Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

An Approach to Elicit Human-Understandable Robot Expressions to Support Human-Robot Collaboration

Jan Leusmann, Steeven Villa Salazar, Thomas Liang, Chao Wang, Albrecht Schmidt, Sven Mayer

2025

Preprint:

Copyright ACM 2025. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in CHI 2025, https://doi.org/10.1145/3706598.3713085.

Check for updates

ACM DIGITAL LIBRARY | Association for Computing Machinery | acm open

Latest updates: https://dl.acm.org/doi/10.1145/3706598.3713085

RESEARCH-ARTICLE

An Approach to Elicit Human-Understandable Robot Expressions to Support Human-Robot Interaction

JAN LEUSMANN, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

STEEVEN VILLA, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

THOMAS LIANG, University of Illinois Urbana-Champaign, Urbana, IL, United States

CHAO WANG, Honda Research Institute Europe GmbH, Offenbach, Hessen, Germany

ALBRECHT SCHMIDT, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

SVEN MAYER, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

Open Access Support provided by:

Ludwig-Maximilians-University Munich

University of Illinois Urbana-Champaign

Honda Research Institute Europe GmbH

PDF Download
3706598.3713085.pdf
05 January 2026
Total Citations: 3
Total Downloads: 1179

Published: 26 April 2025

Citation in BibTeX format

CHI 2025: CHI Conference on Human Factors in Computing Systems
April 26 - May 1, 2025
Yokohama, Japan

Conference Sponsors:
SIGCHI

https://doi.org/10.1145/3706598.3713085
ISBN: 9798400713941

An Approach to Elicit Human-Understandable Robot Expressions to Support Human-Robot Interaction

Jan Leusmann
LMU Munich
Munich, Germany
jan.leusmann@ifi.lmu.de

Steeven Villa
LMU Munich
Munich, Germany
villa@posthci.com

Thomas Liang
University of Illinois Urbana-Champaign
Champaign, Illinois, USA
LMU Munich
Munich, Germany
ttliang2@illinois.edu

Chao Wang
Honda Research Institute Europe
Offenbach/Main, Germany
chao.wang@honda-ri.de

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

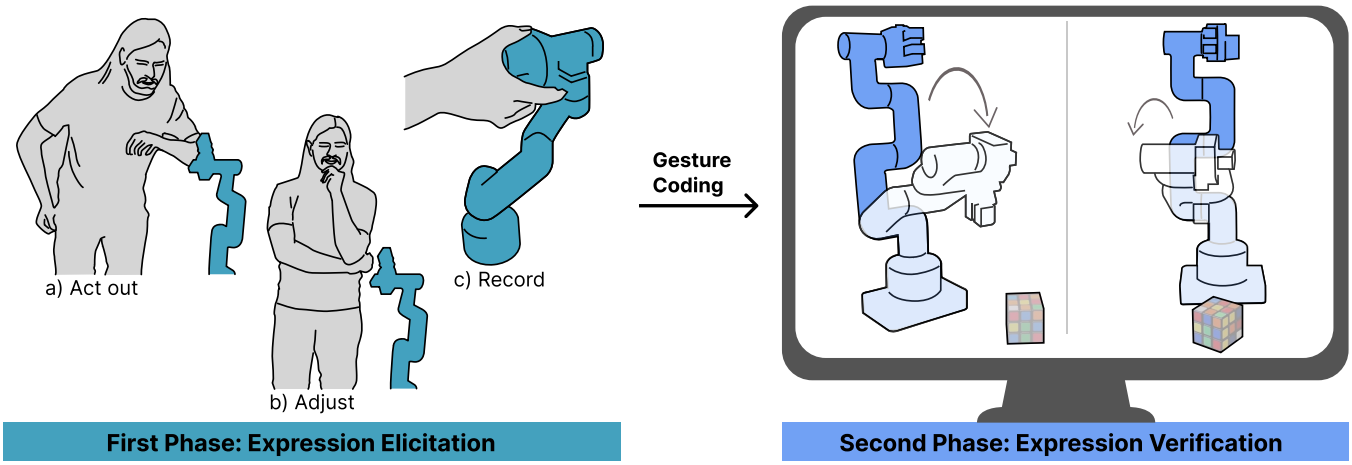


Figure 1: The two-phase process for eliciting and verifying gestures.

Abstract

Understanding the intentions of robots is essential for natural and seamless human-robot collaboration. Ensuring that robots have means for non-verbal communication is a basis for intuitive and implicit interaction. For this, we describe an approach to elicit and design human-understandable robot expressions. We outline the approach in the context of non-humanoid robots. We paired human mimicking and enactment with research from gesture elicitation in two phases: first, to elicit expressions, and second, to ensure they are understandable. We present an example application through two studies (N=16 & N=260) of our approach to elicit expressions for a simple 6-DoF robotic arm. We show that the approach enabled us to design robot expressions that signal curiosity and interest in getting attention. Our main contribution is an approach to generate and validate understandable expressions for robots, enabling more natural human-robot interaction.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**.

Keywords

Human-Robot Interaction, Gesture Elicitation, Understandable Robot Expressions, Approach

ACM Reference Format:
Jan Leusmann, Steeven Villa, Thomas Liang, Chao Wang, Albrecht Schmidt, and Sven Mayer. 2025. An Approach to Elicit Human-Understandable Robot Expressions to Support Human-Robot Interaction. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3706598.3713085

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713085

1 Introduction

As robots become everyday companions, assisting with various tasks, including those currently beyond our capabilities, this propels us into a new era of human-robot coexistence, where robots implicitly become integral parts of our lives. Central to this is establishing effective communication between humans and robots [53, 90, 100]. Verbal communication is a cornerstone of human-human interaction [33], which can either replace or support non-verbal communications [73]. Humans use non-verbal cues for conveying emotions, intents, and information [14, 77]. For robots, there are many existing approaches to creating expressions, yet, they focus on the expression creation process rather than the understandability of the expressions [48], which is essential for successful communication. In human-robot interaction (HRI), the expressions performed by the robot must be understandable by users to support interaction. Yet, there is no approach that focuses on creating human-understandable expressions to support human-robot interaction.

Previous research has shown that matching verbal and non-verbal communication from robots can improve users' perception of robots [87]. Moreover, non-verbal communication aids in conveying intentions, preferences, and emotional states, facilitating more meaningful and intuitive interactions [89] and positively influences both efficiency and robustness in collaborative tasks [13, 117]. Today, many expressions are in the domain of emotions [30, 51, 89, 92]. However, Jung [48] argues that when designing expressions for robots, there needs to be a shift from making the robot "emotional" to an understanding of how people interpret that behavior. While we see that expressions are created with a wide range of different approaches (experts in specific fields [3, 76], motion tracking [44, 115], or animation theory [44]), they are rarely understanding-centric, applying them in rapid human-centered design cycles is restrictive and non-practical, and ensuring human understandability not directly possible. On the other hand, gesture elicitation studies allow the rapid generation of user-defined gestures by ordinary people [111]. This approach is widely used in human-computer interaction (HCI) [84, 93, 104, 105] and effective in creating intuitive user-to-device gestures by having participants create novel gestures based on given referents. In contrast, others have shown the applicability of gesture elicitation studies to create device-to-user expressions [9, 34, 35, 88]. In this work, we build on prior practices to outline a standardized approach that future research can adopt, ensuring replicability in creating human-understandable robot expressions.

We propose an approach with two phases (*Expression Elicitation* and *Expression Verification*) to create validated human-understandable robot expressions based on prior approaches. We envision that in the *Expression Elicitation* phase, participants first use their own bodies to imagine how they would express something given *referents* through non-verbal cues and second to relay their human cues to a specific robot by performing, recording, and re-watching the newly created expressions (given the robot's form and size variations). Inspired by Wobbrock et al. [111], we then propose to cluster the resulting expressions, describe them using a taxonomy, and calculate occurrence scores for the resulting *expressions* to generate potential *expressions* for the selected robot configuration. In the *Expression Verification* phase, we then use a confirmatory approach to verify the human understandability of the expressions through participants' interpretation of the displayed robot expression. We showcase our approach in an example study. By applying our proposed approach, we created human-understandable expressions allowing a 6-DoF robotic arm to perform expressions, which users identified as curious behavior. In the first phase (N=16), we elicited 128 robotic expressions based on eight *referents*. Using open and axial coding, we found 13 distinct expressions. In the second phase, we ran an online survey (N=260) to verify whether the 13 expressions are human-understandable using videos.

In this work, we present our approach to eliciting and validating human-understandable robot expressions. We outline the approach and showcase it with an example study. Most importantly, our example study revealed that having users create robot expressions for a robot arm leads to human-understandable expressions even without humanoid traits. Thus, we conclude that this approach is suited to design non-verbal expressions for robot arms. By applying our approach in future work, we support expression generation for different robots, contexts, and modalities. In return, this highlights the degree to which this approach is generalizable. Thus, applying this approach to different robots in different contexts will provide deeper insights into how robots can express themselves understandably and which commonalities and differences exist in understandable non-verbal human-robot communication.

2 Related Work

Here, we assess existing literature on non-verbal communication in human-human and human-robot interactions, as well as gesture-elicitation studies.

2.1 Non-verbal communication in Human-Human Interaction

Non-verbal communication is a vital component of interaction, as 66% to 87% of information in face-to-face interaction is conveyed this way and often overshadows verbal communication [14]. We use it to communicate emotions, convey interpersonal attitudes, and manage conversations [60]. It also plays a significant role in maintaining long and effective interpersonal relationships [98]. Non-verbal communication includes all bodily and facial expressions [26]. Gestures are one vital part of non-verbal communication [37]. Non-verbal communication is an essential part of human-human interaction, and especially gestures have a huge potential to express intentions and to show attention while listening [36] and can often convey even unspoken thoughts [37]. Additionally, humans are also good at understanding the gestures from other people subconsciously [23, 26, 108].

However, interpreting the meaning of others' gestures is subjective and multifaceted [67]. Often, specific gestures only make sense given a specific context and can be interpreted differently in a different context [62]. For example, nodding can be used to confirm, agree, submit, or give permission [75] and, thus, also be interpreted to have these different meanings given a different context. Consequently, when asked what a gesture means without giving context multiple answers can be correct.

Non-verbal expressions enhance conversational flow, offering a simple, non-intrusive, and easily understandable means of communication. People interpret gestures within the context of ongoing conversations and recognize their nuanced meanings. Given the significance of non-verbal communication in human-human interactions, we advocate for a similar level of importance in human-robot communication.

2.2 Human-Robot Communication

In the context of HRI, it also has been shown that verbal and non-verbal communication is possible in both directions [66]. Similar to humans, robots can use expressions to convey their emotional state [30]. For example, Deshmukh et al. [28] found that changing the speed and amplitudes of these gestures affect user perception in terms of Animacy, Anthropomorphism, Likeability, and Perceived Safety, measured through the Godspeed scale [6]. Cabibihan et al. [16] conducted a study with a humanoid robot mimicking humans and found that humans, in general, understand what the robot tried to express. Robots using gestures increase the willingness of users to engage with them [91]. It has been found that matching gestures and verbal output increased participants' likeability of the robot and willingness to interact [54, 87]. Humans are more willing to approach robots that use gestures [91]. In collaborative tasks, robot gestures increase communication and understanding between the two parties and can even lead to the users perceiving a lower workload [63].

Researchers have investigated how to improve interaction with non-humanoid robots. Although humanoid robots are often the way how we envision future robots to look like due to their higher levels of anthropomorphism, non-humanoid robots will be an essential part of future interaction. Due to the various areas of application for robots, it is often necessary to choose a robotic design fitting into the environment, e.g., a humanoid robot supporting us with kitchen tasks would stand in our way, but a robotic arm from the ceiling could achieve the same results without being in the way [72]. Even slight changes in the appearance of robotic arms can lead to higher levels of anthropomorphism [97]. Furthermore, humans perceive even inanimate objects as social entities, attributing emotions, inner states, and personality to them [58, 78]. Non-humanoid robots are able to express emotions [47], the will to collaborate [20], and can be used to support interaction between humans [46, 76] using gestures. It is crucial for robots in the various aspects of our lives to communicate their intent so that humans can make more informed decisions about their interaction with them [118].

Curiosity, a foundational element of cognition [52], drives exploration, learning, and social connections in humans [74, 113]. It serves as an intrinsic motivator to reduce uncertainty and novelty [94]. Humans express curiosity through various different expressions and behaviors [52], and also robots behavior can be perceived as curious [18, 30]. While these studies offer solutions for expressive robotic gestures, not much is directed toward curiosity, and many of the gestures are developed based on existing research on human gestures or even ambiguously [20, 30, 54, 61, 76], which is generally harder to implement on a non-humanoid robot. In our example study, we focus specifically on the expression of curiosity by a robot and generate our gestures through an elicitation phase.

Effective communication is essential in HRI, particularly in collaborative scenarios where mutual understanding is crucial. Robots that employ gestures positively impact collaborative tasks. Therefore, enhancing HRI involves using non-verbal communication that humans can implicitly understand.

2.3 Expressions in Human-Robot Interaction

Expressions are a means of communication between humans and robots, as they play a vital role in facilitating effective and intuitive interaction between two parties [89, 92]. They can convey information, intentions, and emotional states to humans. By mimicking and utilizing human-like gestures, robots can enhance their ability to communicate and engage with humans in a more natural and intuitive manner. Non-verbal communication aids in conveying intentions, preferences, and emotional states, facilitating more meaningful and intuitive human-robot interactions [89]. They can make robots appear more polite [59], improve task performance in human-robot collaboration [13], and convey affection through changes in proxemics, color, sound, and kinematic properties, which affects the emotional valence of users [8, 96]. Most current studies focus on designing movements that convey emotions and whether humans can correctly interpret these emotions rather than examining the gestures classified by Kendon [51].

The creation of expressions for robots is currently performed with various different methods [112]. Currently, expressions are being created through automatic pipelines [27, 55, 95], through combination and mapping of predefined motion sets [101, 116], the usage of motion capturing data [12, 43, 71, 82, 83], in applicable fields (e.g., humor, animation, choreography) expert knowledge and design processes [3, 76], a combination out of experts and motion capturing data [79], or animation design theory [46, 81]. Especially methods using motion-capturing data often imitate actual human gestures, proposing that robots should communicate non-verbally exactly like humans. Generating this data or gaining knowledge from experts is often costly or unfeasible. Many also use LABANs notation of human dance movement [106] to describe and generate movement for mostly humanoid robots [1, 15, 56, 86], segmenting expressions into primitives. Automatic labeling approaches show a strong correlation with expert analysis [15], making this a suitable approach to generate anthropomorphic expressions.

In summary, expression creation typically relies on motion tracking data, expert knowledge, or animation theory, which can be time-consuming and costly. Furthermore, these methods may not guarantee that humans easily understand the resulting expressions. In general, qualitative studies are underutilized in HRI research [45].

2.4 Gesture Elicitation

With the seminal work by Wobbrock et al. [111], user-defined gestures for human-computer interaction became the norm for designing gesture inputs for today's computing systems. Recent literature reviews found that over 267 gesture elicitation studies have been conducted [103, 104]. Although originally planned for eliciting hand gestures as input for interactive systems, the core idea of gesture elicitation studies can be used with different body parts [21, 32] and has also been used to create gestures for drones [9, 34]. In gesture elicitation studies, many participants create multiple single

gestures, given a referent. Originally, a referent is one feature for which the participants should create a new gesture. The method then proposes to merge similar gestures, calculate an agreement score for the gestures for each referent, and then choose one different gesture with the highest agreement score for each referent. On average, studies use 16 referents and collect 723 gestures from 25 participants [103].

So far, gesture elicitation studies have mainly been used to map gestures to device functionalities. In the following, we discuss and present how we apply knowledge from gesture elicitation studies to enable researchers to create robot expressions through a user-centered approach. This could spark a new direction for expression elicitation in HRI.

3 Research Approach to Elicit Robot Expressions: A Research Plan

We envision a research plan to create human-understandable expressions for robots easily and in a user-centered way, see Figure 2. This method uses a two-phase approach and can be applied to multiple domains where creating understandable robotic expressions is required.

3.1 Approach Overview

Humans' expressions are mostly intuitively human-understandable [51]. However, not all robot form factors can directly mimic human gestures as they do not resemble typical human bodies. To bridge this gap, we employ a two-phase approach containing an *Expression Elicitation* and *Validation* phases.

First, in the *Expression Elicitation* phase, participants enact expressions with their own bodies, explaining their thought processes.

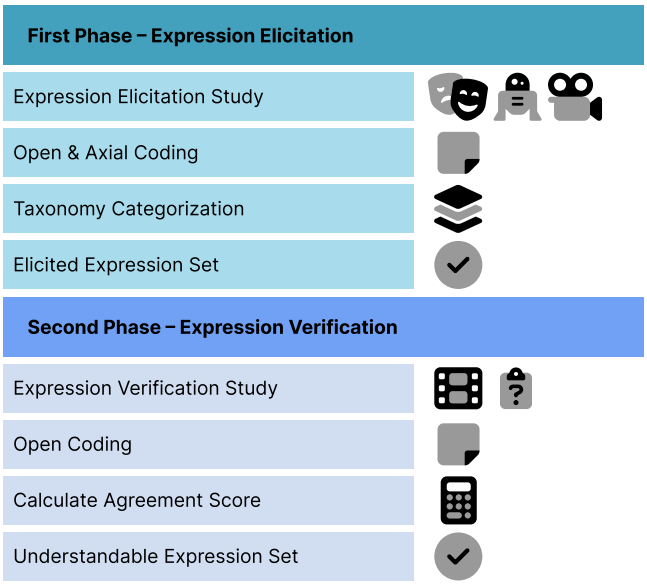


Figure 2: Procedure of creating human-understandable robot expressions, divided into two phases; *Expression Elicitation* and *Expression Verification* leading to a final understandable expressions set.

This phase is strongly influenced by the well-established method of gesture elicitation from Wobbrock et al. [111] and others [34, 84]. Next, participants should create corresponding robot expressions, keeping their physical gestures in mind. We propose to ensure that robot movement manipulation is user-friendly, allowing participants to adjust joint configurations manually and easily. Afterward, participants can view and iterate over their newly created robot expression until they match their envisioned expression. However, in contrast to most gesture-elicitation studies, we do not need to create a different expression for every *referent*; humans can map (and do map) singular expressions to different meanings but differentiate them with contextual information.

For the expression *Expression Verification* phase, we suggest asking for the participants' unbiased opinions on what they think the robot expressed, similar to other practices [9, 35, 88]. This second phase is to confirm that an independent sample of participants correctly matches the expressions created in the first phase to their meaning. Thus, we suggest asking participants to interpret the expression as the first thing of the phase without the possibility of them seeing other information from the subsequent questions. In order to mitigate potential biases in the measurements, we recommend choosing a between-subject study design [22, 39]. Finally, we can calculate which expressions were generally perceived to express the same things as the original *referent* asked about.

3.2 Goals and Requirements

As robots become companions for different tasks in daily life, it is important to understand how users understand and perceive the interaction with robots. Yet, a general understanding of human-robot interaction is affected by the wide range of robot configurations and form factors. This heavily impacts communication types that are typically linked to a specific body configuration. However, it is unclear how this bodily communication would extend to the different robot form factors. To address these challenges, three authors, with experience in HRI discussed and agreed on the following set of goals and requirements. These requirements were grounded in existing literature on gesture elicitation and HRI, for developing human-understandable robot expressions, the foundational literature for the requirements is cited in each goal. With these goals, we aim to facilitate the creation of human-understandable expressions across diverse robot configurations. While there are also additional important criteria (e.g., explainability, transparency), these might relate more to the evaluation phase of the generated gestures or are more context-dependent.

(1) **Universal**: The method should be universally applicable, working seamlessly with different robot types. The need for universality in robot design stems from the wide variety of robot morphologies and applications, requiring adaptable communication methods [44, 99]. Universality ensures that expressions are not constrained by specific physical configurations or cultural contexts [99].
(2) **Generalizable**: The method should enable the generation of expressions for any topic or domain. Similarly to gesture elicitation studies, where it is beneficial to have the same gesture set across different devices [29], robots are being used in many different domains, where humans interact

with various robots using different mental models [114]. To ensure effective communication, it is essential to design robot expressions that are understandable across contexts [42]. Gesture elicitation's referents are topic agnostic and allow to create expressions for any topic based on user preferences [103].

(3) **Accessible**: Researchers, regardless of their available resources, should find the expression creation process straightforward, efficient, and capable of yielding human-understandable expressions. Prior work suggests to choose a representative user participant group to avoid sample bias, ensure replicability for HRI studies, and highlights the value of user-friendly interfaces and intuitive methods for behavior programming, which empower diverse users, including those with limited resources, to engage in meaningful robot design processes [7, 17].

(4) **Comparable**: The results should be standardized and comparable across all studies utilizing this approach, facilitating a comprehensive analysis of the key elements in an expression's effect [31].

(5) **Validated**: The created robot expressions should be easily and accurately interpreted by a broad and diverse user group, ensuring their effectiveness in communication. Validation is a critical step in HRI design, as it ensures that the developed expressions align with the intended meaning and are accessible to diverse users [49, 64].

3.3 Blending Existing Approaches

To establish a foundation for generating and evaluating expressions or gestures, we explored several existing approaches. For expression generation, we considered methods such as predefined motion sets [101, 116], motion capturing data [12, 43, 71, 79, 82, 83], expert knowledge [3, 76], or animation design theory [46]. In contrast, HCI research often favors user-centered techniques like gesture elicitation studies [104], focus groups [50], mimicking [19], or role-play [4, 68]. We draw from their insight to design the first phase of our approach: *Expression Elicitation*. In traditional gesture elicitation studies, users are asked to generate gestures they want to perform on a device to get a certain output. If many users come up with a similar idea of how a gesture for one action should look like, this is an intuitive gesture. We use this idea to let users create gestures that they want the robot to perform, instead of themselves. The intuition is the same; if many users agree that one robotic gesture means a certain thing, that gesture is an intuitive gesture.

As Wobbrock et al. [111] highlighted in their seminal paper, validation is an important next step after elicitation. This aligns with other methods like scale development [10]. However, it is not always common practice to validate new sets [104]. Consequently, we build the verification set directly into our approach as the second phase: the *Expression Verification*. Here, we propose to use a large set of participants and ask, without priming them, what they think a robot wants to express doing one certain expression. If the large majority of participants understand one expression with the same meaning, this expression is a reliable way to communicate the goal, with which the expression was originally created.

3.4 Choice of Participants

As bodily expressions are often part of implicit memories, which the individual cannot explicitly articulate in words or descriptions, We envision that during the *Expression Elicitation*, participants interact physically with the robots, allowing them to manipulate the body of the robot, or to use their own bodies as a tool to articulate the expressions. Thus, the phase should ideally be conducted as a lab or in-situ study. Depending on the explicit context for which our method should be used to create new expressions, both domain experts and regular users could be included in the elicitation process. Domain experts have deeper knowledge and background, leading to more informed opinions. However, non-verbal expressions are inherently human; thus, regular users also have an understanding of how they would express certain concepts.

We propose to counteract the limited participant pool from the first phase real-world study by conducting the the second phase as a large scale online survey. In an online survey, video recordings of the expressions will be necessary to convey the newly designed expressions to the participants.

3.5 Choice of Referents

In gesture elicitation studies, a *referent* is the target or subject of the designed gestures [111]. In these traditional studies, researchers aim to understand the relationship between a *referent* and gestures to design more effective and user-friendly interfaces. In traditional gesture elicitation studies, each *referent* has to have one distinct gesture associated with it. This contrasts our case; here, the referents are used to generate expressions for specific topics. However, they do not need to be unique, as in human-human communication, we often use the same non-verbal communication patterns for multiple different interactions, and the conversation partner can infer from the context what is meant by that expression. Thus, the choice of referents does not need to reflect the exact set of expressions one wants to create with this approach. Yet, the referents serve as a possibility for designers to instruct users to allow two design expressions with a use case in mind. We propose to first define the construct for which expression should be designed and create the referents based on factors of that construct. In detail, each referent should prompt the user to perform a non-verbal expression for a certain topic. We suggest using additional control referents. The aim is to create expressions that are unrelated or contrasting to the domain of created expressions for verification.

3.6 Analysis and Taxonomy: Finding Commonalities in Expressions

To condense findings from the expression elicitation phase, we suggest a two-step coding process for the robot expressions created by participants in the first phase. First, we recommend using open coding to identify commonalities and subtle differences among expressions, preserving as much expression information as possible. Next, we propose clustering similar expressions by referent through axial coding, with the possibility of similar clusters emerging across different referents. Initially, this will yield numerous unique expressions, each described in detail using our taxonomies. However, in subsequent stages, we advocate for condensing these

Table 1: Taxonomy for human-understandable robot *expressions*. The taxonomy includes six dimensions.

Dimension	Category	Description	Example
Speed	Slow Normal Fast	Subjective assessment of expression speed. The default speed is labeled as **normal**, while expressions significantly faster or slower are classified as **fast** or **slow**.	An expression is fast if the movement is noticeably faster than other expressions (e.g., the end-effector rapidly turns from left to right).
Complexity	Single Compound	Describes whether an expression consists of a **single** motion or a combination of multiple motions (**compound**), which can either involve multiple sequential movements or simultaneous combinations of different movements.	Vertical end-effector movement while simultaneously moving the body from left to right would be a compound expression, while only vertical movement with the end-effector would classify as single..
Flow	Continuous Discrete Combined	Classifies whether an expression involves **continuous**, uninterrupted motion or **discrete**, non-continuous motion. **Combined** expressions combine both continuous and discrete movements.	A discrete expression involves the robot stopping at multiple points (e.g., the end-effector moves around using the full body, stops at multiple points, to then rotate the end-effector).
Binding	Environment Object Person	Assesses whether the robot's face is anchored toward a specific **object** or **person** or moves independently (**environment**).	Classifies the point of interest of the expression. A nodding expression can be targeted towards an object or person, or it is a general expression with no direct point of interest.
Dynamics	Dynamic Static	The robot's movement can either be **static** or **dynamic**. **Dynamic** should be further specified in the context of the expression's target.	An expression is static if all joints, except for the ones needed for the expressions, are not moving. Dynamic describes expressions with movement involved (e.g., moving away from a target, moving towards a target).
Focus	Focused Unfocused	Characterizes the relationship between the communication node and the target, which can be either **focused** on the target, tracing it, or **unfocused**.	The communication node refers to the point a user would identify as the robot's "face" (e.g., the end-effector for robot arms).

expressions into their core ideas. To achieve this, we suggest clustering groups by their primary components into what we term *expressions*. Once again, the same *expressions* may appear across different *referents*. From this point forward, we suggest analyzing the *expressions* independently of the *referents*. These *expressions* should clearly convey some internal state of the robot. To describe and distinguish the emerged *expressions*, we propose to use our taxonomy with six dimensions: (1) **speed** provides a subjective assessment of the expression speed, (2) **complexity** distinguishes single motion or combination of motion expressions, (3) **flow** describes whether a gesture is continuous or discrete, (4) **binding** describes towards what the robot's face is anchored, (5) **dynamics** describes the movement of the robot, and (6) **focus** characterizes the relationship between the robots face ands target. Table 1 shows all dimensions with its categories, descriptions, and examples. Depending on the robot, more dimensions can be added to further describe expressions, e.g., using the LABAN notation to describe dynamics more fine grained [106]. This taxonomy makes it easier and more transparent to decide which expressions are the same and, thus, need to be merged or which are different from each other.

3.7 Measurements

We recommend employing established domain-specific questionnaires whenever feasible in both study phases. In the first elicitation phase, these questionnaires serve to gauge participant satisfaction with the created expressions and identify anomalies. More importantly, the questionnaire should be integrated and statistically evaluated in the second expression verification phase. Additionally, we suggest asking about the user perception of the chosen expressions using a subset of questions proposed by Rzayev et al. [85], as this can provide additional properties of each expression. Lastly, and most importantly, participants of the second phase should be asked what they think the robot expressed without biasing them with any previous domain information.

3.8 Metrics

We propose one metric each for the *Expression Elicitation* and *Expression Verification*. In the first phase of our approach, we want to determine which gestures are suggested by how many participants. With the *occurrence score*, we can later evaluate whether expressions suggested by more participants are also better understood. In the second phase, we want to measure how much participants agree on one expression expressing one specific thing. For this, we propose calculating the *qualitative response accuracy*.

To measure the agreement for the *Expression Elicitation*, the agreement score (A) [111] and agreement rate ($\mathcal{AR}$) [102] do not apply to our case. The reasoning for this is three-fold: (1) We do not need to find one extremely well-fitting expression for one *referent* but rather find out which expressions are understood in which way. The same expression can be the best fit for multiple *referents*, and we encourage diversity in expressions. (2) The results of this method should state how confident humans understand each expression. (3) Users can propose multiple candidates (this is a common issue [69]). However, $\mathcal{AR}$ penalizes diversity due to squaring the ratio. For this reason, we favor a linear relationship to calculate a quality metric. For this, we propose to rate each expression in a given set of expressions of one *referent* by an occurrence score (*OS*):

$$OS_{R_i,E_j} = \frac{|E_j|}{|R_i|}, E_j \in R_i, \tag{1}$$

With this, we can calculate the OS for all expressions E_j per *referent* R_i. An *OS* of 1 indicates that only one expression was envisioned for a given *referent* by participants, while $\frac{1}{N}$ denotes the lowest possible score, where N distinct expressions were proposed once.

An additional approach to understanding the diversity and elicitation saturation is the *consensus-distinct ratio*, which was introduced by Morris [69] for gestures. Here, the idea is to count all gestures that are named more often than a given threshold. Here Morris [69] proposes a default threshold of ≥ 2. We argue that this metric scoring low could have two potential reasons. (1) The saturation point of gestures has not been reached. Thus, more participants are needed to fully understand potential expressions for this *referent*. (2) The *referent* itself lacks specificity, and thus, consensus is hard to reach. Other metrics, as *max-consensus* [69] or $C\mathcal{R}$ [102] are not applicable to evaluate our case of expressions as they tend to find well-separable outcomes, which we do want to force onto our resulting set of expressions.

In line with Equation 1, we propose to quantify the perceived *qualitative response accuracy* (*QRA*) for the *Expression Verification*. Here, to calculate *QRA* by textual responses from participants who were asked to describe the expression. Through open coding, the researchers then need to determine if the response matches the intended meaning of the *expression* while respecting the meaning of the original *referent*. Here, we count the fitting labels as C^+, and the not-fitting ones as C^-. As a result of respecting the initial *referent*, an *expression* can have multiple *QRA*'s. Thus, we calculate *QRA* as:

$$QRA_{R_i,E_j} = \frac{|C^+_{E_j}|}{|C^+_{E_j}| + |C^-_{E_j}|}, E_j \in R_i, \tag{2}$$

A high *QRA* means the human-created *expression* leads to a human-understandable *expression*.

4 Example Process of the Two Phases

We present a full cycle of this approach. First, we conducted an expression elicitation phase to create expressions to signal curiosity. Secondly, we conducted an expression verification phase to confirm whether the created expressions were understandable.

4.1 Example Expression Elicitation Study

4.1.1 Study Design. We conducted a lab-based expression elicitation study with 16 participants, allowing them to create and refine the robot expressions directly. Participants were provided with eight different *referents*. They first acted out the gestures with their bodies, explaining their thought process, and then translated these expressions onto the robot. This approach helped them mentally align with how they would express each referent. We used Latin-square ordering to counterbalance the *referents*' presentation order.

4.1.2 Apparatus. For our study, we used the MyCobot 280 robotic arm from Elephant Robotics. We chose this specific robot because it is lightweight and small, making it easy for participants to manipulate its movement with their hands. Its lack of anthropomorphic traits offers a wide range of possible degrees of freedom with its movement and does not prime participants into just copying their own, e.g., head movement.

Table 2: The eight referents we used to prompt users to create expressions. We showed these referents to users and then asked them to create an expression the robot would do to match the descrption.

Ref.	Description
R1	Imagine a non-movable object on the table in front of you. Without touching it, how would you visually observe the object to intake more information?
R2	Imagine a constant sound in the corner of the room. How would you listen to the sound to intake more information?
R3	Imagine a non-movable object on the table in front of you, but you don't understand how it works (e.g., why a computer screen isn't black). How would you reduce uncertainty about this object?
R4	Imagine someone is explaining something to you. How would you show understanding and acknowledge the information?
R5	Imagine someone starts talking to you. How would you use an open posture to signal that you are listening?
R6	Imagine someone is talking to you. How would you express that you are engaged and attentive?
R7	Imagine you see something scary. How would you back away in fear?
R8	Imagine someone starts talking to you. How would you use a closed posture to signal that you are rejecting what they are saying?

To record and play the robot's joint positions and movements, we created a simple Python application using TKinter for the GUI, the MyCobot Python library, and ROS Noetic. The application allowed the study conductor to record key points of the robot's joint positions, shown by the participants, into one motion. The created motion could then directly be played back to the user and allowed for direct refinements. The study conductor handled the creation tool while the participants gave instructions on how they wanted the robot to move. Furthermore, we video-recorded the participants' movements and explanations for later analysis.

4.1.3 Referents. Based on core ideas from related work on curious behavior, we selected six curiosity-related *referents* and two control *referents* for our study. Here, we looked into research on observable curious behavior for adults, children, and animals, as these behaviors are usually well-known to users. We then formulated the referents in a way that users would be prompted to show these behaviors, see Table 2. (1) Visually intake more information in unfamiliar environments [38], (2) Adapt sensor positioning to intake sound better [2, 25], (3) Construct knowledge about unknown things [11, 74, 107], (4) Express interest in other people to learn more about them [2], (5) Express will to listen [2], and (6) Show attention and engagement to express interest [24]. From this, we constructed the six curiosity *referents* (R1-R6) and two control *referents* (R7, R8).

4.1.4 Task. For each of the eight *referents*, we instructed participants to create an expression for the robot by first acting it out

Table 3: Description of the 13 resulting robot *expressions*. In referents shows which *referents* created each *expression*.

Exp.	In Ref.	Description
E01	R1, R3	Moving closer with end-effector and *scanning*[1] the object in one axis.
E02	R1, R3	Moving closer with end-effector, continuously moving around the object and *looking*.
E03	R1, R3	See E02, but doing the movement multiple times sequentially *looking* at the object from two sides.
E04	R2	Moving forward, scanning environment by turning the end-effector horizontally to both sides.
E05	R2	Moving forward, facing one spot in environment.
E06	R2	Repeatedly moving forward, *looking* at one spot in the environment.
E07	R4, R5, R6	Nodding with the end-effector
E08	R4, R5, R6	Nodding with the end-effector with body/end-effector movement in between.
E09	R5, R6	Showing attention through body movement.
E10	R7	Body moving back, end-effector focused on a target.
E11	R7	Body first moving forward then back, end-effector focused on a target.
E12	R8	End-effector turning horizontally to both sides multiple times (shaking head).
E13	R8	Body turning and moving away from the target.

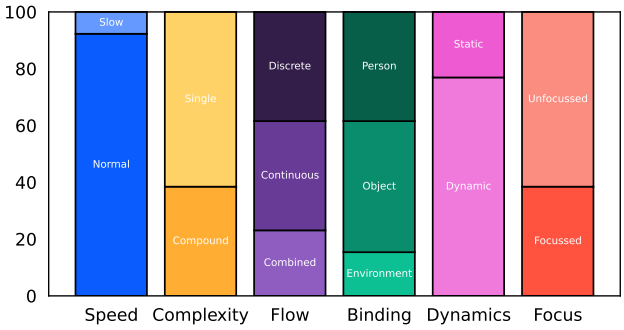


Figure 3: Distribution of expressions in six taxonomy categories (c.f., Table 3).

themselves and then replicating it on the stationary robot. The only constraint was that they could not move away from their position, as the robot was also mounted stationary. After acting out the referent, participants physically positioned the robot's joints to certain key points and then asked the study conductor to save the position when they were satisfied, stating their preferred duration, speed, and pauses for each movement. The tool allowed participants to preview, modify, and refine their expressions, including undoing points or adjusting the speed.

4.1.5 Procedure. The study began by welcoming the participants, introducing the study objectives to them, a request for their consent, and a demographics questionnaire. Then the study began with two tutorial referents (the emotions happy and sad) to explain and show participants the process of recording gestures with the robot. For the first tutorial referent, the study conductor demonstrated the process while explaining their thought process. In the second, participants performed the task independently with guidance. After completing the tutorials, participants moved on to the eight study referents, creating and refining one expression for each until they were satisfied. After recording the expressions, we played them back to participants in a random order, asking them to rate each one using five questions. We formulated these questions based on curiosity literature, reflecting the pillar from which we created our *referents*: "I perceived the system to be very [engaged, attentive, explorative, information seeking, curious]". Participants could watch the robot's performance as many times as they wanted. After rating each expression, participants moved to the next without seeing the original referents. Finally, we gathered feedback, addressed any questions, and thanked them for participating.

4.1.6 Participants. We recruited 16 participants (male = 10, female = 6) via convenience sampling. Their education levels included master's degree (11), bachelor's degree (3), and some college but no degree (2). Their age ranged from 20 to 35 (M = 27.50, SD = 4.26). Our participants had three different nationalities: German (12), USA (3), and Canada (1). The mean Affinity for Technological Interaction (ATI) [109] score across all participants was 4.61 (SD = 0.68, Cronbach's α = 0.85). On average, the study took 45 minutes, and we reimbursed participants with 10€.

4.1.7 Analysis. Three authors used open coding to identify and merge similar expressions, initially coding 128 expressions and condensing them into 37 distinct expressions. We then applied axial coding to further consolidate these into 13 unique expression categories, based on how the robot moved and what it conveyed. The final expressions are documented in the Supplemental Material.

In detail, first, three researchers iteratively coded the robot gestures, observing that participants often perceived the robot's end-effector as a face, as reflected in their reactions and think-aloud comments. We established the front as representing the direction of the object in the referent, while other movements were categorized as body gestures (e.g., extension, contraction, leaning). This open-coding process yielded 37 *expression*, grouped by the *referents*, though the same expression could appear across multiple *referents*.

In the second step, we condensed 37 expressions into 13 *expression categories* using axial coding. In detail, we categorized expressions based on their primary components using our proposed taxonomy. This process unified variations in movement, which appeared repeatedly across different referents, see Table 3. In this axial coding phase, we focused more on the "core idea" of each created expression instead of coding every single movement (e.g., the core idea could be that participants used the end-effector to let the robot do a nodding motion. The axial coding phase would combine all nodding expressions, independently whether the expression was performed with the robot straight or contracted).

[1]The robot did not have a camera or eyes; however, every single participant in our first study phase assumed the end-effector of the robot to be a head.

4.1.8 Results. Each *referent* had 16 resulting expressions. This results in $8 \times 16 = 128$ expressions. Which we reduced into the final expression set of 13 *expressions*, see Table 3. Moreover, we classified each resulting expression based on our taxonomies into six categories. Figure 3 displays the distribution.

Occurance Score (OS). We calculated the *OS* for every expression, see Table 4. The mean *OS* is 36.5 ($SD = 15.42$). E10 received the highest *OS* with 69% while E08 received the lowest *OS* with 12%.

Gesture Ratings. We found that all six created curious expressions were perceived to be significantly more curious than the two control expressions and did not find any anomalies (See Section 7 and the supplemental material for results).

4.2 Example Expression Verification Study

We continued the second study with our set of 13 *expressions* (E01–E13). In the following, we describe the process of the expression verification phase. Please see videos of all expressions in the Supplementary Material.

4.2.1 Study Design. We conducted a between-subjects design online survey to confirm the expressions are understandable (N=260). Thus, each participant saw one video of the robot performing one expression, interpreted it, and completed a questionnaire.

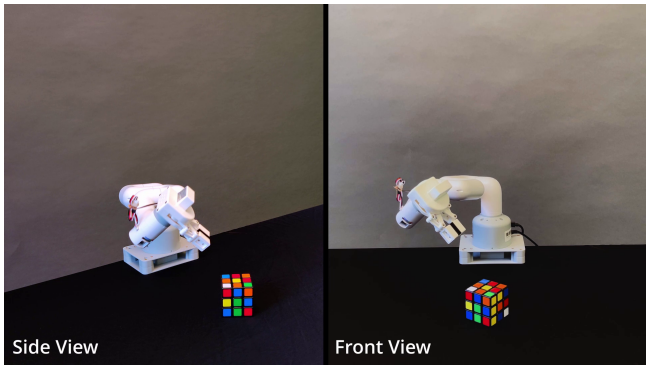
4.2.2 Apparatus. To display the robotic expressions, we filmed each of the 13 expressions from two angles: the front, to simulate the participant's view, and the side, to better illustrate the depth of movement (see Figure 4). For expressions E01, E02, and E03, we included a Rubik's cube to show in which object the robot is interested. For the remaining expressions, the robot interacted directly towards the camera.

4.2.3 Questionnaire. We first asked participants to read an introduction to the robot's non-verbal communication and then watch a video of the robot performing an expression. We then asked them to describe what they believed the robot expressed briefly. We ensured participants had to watch the video by removing the button to continue the survey until the video ended. Next, participants rated the robot's movement using VAS sliders for ten questions, with sliders ranging from 0 (Strongly Disagree) to 100 (Strongly Agree) and no numbered ticks, as these have been shown to yield more precise responses [65, 80]. Participants could rewatch the video as needed. We then asked to which extent the participant agreed with the following questions. (1) I perceived the gesture to be very engaged. (2) I perceived the gesture to be very attentive. (3) I perceived the gesture to be very explorative. (4) I perceived the gesture to be very information-seeking. (5) I perceived the gesture to be very curious. (6) This gesture is very understandable. (7) This gesture is very effective in communicating the robots' intent. (8) This gesture is very intrusive. (9) This gesture is very noticeable. (10) This gesture is very disturbing.

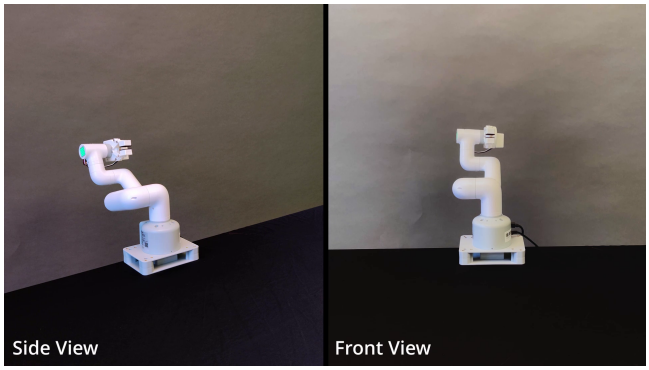
4.2.4 Participants. We recruited 289 participants via Prolific, excluding 3 for failing attention checks and 26 for describing only the robot's movement instead of its interpretation[1], leaving 260 participants (female = 131, male = 128, non-binary = 1). We ensured gender and nationality balance in our between-subject design, with each

Table 4: Calculated *OS* (c.f., Section 3.8) in rounded percent (%) for each *expression-referent* combination. The same *expression* could occur over different *referents*. For example, from all the proposed expressions in R1 25% were E01.

Expression	Referents (*OS*)
E01	R1=25, R3=38, R4=19
E02	R1=38, R3=25
E03	R1=38, R3=38
E04	R2=38
E05	R2=44
E06	R2=19
E07	R4=62
E08	R4=19, R5=38, R6=50
E09	R5=50, R6=12
E10	R7=69
E11	R7=31
E12	R8=56
E13	R8=44



(a) E02: Displaying the robot showing interest in an object.



(b) E10: Displaying the robot backing away in fear.

Figure 4: Two example expressions from our setup to demonstrate the 13 expressions to the participants of the verification study in video form.

[1]The participants were clearly instructed to state their interpretation and not describe movement

Table 5: Calculated qualitative response accuracy (*QRA*) (c.f., Section 3.8) in rounded percent (%) for each *expression-referent* combination. Each value is the percentage of how often an *expression* was described with a matching label for a *referent*. As multiple *expressions* could arise from one *referent*, one *expression* has a separate matching score for each *expression-referent* combination. 100% means that every single label given by participants in the second study is a fitting label for its referent. One column can be read as follows: To express R1, the *expression* E01 was correctly detected by 75% of participants.

Expression	Referents (*QRA*)
E01	R1=75, R3=75
E02	R1=86, R3=86
E03	R1=94, R3=94
E04	R2=3
E05	R2=7
E06	R2=15
E07	R4=96, R5=96, R6=100
E08	R4=88, R5=88, R6=88
E09	R5=80, R6=80
E10	R7=29
E11	R7=13
E12	R8=73
E13	R8=21

participant rating one of 13 expressions, resulting in 20 responses per expression. Participants, aged 19 to 68 ($M = 34.60, SD = 9.90$), included 209 full-time and 51 part-time employed individuals. They came from over 45 countries, with the highest numbers from the UK (18), Portugal (16), South Africa (15), Spain (14), Netherlands (13), Italy (12), Hungary (12), Ireland (11), Australia (10), Canada (10), Greece (10), and Mexico (10). On average, participants took 3.99 minutes ($SD = 1.77$) to complete the survey and watched the video 2.98 times ($SD = 3.33$).

4.2.5 Analysis. We utilized open coding for each expression description, iteratively refining the main component for each statement and reaching a consensus on a single label. However, some participants suggested that the robot expression could convey multiple meanings, leading to multiple labels within the final code for one expression. This process yielded 68 unique labels and 333 in total across 260 descriptions. Subsequently, we grouped these labels into 12 distinct code groups for further analysis. Documentation of this coding process is available in the Supplemental Material.

4.2.6 Results. In the following, we report our results for the second study phase. This includes results regarding the expression descriptions and results of the questionnaires.

Expression Descriptions. In total, we collected 333 labels from the 260 gesture descriptions across 68 unique labels. We found 12 overarching code groups, further categorized into four themes. Figure 5 displays the distribution of the code groups for each expression. We then marked each code group as either *matching* or *not matching*, depending on whether the understood intent matches the originally proposed intent by the referent, to later calculate the *QRA*.

The main themes we found are (1) Exploratory Expressions, (2) Negative Expressions, (3) Positive Expressions, and (4) Interactive Expressions. The theme Exploratory Expressions includes the labels *task evaluation* (52) and *curiosity and interest* (31). These groups include labels like "assessing task", "examining", or "curiosity". The Negative Expressions theme include the labels *negative emotion* (6), *negative social interaction* (13), *negative consensus* (8), *fearful reactions* (17), and *uncertainty and confusion* (14). *Negative emotion* includes labels like "sadness," "anger," or "annoyment," *negative social interaction* "ignoring" or "distancing," and *negative consensus* "disagreement" or "unwillingness." The theme Positive Expressions includes the labels *positive social interaction* (59), *positive emotion* (6), and *positive consensus* (48). *Positive social interaction* expressions include labels like "approaching," "available," or "submissive," positive emotions "empathy" or "happiness", and *positive consensus* "acceptance," "agreement," or "confirmation." Lastly, the theme Interactive Expressions includes the two labels *active social interaction* (7) and *indicative social interaction* (35). *Active social interactions* include labels like "dancing" or 'laughing" and *indicative social interactions* like "commanding," "taking," or "pointing."

We calculated *QRA* between each expression and the goals of the *referents*, i.e., the ratio between the *matching* and *not matching* labeled expressions, displaying the results in Table 5. For instance, expression E01, which was created for referents R1 and R2, achieved a 75% qualitative response accuracy, indicating that 75% of labels matched the goals of the original referents. Most participants accurately identified expressions related to visual observation (R1, R3) and attentive listening (R4, R5, R6). However, expressions for observing an unknown sound (R2) were correctly described by only 3% and 7% of participants, indicating poor suitability for expressing interest in sound. For the control referent expressing fear (R7), neither of the created expressions was correctly identified by most participants. Conversely, one expression (E12) for signaling disinterest performed well, while another (E13) did not.

Curiosity and Perception Questionnaire. The questionnaire results are shown in Figure 6. Normality tests using the Shapiro-Wilk test revealed that none of the ten questions (five on user perception and impact, five on curiosity) followed a normal distribution ($p < 0.01$). Therefore, we used Kruskal-Wallis tests to detect significant effects among the expressions. We observed significant effects for all five curiosity-related questions and three out of five questions related to perception. Subsequently, we conducted Mann-Whitney U post hoc tests to determine the significant differences between expressions, which are presented in Figure 7.

Visual Exploration: For expressions E01, E02, and E03 (arising from R1 and R3), we found no significant difference for attentive, explorative, and understandable. We found E02 to be significantly more engaged and information-seeking than E01. We found E03 to be significantly more curious and effective than E01. And lastly, we found E02 and E03 to be significantly more noticeable than E01.

Auditory Exploration: For E04, E05, and E06 (arising from R2), we found no significant difference for engaged, attentive, explorative, information-seeking, understandable, and effective. We found E06 to be significantly more curious and noticeable than E04.

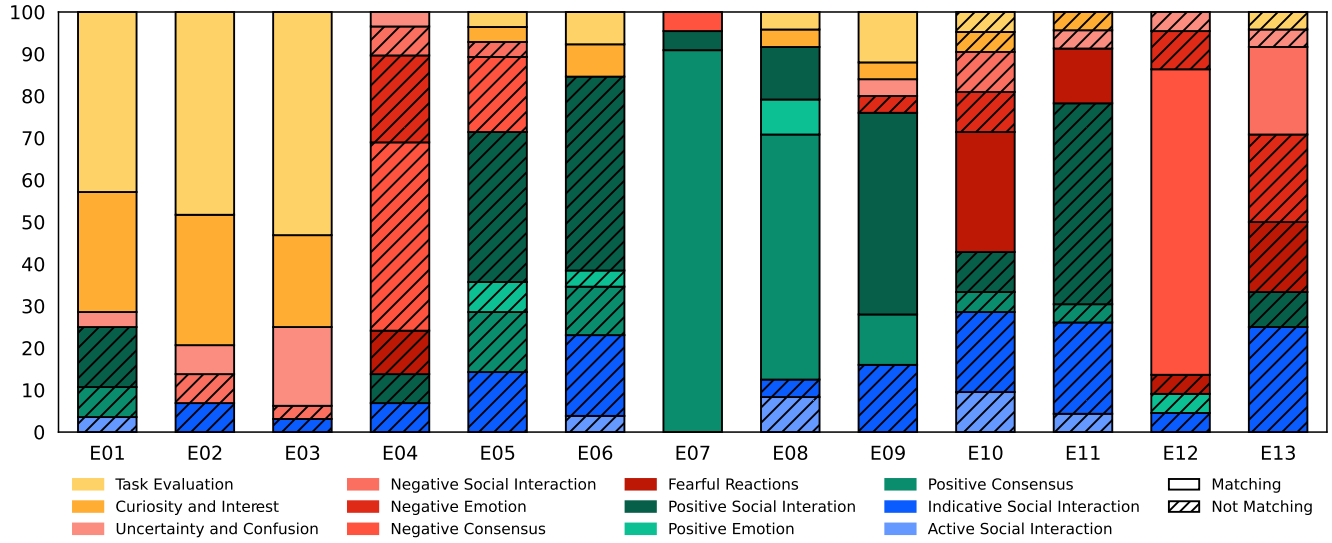


Figure 5: Distribution of groups of different interpretations for each expression. We striped non-matching expressions, c.f., Table 5. An expression is matching if the majority of users in the second study phase understood the originally, in the first study phase, planned intent of the expression.

Listening Attention: For E07, E08, and E09 (arising from R4, R5, and R6), we found no significant difference for engaged, attentive, curious, or noticeable. We found E08 to be significantly more information-seeking than E07, and E09 significantly more explorative and information-seeking than E07. We found E09 to be significantly less understandable and effective than E07 and E08.

Scared: For E10 and E11 (arising from R7), we found no significant differences for engaged, attentive, information-seeking, understandable, and effective. We found E11 to be significantly more curious and noticeable than E10.

Auditory Disengagement: For E12 and E13 (arising from R8), we found no significant differences for attentive, explorative, information seeking, and curious. We found E13 to be significantly less engaged, understandable, effective, and noticeable than E12.

5 Discussion

We described an approach to create human-understandable robot expressions via two phases. We then showed the applicability of this approach through an example study to create curious robotic expressions In the following, we discuss this approach.

5.1 Towards a Robust Approach for Expression Elicitation and Validation

Traditionally, developing expressive expressions for robots in HRI required laborious processes, including creating expressions from scratch, relying on domain experts, or using motion data of human users. However, these methods often lacked efficiency and were not suitable for all HRI scenarios. Recognizing the growing importance of HRI research, we set out to establish a rapid and validated method for generating expressive robot expressions, addressing the need for efficiency and flexibility. In designing our approach, we established five key goals (see Section 3.2).

(1) **Universal:** While we showed that our approach works on a non-anthropomorphic robot arm through our example study, we could not prove that our approach works for any robot. However, a key strength of our example study was allowing participants to directly move the robot, and future studies should continue to prioritize ease of expression creation. Participants can manipulate the robot by hand, with tools, or by instructing a conductor, depending on the robot used. Thus, we hypothesize that our method is also applicable to robots where easy manipulation is possible, which needs further investigation in future work. Regardless of the method, the expression generation process should be iterative, allowing users to refine expressions until they are satisfied.

(2) **Generalizable:** Our approach was designed to be generalizable by using flexible referents, allowing participants to create expressions for any topic. This topic-agnostic method is adaptable to different HRI contexts without relying on specific datasets or subject knowledge, making it applicable across various robots and scenarios. The use of *referents* ensures consistency while allowing for diverse, transferable expressions that can be compared across studies.

(3) **Accessible:** Our approach intentionally excluded the need for external datasets or domain experts, ensuring accessibility to all researchers. However, experts can still be involved in the expression elicitation phase. By focusing on low-resource requirements and streamlined processes, accessible methods democratize participation in HRI research and facilitate reproducibility across varied contexts.

(4) **Comparable:** By employing taxonomies and *QRA*, we aimed to provide a clear description of the created expressions

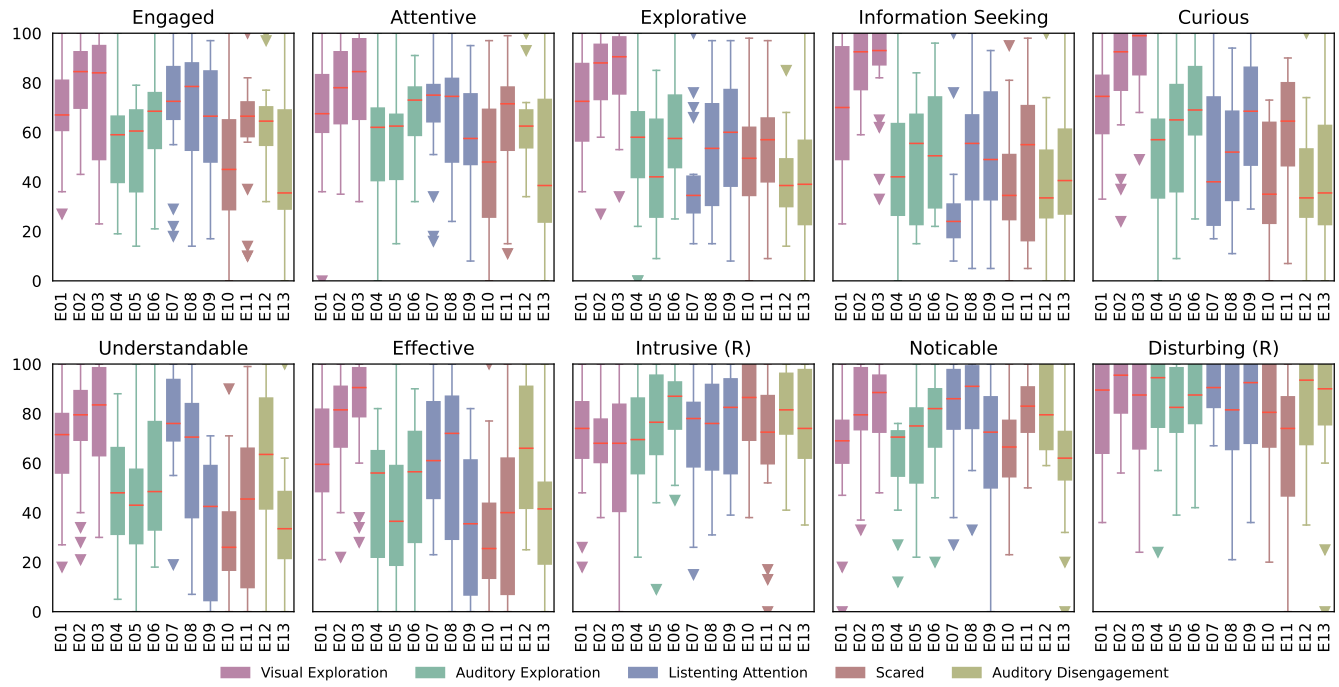


Figure 6: Boxplots of the ten questions we asked for every *expression*. The top row displays the results for the curiosity questions. The bottom row displays the assessment of perception and impact questions. The scales for intrusive and disturbing are reversed such that 100 always means good.

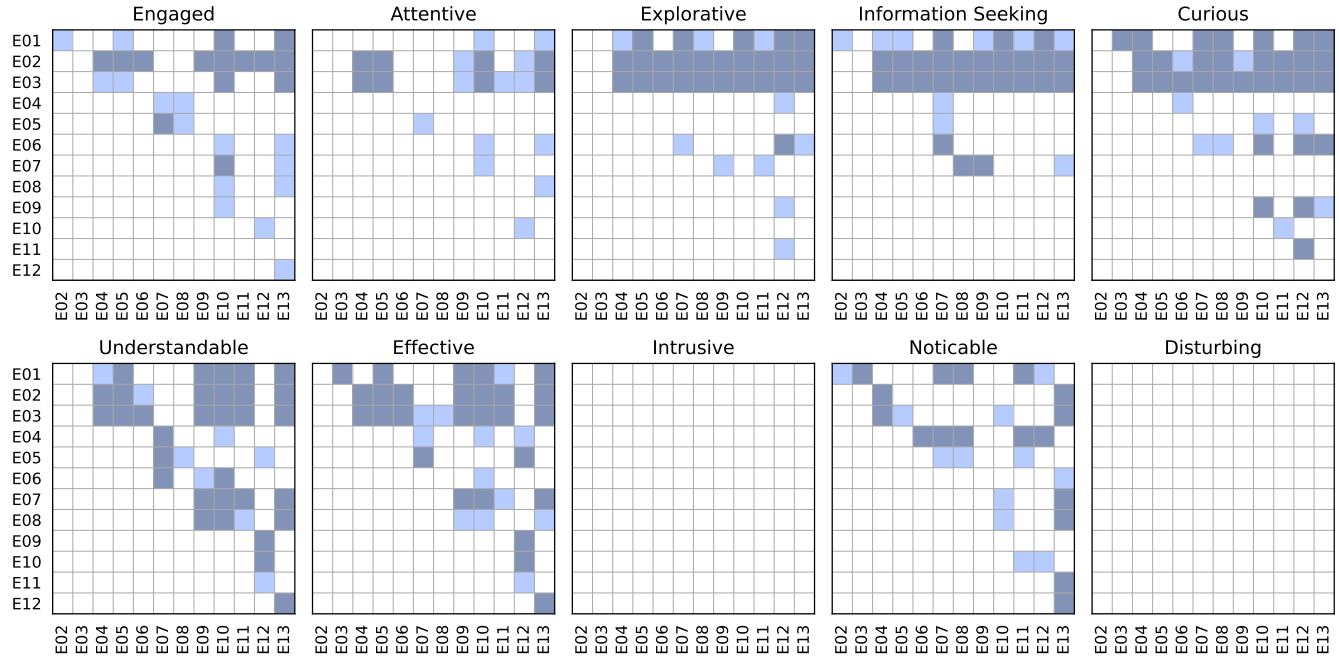


Figure 7: Significant differences calculated through a Mann-Whitney U test. Darker blue values show a significant difference with $p < 0.01$, and lighter blue $p < 0.05$. Empty grids show that this measurement had no significant effect.

and their comprehensibility, making it possible to compare results across studies and generalize findings.

(5) **Validated:** To confirm the understandability of the expressions, we propose conducting an online study with a large participant pool. This approach ensures broad feedback, helping confirm that the created expressions are comprehensible to a diverse audience.

In summary, our approach successfully met all five goals, offering a versatile and efficient method for generating robot expressions in HRI research.

5.2 Our Approach Successfully Uses Human Creativity to Design Understandable Expressions

For the expression creation part, we used ideas from gesture elicitation studies and combined them with enactment. We found that this second part got participants in our example study into the correct thinking mood in which they would express the different *referents* and, thus, how they would want the robot to express these things. This worked well in our case, even though we chose a non-humanoid robot, using the difficult part of the spectrum of user-created expressions as there is no direct 1-to-1 mapping between their own expressions and how the robot should express that, and some abstraction is needed. We propose that as this approach worked for a non-humanoid robot, using this approach for a robot with more humanoid or animalistic traits will likely work, as users can use more previous knowledge to create expressions.

We aimed to create expressions in the domain of curiosity and successfully created six expressions. With the knowledge of which core ideas from expressions lead to participants assuming the robot expresses something designers can now take these ideas and design choreographies for robots. In our expression validation phase, we did not give the users any context about the robot's current routine. However, as human expressions are interpreted differently, depending on the context [67, 75], it is important to note that also different interpretations of the same expression should be labeled as correctly understood. By giving users more context, the potential interpretation of the robot's expression can also be limited, which might be desirable or undesirable.

5.3 Generalizablility Comes Across Studies Not Within

Through our example study, we found that our approach works for non-humanoid robot arms. Furthermore, a similar approach has already been used to create understandable drone gestures [9, 34]. With our proposed key goals for our approach, we aimed to create a method that future work can apply to various robots. However, showing how generalizable our approach is can only be done by subsequent studies, using this approach to create expressions for different robots, different modalities, different contexts, and in different topic domains. Many now-established HCI methods have gone through this iterative process of getting improved and also verified over time through the HCI community, e.g., general gesture elicitation studies [111], the NASA-TLX questionnaire [40, 41], or ART Anovas [110]. We aimed to create a generally applicable approach to create human-created robotic expressions and validate their understandability in a rapid way, as many HRI studies need robotic behavior. We also note that our approach only generates verified general expressions that are understandable. In contrast to other works, we do not focus on the design process of these gestures, which focus on generating "crisp" looking expressions [57]. We propose that this should be done by designers when applied afterward.

5.4 How and Who Does Our Approach Help

Non-verbal communication holds immense importance in both human-human interactions and human-robot interactions. Implicitly conveying one's current state and needs through non-verbal cues is particularly valuable in human-robot communication because it seamlessly integrates with ongoing tasks, unlike intrusive notifications from smart devices, which can disrupt the user's focus. With robots, we have the unique opportunity to redefine how we interact with advanced technical systems capable of providing comprehensive support in our daily lives. Achieving seamless and effective communication is pivotal for robots to fulfill their potential as supportive companions.

However, traditional methods of developing non-verbal communication for robots have limitations and are mostly data-driven. Here, we want to add the the options of how future robot expressions can be created through a more human-centered approach. Our approach is usable with both expert and non-expert users for the expression elicitation phase, depending on the domain for which the expressions should be created for. Through the validation phase with a large-scale sample of non-experts, it becomes clear which expressions will be understood by the general population.

5.5 Insights from our Example Study

In the following, we will discuss our findings from our example study to create human-understandable robotic expressions expressing curiosity.

5.5.1 People Elicit Convergent Expressions. In our expression elicitation phase, we observed that participants often shared similar ideas for expressing various *referents*. Out of 128 recorded expressions, we identified 13 unique expressions, averaging 9.84 expressions per type ($SD = 5.35$). This similarity highlights that a large participant pool may not be necessary due to the low variation in human expressions. Providing similar referents allowed participants to be creative while addressing comparable ideas. Human expressions tend to exhibit similarities across various situations, with contextual cues playing a vital role in interpreting the intended meaning of an expression. This contextual influence is further evidenced by the consistently higher scores for expressions E01, E02, and E03 across all measurements in the second phase, as these expressions were the only ones framed within the context of an object.

5.5.2 Human-Understandable Curious Expressions for Visual Exploration and Listening Attention. Our study revealed that of the 13 human-created expressions, seven were well understood by participants, while six were not. This indicates that we successfully created at least one well-understood and, thus, effective expression

for six of the eight referents (five of six curiosity referents). We can now compare these effective expressions with the quantitative questionnaire results.

For referents R1 and R3, expressions E01, E02, and E03, which had high qualitative response accuracies, were all suitable for conveying visual interest. However, E03 had the highest accuracy of 94%, making it the best match for R1 and R3. It combines both continuous and discrete flow, distinguishing it from E01 and E02, which use only one flow type. For referents R4, R5, and R6, expressions E07, E08, and E09 were applicable. E07 had the highest accuracy (96% for R4 and R5, 100% for R6), while E08 and E09 were deemed significantly more information-seeking than E07, though E09 was less understandable. E08, with an 88% accuracy, and E07 are both suitable for these referents. Given the similarity in their movements (nodding with varying body involvement), these results align with expectations. Referent R8 had two matching expressions (E12 and E13). E12 had a higher qualitative response accuracy (73%) compared to E13 (21%) and was significantly more engaged, understandable, effective, and noticeable. Consequently, E12 emerged as the superior choice for expressing robot rejection of a person's statement. Thus, head shaking is more understandable than turning the body away.

This shows that we successfully created human-understandable expressions for visual curiosity (E1, E2, E3) and conversational curiosity (E07, E08, E09). These expressions effectively conveyed the robot's interest in visual objects and attentiveness to spoken conversation. However, auditory curiosity expressions faced challenges and were poorly understood by participants. The effective visual curiosity expressions involved the robot approaching the object, inspecting it from various angles, and using its entire body for movement. In contrast, effective listening attention expressions primarily featured nodding with the robot's end-effector, while another expression involved body movement, shifting from facing away to facing and approaching the users. We attribute the success of these expressions to (1) the contextual relevance of the target object and (2) the clarity of expressions associated with listening attention, such as nodding. Auditory exploration expressions lacked these advantageous properties, contributing to their lower participant comprehension rates.

5.5.3 Obvious Expressions are Understandable and Context Matters. Our study demonstrated that nodding and head shaking effectively signal interactions with the robot, whether inviting or rejecting conversations. These findings align with human behavior, where a limited set of expressions is employed to convey various emotional states [5], emphasizing the importance of familiar and comprehensible expressions. Additionally, we found that even with a non-humanoid robot, participants generated human-like expressions, and visual curiosity gestures—such as examining objects—were well understood. Expressions indicating attentiveness to spoken conversation also led to high participant comprehension.

However, the study revealed challenges in participants' understanding of auditory curiosity expressions. We attributed this confusion to the absence of context in the gestures. By filming the robot's expressions without providing contextual information, participants struggled to interpret the intended meaning behind auditory curiosity expressions. This highlights the significance of incorporating context into expressions, which can substantially enhance their interpretability in human-robot interactions.

5.6 Limitations and Legacy Bias

In our example study, we intentionally refrained from giving participants too much context in the expression creation process. Here, we aimed to maximize generalizability and avoid priming participants. However, we observed that expressions tied to direct contextual cues (e.g., visual exploration vs. auditory exploration) tended to perform better (see Figure 6). Giving participants more context and limitations in the expression creation process might increase the robustness of the resulting expressions. However, this then leads to more specific expressions, taking away some generalizability. Thus, this is a trade-off that each researcher needs to evaluate for their specific context when applying our approach.

We tested our approach in one example study, creating expressions for the trait curiosity. While we created our five goals of the approach to ensure generalizability, we, for now, can not prove how applicable it is for other robots. Our approach allows future work to create expressions for other traits and robots, showing whether it is generalizable over time.

The issue of "legacy bias" in gesture elicitation studies, as highlighted by Morris et al. [70], raises a valid point. It suggests that when participants are asked to create gestures for novel interaction capabilities, well-known and familiar gestures tend to emerge as the preferred choices. For instance, when prompted to create a gesture for zooming, many participants might instinctively choose a pinching motion, reflecting their familiarity with touchscreens. While this legacy bias might initially seem limiting, it does not necessarily have to be viewed as a negative aspect. Human-computer interaction has evolved over decades, and users have become accustomed to certain interaction concepts. Requiring users to relearn entirely new ways of interacting with technical devices each time a novel interaction capability emerges could be cumbersome and counterproductive. In this context, our method, which prioritizes well-known and obvious expressions, can be seen as a valuable tool. It allows researchers to systematically identify and leverage the most common and widely understood non-verbal communication cues. These cues are already ingrained in users' minds due to their experience with various technical devices. Thus, our approach enables researchers to tap into this existing knowledge base and build upon well-established interaction concepts, making the transition to new technology more seamless and intuitive for users.

6 Conclusion

In this work, we presented an approach to elicit and generate human-understandable robot expressions. This helps to create better human-robot interaction as a better understanding of robots' intentions leads to more seamless collaboration. Our approach uses ideas from gesture elicitation studies, human creativity and intuition, and wide verification. We showed our approach with an example case of creating understandable, curious robot expressions with a non-humanoid robot. By employing our approach, future work can show whether our approach is also applicable to different robot types. We have successfully created expressions to deliver five out of six referents correctly. We found that robot expressions,

in contrast to interaction gestures, do not have to be distinctly different from each other, and the same expressions can be the best way to convey different robot states. Through our study, we found that familiar expressions tend to be understood better than novel expressions. With this approach, we aim to support future work for HRI; through a way to easily and rapidly fabricate understandable robotic expressions.

7 Open Science

We encourage readers to review, reproduce, and extend our results and analysis methods. To achieve this goal, we make available our coding process and taxonomy classification for the expression elicitation phase and study results, material, coding process, referent matching, and grouping for the expression verification phase at this link https://osf.io/3sywc.

References

[1] Naoko Abe, Jean-Paul Laumond, Paolo Salaris, and Florent Levillain. 2017. On the Use of Dance Notation Systems to Generate Movements in Humanoid Robots: The Utility of Laban Notation in Robotics. 56, 2 (2017), 328–344. doi:10.1177/0539018417694773

[2] Mustafa Altun. 2019. An Underestimated Tool: Body Language in Classroom during Teaching and Learning. *International Journal of Social Sciences & Educational Studies* 6, 1 (2019). doi:10.23918/ijsses.v6i1p155

[3] Lucy Anderson-Bashan, Benny Megidish, Hadas Erel, Iddo Wald, Guy Hoffman, Oren Zuckerman, and Andrey Grishko. 2018. The Greeting Machine: An Abstract Robotic Object for Opening Encounters. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 595–602. doi:10.1109/roman.2018.8525516

[4] P. Asensio, R. Arbós, A. Gomà, and A. Pérez. 2011. The Use of Role Playing in Engineering Curricula: A Case Study in Human-Automation Systems. *International Journal of Engineering Education* (2011).

[5] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest* 20, 1 (July 2019), 1–68. doi:10.1177/1529100619832930

[6] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. doi:10.1007/s12369-008-0001-3

[7] Paul Baxter, James Kennedy, Emmanuel Senft, Severin Lemaignan, and Tony Belpaeme. 2016. From Characterising Three Years of HRI to Methodology and Reporting Recommendations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch, 2016-03). IEEE, 391–398. doi:10.1109/HRI.2016.7451777

[8] Cindy L. Bethel and Robin R. Murphy. 2008. Survey of Non-facial/Non-verbal Affective Expressions for Appearance-Constrained Robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38, 1 (Jan. 2008), 83–92. doi:10.1109/tsmcc.2007.905845

[9] Alisha Bevins and Brittany A. Duncan. 2021. Aerial Flight Paths for Communication: How Participants Perceive and Intend to Respond to Drone Movements. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder CO USA, 2021-03-08). ACM, 16–23. doi:10.1145/3434073.3444645

[10] Godfred O. Boateng, Torsten B. Neilands, Edward A. Frongillo, Hugo R. Melgar-Quiñonez, and Sera L. Young. 2018. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health* 6 (2018).

[11] Elizabeth Baraff Bonawitz, Tessa J.P. van Schijndel, Daniel Friel, and Laura Schulz. 2012. Children Balance Theories and Evidence in Exploration, Explanation, and Learning. *Cognitive Psychology* (2012).

[12] Luc Boutin, Antoine Eon, Said Zeghloul, and Patrick Lacouture. 2010. An Auto-Adaptable Algorithm to Generate Human-like Locomotion for Different Humanoid Robots Based on Motion Capture Data. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1256–1261. doi:10.1109/iros.2010.5652230

[13] C. Breazeal, C.D. Kidd, A.L. Thomaz, G. Hoffman, and M. Berlin. 2005. Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 708–713. doi:10.1109/iros.2005.1545011

[14] Judee K Burgoon, Valerie Manusov, and Laura K Guerrero. 2016. *Nonverbal Communication*. Routledge, New York. doi:10.4324/9781315663425

[15] Sarah Jane Burton, Ali-Akbar Samadani, Rob Gorbet, and Dana Kulić. 2016. Laban Movement Analysis and Affective Movement Generation for Robots and Other Near-Living Creatures. In *Dance Notations and Robot Motion*, Jean-Paul Laumond and Naoko Abe (Eds.). Vol. 111. Springer International Publishing, 25–48. doi:10.1007/978-3-319-25739-6_2

[16] John-John Cabibihan, Wing-Chee So, and Soumo Pramanik. 2012. Human-Recognizable Robotic Gestures. *IEEE Transactions on Autonomous Mental Development* 4, 4 (Dec. 2012), 305–314. doi:10.1109/tamd.2012.2208962

[17] Maya Cakmak and Leila Takayama. 2014. Teaching People How to Teach Robots: The Effect of Instructional Materials and Dialog Design. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*. ACM, Bielefeld Germany, 431–438. doi:10.1145/2559636.2559675

[18] Jessy Ceha, Nalin Chhibber, Joslin Goh, Corina McDonald, Pierre-Yves Oudeyer, Dana Kulić, and Edith Law. 2019. Expression of Curiosity in Social Robots: Design, Perception, and Effects on Behaviour. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Chi '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300636

[19] Ergene Mehmet Celalettin, Durdu Akif, and C. Halil. 2016. Imitation and Learning of Human Hand Gesture Tasks of the 3D Printed Robotic Hand by Using Artificial Neural Networks.

[20] Elizabeth Cha, Maja Matarić, and Terrence Fong. 2016. Nonverbal Signaling for Non-Humanoid Robots during Human-Robot Collaboration. *The eleventh ACM/IEEE international conference on human robot interaction* (2016). doi:10.5555/2906831.2907001

[21] Michael Chamunorwa, Mikolaj P. Wozniak, Susanna Krämer, Heiko Müller, and Susanne Boll. 2023. An Empirical Comparison of Moderated and Unmoderated Gesture Elicitation Studies on Soft Surfaces and Objects for Smart Home Control. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (Sept. 2023), 198:1–198:24. doi:10.1145/3604245

[22] Andy Cockburn and Carl Gutwin. 2019. Anchoring Effects and Troublesome Asymmetric Transfer in Subjective Ratings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. doi:10.1145/3290605.3300592

[23] Nele Dael, Martijn Goudbeek, and K R Scherer. 2013. Perceived Gesture Dynamics in Nonverbal Expression of Emotion. *Perception* 42, 6 (June 2013), 642–657. doi:10.1068/p7364

[24] Nele Dael, Marcello Mortillaro, and Klaus R. Scherer. 2012. Emotion Expression in Body Action and Posture. *Emotion* 12, 5 (2012), 1085–1101. doi:10.1037/a0025737

[25] Charles Darwin. 1872. *The Expression of the Emotions in Man and Animals*.

[26] B de Gelder, A W de Borst, and R Watson. 2014. The Perception of Emotion in Body Expressions. (2014).

[27] Eric Deng. 2018. Generative Methods for Object-Based Gestures for Socially Interactive Robots.

[28] Amol Deshmukh, Bart Craenen, Alessandro Vinciarelli, and Mark Foster. 2018. 'shaping Robot Gestures to Shape Users' Perception: The Effect of Amplitude and Speed on Godspeed Ratings'. (2018). doi:10.1145/3284432.3284445

[29] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. 2018. Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. doi:10.1145/3173574.3173993

[30] Stephanie Embgen, Matthias Luber, Christian Becker-Asano, Marco Ragni, Vanessa Evers, and Kai Arras. 2012. Robot-Specific Social Cues in Emotional Body Language. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. 1019–1025. doi:10.1109/roman.2012.6343883

[31] Frank Esser and Rens Vliegenthart. 2017. Comparative Research Methods. In *The International Encyclopedia of Communication Research Methods* (1 ed.), Jörg Matthes, Christine S. Davis, and Robert F. Potter (Eds.). Wiley, 1–22. doi:10.1002/9781118901731.iecrm0035

[32] Mingming Fan, Yizheng Ding, Fang Shen, Yuhui You, and Zhi Yu. 2017. An Empirical Study of Foot Gestures for Hands-Occupied Mobile Interaction. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers (Iswc '17)*. Association for Computing Machinery, New York, NY, USA, 172–173. doi:10.1145/3123021.3123043

[33] Simona Fer. 2018. Verbal Communication As a Two-Way Process in Connecting People. doi:10.2139/ssrn.3128115

[34] Justin W. Firestone, Rubi Quinones, and Brittany A. Duncan. 2019. Learning from Users: An Elicitation Study and Taxonomy for Communicating Small Unmanned Aerial System States Through Gestures. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu, Korea (South), 2019-03). IEEE, 163–171. doi:10.1109/HRI.2019.8673010

[35] Brian Gleeson, Karon MacLean, Amir Haddadi, Elizabeth Croft, and Javier Alcazar. 2013. Gestures for Industry: Intuitive Human-Robot Communication from Human Observation. In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction (Hri '13)*. IEEE Press, Tokyo, Japan, 349–356.

[36] Susan Goldin-Meadow. 1999. The Role of Gesture in Communication and Thinking. *Trends in Cognitive Sciences* 3, 11 (Nov. 1999), 419–429. doi:10.1016/

s1364-6613(99)01397-2
[37] Susan Goldin-Meadow and Martha Wagner Alibali. 2013. Gesture's Role in Speaking, Learning, and Creating Language. *Annual Review of Psychology* 64, 1 (2013), 257–283. doi:10.1146/annurev-psych-113011-143802
[38] Jacqueline Gottlieb, Mary Hayhoe, Okihide Hikosaka, and Antonio Rangel. 2014. Attention, Reward, and Information Seeking. *The Journal of Neuroscience* (2014).
[39] Anthony G. Greenwald. 1976. Within-Subjects Designs: To Use or Not to Use? *Psychological Bulletin* 83, 2 (March 1976), 314–320. doi:10.1037/0033-2909.83.2.314
[40] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (Oct. 2006), 904–908. doi:10.1177/154193120605000909
[41] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. doi:10.1016/S0166-4115(08)62386-9
[42] Thomas Hellström and Suna Bensch. 2018. Understandable Robots - What, Why, and How. 9, 1 (2018), 110–123. doi:10.1515/pjbr-2018-0009
[43] Alexis Heloir, Sylvie Gibet, Franck Multon, and Nicolas Courty. 2006. Captured Motion Data Processing for Real Time Synthesis of Sign Language. In *Gesture in Human-Computer Interaction and Simulation (Lecture Notes in Computer Science)*, Sylvie Gibet, Nicolas Courty, and Jean-François Kamp (Eds.). Springer, Berlin, Heidelberg, 168–171. doi:10.1007/11678816_19
[44] Guy Hoffman and Wendy Ju. 2014. Designing Robots with Movement in Mind. *Journal of Human-Robot Interaction* 3, 1 (Feb. 2014), 91–122. doi:10.5898/JHRI.3.1.Hoffman
[45] Guy Hoffman and Xuan Zhao. 2021. A Primer for Conducting Experiments in Human–Robot Interaction. *ACM Transactions on Human-Robot Interaction* 10, 1 (March 2021), 1–31. doi:10.1145/3412374
[46] Guy Hoffman, Oren Zuckerman, Gilad Hirschberger, Michal Luria, and Tal Shani Sherman. 2015. Design and Evaluation of a Peripheral Robotic Conversation Companion. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (Hri '15)*. Association for Computing Machinery, New York, NY, USA, 3–10. doi:10.1145/2696454.2696495
[47] Marius Hoggenmueller, Jiahao Chen, and Luke Hespanhol. 2020. Emotional Expressions of Non-Humanoid Urban Robots: The Role of Contextual Aspects on Interpretations. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays (PerDis '20)*. Association for Computing Machinery, New York, NY, USA, 87–95. doi:10.1145/3393712.3395341
[48] Malte F. Jung. 2017. Affective Grounding in Human-Robot Interaction. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 263–273. doi:10.1145/2909824.3020224
[49] Peter H. Kahn, Brian T. Gill, Aimee L. Reichert, Takayuki Kanda, Hiroshi Ishiguro, and Jolina H. Ruckert. 2010. Validating Interaction Patterns in HRI. In *Proceeding of the 5th ACM/IEEE International Conference on Human-robot Interaction - HRI '10* (Osaka, Japan, 2010). ACM Press, 183. doi:10.1145/1734454.1734534
[50] J. Karat and C. M. Karat. 2003. The Evolution of User-Centered Focus in the Human-Computer Interaction Field. *IBM Systems Journal* 42, 4 (2003), 532–541. doi:10.1147/sj.424.0532
[51] Adam Kendon. 2004. *Gesture: Visible Action as Utterance.* Cambridge University Press, Cambridge. doi:10.1017/cbo9780511807572
[52] Celeste Kidd and Benjamin Y. Hayden. 2015. The Psychology and Neuroscience of Curiosity. *Neuron* 88, 3 (Nov. 2015), 449–460. doi:10.1016/j.neuron.2015.09.010
[53] S. Kiesler. 2005. Fostering Common Ground in Human-Robot Interaction. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* 729–734. doi:10.1109/roman.2005.1513866
[54] Aelee Kim, Hyejin Kum, Ounjeong Roh, Sangseok You, and Sukhan Lee. 2012. 'Robot Gesture and User Acceptance of Information in Human-Robot Interaction'. *Association for Computing Machinery* (2012). doi:10.1145/2157689.2157793
[55] Heon-Hui Kim, Hyong-Euk Lee, Yong-hwi Kim, Kwang-Hyun Park, and Z. Zenn Bien. 2007. Automatic Generation of Conversational Robot Gestures for Human-friendly Steward Robot. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication.* 1155–1160. doi:10.1109/roman.2007.4415254
[56] Heather Knight, Ravenna Thielstrom, and Reid Simmons. 2016. Expressive Path Shape (Swagger): Simple Features That Illustrate a Robot's Attitude toward Its Goal in Real Time. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Daejeon, South Korea, 2016-10). IEEE, 1475–1482. doi:10.1109/IROS.2016.7759240
[57] Amy Koike and Bilge Mutlu. 2023. Exploring the Design Space of Extra-Linguistic Expression for Robots. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference.* ACM, Pittsburgh PA USA, 2689–2706. doi:10.1145/3563657.3595968
[58] Markus Koppensteiner. 2011. Perceiving Personality in Simple Motion Cues. *Journal of Research in Personality* 45, 4 (Aug. 2011), 358–363. doi:10.1016/j.jrp.2011.04.003
[59] Shikhar Kumar, Eliran Itzhak, Yael Edan, Galit Nimrod, Vardit Sarne-Fleischmann, and Noam Tractinsky. 2022. Politeness in Human–Robot Interaction: A Multi-Experiment Study with Non-Humanoid Robots. *International Journal of Social Robotics* 14, 8 (Oct. 2022), 1805–1820. doi:10.1007/s12369-022-00911-z
[60] Marianne LaFrance and Clara Mayo. 1978. Cultural Aspects of Nonverbal Communication. *International Journal of Intercultural Relations* 2, 1 (March 1978), 71–89. doi:10.1016/0147-1767(78)90029-9
[61] Javier Laplaza, Nicolas Rodriguez, J. E. Dominguez-Vidal, Fernando Herrero, Sergi Hernandez, Alejandro Lopez, Alberto Sanfeliu, and Anais Garrell. 2022. IVO Robot: A New Social Robot for Human-Robot Collaboration. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Ieee, Sapporo, Japan, 860–864. doi:10.1109/hri53351.2022.9889458
[62] A. Lascarides and M. Stone. 2009. A Formal Semantic Analysis of Gesture. *Journal of Semantics* 26, 4 (Nov. 2009), 393–449. doi:10.1093/jos/ffp004
[63] Manja Lohse, Reinier Rothuis, Jorge Gallego-Pérez, Daphne E. Karreman, and Vanessa Evers. 2014. Robot Gestures Make Difficult Tasks Easier: The Impact of Gestures on Perceived Workload and Task Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Acm, Toronto Ontario Canada, 1459–1466. doi:10.1145/2556288.2557274
[64] Jeremy A. Marvel, Shelly Bagchi, Megan Zimmerman, Murat Aksu, Brian Antonishek, Yue Wang, Ross Mead, Terry Fong, and Heni Ben Amor. 2019. Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu, Korea (South), 2019-03). IEEE, 696–697. doi:10.1109/HRI.2019.8673149
[65] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. 2016. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (Chi '16)*. Association for Computing Machinery, New York, NY, USA, 5421–5432. doi:10.1145/2858036.2858063
[66] Nikolaos Mavridis. 2015. A Review of Verbal and Non-Verbal Human–Robot Interactive Communication. *Robotics and Autonomous Systems* 63 (Jan. 2015), 22–35. doi:10.1016/j.robot.2014.09.031
[67] David McNeill. 2005. *Gesture and Thought.* University of Chicago Press, Chicago.
[68] Kristina Moroz-Lapin. 2009. Role Play in HCI Studies. In *HCI Educators 2009 - Playing with Our Education.* BCS Learning & Development. doi:10.14236/ewic/HCIED2009.12
[69] Meredith Ringel Morris. 2012. Web on the Wall: Insights from a Multimodal Interaction Elicitation Study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (Its '12)*. Association for Computing Machinery, New York, NY, USA, 95–104. doi:10.1145/2396636.2396651
[70] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *Interactions* 21, 3 (2014), 40–45. doi:10.1145/2591689
[71] A. Nakazawa, S. Nakaoka, K. Ikeuchi, and K. Yokoi. 2002. Imitating Human Dance Motions through Motion Structure Analysis. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 3. 2539–2544 vol.3. doi:10.1109/irds.2002.1041652
[72] Carl Oechsner, Sven Mayer, and Andreas Butz. 2022. Challenges and Opportunities of Cooperative Robots as Cooking Appliances. (2022), 7.
[73] Deepika Phutela. 2015. The Importance of Non-Verbal Communication. *IUP Journal of Soft Skills* 9, 4 (Dec. 2015), 43–49.
[74] Jean Piaget. 1945. *Play, Dreams and Imitation in Childhood.* The International Library of Psychology.
[75] Isabella Poggi, Francesca D'Errico, and Laura Vincze. 2010. Types of Nods. The Polysemy of a Social Signal. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (19-21), Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias (Eds.). European Language Resources Association (ELRA), Valletta, Malta.
[76] Viva Sarah Press and Hadas Erel. 2023. Humorous Robotic Behavior as a New Approach to Mitigating Social Awkwardness. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* Acm, Hamburg Germany, 1–16. doi:10.1145/3544548.3580821
[77] B. Balaji Reddy and Dr. S. Krishnaiah Dr. S. Krishnaiah. 2012. Non Verbal Communication Cues: An Urgent Required Skill for Achieving Effective Communication. *International Journal of Scientific Research* 2, 11 (June 2012), 212–213. doi:10.15373/22778179/nov2013/68
[78] Byron Reeves and Clifford Nass. 1996. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Pla. *Bibliovault OAI Repository, the University of Chicago Press* (Jan. 1996).
[79] Carol E. Reiley, Erion Plaku, and Gregory D. Hager. 2010. Motion Generation of Robotic Surgical Tasks: Learning from Expert Demonstrations. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology.* 967–970. doi:10.1109/iembs.2010.5627594
[80] Ulf-Dietrich Reips and Frederik Funke. 2008. Interval-Level Measurement with Visual Analogue Scales in Internet-based Research: VAS Generator. *Behavior Research Methods* 40, 3 (Aug. 2008), 699–704. doi:10.3758/brm.40.3.699

[81] Tiago Ribeiro and Ana Paiva. 2020. The Practice of Animation in Robotics. In *Modelling Human Motion: From Human Perception to Robot Design*, Nicoletta Noceti, Alessandra Sciutti, and Francesco Rea (Eds.). Springer International Publishing, Cham, 237–269. doi:10.1007/978-3-030-46732-6_12

[82] José Rosado, Filipe Silva, and Vítor Santos. 2014. A Kinect-Based Motion Capture System for Robotic Gesture Imitation. In *ROBOT2013: First Iberian Robotics Conference: Advances in Robotics, Vol. 1*, Manuel A. Armada, Alberto Sanfeliu, and Manuel Ferre (Eds.). Springer International Publishing, Cham, 585–595. doi:10.1007/978-3-319-03413-3_43

[83] José Rosado, Filipe Silva, and Vítor Santos. 2014. Using Kinect for Robot Gesture Imitation. *Procedia Technology* 17 (Jan. 2014), 423–430. doi:10.1016/j.protcy.2014.10.250

[84] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-Defined Motion Gestures for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Chi '11)*. Association for Computing Machinery, New York, NY, USA, 197–206. doi:10.1145/1978942.1978971

[85] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (Chi Play '19)*. Association for Computing Machinery, New York, NY, USA, 199–211. doi:10.1145/3311350.3347190

[86] Paolo Salaris, Naoko Abe, and Jean-Paul Laumond. 2016. A Worked-Out Experience in Programming Humanoid Robots via the Kinetography Laban. In *Dance Notations and Robot Motion*, Jean-Paul Laumond and Naoko Abe (Eds.). Vol. 111. Springer International Publishing, 339–359. doi:10.1007/978-3-319-25739-6_16

[87] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To Err Is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics* (2013).

[88] Maha Salem, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing, and Frank Joublin. 2012. Generation and Evaluation of Communicative Robot Gesture. *International Journal of Social Robotics* 4, 2 (April 2012), 201–217. doi:10.1007/s12369-011-0124-9

[89] Shane Saunderson and Goldie Nejat. 2019. How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction. *International Journal of Social Robotics* 11, 4 (Aug. 2019), 575–608. doi:10.1007/s12369-019-00523-0

[90] T.B. Sheridan. 1997. Eight Ultimate Challenges of Human-Robot Communication. In *Proceedings 6th IEEE International Workshop on Robot and Human Communication. RO-MAN'97 SENDAI*. 9–14. doi:10.1109/roman.1997.646944

[91] Mei Si and Joseph Dean McDaniel. 2016. Using Facial Expression and Body Language to Express Attitude for Non-Humanoid Robot: (Extended Abstract). In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (Aamas '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1457–1458.

[92] Candace L. Sidner, Christopher Lee, Cory Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in Engagement for Humans and Robots. doi:10.48550/arXiv.cs/0507056 arXiv:cs/0507056

[93] Chaklam Silpasuwanchai and Xiangshi Ren. 2015. Designing Concurrent Full-Body Gestures for Intense Gameplay. *International Journal of Human-Computer Studies* 80 (Aug. 2015), 1–13. doi:10.1016/j.ijhcs.2015.02.010

[94] Paul J. Silvia. 2012. Curiosity and Motivation. In *The Oxford Handbook of Human Motivation*, Richard M. Ryan (Ed.). Oxford University Press, 0. doi:10.1093/oxfordhb/9780195399820.013.0010

[95] Michael Suguitan, Nick DePalma, Guy Hoffman, and Jessica Hodgins. 2023. Face2Gesture: Translating Facial Expressions Into Robot Movements Through Shared Latent Space Neural Networks. *ACM Transactions on Human-Robot Interaction* (Oct. 2023), 3623386. doi:10.1145/3623386

[96] Haodan Tan, John Tiab, Selma Šabanović, and Kasper Hornbæk. 2016. Happy Moves, Sad Grooves: Using Theories of Biological Motion and Affect to Design Shape-Changing Interfaces. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (Dis '16)*. Association for Computing Machinery, New York, NY, USA, 1282–1293. doi:10.1145/2901790.2901845

[97] Yunus Terzioğlu, Bilge Mutlu, and Erol Şahin. 2020. Designing Social Cues for Collaborative Robots: The Role of Gaze and Breathing in Human-Robot Collaboration. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Acm, Cambridge United Kingdom, 343–357. doi:10.1145/3319502.3374829

[98] Ashish Tiwari. 2016. Non-verbal Communication-an Essence Of Interpersonal Relationship At Workplace. *Management Insight* (Jan. 2016).

[99] Gabriele Trovato, Tatsuhiro Kishi, Nobutsuna Endo, Kenji Hashimoto, and Atsuo Takanishi. 2012. A Cross-Cultural Study on Generation of Culture Dependent Facial Expressions of Humanoid Social Robot. In *Social Robotics*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, and Mary-Anne Williams (Eds.). Vol. 7621. Springer Berlin Heidelberg, Berlin, Heidelberg, 35–44. doi:10.1007/978-3-642-34103-8_4

[100] Vaibhav Unhelkar and Jessie X. Yang. 2017. Challenges for Communication Decision-Making in Sequential Human-Robot Collaborative Tasks.

[101] Greet Van de Perre, Michaël Van Damme, Dirk Lefeber, and Bram Vanderborght. 2015. Development of a Generic Method to Generate Upper-Body Emotional Expressions for Different Social Robots. *Advanced Robotics* 29, 9 (May 2015), 597–609. doi:10.1080/01691864.2015.1031697

[102] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Acm, Seoul Republic of Korea, 1325–1334. doi:10.1145/2702123.2702223

[103] Santiago Villarreal-Narvaez, Arthur Sluÿters, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2024. Brave New GES World: A Systematic Literature Review of Gestures and Referents in Gesture Elicitation Studies. *Comput. Surveys* 56, 5 (May 2024), 1–55. doi:10.1145/3636458

[104] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Acm, Eindhoven Netherlands, 855–872. doi:10.1145/3357236.3395511

[105] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 2018. Gesture Elicitation Studies for Mid-Air Interaction: A Review. *Multimodal Technologies and Interaction* 2, 4 (Sept. 2018), 65. doi:10.3390/mti2040065

[106] Rudolf von Laban. 1972. *Principles of dance and movement notation*. A Dance Horizons republication.

[107] Jinjing (Jenny) Wang, Yang Yang, Carla Macias, and Elizabeth Bonawitz. 2021. Children With More Uncertainty in Their Intuitive Theories Seek Domain-Relevant Information. *Psychological Science* 32, 7 (July 2021), 1147–1156. doi:10.1177/0956797621994230

[108] Stephanie Wermelinger, Anja Gampe, Natascha Helbling, and Moritz M Daum. 2020. Do You Understand What I Want to Tell You? Early Sensitivity in Bilinguals' Iconic Gesture Perception and Production. *Developmental Science* (2020).

[109] Daniel Wessel, Moreen Heine, Christiane Attig, and Thomas Franke. 2020. Affinity for Technology Interaction and Fields of Study: Implications for Human-Centered Design of Applications for Public Administration. In *Proceedings of Mensch Und Computer 2020 (MuC '20)*. Association for Computing Machinery, New York, NY, USA, 383–386. doi:10.1145/3404983.3410020

[110] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

[111] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Acm, Boston MA USA, 1083–1092. doi:10.1145/1518701.1518866

[112] Pieter Wolfert, Nicole Robinson, and Tony Belpaeme. 2022. A Review of Evaluation Practices of Gesture Generation in Embodied Conversational Agents. *IEEE Transactions on Human-Machine Systems* 52, 3 (June 2022), 379–389. doi:10.1109/thms.2022.3149173 arXiv:2101.03769 [cs]

[113] Qiong Wu and Chunyan Miao. 2013. Curiosity: From Psychology to Computation. *Comput. Surveys* 46, 2 (Dec. 2013), 18:1–18:26. doi:10.1145/2543581.2543585

[114] Yaqi Xie, Indu P Bodala, Desmond C. Ong, David Hsu, and Harold Soh. 2019. Robot Capability and Intention in Trust-Based Decisions Across Tasks. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu, Korea (South), 2019-03). IEEE, 39–47. doi:10.1109/HRI.2019.8673084

[115] Katsu Yamane, Yuka Ariki, and Jessica Hodgins. 2010. *Animating Non-Humanoid Characters with Human Motion Data*. The Eurographics Association. doi:10.2312/sca/sca10/169-178

[116] Jeong-Yean Yang and Dong-Soo Kwon. 2014. Robotic Gesture Generation Based on a Cognitive Basis for Non-Verbal Communication. In *2014 11th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. 683–687. doi:10.1109/urai.2014.7057497

[117] Yiju Yang and Andrew B. Williams. 2021. Improving Human-Robot Collaboration Efficiency and Robustness through Non-Verbal Emotional Communication. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 354–356. doi:10.1145/3434074.3447191

[118] Tom Ziemke. 2020. Understanding Robots. *Science Robotics* 5, 46 (Sept. 2020), eabe2987. doi:10.1126/scirobotics.abe2987