighting a human tendency to project social meaning onto expressive movement patterns. This reinforces existing HRI research suggesting that human's default to social reasoning when motion lacks clear task structure.

These cross-referenced insights demonstrate that the themes in this study are grounded in systematically coded evidence rather than narrative description. The integration of the codebook excerpt Table 4, the triangulation figures Figure 6, Figure 7 and Figure 8, and the motion-type coding distribution Figure 9

confirms that motion-only communication can be legible, interpretable, and reliable which provided the movements align with human expectations of hand-like action, contextual relevance, and temporal expressiveness. By isolating motion intent as the sole communication channel, this study offers human-centred design principles for collaborative robots that communicate naturally and intuitively, without reliance on lights, sounds, or external signalling devices.

9 Sustainability

Sustainability has become a central focus in modern research and industrial development, aiming to balance technological progress with environmental protection, social well-being, and long-term economic stability. The idea focuses on providing for current needs without sacrificing the capacity of future generations to provide for their own. The Sustainable Development Goals (SDGs), a framework of 17 interrelated objectives that support equitable growth, innovation, and responsible resource use, were introduced by the UN to direct global initiatives. Sustainability in the context of industrial automation and human-robot cooperation entails creating systems that minimize environmental impact while being safe, effective, and supportive of human workers. Robotics research that incorporates sustainability guarantees that technological developments will contribute to a robust, fair, and ecologically sensitive industrial future in addition to increasing productivity.

The social, economic, and technological facets of sustainable industrial development are all enhanced by this research. It promotes the development of safer, more effective, and human-centered work environments fundamental principles of sustainable manufacturing by improving the readability and interpretability of robot actions in HRC. This study covers a number of important goals in line with the Sustainable Development Goals (SDGs) of the United Nations. By encouraging safer and more ergonomic human-robot collaboration, it helps achieve SDG 8 (Decent Work and Economic Growth), which lessens physical strain and increases worker confidence, job satisfaction, and productivity. By promoting innovation in intelligent manufacturing systems through motion-based intent signalling, it supports SDG 9 (Industry, Innovation, and Infrastructure) and aids in the development of flexible, dependable, and secure collaborative robots for upcoming industries. By enabling more effective workflows that decrease production errors, reduce material waste, and optimize resource utilization, the research also supports ecologically responsible manufacturing practices, which is in line with SDG 12 (Responsible Consumption and Production). Additionally, by lowering worker stress, hesitancy, and exhaustion in industrial settings through predictable and transparent robot behaviour, it indirectly supports SDG 3 (Good Health and Well-Being), resulting in safer and healthier workplaces. All things considered, the study enhances sustainability through better human-machine cooperation promoting environmental efficiency by cutting waste and errors, social sustainability by improving safety and inclusivity, and economic sustainability by spurring innovation in dependable, affordable, and adaptable automation systems.

10 Limitations

A number of limitations should be noted, despite the fact that this study offers valuable insights into how people interpret robot motion intent. Although the small sample size is appropriate for a qualitative study with an emphasis on depth, it restricts how broadly the results may be applied. Broader viewpoints and greater confidence in the themes found could be obtained with a larger and more varied sample. Additionally, the Dobot Magician was the only robot model used in the study, and it only looked at motion-based cues rather than additional intent communication techniques like lights or sounds. Because of this, the results may not accurately reflect the more advanced multimodal systems found in actual collaborative robots and instead reflect motion-only interpretations. Another limitation is that the experimental setup was carried out in a controlled tabletop setting, which does not accurately reflect the loud and dynamic characteristics of actual industrial settings where robot intent can be interpreted differently. In contrast to skilled operators or technicians, the majority of participants had little prior experience with robots, which might have affected their interpretations. Furthermore, the range of motions studied was limited to simple gripper operations such as opening, shutting, lifting, rotating, and lateral movements, rather than more complicated task-oriented behaviours. Notwithstanding these drawbacks, the study's objective of examining human perception and interpretation of motion-based robot signals was effectively achieved. In order to improve our understanding of human–robot collaboration, these limitations also point to opportunities for future research, such as examining expert users, incorporating more communication modalities, and testing in more realistic settings.

11 Conclusion

This chapter concludes by summarising the key findings derived from the qualitative analysis of participant interpretations of robot motion. Drawing on thematic patterns identified across think-aloud protocols, interviews, and video elicitation, the results highlight how humans make sense of robot actions in collaborative settings. Rather than focusing on performance outcomes, these conclusions reflect the underlying cognitive and perceptual mechanisms that shaped participants' understanding of motion intent.

- Human interpretation of robot motion is shaped by familiarity, context, and timing. Across all data sources, participants consistently relied on human-like reasoning to understand the robot's actions.

- Hand-like motions are naturally legible. Grasping, opening, closing, lifting, and placing produced the highest levels of clarity, confidence, and intuitive understanding.

- Non-standard motions require contextual support. Rotations and lateral sweeps frequently caused uncertainty unless objects or environmental cues were present.

- Motion parameters matter as much as the motion itself. Speed, pauses, smoothness, and timing strongly influenced predictability and perceived intention.

- People naturally project social meaning. Ambiguous or expressive motions triggered anthropomorphic interpretations such as “waving” or “gesturing.”

These findings demonstrate that robot motion is not interpreted solely as a physical trajectory, but as a communicative signal embedded within human cognitive and social frameworks. Participants consistently drew on prior experience, contextual cues, and familiar human actions when inferring intent, particularly in the absence of explicit signalling. These insights reinforce the importance of designing robot motions that are not only technically correct but also cognitively legible and context-aware, supporting more intuitive and trustworthy human–robot collaboration.

References

Abe, N., Hu, Y., Benallegue, M., Yamanobe, N., Venture, G., & Yoshida, E. (2024). Human Understanding and Perception of Unanticipated Robot Action in the Context of Physical Interaction. *ACM Transactions on Human-Robot Interaction*, *13*(1), 1–26. https://doi.org/https://dl.acm.org/doi/full/10.1145/3643458

Adams, A., Lunt, P., & Cairns, P. (2008). A qualitative approach to HCI research. *Research Methods for Human-Computer Interaction. Cambridge University Press*, 138–157. https://www.academia.edu/87062871/A_qualitative_approach_to_HCI_research

Bonarini, A. (2020). Communication in Human-Robot Interaction. *Current Robotics Reports 2020 1:4*, *1*(4), 279–285. https://doi.org/10.1007/S43154-020-00026-1

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706QP063OA

De Santis, A., Siciliano, B., De Luca, A., & Bicchi, A. (2008). An atlas of physical human–robot interaction. *Mechanism and Machine Theory*, *43*(3), 253–270. https://doi.org/10.1016/J.MECHMACHTHEORY.2007.03.003

Dragan, A. D., Lee, K. C. T., & Srinivasa, S. S. (2013). Legibility and predictability of robot motion. *ACM/IEEE International Conference on Human-Robot Interaction*, 301–308. https://doi.org/10.1109/HRI.2013.6483603

Gielniak, M. J., & Thomaz, A. L. (2012). Enhancing interaction through exaggerated motion synthesis. *HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 375–382. https://doi.org/10.1145/2157689.2157813

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough? *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hoda, R. (2025). Socio-Technical Qualitative Research in Human Robot Interaction. *ACM/IEEE International Conference on Human-Robot Interaction*, 2003–2005. https://doi.org/10.1109/HRI61500.2025.10973908

Kirschner, R. J., Mayer, H., Burr, L., Mansfeld, N., Abdolshah, S., & Haddadin, S. (2022). Expectable Motion Unit: Avoiding Hazards from Human Involuntary Motions in Human-Robot Interaction. *IEEE Robotics and Automation Letters*, *7*(2), 2993–3000. https://doi.org/10.1109/LRA.2022.3144535

Lemaignan, S., Ros, R., Akin Sisbot, E., Alami, R., & Beetz, M. (2012). Grounding the Interaction: Anchoring Situated Discourse in Everyday Human-Robot Interaction. *International Journal of Social Robotics*, *4*(2), 181–199. https://doi.org/10.1007/s12369-011-0123-x

Lemasurier, G., Bejerano, G., Albanese, V., Parrillo, J., Yanco, H. A., Amerson, N., Hetrick, R., & Phillips, E. (2021). Methods for Expressing Robot Intent for Human-Robot Collaboration in Shared Workspaces. *ACM Transactions on Human-Robot Interaction*, *10*(4). https://doi.org/10.1145/3472223

Matheson, E., Minto, R., Zampieri, E. G. G., Faccio, M., & Rosati, G. (2019). Human–Robot Collaboration in Manufacturing Applications: A Review. *Robotics 2019, Vol. 8, Page 100*, *8*(4), 100. https://doi.org/10.3390/ROBOTICS8040100

Pascher, M., Gruenefeld, U., Schneegass, S., & Gerken, J. (2023). How to Communicate Robot Motion Intent: A Scoping Review. *Conference on Human Factors in Computing Systems - Proceedings*, *1*(23). https://doi.org/10.1145/3544548.3580857

Ramachandran, A., Huang, C.-M., Gartland, E., Scassellati, B., & Scas, B. (2018). Thinking Aloud with a Tutoring Robot to Enhance Learning. *13th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Chicago, IL, USA*, 59–68. https://doi.org/10.1145/3171221.3171250

Szafir, D., Mutlu, B., & Fong, T. (2015). Communicating Directionality in Flying Robots. *ACM/IEEE International Conference on Human-Robot Interaction*, *2015-March*, 19–26. https://doi.org/10.1145/2696454.2696475

Villani, V., Pini, F., Leali, F., & Secchi, C. (2018). Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, *55*, 248–266. https://doi.org/10.1016/J.MECHATRONICS.2018.02.009

Zappa, I., Vignali, S., Zanchettin, A. M., & Rocco, P. (2026). Symbolic representation of objects relative poses for robotic manipulation tasks. *Engineering Applications of Artificial Intelligence*, *163*, 112882. https://doi.org/10.1016/J.ENGAPPAI.2025.112882

Appendix 1: Participant Questionnaire – Motion intent Perception Study

Section A: Intent Recognition and Clarity

1) Which intent was identified?
   (e.g., Gripper Open/Close, Rotation, up–Down, forward–Backward, Left–Right)

2) Did you clearly understand what the robot was trying to do?
   ☐ Yes ☐ No ☐ Not sure

3) How confident are you in interpreting the purpose of this motion?
   (1 = Not confident 2 3 4 5 = Very confident)

4) Do you think this motion could be misinterpreted?
   ☐ Yes ☐ No
   If Yes, please explain briefly:

5) Was the intent of this motion intuitive, or did it require prior explanation?
   ☐ Intuitive ☐ Required explanation

6) How easy was it to visually follow the robot's movements?
   (1 = Very hard 2 3 4 5 = Very easy)

7) Did the intent motion stand out clearly among the other motions?
   ☐ Yes ☐ No ☐ Somewhat

8) Did any part of the movement seem too subtle or too exaggerated?
   ☐ Yes ☐ No
   If Yes, which part?

Section B: Improvement Suggestions

1) What changes (if any) would make this motion more understandable?

2) Do you have any suggestions for improving robot–human interaction using motion?

Section C: Participant Review

Participants were encouraged to provide a short-written review describing their general impressions or observations, including:

1) How the robot moved during the demonstration.
2) Which motions felt clear, confusing, or exaggerated.
3) Any ideas for improving the robot's movement or communication.

Appendix 2: Codebook

Robot Motion Intent Study

Codes

Name	Description	Files	References
Theme 2 Ambiguity in Non-Standard Motions		1	21
Left - Right		1	11
Confusion or ambiguity	"This sweep is confusing." "It doesn't convey a clear goal even after rewatching."	1	2
Searching or Scanning	"Why is it moving side-to-side? Looks like it's searching." "It looks like scanning or exploring." "The side-to-side sweep was confusing — it looked like it was searching or maybe cleaning."	1	9
Rotation		1	10
Alignment	"Slow rotation feels like alignment." "When the gripper rotated slowly, I thought it was aligning with something." "Slow rotation clearly aligns with the object on replay."	1	8
Confusion or ambiguity	"The rotation just looked like spinning for no reason." "Rotation seems playful or random." "Intent remains unclear."	1	2
Theme 3 Influence of Motion Parameters		1	40

Name	Description	Files	References
Going Backward	"The backward return seems like a home-position reset." "Backward motion is definitely a return to home." "The backward motion seemed like it was done after completing a task."	1	10
Move Forward	"It moved forward… looks like it's reaching toward something or me." "Forward looks like reaching out."	1	9
Pause-as-Intention	"When it paused before moving forward, I could guess what it was about to do." "The pause before moving forward is clearer on playback."	1	2
Reset Motion	"The backward return seems like a home-position reset." "Backward motion is definitely a return to home." "The backward motion seemed like it was done after completing a task."	1	3
Theme 5 Anthropomorphic Interpretations		1	11
Left - Right		1	11
Confusion or ambiguity	"Rotation seems playful or random." "Slow open-close feels communicative."	1	2
Searching or Scanning	"This feels like a signal or gesture." "The sideways movement felt like it was gesturing to me, maybe waving." "On replay it still looks like a wave-like gesture."	1	9
Theme1 Clarity in Hand-Like Actions		1	45
Gripper Open - Close		1	17

Name	Description	Files	References
Grasp Intention	"The gripper opened — maybe it wants to grab." "When the gripper opened, I immediately thought it wanted to grab something." "Close means grabbing or holding."	1	9
Holding	"It closed like it's holding something." "When the gripper closed after touching the block, I assumed it was holding it."	1	7
Open or releasing	"The gripper opened — maybe it wants to grab." "When the gripper opened, I immediately thought it wanted to grab something." "Close means grabbing or holding."	1	1
Up - Down		1	28
Lifting	"It's moving up slowly — like lifting something." "Upwards feels like picking up."	1	8
Placing	"Now it's going down — maybe placing an item." "When it moved down, I thought it was placing something on the table."	1	10
Precision Near Object	"When it slowed down near the object, it looked more careful and precise." "The upward slowdown near the object suggests precision."	1	10

Name	Description	Files	References
Rotation		1	10
Alignment	"Slow rotation feels like alignment." "When the gripper rotated slowly, I thought it was aligning with something." "Slow rotation clearly aligns with the object on replay."	1	8
Confusion or ambiguity	"The rotation just looked like spinning for no reason." "Rotation seems playful or random." "Intent remains unclear."	1	2
Up - Down		1	30
Lifting	"It's moving up slowly — like lifting something." "Upwards feels like picking up."	1	8
Pause-as-Intention	"When it paused before moving forward, I could guess what it was about to do." "The pause before moving forward is clearer on playback."	1	2
Placing	"Now it's going down — maybe placing an item." "When it moved down, I thought it was placing something on the table."	1	10
Precision Near Object	"When it slowed down near the object, it looked more careful and precise." "The upward slowdown near the object suggests precision."	1	10
Theme 4 Context-Based Interpretation		1	24
Forward - Backward		1	24

Appendix 3: Data Collected

Participant	Motion Type	Method	Verbatim Response
P1	Forward–Backward	Think-Aloud	"It moved forward… looks like it's reaching toward something or me."
			"Now it's going backward — maybe it's finishing the task."
			"The backward return seems like a home-position reset."
		Interview	"When it moved forward, I felt like it was reaching toward me."
			"The backward motion seemed like it was done after completing a task, like pulling back."
		Video-Elicitation	"Seeing it again, the forward movement looks intentional."
			"The backward motion clearly looks like a reset."
			"The approach toward the object becomes more obvious on replay."
P2	Forward–Backward	Think-Aloud	"The forward movement feels like preparation for doing something."
			"Now it's going back — resetting maybe."
			"Forward looks like reaching out."
		Interview	"When it paused before moving forward, I could guess what it was about to do."
			"I noticed that when it slowed down near the object, it looked more careful and precise."
		Video-Elicitation	"The pause before moving forward is clearer on playback."

			"It slows down exactly before interacting — looks intentional."
			"Backward motion is definitely a return to home."
P3	Left–Right	Think-Aloud	"Why is it moving side-to-side? Looks like it's searching."
			"This sweep is confusing."
			"It looks like scanning or exploring."
		Interview	"The side-to-side sweep was confusing — it looked like it was searching or maybe cleaning."
		Video-Elicitation	"The sweeping motion still looks like searching behaviour."
			"It doesn't convey a clear goal even after rewatching."
			"The left-right gesture feels exploratory."
P4		Think-Aloud	"Left-right makes it look unsure."
			"This feels like a signal or gesture."
			"Is it waving?"
		Interview	"The sideways movement felt like it was gesturing to me, maybe waving."
		Video-Elicitation	"On replay it still looks like a wave-like gesture."
			"The sweeping is still ambiguous."
			"It resembles a communicative gesture more than a task."
P5	Up–Down	Think-Aloud	"It's moving up slowly — like lifting something."
			"Now it's going down — maybe placing an item."

			"That was too quick; can't tell what it's doing."
		Interview	"The robot moved its gripper up slowly, like it was lifting something carefully."
			"When it moved down, I thought it was placing something on the table."
		Video-Elicitation	"The lifting gesture looks smoother than I first noticed."
			"Downward movement looks like a placing action on replay."
			"The fast part still makes interpretation difficult."
P6	Up–Down	Think-Aloud	"Upwards feels like picking up."
			"Downwards feels like drop-off."
			"Hard to tell because the speed changes."
		Interview	"The up-and-down motion was too quick, so I couldn't tell if it was lifting or dropping."
		Video-Elicitation	"Speed is the main issue — still hard to interpret."
			"The upward slowdown near the object suggests precision."
			"Downward movement seems deliberate."
P7	Rotation	Think-Aloud	"It's rotating… but why?"
			"Slow rotation feels like alignment."
			"Fast rotation looks random."
		Interview	"The rotation just looked like spinning for no reason."
			"When the gripper rotated slowly, I thought it was aligning with something."

		Video-Elicitation	"Slow rotation clearly aligns with the object on replay."
			"Fast spin still looks purposeless."
			"Rotation lacks context."
P8	Rotation	Think-Aloud	"Rotation seems playful or random."
			"Slow rotation feels careful."
			"I do not see a clear task."
		Interview	"The fast rotation made it look playful or random, not goal-directed."
		Video-Elicitation	"Fast rotation still seems random when re-watched."
			"Slow rotation looks more deliberate."
			"Intent remains unclear."
P9	Gripper Open–Close	Think-Aloud	"The gripper opened — maybe it wants to grab."
			"It closed like it's holding something."
			"It opened again but didn't pick anything."
		Interview	"When the gripper opened, I immediately thought it wanted to grab something."
			"When the gripper closed after touching the block, I assumed it was holding it."
			"It opened but didn't pick anything up, so I wasn't sure what it was trying to do."
		Video-Elicitation	"Gripper opening exactly matches the start of a grasping action."
			"Closing gesture clearly looks like holding."
			"The confusing open-without-picking is visible."
P10			"Open means preparing."

	Gripper Open–Close	Think-Aloud	“Close means grabbing or holding.”
			“But when it opened without doing anything, I got confused.”
		Interview	“The open–close movement was hand-like and easy to understand.”
			“Some parts were confusing when it opened but didn’t act.”
		Video-Elicitation	“The opening without grasping is more obvious in replay — still confusing.”
			“Closing looks definitive, like a grip.”
			“Slow open-close feels communicative.”