GRADUATION PROJECT

BODY LANGUAGE FOR BOTS:
Adapting a virtual robot's gestures by means of an Interactive Evolutionary Algorithm

August 31, 2017

Helena Frijns
Thesis advisors: Maarten H. Lamers and Tessa Verhoef

Media Technology MSc programme, Leiden University
mail@helenafrijns.com

Contents

Abstract

Robots find application in contexts as diverse as retail, factory work and elderly care. In a social context, it is important that robots can communicate with humans and adapt their behaviour to a changing environment. We raised the question whether we can adapt a robot's gesture behaviour based on the effectiveness of its communication towards humans. We developed a system that made use of an interactive evolutionary algorithm to develop gestures for a virtual robot. In addition, we investigated if and how interactive evolutionary algorithms can be suitable for the development of gestures for robots. In the experiment we performed (N=16), we found indications that it may well be possible to develop gestures that can be interpreted by other people this way. Additionally, a majority of participants indicated to perceive improvements in the gestures over successive generations of the evolutionary algorithm. However, improvements of the system are necessary, as well as more experiments to confirm the results. To investigate whether gestures that are based on the interaction of multiple people with the system are easier to interpret than gestures developed by an individual, we propose an experiment in which gestures are developed by multiple people by means of a transmission chain experiment.

1 Introduction

Robots are developed for tasks as diverse as building houses (Fastbrick Robotics), keeping track of stock in retail stores [73] and supporting children with diabetes [59]. As they can perform physically heavy and dangerous tasks, they are increasingly employed in factories and warehouses, and their use in disaster response is investigated. The TRADR[1] project investigates the exploration of disaster areas by human-robot teams of unmanned aerial and ground vehicles and the humans that teleoperate them [44]. Amazon expanded its use of robots that work alongside its 230 000 human employees to 45000 in 2016 [72].

Robots no longer operate solely behind fences in factories. In these new situations, whether it is in a human-robot team or in the context of social robotics, it is important that robots can communicate with humans, indicate that they understood messages and adapt their behaviour to new situations.

The effectiveness of social interactions can be quite reduced when mediated by naïve technological interfaces. We are already familiar with technologies that mediate our social interactions. We send each other messages via our mobile phones and e-mail. However, the affective expressiveness of messages sent this way is greatly reduced, compared to speaking to each other face to face.[2] Likewise, a limited set of responses in a robot will make long-term interaction less interesting [45].

Robot faces are often designed in a humanoid fashion, as they can give us affective feedback in a way we are already used to. In a sense, they afford interaction [7]. Smiling means something to us immediately. When a robot communicates a certain affective state to humans, humans can recognize the mood that is being communicated and the affective communication can have an effect on human task performance [87].

A number of existing models for affective social robot feedback contain a limited number of iconic expressions, for instance happiness, anger and surprise [7, 53]. These expressions are often based on a two- or three-dimensional affective space with dimensions such as valence, arousal and stance. However, not all emotions can be convincingly expressed by a limited number of facial expressions. Additionally, the way expressions are interpreted depends on the context. Hard-coding every single possible interaction is not practical.

Body language and gestures can aid the understanding of a message by (human) interaction partners. Gestures can take on a language-like role when it is performed by itself. Gestures that accompany speech can convey additional or different information, thereby enriching communication [31].

Learning from humans by imitation is a possible solution that could help accommodate the expression of complex social information. Breazeal and Scassellati describe imitation as a powerful tool for social learning in robots [10]. However, this implies a humanoid or android body, and these have their problems. Humanoid robots and androids are costly and complex to develop. The uncanny valley hypothesis poses that humanoids and androids can appear a bit off, or even downright frightening, as their realism approaches their human model [56]. Modelling a robot on the human form also poses many restrictions and brings up expectations and requirements for the way the robot can and is supposed to move and interact.

Is it necessary to model robots after humans? Much simpler robots could also be effective. Humans tend to project complex intentions and behaviours onto machines that are not human-like at all. The simple vehicles Braitenberg describes serve as an illustration of these projective inclinations [8]. Human beings tend to ascribe characteristics such as being able to feel emotions to non-human animals and even objects or moving shapes [71].

What should the shape of such a simple robot be? What are the modalities that it can use to communicate? If a robot could not only alter its behaviour, but adapt its capabilities and body, based on the success or failure of its attempts at communication, these questions could be explored by the robot itself, in context. Additionally, if robots could evolve behaviours in the context of interacting with humans, this might create behaviour that is more "meaningful": behaviour that is grounded in interaction. The change of interaction over time would give the human and the system a shared history, and behaviour could become personalized.

In humans, language acquisition takes place in the context of interaction, as well as the conversational use of language. Another motivation for letting communicative behaviours adapt and evolve within the context of interaction, is that work on human-human communication shows that cultural transmission can lead to increased structure and transmissibility of language. The language can appear designed,

[1] TRADR is an acronym for Long-Term Human-Robot Teaming for Robot Assisted Disaster Response.

[2] Attempts have been to mediate affective content in the form of emoticons and emoji. However, interpretations tend to vary within- and across-platforms [52].

without conscious interference of a designer [42].

This raises the following questions. How can we create the embodiment and behaviour of a robot without directly modelling these on humans? Would using a system that can adapt its behaviour result in more engaging interaction in the long term? *Could we adapt a robot's behaviour based on the effectiveness of its communication towards humans?*

We will try to answer this last question by developing a system that makes use of an interactive evolutionary algorithm to suggest gestures to a human experiment participant, and to subsequently adapt those gestures. We will first review the literature on current designs of (social) robot embodiments and potential shortcomings. We will describe some effects of robot motion and behaviour, and discuss ways these have been achieved within a "social" context, in which robots can learn from one another. An approach is proposed that involves adapting a robot's gestures based on interaction with humans. We describe the development of a system that intends to achieve this, and an experiment in which the developed system is tested. Developing suitable gestures is not the only goal of the system. We will also consider the way human participants interact with the developed system. From this we hope to infer what the requirements are for a system that makes use of an interactive evolutionary algorithm to develop gestures. The terminology used in this paper is clarified in the Appendix.

2 Literature Review: Designing Robots for Interaction

2.1 Embodiment design of social robots

2.1.1 Anthropomorphic designs

Anthropomorphism is a frequently applied strategy when it comes to designing a social robot's embodiment or behaviour. Humanoid social robots can share human physical characteristics and behaviours, varying from more realistic to more abstract. Examples of humanoid social robots are the NAO and Pepper robots by Aldebaran robotics [2, 3], MIT Media Lab's Nexi and Kismet [53, 54], Honda's ASIMO [37] and Waseda University's KOBIAN [85]. Androids are developed to be as similar to humans as possible, and are often modelled after individual human beings. Robot Nadine from Nanyang Technological University Singapore was developed to take on the role of a receptionist [57]. Geminoids have also been developed by Hiroshi Ishiguro [35] and Hanson Robotics [36].

Robots that were not intended as social robots are also frequently designed using humanoid forms. This can allow for more intuitive control by human operators. Stanford´s diver robot Ocean One can give touch feedback to the hands of a remote human teleoperator based on objects touching the robot's hands [13]. Most of the robots that competed in the 2015 DARPA robotics challenge were humanoid. The robot Handle, which was recently developed by Boston Dynamics, has been described as a chimera merging the human form with wheels [74].

An argument for anthropomorphism is an increase in familiarity and as a result being able to relate to the robot and understand the actions it performs. However, anthropomorphic embodiments in social contexts may create large expectations in users, which can lead to disappointments if these expectations are not met [19].

Another risk factor is that anthropomorphic designs can be perceived as uncanny or threatening. The concept of the uncanny valley is a nonlinear relationship proposed by Mori between the human likeness of an entity (f.e. a robot, zombie or prosthetic hand) and the affinity humans feel for the entity. Humans tend to feel more affinity when a design has some human characteristics. However, when the human likeness of an entity reaches human levels, the affinity drops and humans can start to feel uncomfortable around the entity. According to Mori, this feeling of uncanniness likely stems from an instinct to avoid proximal sources of danger such as different species and corpses [56].

Broadbent *et al.* discuss that mixing human and robotic features may increase the perceived sense of eeriness, if done poorly. In an experiment with a Peoplebot healthcare robot, they compared a condition in which it showed a humanlike face on its display to a condition in which the face was made silver coloured, and a condition without a displayed face. Participants preferred the humanlike face most, and rated the robot with the silver face on its display as most eerie [12].

Ferrari *et al.* propose a related thesis to explain the fear of humanoid social robots: the *threat to distinctiveness hypothesis.* Social robots that are perceived as very similar to humans might be seen as damaging to the identity of humans and of humans as a group. In two studies they found that people perceive robots with a more anthropomorphic appearance as more threatening than robots with no anthropomorphic features. Androids were rated as most anthropomorphic and potentially most threatening to human identity, followed by humanoid robots and then mechanical ones. They suggest designing expressions that are familiar yet not threatening to the psychological distinctions that humans construct between robots and themselves [26].

A possible explanation for the uncanny valley phenomenon, based on neural activity in response to action perception, was suggested by Saygin *et al.* They performed functional magnetic resonance imaging (fMRI) while experiment participants watched short clips of a robotic agent, an android, and a human. They found stronger suppression effects for the android. They speculated that this could indicate that an increase in prediction error occurs when the brain perceives a human form that displays non-biological, unfamiliar movement [70].

Mori predicted that *"it is possible to create a safe level of affinity by deliberately pursuing a nonhuman design"* [56]. Instead of trying to approach biological human movement with a humanoid or android robot, we could start with a body that does not raise expectations of close-to-human behaviour, and develop behaviour from the ground up. Familiarity can arise in different ways than one-to-one modelling.

A lot can be gained from developing humanoid and android robots in terms of knowledge and building an infrastructure for robot development, and benefits for specific target groups and to specific contexts. However, Veruggio and Operto expect that the high cost of humanoid robots will mean that they will be deployed in situations where the human form is necessary to the task, such as in elderly care, child care, or in sensitive military operations. Concerns have been expressed about deploying humanoid robots in these contexts, ranging from possible development of psychological problems in humans, such as problems of attachment, to displacement of humans in the work force, and robots being used to manipulate and control people [83]. We should take care to engage in discussion around these topics when it comes to the deployment of these robots, so potential benefits outweigh the costs and risks.

Perhaps we should not aim to create social robots that are perfectly modelled on humans. Rather, we could create possibilities for interactive technologies to develop and adapt. If a robot were to “learn” how to communicate based on interaction with a human, it could develop a way of communicating that is not modelled on existing (human) behaviour, yet understandable for humans. This is not to say that the behaviour and body that will develop will not be humanoid at all, perhaps they will be. But in that case we have more reason to believe that the humanlike characteristics that have developed, are sufficient and suitable to the interaction capabilities of the system. Expanding even further the idea of creating an adaptive system: perhaps there is a way to design social robots that would redefine them as their own species, with their own ways of communicating. After all, we are perfectly capable of interacting with dogs and horses, provided they were raised by humans.

2.1.2 Alternative designs

Zoomorphic embodiment designs have been applied to social robots, for instance the iCat robot, robotic seal PARO, robot dog AIBO and Leonardo by MIT. It is suggested that the animal form of these robots evoke lower levels of expectations in humans than anthropomorphic ones [45]. However, one could argue that the iCat robot and some of the other zoomorphic robots are in a sense quite humanoid the ways they communicate, for example by means of certain facial expressions.

Rather than making robots more and more humanlike, we could develop robots based on a different paradigm. De Rooij *et al.* propose an approach they call *abstract expressions of affect.* They propose an alternative to the commonly chosen approach of mimicking human and animal forms when designing social robots. Instead, we could adopt features from abstract art. This frees up some of the technological requirements that come with a humanoid body. Such an approach could merge affect and technological design on a more fundamental level. Rather than designing expressions similar to existing ones, we can choose ones that are effective instead [65].

Another possibility is to take the approach of capitalizing on the communicative potential some readily available technologies already possess. Song and Yamada discuss the design of multimodal affective expressions for an expression-constrained robot. Designing affective expressions with humanoid robots can be costly and technically complex. Therefore, they chose to investigate a more low cost, simple approach. They combined colour, sound and vibration to convey the emotions happiness, sadness, anger and being relaxed. They found that using multiple modalities conveys the emotions better than single modalities, and especially the colour modality helps [76].

Duffy argues for rendering interaction with a social robot “transparent”, by facilitating the interaction to the point people are no longer aware of the interface. Rather than trying to make robots all-purpose, or approach human-likeness as close as possible, Duffy suggests applying only those capabilities that facilitate social interaction and using already familiar communication cues. Additionally, he suggests that robots could gather information about us and their environment through sensors we do not possess, thereby diverging from the course of being designed more like humans [22].

2.2 Adding the time dimension: robot motion

Blow *et al.* describe a design space for social robot faces, from more realistic to more iconic and abstract, adapted from the design space of faces for comics by Scott McCloud [7]. We would like to extend this design space to include the time dimension as well. Movement and shape change add information. We argue here that in addition to focusing on visual characteristics and similarities to existing models and systems, we should take gesture and behaviour into account as well. The quality and type of movement are important characteristics when it comes to communication.

Rasmussen *et al.* describe expressive parameters for shape-changing interfaces, such as orientation, form and volume. Kinetic parameters (speed, tempo, frequency) can change the perception of the interface, as well as direction. Expressive parameters can be used to describe such an interface. Contrast for instance fast movements and slow, flowing movements [64]. A study by Saerbeck and Bartneck provides indications that humans perceive affective signals in robot motion and the perception of affect holds across the different embodiments they tested [67].

Salem *et al.* carried out an experiment with robot ASIMO (the Honda humanoid robot). They compared the conditions in which the robot gave verbal instructions accompanied by congruent co-verbal gesture to a condition with no gesture and a condition with incongruent co-verbal gesture. As they had anticipated, people in the co-verbal gesture condition perceived the robot as more likable and more humanlike. They scored higher on measures of having a sense of shared reality with the robot, as well as the intention to have contact with it in the future. Surprisingly, in the incongruent co-verbal gesture condition these measures were rated even higher, although task performance was lower. The researchers discuss this could be the result of the unpredictability of the robot behaviour being perceived as more alive or humanlike. Lower task performance is not desirable in situations in which task performance is vital. At the same time, imperfection can be perceived as more likeable [69].

The results by Salem *et al.* suggest that humans attribute perceptions of affect to robot motion. This was also found by Saerbeck and Bartneck, who investigated the perception of affect of robot motion for the iCat and Roomba robots. By varying the acceleration and curvature of robot movements, they found that the acceleration parameter can convey different levels of arousal for both embodiments. They also found a relationship between valence and the combination of acceleration and curvature [67].

2.3 From motion to behaviour

Instead of having a human designer predetermine a robot's embodiment and behaviour, it is possible to evolve a robot to fit a task it is supposed to fulfil. This was investigated in aLife contexts before.

In a now classic paper, Sims describes the evolution of three-dimensional virtual creatures. Their fitness was determined by competition over a resource. He suggests it might be easier to evolve systems that appear to display intelligent behaviour, rather than attempting to design this behaviour [75]. Lund *et al.* argue for making the entire morphology of a robot evolvable, including control system, sensors, motors and physical structure, as performance is not only influenced by the control circuit, but by the body plan as well [47]. Cheney *et al.* describe an approach in which they evolved soft three-dimensional virtual robots that were made up of voxels of different materials. Their fitness was evaluated based on locomotion speed in four environments with different restrictions [14].

These examples show that it is possible to evolve based on fitness. In the examples given before, fitness is based on abilities such as locomotion. In the context of Human-Robot Interaction (HRI), fitness could be determined by the robot's communicative abilities towards humans. In the field of HRI, the effects of implementing specific behaviours in robots to facilitate interaction have been studied extensively. Effects of nonverbal communication on human task performance by robots have been found [11, 40]. Ende *et al.* found that participants in their experiment could infer instructions from gestures performed by a humanoid robot and even a single arm manipulator (SAM) [24]. Indications have been found that people experience social interaction with a robot as more meaningful, enjoyable and engaging if the robot performs gestures while talking [41].

2.4 Social learning

Quinn argues for allowing communicative behaviours to evolve from existing functional behaviours, rather than implementing isolated communication channels in Artificial Life agents. This would allows us to better study the origins of communicative behaviour [62].

The emergence of communication behaviour between robots has been investigated in the contexts of Artificial Life and evolutionary robotics research [34, 49]. For example, Floreano *et al.* found that the communication behaviour of signalling close to a source of food arises in genetically similar populations. When unrelated individuals that performed best were selected across populations, deceptive behaviours developed, such as signalling away from the food, which attracted non-related robots [27]. Marocco *et al.* found that when the agents in their study established communication from parents to offspring, the language that emerged increased fitness [49].

Social learning in robot-robot interaction scenarios has also been used to investigate the emergence of vocabulary. In a Talking Heads experiment, two agents play a language game in which they have to indicate a shape within their visual field with a linguistic utterance. Over time, a linguistic structure emerges. Agents start using certain words to indicate specific shapes more and more often.

Steels and his team performed this experiment with two robots equipped with pan-tilt cameras. The listening robot had to decipher which one of the coloured geometric figures the speaker robot was trying to indicate. One agent points out an object to another, accompanied by an utterance, such as *"pargalup"*. The other agent will then associate the object with that specific word. The robots were connected to the internet and people could create agents that could play the game online. During a period of three months, agents played almost half a million games together and developed a contextual core vocabulary consisting of 300 words [78]. Spranger describes a more recent implementation of a language game in robots. He describes a scenario in which two robots engage in a spatial language learning task [77]. This development of vocabulary is interesting because it suggests that we could develop verbal and gestural vocabularies in robots, which obtain a meaning in the context of human-robot and robot-robot interaction scenarios.

2.5 Key findings

Here, we briefly summarize key findings from the literature review. These findings inform the project that is described in the next sections.

F1 Pursuing a social robot design that is nonhuman, or humanoid only to a certain extent, is likely to help people feel affinity for the robot, and not feel threatened by it [22, 26, 56].

F2 We can design multimodal robot behaviour by using existing technologies [76].

F3 We can choose to implement only those behaviours and expressions that are functional, instead of trying to model all behaviours and peculiarities of humans and the interactions between them [22, 65].

F4 Instead of *designing* behaviour that appears intelligent, we can *evolve* behaviour that appears intelligent from the ground up [27, 49, 62, 75, 78].

F5 Evolving a robot's morphology can improve task performance [14, 47, 75].

F6 A robot's nonverbal communication has an impact on human task performance [11, 40].

F7 If a robot performs co-verbal gesture, human interaction partners can experience the interaction as more meaningful, enjoyable and engaging, [41] and rate the robot as more likeable [69].

F8 People can infer instructions from robot gesture [24].

F9 People attribute perceptions of affect to robot movement [67].

3 Problem Statement

As we have seen, giving robots designs that are nonhuman or humanoid only to a certain extent may increase the affinity people feel for them [Finding F1]. Instead of designing behaviour that appears intelligent, we can evolve it [F4]. A robot's nonverbal communication has an effect on human task performance [F6], and human interaction partners can infer instructions [F8] and affective signals [F9, F7] from robot gestures. If we could adapt or evolve a social robot's embodiment and its social interaction behaviour based on interaction with humans, it may be possible to create a robot that humans perceive as likeable, not threatening, and that improves task performance in human-robot teams, *without hard-coding every possible interaction.* We could end up with new effective ways of communication between robots and humans that do not stem from the mind of a human designer but from what users find convenient forms of communication. Personalization of behaviour becomes an option as well. Alternatively, we could assist a human designer by making different parts of the search space apparent to the designer.

In this paper, we will focus on the development of a system in which a virtual robot performs gestures. We have two main research questions:

Research Question 1. *Can we develop a system in which a robot (simulation) adapts its nonverbal (gesture) behaviour based on human feedback, in such a way that humans not familiar with the robot can obtain information the robot has access to?*

Research Question 2. *Is it feasible to develop gestures using an interactive system, and what are requirements for such a system?*

3.1 Approach

We propose to create gestures for a virtual robot that are the result of an adaptation process. The adaptation process takes place based on interaction with humans. We will run experiments during which human experiment participants will provide feedback to the robot's gestures. The robot, which is simulated on screen, will use the feedback to adapt its gestures. The aim is to improve the gesture behaviour over successive interactions. The idea is that the way a robot can perform gestures has a large effect on the way it will be perceived and on the way its body will be designed. Therefore, we will start with simple gesture behaviour that the robot can perform by moving its "limbs". The addition of other modalities that are part of the adaptive body and behaviour may follow later.

From now on, we will refer to the simulated robots used in the experiment as *bots*. We suppose that the label *bot* communicates that the robot we are talking about could in principle be digital or simulated. The bot is made up of joints and connecting limbs, and is able to move. We see this as an opportunity for nonverbal communication by the bot. The bot can communicate with its body only, by means of gestures.

Hypothesis 1. *It is possible to adapt a simulated robot's gestures or movement patterns based on human feedback that indicates whether instructions were correctly conveyed, in such a way that the adapted movement patterns are easier to understand by humans.*

This study is intended as a proof of concept. It does not include developing a full perceptual system or construction of visual categories by the robot. The body and behaviour that will be developed, will likely only suit the context of the experiment task described here. It is possible that it cannot be generalized to different contexts. The requirements for the design of the gesture system are suitability to the context of the experiment task. We will now describe methods that have been employed for design problems before, which are suitable to the context of our experiment.

3.1.1 Interactive evolutionary algorithms

Interactive evolutionary algorithms allow users to select solutions interactively, based on what they perceive as solutions with high fitness. Evolutionary algorithms with humans in the loop have advantages in situations that do not have a clear fitness function. Search ability can be improved, and exploration and diversity increased [23]. Takagi provides an overview of the diverse applications of interactive evolutionary algorithms [81]. Suga *et al.* implemented an interactive evolutionary algorithm in robot WAMOEBA-3. They encoded the connection weights between motor control agents into a genome, which allowed people interacting with the robot to adapt the strength of a set of predefined reactive behaviours [79].

The possible drawbacks of the use of interactive evolutionary algorithms are the limited attention span of humans and the development of fatigue. Quick improvements are important to strive for, instead of evaluating large populations over thousands of generations. Expectations are also likely to rise as humans are confronted with more successful solutions. Using a surrogate fitness function can help reduce the number of cycles that require human feedback [23]. An example of a surrogate fitness function is a neural network that learns from humans, and estimates which instruction humans would choose for a particular embodiment and behaviour.

3.1.2 Developing social skills for robots

Dautenhahn argues for social intelligence as a key aspect to making robot behaviour appear smarter, based on the social intelligence hypothesis. This hypothesis suggests that primate intelligence arose as a response to socially complex situations. Rather than first focusing on other skills such as navigating, planning and reasoning, we could treat social skills as just as fundamental to intelligent behaviour [21].

Dautenhahn proposes a model in which robots are socialized in three phases. First, background knowledge on particular tasks and social behaviour appropriate for specific environments is gathered. Next, a prototype robot is developed in the laboratory, and general behavioural parameters are set. During the third phase, the robot interacts with a human and behaviour is personalized [20]. Personalization of robot behaviour can make a robot's behaviour more suitable for a particular context and interaction with specific individuals. Personalization has been implemented by Mason and Lopes, who describe the development and testing of a robotic system that is able to learn personal profiles of desirable world states [48].

One way to get more grip on a design problem is by decomposing it. Wang and Terpenny describe a method for engineering design, which they call interactive evolutionary design synthesis. It enables human designers to complement their intuitive sense of correct solutions with the storage and generative possibilities of computers. With this approach, the design problem is decomposed into subproblems, thereby reducing the complexity of the design task [84].

We will look at a way to decompose our design problem into sub problems in section 4.2 and run a user test with this system. In section 4.4 we will describe an interactive evolutionary system. The aim is to develop personalized gestures with this system, and to find out what the requirements are for such a system. Before we do those things, we will take Dautenhahn's advice to gather background knowledge on our particular task: we will look at gesture in more detail. This will allow us to simplify our problem and to specify parameters that will describe the most relevant parts of gesture.

3.2 Describing gesture

It has been hypothesized that the origins of human language lie in gesture, and we can conceptualize speech as a gestural system [16]. In animals (including humans), behaviour by other animals may function as a signal even if it is not intended to be one, such as attack-response behaviour.

How can we describe gesture? McNeill argues for the thesis that gesture and speech share the same psychological basis: they are components of the same psychological process. He distinguishes iconic gestures, metaphoric gestures and beats. Iconic gestures show similarity to the linguistic concept that is expressed by the speaker who is performing the gesture cotemporally. For example, an iconic gesture could be the speaker's hand moving up diagonally, accompanying the sentence "He was walking up the hill". Metaphoric gestures have a more indirect relationship to the linguistic meaning of the sentence. Examples are mathematics gestures, referring to mathematical concepts, and conduit metaphors (e.g. holding the hand as if holding something, while talking about an abstract concept). As metaphoric gestures refer to abstract concepts, not every culture shares the same types of metaphoric gestures. Beats serve a pragmatic function. They can, for example, indicate that what the speaker is talking about is not the main topic [51].

Mitra and Acharya note that a gesture's meaning can be influenced by its affective, symbolic and spatial information, as well as by its path. They distinguish gesticulation, language-like gestures, pantomimes, emblems and sign languages. Gesticulation is movement that occurs spontaneously during speech. Language-like gestures can take the place of a word in a sentence. Pantomimes depict objects or actions visually. Emblems are particular, culturally-bound signs, while sign languages are entire systems on themselves. These categories have different levels of spontaneity and are regulated by culture to a different extent [55].

Salem *et al.* describe the organization of the production of speech and gestures in successive chunks. Gesture can be organized in phases: preparation, stroke, retraction and hold. One gestural phrase can contain one or more of such phases [68]. When developing robotic gesture behaviour, it could be a good idea to organize this behaviour in different chunks. The way a gestural phrase is interpreted, is highly dependent on its organization. Compare a gesture in which you first move your hand forward, pause, then stretch out your fingers, to a gesture that involves moving your hand forward and stretching out your fingers at the same time. The second one can appear more urgent than the first.

Here, we will focus on iconic gestures without co-occurring speech, which are described as language-like gestures in Mitra and Acharya's classification [55]. Existing gesture annotation schemes can help shed a light on how other researchers have classified and described gesture for practical purposes. These schemes are likely to include the items that describe the most informative aspects of gesture. Those features of gesture will help people communicate and infer information from the gesture. Gesture annotation schemes are designed in such a way that after encoding, the (essence of) the gesture can be reproduced. This could help us specify which features are most important to describe, and to build a gesture language from the selected elementary building blocks.

Various gesture annotation schemes have been developed, such as FORM, CoGesT, and MURML. Annotation systems for sign languages exist as well, such as HamNoSys (the Hamburg Notation System for Sign Languages), SignWriting and the Stokoe notation.

3.2.1 Gesture annotation schemes

The annotation scheme FORM describes kinetic information of gestures with different tracks. One track contains *Location*, *Shape* and *Orientation*, the other track contains descriptions of *Movement* [50].

CoGesT is a transcription scheme for hand and arm gestures. Gestures are encoded in so-called *Simplex Gestures*, which are described by feature vectors that consist of the start and target locations, shape of the hand and movement, size of the gesture and speed during the gesture [29, 82].

In MURML, gesture is described by its stroke, which is defined by three features for a hand or arm configuration: location and orientation of the wrist and the shape of the hand. If the gesture is dynamic, constraints for the start and end time and the moment of peak velocity have to be described. Complex trajectories can be described by assigning descriptors "linear", "curve" or "circle" to segments of the trajectory [43].

3.2.2 Sign language annotation schemes

HamNoSys (*the Hamburg Notation System for Sign Languages*) is based on the Stokoe notation. It was designed with the aims of being suitable for international use, iconic in form and extensible. A sign is encoded by describing initial posture and actions changing the posture. Posture consists of nonmanual features, and the shape, orientation and location of the hand(s). Path movements, in place movements and nonmanual movements are combined sequentially or co-temporally in actions. Path movements can be described by straight, curved or zigzag lines, circles and other forms. Posture and actions are quantized with steps of 45 degrees, and can be visualized from three perspectives to adequately capture their 3D orientation and trajectory [33].

SignWriting was developed by Valerie Sutton in 1974. The aim of the writing system is to represent the gestures in sign languages as they are perceived visually. Sign-boxes contain icons that may for example refer to hands, the head, movements and facial expressions [60]. Movement takes place in the "Wall Plane" or the "Floor Plane". Movement can be expressed by directions away from the body and towards the body. An example of a specification of movement that takes place in both dimensions is the icon for "Up-Forward-Diagonal". Differently shaped arrows can be used to indicate the shape of the movement (zigzag, curved, circular, axial, etc.), as well as their speed. Movement Dynamics are small symbols that express characteristics of the movement, such as fast, slow, tense, or relaxed [80].

3.2.3 Takeaways

We intend to develop simplified bots that are only made up of limbs and joints. Therefore, it is not necessary or possible to include such features as facial expressions and hand shapes. Most of the described annotation systems contain possibilities for encoding location, start and target rotation, movement, and speed of the gesture, as well as path characteristics (moving in a straight line, versus curved or zigzag shaped paths). The use of steps of 45 degrees is convenient to reduce the number of possible visualisations.

Movement in the SignWriting system is described as movement away from and towards the body, which captures one of the aspects of gestural communication: it is performed by an individual, a sender, and hence movement takes place in the space relative to the body of this sender. We could capitalize on this in our encoding of gesture.

3.3 Gesture lexicon

Aigner *et al.* investigated gestures for HCI. They departed from the notion that we should first classify which effects the gestures supposedly seek, and then look for a gesture that best supports achieving the desired effect [1]. This is the approach we will follow. We will first define a gesture lexicon. During the experiment we will describe later, a virtual bot will perform gestures. The aim is to adapt the bot's gestures in such a way, that they will better fit the instruction based on participant feedback.

The instructions the bot provides to human participants is based on the gesture lexicon developed by Ende et al [24]. They developed a gesture lexicon based on human-human interactions and tested a subset for a SAM (single arm manipulator), by showing a video of a SAM to participants via an online survey. Some of their results are included in Table 1. They compared these results to a human arm and a humanoid robot arm performing the same gestures.

Instructions *"Come here!"* and *"Come closer!"* could be interpreted as very similar to each other. *"This one!"*, *"From here, to there!"*, *"Display object!"*, *"To give something!"*, and *"Give it to me!"* imply access to an existing object at a shared location. We chose to only include those instructions for which Ende *et al.* obtained an identification rate of at least 30%, which did not imply an existing object, and could not be easily confused with a different instruction. This leaves us with the following subset of instructions:

- Wave one's hand!
- Come here!
- Go away!
- No!
- Caution!
- Stop!

We will refer to items in the set of instructions as *gesture classes*. Because the gesture lexicon has already been tested, we will be able to compare our results to scores obtained by Ende *et al.* [24].

3.4 Key findings

Here, we summarize the key findings from this section.

F10 Interactive evolutionary algorithms can be used to select solutions interactively, which can improve search ability, exploration and diversity of solutions [23].

F11 Systems that use human feedback should be designed in a way that limits the required human feedback, to avoid the development of fatigue in the person giving feedback [23].

F12 A surrogate fitness function can be used to estimate the choices a person interacting with the system would make, thereby reducing the amount of feedback that is necessary from the person [23].

F13 Gesture can be organized in phases: preparation, stroke, retraction and hold. One gestural phrase can contain one or more of such phases [68].

F14 Common ways to describe gesture in gesture and sign language annotation schemes include location, rotation, movement, speed of the gesture, and path characteristics. See sections 3.2.1 and 3.2.2.

F15 Dividing the space in steps of 45 degrees when depicting gestures is a convenient means to reduce the number of possible visualisations [33, 60].

F16 Gesture movement can be described as movement away from and towards the body, which captures one of the aspects of gestural communication: it is performed by an individual and movement takes place in the space relative to the body of this sender [60].

F17 Determining a gesture lexicon can aid in classifying the desired effect of gestures and looking for a gesture that best supports achieving the desired effect [1, 24].

Stop!	92%
From here, to there!	90%
Go away!	84%
This one!	84%
To give something!	71%
Caution!	68%
Display object!	63%
No!	58%
Come here!	47%
Give it to me!	39%
Wave one's hand!	30%
Come closer!	26%
No idea!	17%
Slow down!	04%

Table 1: Experimental results obtained by Ende *et al.* of the recognition rate of gestures executed by a single arm manipulator by human experiment participants (N=40) [24]

.

4 Design of an Interactive Gesture System

As described in section 3.1, we aim to adapt a virtual bot based on interaction with humans. The bot's gestures are adapted with the aim of developing gestures that better fit the gestures from the gesture lexicon. During each interaction cycle, the human-robot team executes a task. In order to avoid the development of fatigue in the human participants in the experiment, the amount of information presented to them and the duration of the session should be limited [F11]. One way to limit the required amount of information is to develop a surrogate fitness function [F12].

We can identify the following task decomposition:

1. Definition of the embodiment and gesture behaviour of the bot. These need to be limited in terms of complexity.
2. Definition of the ways embodiment and behaviour can be adapted.
3. Implementation of the embodiment and behaviour.
4. Development of an interface that provides a means to respond to the bot's gesture behaviour.
5. A means of generating new solutions based on previously provided feedback.
6. Development, implementation and training of a surrogate fitness function that can be used to take over the role of the human experiment participant for cycles of the evolutionary algorithm.
7. Running experiments and finding solutions.
8. Evaluation of the solutions.

Tasks 1 to 5, task 7 and task 8 fall within the scope of this paper. Section 4.1 describes an implementation of task 1. In section 4.1.2, a way of defining gesture behaviour based on [F-13]-[F16] is described. Here, we choose to write an implementation in C# for Unity. The human experiment participant will need to interact with the system for quite some time. Therefore, the amount of information that is displayed to the human interacting with the bot should not further exhaust the participant. Different ways of presenting information are tested during the user test described in section 4.3 and the experiment described in section 5. During the user test, chosen solutions are adapted by varying predetermined parameters. The experiment makes use of an implementation with an evolutionary algorithm.

4.1 Embodiment and behaviour of bots

We will describe one bot as a tuple of structure and behaviour:

$$bot := (structure, behaviour)$$

4.1.1 Structure

The bot's structure resembles a tree structure. Because we expect the reader to be familiar with the terminology that is commonly used to describe tree structures, we will use that terminology to describe the bots.

Bots consist of three-dimensional simulated objects. **One bot is an interconnected structure of three-dimensional node and edge objects**, which could also be described as "joints" and "limbs". See Figure 1 for reference. One of the nodes, the start node, has a fixed position in world space to ensure the visibility of the performed gestures. We will refer to this node object as the *root node*. Edges are cuboid-shaped, whereas nodes will be depicted by spheres. An edge is a connecting element between two nodes. Node objects can move and rotate and thereby move the connecting edge objects. If a node object does not have any children, we will refer to this joint as a *leaf node object*. This leaf node object consists of a sphere that is bigger than other node objects, for purposes of making the 3D orientation of the bot more visible to experiment participants. The larger sphere makes it easier to estimate the simulated distance of the virtual bot relative to the viewer.

The simulation of the bot can be adapted by changing one of the nodes so the structure of the body changes. We will limit the maximum number of edges, and the maximum number of edges that can be attached to one node. During the experiment, the structure of the bot will consist of the root node object, an edge object, a node object, an edge attached to this node, and a leaf object at the end of it (as displayed in Figure 1A). For more complicated bots, another edge object or sub

	$\varphi = \frac{1}{4}\pi$	$\varphi = \frac{1}{2}\pi$	$\varphi = \frac{3}{4}\pi$
$\theta = \frac{1}{4}\pi$	$(\frac{r}{2}, \frac{r}{2}, \frac{r}{\sqrt{2}})$	$(0, \frac{r}{\sqrt{2}}, \frac{r}{\sqrt{2}})$	$(-\frac{r}{2}, \frac{r}{2}, \frac{r}{\sqrt{2}})$
$\theta = \frac{1}{2}\pi$	$(\frac{r}{\sqrt{2}}, \frac{r}{\sqrt{2}}, 0)$	$(0, r, 0)$	$(-\frac{r}{\sqrt{2}}, \frac{r}{\sqrt{2}}, 0)$
$\theta = \frac{3}{4}\pi$	$(\frac{r}{2}, \frac{r}{2}, -\frac{r}{\sqrt{2}})$	$(0, \frac{r}{\sqrt{2}}, -\frac{r}{\sqrt{2}})$	$(-\frac{r}{2}, \frac{r}{2}, -\frac{r}{\sqrt{2}})$

Table 2: The relative target positions we will consider, converted from spherical (θ, φ, r)-coordinates ($r = 1$) to (x, y, z) coordinates

tree of two edge objects can be connected to the root node. A limited number of node objects will be selected (minimum 1, maximum 2). For these node objects, the behaviour scripts can change the target positions.

If desired at a later stage, the width, height and length of the cuboid segments as well as the radii of the spherical segments and other parameters, can be adapted to give the bot the impression of, for example, having a body with limbs.

Figure 1: A 2D sketch of some possible bot structures. The red dot denotes the root node object. During the user test (section 4.3) and the experiment (section 5), bot A is displayed and adapted.

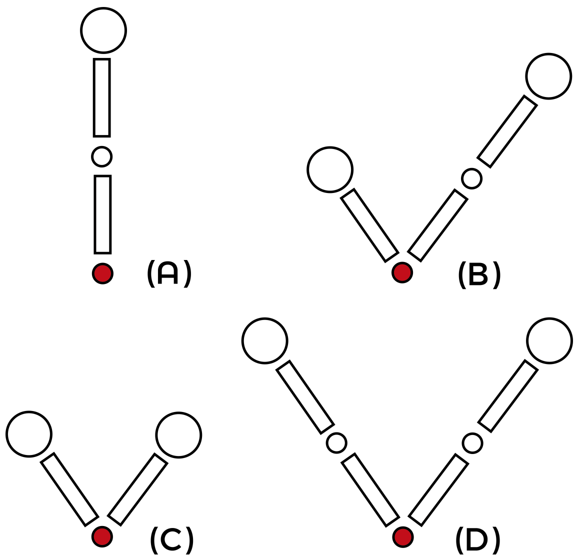
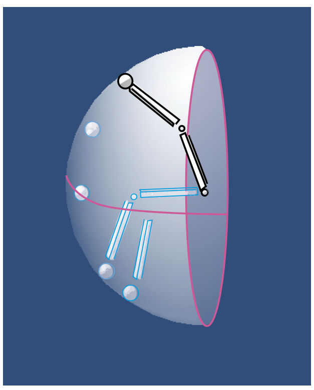


Figure 2: A 3D sketch of some of the possible target locations the bot can reach.



4.1.2 Gesture behaviour

We propose to describe the gesture behaviour of the bots by the following properties:

- Leaf node object that will be moved.
- Target position of the leaf node object, conceptualized as position relative to the root node object, described by spherical coordinates.
- Speed of moving from one position to the next.

The target positions are stored as positions on the half-sphere with radius r around the bot's root node object oriented towards the viewer. See Figure 2. Not every position on this half-sphere can be reached: they are separated by 45 degree rotations of a point that lies on one of the $x, y, z-$axes at a fixed distance r from the root node around each axis. See Table 2. During the experiment, we will only consider ϕ and θ values ranging from $\frac{\pi}{4}$ to $\frac{3\pi}{4}$. This yields a total of 9 possible target positions. This allows us to easily propose sufficiently diverse solutions, thereby limiting the number of solutions we need to present to the participant. We include one extra position: the start position of the leaf node object, which is located closer to the root node of the bot.

Gesture behaviour is described by the leaf node object(s) that will be moved, the target position(s) of these limbs, and the speed of movement. In case one limb can be moved, it can be described the following way.

$$\text{gesture}_{\text{class 1}} := (o_x, v_1, t_1, v_2, t_2, \cdots, v_n, t_n) \text{ for } n \geq 1$$

Here n denotes the number of target positions for the gesture behaviour associated with one leaf node object and o_x indicates the leaf node object that will be moved. Vector v_i denotes a (θ, φ) vector, which will be converted to a (x, y, z) vector so Unity can perform the translation, and t_i denotes the time to move from v_i to v_{i+1} (or to v_0 in the case of v_n).

The Unity Engine will linearly interpolate between the current position and the target position.[3]

4.1.3 Limitations

We realize that the design space has its limitations. Despite the aim of limiting human design choices, a number of design assumptions were made in the process of defining and programming the bot. We will name some of them. The bot was supposed to:

- exist in 3D space

[3]Because the bot's structure is interconnected, this will alter the position and rotation of each of the bot's node and edge objects on the shortest path from the root node object to the leaf node object, with exception of the root node. Objects that are connected to other moving objects, will move as well, unless the mass of other connecting objects that do not move is too high.

- have fixed-length, fixed-width, fixed-height edge objects
- be able to move
- consist of components that are rigid bodies
- exist of node objects that are connected by edge objects
- have one root node object that is fixed in its connection to the world, for purposes of visibility and ability to communicate information from one spot
- have a limited number of node and edge objects

Because calculations for moving between gesture positions are left to the Unity Physics Engine, gestures may at times seem unpredictable or unnatural. However, we expect that this will not be too large of an issue, as the embodiment is unlikely to evoke expectations of biological human or animal likeness in terms of its movement at the start of the interaction.

4.2 Generating bots and behaviour

We started by building a system that generates and adapts gestures. We will refer to a gesture that has been developed for a gesture class as a *solution*.[4]

The fitness of a solution can, in our case, be described as the preference of the human designer for a particular solution. For the user test described later, gestures are generated based on participant choices and on the functions that the system has at its disposal to generate new combinations.

4.2.1 Initialization

First, we will perform an initialization step in our experiment. A bot first displays simple behaviour: it moves towards a certain target position, and then it moves back. The participant is asked to indicate on a scale from 1 to 5 how strongly s/he associates the movement with each of the six gesture classes. See section 4.3. These ratings are stored in an array containing 9 scores for associated (θ, φ) target positions.

Using the array, a gesture design agent can select target positions that score well for the simple task, combine them and propose the new combination to the participant for evaluation. The participant then evaluates how well each of the proposed gestures fits the current gesture class. The participant now evaluates gestures for one gesture class at a time. The next section describes the adaptation possibilities we implemented in the system.

4.2.2 Adaptation

The participant can select one of three gestures, based on how well the gesture conveys the instruction belonging to a particular gesture class. Gestures are constructed and adapted according to the following steps:

A Three different gestures are created by:

1. Combining the target positions from the highest-scoring subspace. [5]
2. Combining the target positions from the second-highest scoring subspace
3. Combining the two highest scoring target positions in a gesture

B Two additional gestures are created by varying the speed of movement of the gesture chosen at A.

C Two additional gestures are created by combining a subset of the target positions of the gesture chosen at B.

D Two additional gestures are created by adapting the gesture chosen at C:

1. A gesture with the same target positions as the gesture chosen at C, in reversed order
2. A gesture that repeats the gesture chosen at C

Steps A-D are repeated for each gesture class. The gesture that has been selected by the participant at step A is displayed again at step B. Similarly, the selected gesture at step B is shown again at step C, and so forth.

Displaying the current best choice has been mentioned as a tactic to reduce frustration in participants, as doing otherwise may induce participants to feel as if the system is forgetting their choices [23] (p.216). By displaying three bots, the gestures can be varied while hopefully not overwhelming the participant with information.

The participant evaluates 9 initial gestures, and 4x3 gestures (or 3+2+2+2 unique gestures) for each of the six gesture classes. This yields a total of 81 gestures that need evaluation.

4.3 User test

4.3.1 Aim

A test with one user was carried out as part of the design process. The user test was intended as a run-through of the system with someone other than the system designer.

The aim of the user test was to test whether the amount of information presented was manageable for the participant, or if solutions that move across the screen place too much of a cognitive burden on the participant. Additional aims were to get an indication of how the bot is perceived, and whether the interface is easy to navigate. The system was tested with one participant.

[4] Mainly for reasons of brevity. One could also refer to a gesture as a sub solution. A complete solution would then be a bot and one or more gestures for each of its gesture classes.

[5] Add up the scores for moving towards a position in different parts of the space. The previously described (θ, φ) array contains scores for moving towards single target positions. Sum up the scores along the rows, columns and diagonals of the 3x3 array. This way, we can construct scores for horizontal movement across the top, middle and lower part of the space, as well as vertical movement across the left, right, and middle of the space. We can also construct scores for diagonal movement across the space, and movement restricted to one of the corners of the space.

The row, column or diagonal with the highest normalized sum gives an indication in which subspace the movement was most successful, for example horizontal movement in the top part of the space. In that case, target positions for horizontal movement in the top part of space are combined in different ways, and options are presented to the participant.

4.3.2 Method

The participant is seated in front of a laptop and receives an explanation on the system. The participant can see a bot move on screen and a menu that contains the instructions from the gesture lexicon as described in section 3.3.

During the first phase, the bot moves from a starting position to one of the target positions and back. The participant is asked to rate the applicability of this movement for each of the gesture classes on a scale from one to five. This is repeated for the other 8 possible target positions. See figure 3. The ratings are used to construct gestures during the next phase.

Figure 3: Interface during initialization phase.

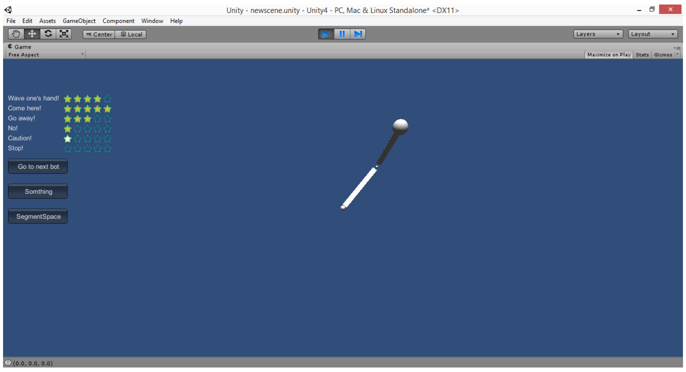
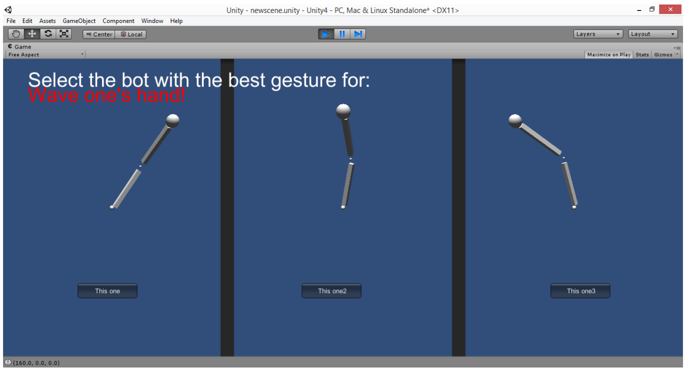


Figure 4: Interface during adaptation phase.



During the second phase, the participant is presented with three bots that each carry out a different gesture. The participant is asked to choose the bot s/he thinks is most suitable to the current gesture class. See figure 4. When the participant chooses a particular gesture as most suitable, this gesture is adapted. The chosen gesture is subsequently displayed on the left, and two adapted versions are displayed in the middle and on the right of the screen.

This way of offering feedback allows for a direct way to select an appropriate gesture from multiple examples and subsequently adapt the selected bot. It is a common way in which interaction is offered in systems that make use of interactive evolutionary algorithms: the human in the loop can for example select solutions for the next generation [47]. A different option would be to ask the participant which gesture the bot was performing, and offering a choice between gesture classes. This second option could result in conflicts, if a participant would feel a gesture describes multiple gesture classes equally well.

4.3.3 Results

The participant was able to carry out the task as instructed.

During the first phase, the participant expressed that she found the task difficult, because she could imagine different situations in which the gestures would fit a number of the gesture classes, although the gesture classes referred to very different concepts. She noted that she found it complicated, because one gesture could mean a number of different things.

Interestingly, the participant asked the following question during one of the gestures: *"How is this anatomically possible?"* Later during the experiment, the participant asked *"Should I look at this as if the person is looking at me?"* This indicates that the participant anthropomorphized the bot.

After the experiment, the participant indicated that she found the task difficult, because the bot was quite abstract and she had to infer a "mood" from it. The participant did think the interface was clear. When asked about the gestures, the participant indicated that they seemed random, and that she did not perceive improvements during the progression of the experiment.

The participant noted that the gesture the bot made was slightly different each time. We acknowledge that this was the case, because only the target position of the large sphere at the end of the bot structure was controlled.

4.3.4 Discussion

The user test indicates that presenting three bots to a participant at the same time, each of them making a slightly different gesture, presents somewhat of a challenge to the experiment participant. The participant noted that she found it *"difficult to translate the movement of a sphere on two sticks"*. This sounds like a large mental effort on the side of the participant, rather than a more intuitive context of a human teacher and a robot student. Presenting two instead of three bots on screen at the same time might make it easier to distinguish between gestures and counter fatigue. The participant would only need to make one comparison between two bots, instead of three. A trade-off between memory use and concentration on multiple gestures at a time by the participant could take place in this case: if bots are displayed on different pages, participants will need to remember if gestures seen on previously visited pages were suitable.

The participant did not perceive improvements during the progression of the experiment. Indeed, improvements did not necessarily occur during the user test: the gesture of the bot was varied based on previously made choices.

In the tested system, gestures were developed for one gesture class at a time: the gesture was adapted a number of times, before the next gesture class was presented. We hypothesize that it might be more rewarding for a participant to adapt the gestures for a particular gesture class one by one: from an initial, seemingly random gesture to one that seems more appropriate for the gesture class. After finishing one gesture class, it could be interesting to present the participant with an initial and the final gesture. Otherwise, improvements might not me noticed if improvement occurred very gradually, which could give the participant less of an incentive to keep engaging with the process. If improvement occurred, this could provide an incentive to engage in a similar process for the next gesture class. Of course, there is a risk if the final gesture is perceived as similar or worse in quality than the original one. An additional advantage of only working on one gesture class at a time, is that the participant does not have to imagine all the different contexts in which a gesture could be suitable for every gesture class.

4.4 Adaptations after the user test: interactive evolutionary design system

After the user test, the movement of the bot was made more uniform based on participant feedback. Necessary improvements included controlling the node object between the two edge objects as well: leaving the movement free results in noticeable differences between what should be similar gestures, and likely in differences in the interpretation of those gestures.

The generation of solutions will now be changed as described below, with the aim of finding better solutions during the progression of the experiment. Previously described strategies for constructing gestures will be replaced by first a randomly generated population of gestures and subsequent adaptation by an evolutionary algorithm. Then, the system is tested for one gesture class at a time: every participant will evaluate bots for a single gesture class (see section 5).

Interactive evolutionary design systems generally consist of a phenotype definition, a genotype representation, a decoder (mapping from genotypes to phenotypes), a solution evaluation facility (so the user can perform selection), and an evolutionary algorithm that recombines and mutates solutions ([23], p. 220). Here the phenotype of a solution is a gesture, which is applied to the bot that was also used during the user test. The genotype is a list/string of properties:

$$g_x := ((v_0, t_0), (v_1, t_1), (v_2, t_2), \cdots, (v_n, t_n))$$

with $n \geq 1$. Vector v_i denotes a (θ, φ) vector for the target position, which will be converted to a (x, y, z) vector so Unity can perform the translation, and t_i denotes the time to move from v_i to v_{i+1} or to v_0 in the case of v_n. Vector v_0 is the same for all gestures, and will be kept fixed.

Ten solutions are generated by combining between one and four target positions, which are randomly selected from the nine possible target positions. Speed of moving from one target position to the next is set to be a random value between 10 and 20 frames, at a frame rate of 30 frames per second.

One-point crossover is chosen as the recombination operator. Mutation occurs by random resetting of a value to a randomly selected one from the set of possible target positions and speeds. We choose to implement two mutation operators that have different parameter adjustments over time.

The following loop takes place while an experiment participant interacts with the system to adapt a gesture for one single gesture class:

A **Evaluation of candidate solutions by human participant.** The participant evaluates 10 solutions, which are displayed in pairs. A counter indicates the number of solutions that has been selected by the participant in the current generation. The participant is asked to select three solutions s/he deems to fit the gesture class best.

B **The solutions that were selected are recombined.** Recombination is performed with single-point crossover on the three selected solutions. The point at which crossover occurs is selected randomly for each string, because strings differ in length. The string can get longer or shorter. Each parent is combined with every other parent 4 times, resulting in 24 children. The solution that was selected first or the solution that has been selected multiple times is kept in the population. The other parents and solutions that were not selected are discarded.

C **Mutation is applied to all the offspring.** Mutation operator M_1 changes target position to different ones, mutation operator M_2 changes t values. The mutation rate of M_1 starts at one per offspring for the first generation, and decreases to one per generation at the tenth generation. The mutation rate of M_2 starts at one per generation for the first generation, and increases to one per offspring by the 10th generation. The mutation operators are changed this way to first vary the target positions. Once suitable target positions are found, the speed of movement in varied more so the participant is able to fine-tune the gesture.

D 9 children are selected randomly as survivors for the next generation. Those solutions that only consist of the initial target position are discarded, as well as duplicates. In that case, different children are selected instead.

E The 9 children plus one solution from the previous round are presented to the participant. The solution from the previous round is presented last.

Loop A-E takes place 9 times (generations). Including the first randomly generated population of solutions, this means that every participant evaluates 10 generations of bots.

Because of the replace-worst strategy we took for fitness-based replacement (the worst 9 solutions are replaced), the small population size and the possibility to select solutions multiple times, there is a risk of premature convergence: the population is quickly dominated by a few solutions that were ranked above the others in the beginning. There is a possibility that we miss out on good solutions that could be found in a different part of the search space. However, converging quickly has an advantage in interactive evolutionary algorithms: as good solutions are kept in the population and varied, the person interacting with the system will likely feel as if the system is converging quicker on good solutions. We implement a no-duplicates policy: the human evaluator's time is valuable, so we do not want to waste his/her time evaluating multiple copies of the same individual. This also avoids the entire (small) population from being filled with copies of the same individual, which lessens the risk of premature convergence.

Parameter control here is deterministic (for instance, mutation rate is not co-evolved). In the future, it could be interesting to make this adaptive, f.e. based on ratings of quality by the participant. Alternatively, the user could be allowed to set the mutation rate and thereby influence whether the search process is taking an explorative or exploitative turn (i.e. looking for new solutions in a different part of the search space, or varying current good solutions slightly).

5 Experiment: Adapting Gestures with an Interactive Evolutionary Algorithm

5.1 Aim

The developed system was tested with 16 participants. Gestures were developed for the gesture classes *Wave one's hand!*, *Come here!*, and *Go away!*. The aims of the experiment were the following:

- Check if participants can use the interface. Measures: can participants complete the task? Can they navigate the interface? This is measured by observation and asking participants to think aloud, which will hopefully incite them to voice any difficulties they encounter.
- Find out if participants experience the task as difficult. Measures: an open question about the difficulty of the task, measuring the time it takes to fulfil the task.
- Check how participants perceive the bot after interacting with it. Do they anthropomorphize the bot? Do they need more humanlike characteristics to fulfil the task?
- Find out if the developed gestures are idiosyncratic; that is, does the interaction result in gestures that only make sense to those people who interacted with the system to develop them, or can the resulting gestures be interpreted by other people?
- Do participants have the sense that the gestures that are proposed by the bot improve? Measures: open question, asking participants to rate the first gesture they selected and the last gesture they selected.

5.2 Method

5.2.1 Participants

We recruited 16 participants, of which 8 female and 8 male (6 Dutch, 2 Chinese, 2 Spanish, 1 Azerbaijani, 1 British, 1 Finnish, 1 Greek, 1 Portuguese, 1 Serbian/Canadian/Dutch), with an average age of M=26.9 (SD=5.1) from Leiden University, Universidade Católica do Porto and the Royal Academy of Art in The Hague. The average of their self-reported computer skills on a scale from 1(=very poor computer skills) to 5 (=very strong computer skills) had a value of M=3.75 (SD=0.68). For the multiple choice question "Do you have any experience with robots?", 7 participants checked the box "No", 6 checked the box "Yes, some" and 3 checked the box "Yes, I have carried out research with one or multiple robots/I have built and/or programmed one or more robots". We note that the participant group included a set of students who may to some extent be familiar with concepts from artificial intelligence, HRI, HCI, and affective computing.

5.2.2 Task

Participants are seated at a table, with a laptop in front of them, in quiet surroundings. The participants are told that the aim of the experiment is to choose gestures that are performed by a simulation of a robot, based on how well the gesture matches the textual instruction that is displayed at the top of the screen.

Each participant is assigned to one of the conditions *"Come here!"*, *"Wave one's hand!"* and *"Go away!"*. Six participants were assigned to the condition *"Come here!"*, five to the condition *"Wave one's hand!"* and five to the condition *"Go away!"*. First, the participant is asked to read the informed consent form and fill out the initial questionnaire. Participants 1 to 3 directly continued with the main part of the experiment. Participants 4 to 16 are first presented with three videos of the virtual robot. The first video demonstrates the bot and the way it can move from its resting position to other positions. Then, the participant is presented two videos that show two of the results of the first three participants (for example, if the assigned condition is *"Come here!"*, the participant is shown the results of the first participants who were assigned the conditions *"Wave one's hand!"* and *"Go away!"*. The participant is asked to write down what s/he thinks the virtual robot is trying to express.

Next, the main part of the experiment starts. The researcher demonstrates how the participant can select gestures, and which actions are allowed. It is allowed to select the same gesture multiple times. The participant is asked to "think aloud" during the experiment. Participants evaluate 10 generations of 10 bots. The evolutionary algorithm described in section 4.4 generates 10 new solutions for each generation. The embodiment of the bot is kept fixed, whereas the gesture behaviour is varied. Figure 5 shows the interface that participants interacted with. Two bots are displayed on screen at a time. Participants can click buttons to move to the next page and the previous page, in order to see all ten bots belonging to one generation. Participants can click a button with the label "This one" to select a gesture. Upon clicking one of the buttons labelled "This one", the counter at the top left of the screen is increased. After selecting three gestures, a different screen is shown (Figure 6) for approximately two seconds. After 10 generations have been evaluated by the participant, the participant is presented with the first gesture s/he selected during the first and during the tenth generation, and asked for feedback on a 5-point scale (Figure 7).

Figure 5: Interface during the experiment.

Figure 6: Interface between generations.

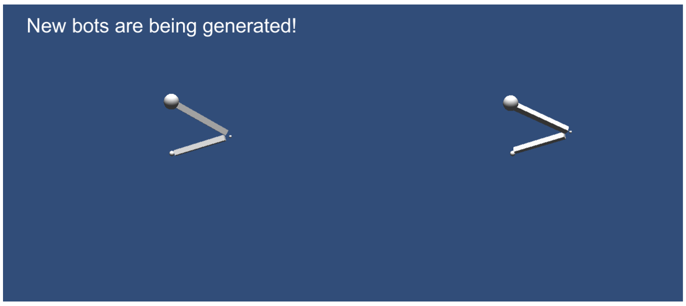


Figure 7: Interface after ten generations.

After completing the interaction with the bot, participants are be asked to fill out the Godspeed questionnaire series and a questionnaire with open questions. See Appendix B for the questionnaire form used in the experiment. The Godspeed questionnaires measure five HRI concepts on semantic differential scales: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety [6]. The questionnaires were developed for evaluating robots. We will consider whether the questionnaires translate well to our virtual bot. As it is hard to determine a ground truth for any of the measures, it would likely be more informative to have two robot conditions and compare the results of the questionnaires for both conditions. We use them here to get a rough indication of the way participants perceive the bot. We expect to find a low value for anthropomorphism. A high value for animacy could indicate that participants perceive the bot as lifelike and responsive, perhaps in the sense of an animal. A high value for perceived intelligence could indicate that participants feel the bot responds to their choices. We measure likeability to determine if participants find the bot pleasant to interact with and/or find the movement of the bot pleasant.

After running the experiment with the first three participants, the following changes were made: the screen would show a black flash after a gesture was finished, based on a suggestion that it would be easier to determine whether a gesture had finished. As mentioned before, participants 4-16 were shown two videos with the results of the previous participants. The videos displayed the gestures belonging to the conditions the current participant was not assigned to. Participants were asked to interpret what the virtual robot was trying to indicate, given two examples (*“Give it to me!"*, *“Slow down!"*). We do not have reason to believe that these changes influenced the results in some way, except possibly making the task slightly easier. As measuring task performance was not the aim of this experiment, we will therefore not consider differences between the first three and last thirteen participants.

5.3 Results

5.3.1 Task completion

After initial explanation by the researcher, participants expressed that they understood the task. All participants could complete the task and made use of provided means of interacting with the interface (buttons to select bots and navigate to other pages, star toggles to rate the first and last selected bot). One of the participants initially misunderstood the task that needed to be completed with the interface. After more explanation and restarting the system, the participant was able to complete the task. Participants interacted in different ways with the system, and interaction also varied between generations. Some participants would watch all 10 gestures in a generation before making a decision, while others would at times choose one or more gestures after watching the first two. Additionally, we note that the way individual participants interacted with the system changed from generation to generation.

5.3.2 Perception of task difficulty

To the open question "Was it difficult to select gestures?" on the final questionnaire, thirteen participants reported some kind of difficulty, for different reasons. The remaining three participants reported no difficulty. Four participants noted that it was difficult to choose between similar gestures. However, one other participant noted that it was easier to choose between options that were more similar, and another noted that even similar gestures have differences in details, which made choosing easy. One participant remarked that there were too many options. Another noted that the task got more difficult towards the end of the experiment, while three others noted that the task was more difficult at the beginning. This was neatly summed up by one participant, who remarked that the task was difficult *“at first, because they were so different and near the end because they were so similar!”*

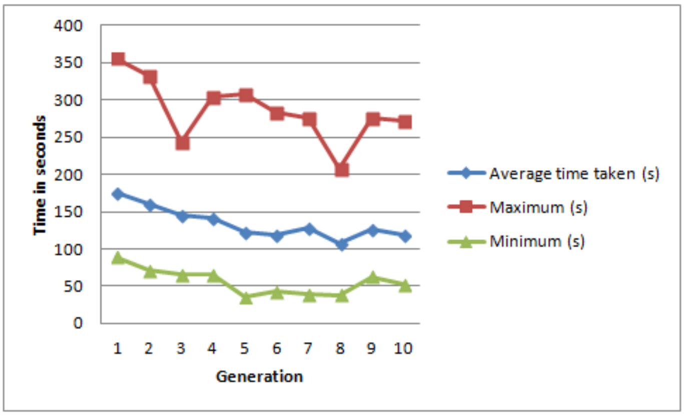
One participant felt she had to get used to it to understand the message. Initially, it was hard to relate the abstract form to a humanlike expression such as *“Go away!”*. Over time she became more familiar with it. Two participants remarked that it was hard to perceive the movement in 3D space, one of them reporting personal difficulties. One participant mentioned she found it hard to focus on the movement, as there were two bots making gestures on screen at the same time. One participant noted that it was hard to compare gestures on different pages. One participant asked for a piece of paper to note down the gestures he found interesting, as a memory aid.

Participants took on average 134.4 seconds (SD=65.4 s, min=36 s, max=356 s) to complete the task of selecting three gestures from the 10 gestures available in one generation. Table 3 shows an overview of the time taken per generation. Figure 8 shows a graphical overview of the average time taken, as well as the minimum and maximum time taken, per generation. The average, minimum and maximum time taken appear to decrease from the 1st to the 10th generation. For the first generation, the average time taken was 174.6 seconds (SD=74.5) and for the 10th generation 119.1 seconds (SD=71.5). A two-tailed paired t-test (N=16) revealed a significant difference (p=0.029) between the 1st and 10th generation.

Generation	Mean	Standard deviation	Minimum	Maximum
1	174.6	74.5	89	356
2	161.3	75.8	71	332
3	145.5	52.0	66	244
4	140.8	67.4	65	304
5	123.2	68.7	36	307
6	118.2	63.0	43	284
7	128.6	59.5	39	275
8	107.0	50.9	38	207
9	125.6	65.8	63	275
10	119.1	71.5	53	272

Table 3: Time taken (in seconds) to complete the task of selecting three gestures from 10 gestures available within one generation (N=16).

Figure 8: Average, minimum and maximum time taken (in seconds) to complete the task of selecting three gestures from 10 gestures available within one generation (N=16).



5.3.3 Perception of the bot

The results of the Godspeed questionnaires can be found in Table 4. Although the Godspeed questionnaires were developed for (physical) robots, the categories *Animacy*, *Likeability* and *Perceived Intelligence* yielded values of Cronbach's alpha that indicated at least an acceptable internal consistency. This means that the questionnaire items appear to be asking the participant questions about the same concept. The categories *Anthropomorphism* and *Perceived Safety* resulted in a questionable internal consistency. The questionable internal consistency of the *Perceived Safety* category could be due to the virtual nature of the bot: physical and virtual robots likely differ in when they are considered safe to interact with. For the *Anthropomorphism* category, the value for Cronbach's alpha might be explained by the limited humanoid nature of our bot: at times it was interpreted as a human arm, but said to miss the body the arm was supposedly attached to. The Anthropomorphism measure from the Godspeed questionnaires was based on a study by Powers and Kiesler, in which they asked participants for ratings of a robot on these measures [61]. The particular robot had humanoid facial characteristics, which was not the case for the bot in our experiment.

As it is hard to determine a ground truth for any of the measures, it would likely be more informative to have two robot conditions. Because we do not consider two different conditions here, we will compare the results to the average score of 3 that can be obtained with the questionnaire, in order to get a rough indication of the way participants perceive the bot. The measure for *Anthropomorphism*, as expected, resulted in a value below average (M=2.81, SD=0.70). The measure for *Animacy* or lifelikeness resulted in a value around average (M=3.02, SD=0.76). The *Likeability* measure resulted in a value above average (M=3.66, SD=0.63). For *Perceived intelligence*, the measured value was a bit above average (M=3.16, SD=0.75). We conclude from the questionnaire that participants generally perceived the bot as likeable, not very humanlike, and somewhat lifelike and intelligent.

Most participants made comments or wrote down answers to the open questions on the final questionnaire that indicated that choices made were related to human movements. Some participants interpreted the bot as having the shape of a human arm, whereas one participant saw it as some kind of worm. Eight participants explicitly responded to the open questions on the final questionnaire that they made their selection based on similarity to human gestures: gestures they themselves or other people would perform. Some participants described movements in terms of how "humanlike" or "natural" they were. Others assigned labels such as aggressive, relaxed, lifelike, empathic and kind to the gestures. Three participants remarked that the bot moved too slowly for them.

5.3.4 Gesture selection

We looked at which gestures participants would select, based on the order in which gestures were displayed. Each generation, ten gestures were displayed on five different pages, which participants could view by pressing navigation buttons on screen. Figure 9 shows a graph in which the number of times a gesture was selected is shown, relative to the position they were displayed at. For example, the gesture displayed at the 9th position was selected 32 times in total (on average M=2.0 (SD=2.0) times per participant). The gesture displayed at the 10th position was selected 66 times in total (on average M=4.1 (SD=2.6) times per participant). A two-tailed paired t-test revealed significant difference between the number of times gestures at the 9th and 10th position were selected, with $p = 0.045$ (N=16).

	Anthropomorphism	Animacy	Likeability	Perceived intelligence	Perceived safety
Cronbach's alpha	0.68	0.85	0.85	0.70	0.67
Internal consistency	Questionable	Good	Good	Acceptable	Questionable
M	2.81	3.02	3.66	3.16	3.54
SD	0.70	0.76	0.63	0.75	0.80

Table 4: Godspeed Questionnaire results (N=16). Measured on a scale from 1 to 5.

Figure 9: Number of times gestures were selected based on the order in which they were displayed.

Looking at the number of times a gesture is selected relative to the position it is displayed at, it appears the rate of selection decreases, with the exception of the gesture at the 10th position. We remind the reader that the gesture displayed at the 10th position was one of the gestures selected by the participant during the previous generation, except for the gesture displayed at the 10th position of the first randomly generated population. We see it as positive that the gesture displayed at the 10th position was selected this many times: although the rate of selection decreases, the high selection rate of the last gesture appears to indicate that participants at least watched the final gestures most of the time.

The number of times gestures were selected based on their position appears to decrease. We can link this effect to one observed by Zhou *et al.*, who note that the position of a video on the related video list on YouTube has an effect on the click through rate(CTR): the video on the first position has a CTR of 5.9%, while the video on the 20th position has a CTR of 1.0% [88].[6]

5.3.5 Perceived improvement

One of the open questions asked on the final questionnaire was *"Did the quality of the gestures that were presented to you get better or worse, or did they stay the same? (Quality here means that the gesture represented the instruction written at the top of the screen well)"*. Thirteen out of sixteen participants explicitly noted that gestures got better or improved. One participant noted that there were good and bad representations. One participant noted that at times the robot appeared to "get tired" and that gestures seemed more apathetic. One participant noted that they stayed roughly the same. In combination with remarks by participants (such as *"You can really see they're adapting, it's really cool."*, we conclude that a majority participants felt that the proposed gestures improved over time, or at least that their choices were taken into account to some extent.

We also included another measure to determine whether participants felt the gestures were qualitatively better during the later cycles. A gesture that was selected during the first generation was presented to the participant, as well as a gesture that was selected during the 10th generation. The presented gesture was the first selected gesture of the three that participants could select each generation. If a participant had selected the same gesture as his/her second and third choice, this gesture would be presented instead. Participants were asked to rate the quality of these gestures on a scale from one to five, with "quality" referring how appropriate the gesture was for signifying the instruction assigned to them.

The gestures were presented side-by-side, so they could be compared by the participant. The gesture selected during the first generation received a score of M=2.3 (SD=1.0) on a scale from 1 to 5. The gesture selected during the 10th generation received a score of M=3.8 (SD=1.0). Shapiro-Wilk tests reveal that the current data set is not normally distributed. Therefore, we will apply the sign test as described by Gibbons and Chakraborti [30]. The null hypothesis is $H_0 : M_D = 0$ and the alternative is $H_1 : M_D > 0$ where M_D is the mean of the differences of ratings of gestures selected during the 10th generation minus the ratings of gestures selected during the 1st generation. The number of positive differences is 12, whereas 2 are negative, and 2 times the differences are zero. The differences are zero when gestures received exactly the same rating by experiment participants. We will treat half of the zero differences as positive and half as negative. This results in 13 positive and 3 negative differences, which is a more conservative estimate than removing the measurements with zero differences entirely from the calculation in this case. The right tail p-value for N=16 with 13 positive differences is $p = 0.0106$, which means we can reject H_0 at significance level $p = 0.05$. This indicates that participants seem to prefer the gesture selected during the 10th generation over the gesture selected during the 1st generation.

We note that the first selected gesture during the 10th cycle is not necessarily the gesture participants would themselves indicate as the "best" gesture during that generation. We also note that participants could be biased to rate the quality of

[6] Zhou et al. do not make explicit whether the ranking of YouTube's video suggestions at the time was completely random or not.

the gesture they selected during the 10th generation as higher than the one selected during the first generation, because they have just seen that gesture.

5.3.6 Developed gestures

The genotype encodings of the gestures selected during the 10th generation were recorded for 12 participants. See Table 5. More experiments would need to be done to see whether similar results would hold across a larger pool of participants.

When we look at the genotypes of gestures developed for a particular gesture class, we can see that there is little consensus among participants about target vectors that belong to a particular gesture. Only two vectors of the *"Wave one's hand"* condition are part of one or more of the gestures for every participant. However, it is difficult to make any substantive claims on the genotype only: a movement towards one target position and back may be interpreted the same way as a movement towards a slightly different target position.

5.3.7 Gesture interpretation

Participants 4 to 16 were asked to interpret two of the gestures developed by the first three participants. One participant had misinterpreted the question and will be excluded from our discussion. Of the remaining 24 interpretations, six appear to be in line with the instruction. Instruction "Wave one's hand!" was interpreted correctly five times as waving or greeting. See Table 6 for details. *"Go away"* was once interpreted as *"You, go over there!"*, and once as *"You must move over there"*. Both interpretations refer to asking a person to go to a location that is at some distance from the sender, but do not quite match the intent or affective content of the expression *"Go away!"*.

The remaining interpretations showed little to no similarity to the actual instruction, or even opposing meanings (e.g. *"Go away!"* being interpreted as *"follow me"*). Although this does not offer conclusive evidence, this gives an indication that it is difficult for participants to correctly identify gestures developed by other people using this system.

5.4 Discussion

5.4.1 Interaction with the system

As we have seen, participants could complete the task, yet most experienced it as difficult. All participants reported some kind of difficulty in interpreting the gestures before selecting the ones of their choice. Participants reported different reasons for this difficulty. Some participants mentioned they found it difficult to interpret the gestures because they were presented in an isolated way. The form of the bot is rather abstract. Some participants noted they missed the body that the "arm" was supposedly attached to.

The task took on average 134.4 seconds per generation to complete, with one participant taking almost six minutes to choose three bots in one of the generations s/he evaluated. The type of interaction we consider desirable would place less of a cognitive burden on the participant, and take less time to complete. The number of choices that have to be made in the current system make the participant's required effort rather large.

We mentioned before that a surrogate fitness function can be used to reduce the number of required interactions with the system by human participants. A very simple surrogate fitness function that reduces the number of solutions that need to be evaluated by the participant, can be achieved by showing only those solutions most different from each other, and estimating the results for the solutions that have not been displayed. This requires the development of a distance measure.

Some people did not finish watching every gesture before making a decision or moving on to different gestures. We already adapted the system to show a short black flash when a gesture is finished. A counter or timer could help give them an indication of when the gesture will be completed, so they realize whether they miss information if they move on to evaluating different gestures. Reducing the number of bots that need to be evaluated by the participant could also help. Additionally, the speed at which gestures are performed can be altered. The range of speeds that is possible now may be wider than required, resulting in a high number of gestures that are performed at a low speed, which can be perceived as annoying by participants. We could also think of allowing participants to set the speed manually during the first generations, and making it co-evolve during later generations.

One participant mentioned she would repeat the instruction *"Come here!"* in her head while she was watching the bot making the movement. The duration of repeating the instruction in her mind and the bot making the movement would often mismatch. This once again stressed the importance of considering the timing and duration of gestures: how long should the gesture take in total to match a spoken utterance or the urgency of the request?

In the current experiment, participants were not told in advance they were dealing with an evolutionary algorithm, or that the decisions they made could alter or improve the gestures that would be displayed next. Due to the computer science background of some participants, some were able to guess that this was the case. It is possible that it makes a difference to the end result if participants are aware they can steer the process, and how they can do so. The amount of information needed to steer the process might not necessarily be that great. Lund *et al.* performed an experiment in which children could evolve robot controllers, and they found that the behaviours that were developed, such as obstacle avoidance, did not differ much from those developed before in the field of evolutionary robotics [47].

The amount of information displayed to participants is another issue. Currently, gestures run at the same time. Lund *et al.* take an approach in which the paths robots take on a 2D plane are displayed one by one, and path trajectories are visualized when the movement is completed [47]. The current set-up might benefit from such an approach.

Another downside of the current system is that gesture movement is highly simplified: only 9 target positions can be reached from the starting position. Additionally, the middle joint is restricted in its movement. This is convenient in limiting the number of possibilities, thereby reducing the search space, but also results in movement that does not satisfy some participants' desires for the movement, as was remarked by two participants. One of them also remarked that the move-

Condition	Wave one's hand!	Come here!	Go away!
Number of recorded gestures	9	15	12
Mean number of target vectors in genotype of a gesture	M=10.3 (SD=5.9)	M=4.5 (SD=1.7)	M=5.3 (SD=3.1)
Minimum number of vectors in genotype	5	3	3
Maximum number of vectors in genotype	23	9	13
Number of unique vectors (excluding start position)	5	5	5
Number of vectors shared by all participants (excluding start position)	2	0	0
Number of vectors used by only one participant	2	2	3

Table 5: Data on the genotype encodings of gestures in the final generations of 12 participants.

Condition	Wave one's hand!	Come here!	Go away!
Recognition rate	62.5%	14 %	0%
Interpretations marked as correct	*"Greeting"* *"Hello"* *"Hi"* *"Waving, saying hello"* *"Hello, I'm here"*	*"You, come over here"*	-
Correct interpretation	5 out of 8	1 out of 7	0 out of 9
Interpretations marked as having some similarity	-	*"I'm here"*	*"You, go over there"* *"You must move over there"*

Table 6: Recognition rates of the three instructions.

ment seemed "boxed-in". This is due to the limited number of target positions that are located at a fixed distance from the bot's root node.

The current system shows rather jumpy movements in straight lines. Rather than bouncing back and forth between target positions, we could make this appear smoother. One solution is to make the bot follow Bézier curves. Initial exploration with rational Bézier curves resulted in smoother movement. If the weights of the rational Bézier curves are part of the genome, the curvature of the movement between target positions could co-evolve.

5.4.2 Results of interaction with the system: developed gestures

Not all gestures developed by the first three participants were interpreted correctly by the last 13 participants. The gesture that was developed by one of the first three participants for condition *"Wave one's hand!"* was interpreted correctly most often, and achieved a recognition rate of 62.5% for our limited sample of 8 participants who were asked to interpret this gesture. This is higher than the 30% rate achieved by Ende *et al.* for this gesture, although the system they developed the gesture for was of course wholly different. We need perform more experiments in which a higher number of different gestures are evaluated, if we want to make claims about how well people can recognize developed gestures.

We might explain the high recognition rate by looking at the type of gesture that achieved it. Emblematic gestures are gestures for which a certain convention exists. An example of such a gesture is the peace sign, and waving one's hand to greet someone may also be described as an emblematic gesture. It is possible that emblematic gestures present a stronger mental model to the participant who is adapting the gesture than other gestures. This could make it easier to know what one is looking for while making a selection. The phrase *"Wave one's hand!"* also implicitly indicates the shape the gesture should take. We did not instruct participants to see the bot as an arm, but it was the way most participants interpreted the bot, based on their answers to the questionnaire and their remarks. Combined with the type of instruction, this may have presented a clearer goal for the participant modelling the gesture for gesture class *"Wave one's hand!"*.

Eight out of sixteen participants explicitly mentioned that they made decisions based on how they themselves or other humans would perform gestures. This indicates that they started out by choosing gestures that conformed to a mental model they had of human gestures. We think the interaction has the potential to change over time, as two participants indicated that they started to choose based on the communication abilities of the bot after interacting with the system after a while. In line with this, we pose that gestures developed by interacting with the current system can become idiosyncratic, due to the fact only one individual interacts with the system for an extended period of time.

The gestures that resulted from interacting with the system could be described as personalized rather than general. This could be different if participants only interact with the system for two cycles per gesture class, and continue with the results of the previous participant. Will this result in gestures that can be interpreted correctly by people not familiar with the system?

6 Experiment Proposal

During the experiment in section 5 we saw that gestures that were developed by an individual participant were not interpreted correctly by other participants most of the time. We propose an experiment in which we investigate whether gestures that are developed by multiple people are easier to interpret by others than gestures developed by an individual who interacts with the system for an extended period of time.

Garrod *et al.* carried out a series of experiments in which participants repeatedly engaged with the same interaction partner through the use of drawings. The interaction partner could choose which meaning out of a set of meanings the drawer was trying to indicate. When the feedback between participants was allowed to be more direct, they found that the graphical complexity of the drawings reduced over subsequent interactions. They pose that iconic graphical signs can evolve over the course of the interaction into more symbolic graphical signs. Garrod *et al.* found that people who were not part of the interaction were significantly less accurate at identifying the meaning of the drawings than those part of the interaction [28].

We expect that gestures developed by an individual interacting with the system for multiple generations have the possibility to become idiosyncratic or particular to the interaction between that individual and the system, whereas gestures developed by a larger pool of participants could potentially become more universal in the sense that they are easier to interpret by other people who are familiar with the culture of those who developed them. We propose to let gestures be developed by multiple people. This could take the shape of a transmission chain experiment during which each participant interacts with the system for one generation per gesture class. For example, if 4 participants take part in one transmission chain experiment, participant 2 continues with the results of participant 1, participant 3 continues with the results of participant 2, and participant 4 continues with the results of participant 3.

In order to be able to compare the gestures developed this way, we would need to carry out a number of separate transmission chains. This would allow us to compare the recognition rates of gestures and make sure a result is not a lucky hit (or miss). Kirby *et al.* describe two diffusion chain experiments in which they observed increased structure and transmissibility in the artificial languages that were learned and reproduced by participants. The diffusion chain started out with an artificial language with a random structure. The first participant had to memorize and reproduce this language. Each of the remaining participants would memorize the artificial language produced by the previous participant, and attempt to reproduce it. Both experiments involved four separate diffusion chains of ten participants each [42]. The interaction with the system we developed is different from a language learning task, but we could use a similarly structured experiment. We could take a similar approach, and ask participants to interact with the system for one generation each, for a number of different gestures classes from the gesture lexicon.

After creating gestures for the gesture lexicon this way, we propose to test whether the gestures developed using this approach are easier to interpret by other people than gestures developed by individual participants. Such an interpretation test could take a similar shape to the interpretation questions used in the experiment in section 5. When this interpretation test is carried out with a larger pool of participants, a measure for similarity of linguistic interpretations to each of the gesture classes should be defined, and the interpretations should be analyzed by two researchers who can compare their results. Alternatively, the interpretation test could take the shape of a multiple choice test in which participants can indicate a gesture class after watching a video of a gesture.

We hypothesize that gestures developed by multiple people will be easier to interpret by people who did not interact with the system, than gestures that were developed over the course of a longer interaction with the system by a single participant. When gestures are developed by multiple people, we propose to only let them select bots for one generation per gesture class. In order to compare the individually developed gestures to the gestures developed by multiple people in an appropriate way, we would need to develop gestures for the individual condition for the same number of gesture classes as the transmission chain experiment.[7] If the gestures that are developed by individuals are to be compared across gesture classes, a much larger sample of participants would be needed for the development of the gestures than the sample of sixteen participants in the current paper.

Developing gestures for the gesture lexicon that is described in section 3.3 would already provide material for comparison. Reconsideration of the gesture classes and ensuring the chosen gesture classes cover a wide enough range of different instructions could improve the validity of the results obtained with the proposed experiment.

7 Conclusion

Our aim was to create body language in the context of interaction between humans and a virtual robot. We defined an embodiment for a virtual robot and described the parameters that could be changed so the virtual robot could perform gestures. In order to investigate our research questions, a system was developed that allowed for interactive selection of gestures via a user interface, and subsequent adaptation of the gestures with an evolutionary algorithm. The system was tested with sixteen participants. The present task was experienced as difficult by a majority of experiment participants. Thirteen out of sixteen participants noted they experienced improvement in the gestures over ten subsequent generations. Gestures selected during the tenth generation were preferred over gestures selected during the first generation. Three gestures that were developed by the first three participants have undergone initial evaluation by the last thirteen experiment participants. The solutions have achieved recognition rates of 62.5% (N=8), 14% (N=7) and 0% (N=9) respectively for the gesture classes *"Wave one's hand!"*, *"Come here!"* and *"Go away!"*, but more evaluations are needed to confirm that these results hold across a larger population.

The developed system needs to be improved and more re-

[7] Preferably with a system that is improved and allows for more possibilities for the gestures.

search needs to be done to confirm Hypothesis 1. We proposed a transmission chain experiment in which gestures are developed by different participants over subsequent generations. We hypothesize that this may lead to gesture language that is more recognizable than gestures that are developed by a single participant.

This paper began by raising the question whether we can adapt a robot's gestures based on the effectiveness of its communication towards humans. The process of developing a system that aims to do just that, emphasized the complexity of this topic. Not only did we need to define an embodiment and a way to encode gestures, we also needed to develop an algorithm and an interface for participants to interact with. All of the choices that have been made in the process influence how people who interact with the system make decisions. The way parameters of the evolutionary algorithm have been set determines how gestures are adapted. Choosing a particular genotype encoding influence the phenotype, determining such properties as the speed of the gesture. This in turn has an effect on the interpretation of the gesture by participants. Human participants are limited in terms of their attention span, and interact with the system in their own particular ways. We would like to encourage more research on the interaction between these variables, so we may one day arrive at robotic systems that successfully adapt the way they communicate in the context of interaction with humans.

8 Future work

The bot's gestures are currently only evolved virtually, not physically. If it would be desirable to implement gestures developed with the system, each instance would need to be tested subject to real-world conditions. For example, pointing at a particular spot is likely perceived differently in the 3D world than in the simulated 3D world on a 2D screen. A partial solution to this could be to use a Virtual Reality environment during the initial design process. Another possibility is to run the gesture adaptation system on hardware, with some changes. Physical embodiments bring along new dimensions that have an impact on the interaction between humans and robotic systems. Rae *et al.* investigated the impact of height of a telepresence system on the dominance participants in their experiment exhibited [63]. By rendering a virtual or physical model of the bot at different scales, we can investigate if there is a difference in the impression the bot makes on an interaction partner.

The system can also be improved to make the task easier for the participant. Reducing the number of gestures a participant needs to review will help counter fatigue. This could be achieved by developing a surrogate fitness function in the form of a distance function, which is used to ensure that only the most different solutions need human evaluation. Another type of surrogate fitness function we may consider are decision trees or other classifiers, which attempt to mimic human choices. The strategies that were previously described for the user test could serve to inform the attributes for the construction of decision trees. The interface can be improved on further, by adding an overview page with all gestures per generation and a counter that indicates how long the currently displayed gesture will take to complete.

We can consider changing the nature of the task. At the moment, gestures are directly selected by the participant, presumably based on how well a gesture matches the participant's mental model of the way the gesture should be executed. Instead, the task could be more focused on *interpreting* the bot's gestures. An example would be to ask the participant to choose which instruction the bot is trying to convey out of a number of possibilities.

Additionally, the current system allows for a limited range of possibilities in the way the bot is rendered. More possibilities to adapt the gestures may be considered. Participants could be given more control over the mutation rate or speed of movement. The effect of allowing participants to determine these properties by themselves needs yet to be determined. Another option is to add more target positions, or to introduce another type of mutation that allows for slightly different target positions. At the moment, the middle joint's movement is restricted, and only allowed to bend in a certain way to ensure movement is replicated the same way each time it is executed. Allowing more freedom could result in different, more suitable gestures. The bot's morphology could be made evolvable. The effect of adding a "body" or torso can be investigated as well. Lastly, we may also consider adding different modalities to the interaction, for example in the form of sound, smells, coloured light, or screen based visuals.

References

[1] Aigner, R., *et al.* (2012) *Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI*, Microsoft Research TechReport MSR-TR-2012-111

[2] Aldebaran Robotics, *NAO robot*, `www.ald.softbankrobotics.com/en/cool-robots/nao`, Retrieved 23/3/2017, 22:29h

[3] Aldebaran Robotics, *Pepper robot*, `www.ald.softbankrobotics.com/en/cool-robots/pepper`, Retrieved 23/3/2017, 22:30h

[4] Rethink Robotics, `http://www.rethinkrobotics.com/baxter/`, Retrieved 24/2/2017, 10:15h

[5] Bartneck, C., Forlizzi, J. (2004) *A Design-Centred Framework for Social Human-Robot Interaction*, Ro-Man 2004, pp. 591 - 594

[6] Bartneck, C., *et al.* (2009) *Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots*, International Journal of Social Robotics 1.1, pp. 71 - 81

[7] Blow, M., *et al.* (2006) *The art of designing robot faces: Dimensions for human-robot interaction*, 1st ACM SIGCHI/SIGART conference on Human-robot interaction, ACM

[8] Braitenberg, V. (1984) *Vehicles*, MIT Press, Cambridge, pp. 1 - 19

[9] Breazeal, C., Scassellati, B. (1999) *How to build robots that make friends and influence people*, Intelligent Robots and Systems, IROS'99

[10] Breazeal, C., Scassellati, B. (2002), *Robots that imitate humans*, TRENDS in Cognitive Sciences 6.11, pp. 481 - 487

[11] Breazeal, C., *et al.* (2005) Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork, IROS 2005, pp. 383 - 388

[12] Broadbent, E., *et al.* (2013) *Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality*, PloS one 8.8, pp. 1 - 9

[13] Carey, B. (2016) *Maiden voyage of Stanford's humanoid robotic diver recovers treasures from King Louis XIV's wrecked flagship*, 27/4/2016, `news.stanford.edu/2016/04/27/robotic-diver-\\recovers-treasures/`, accessed on 23/3/2017, 22:22h

[14] Cheney, N., *et al.* (2013) *Unshackling evolution: evolving soft robots with multiple materials and a powerful generative encoding*, GECCO'13, ACM, pp. 167 - 174

[15] Cambridge English Dictionary, `http://dictionary.cambridge.org/`

[16] Corballis, M.C. (2017) *The evolution of language: sharing our mental lives,* Journal of Neurolinguistics 43.B, pp. 120 - 132

[17] DARPA Robotics Challenge (2015) *Finals 2015, DRC Finals - Qualified Teams*, `http://archive.darpa.mil/roboticschallenge/teams.html`

[18] Dautenhahn, K. (1999) *Embodiment and Interaction in Socially Intelligent Life-Like Agents*, C. Nehaniv (Ed.): Computation for Metaphors, Analogy, and Agents, LNCS 1562, pp. 102 - 141

[19] Dautenhahn, K. (2002) *Design spaces and niche spaces of believable social robots*, Proceedings 11th IEEE International Workshop on Robot and Human Interactive Communication

[20] Dautenhahn, K. (2004) *Robots we like to live with?! - a developmental perspective on a personalized, life-long robot companion*, ROMAN 2004

[21] Dautenhahn, K. (2007) *Socially intelligent robots: dimensions of human–robot interaction*, Philosophical Transactions of the Royal Society B 362, pp. 679 - 704

[22] Duffy, B.R. (2003) *Anthropomorphism and the social robot*, Robotics and Autonomous Systems 42, pp. 177 - 190

[23] Eiben, A.E., Smith, J.E. (2015) *Introduction to Evolutionary Computing,* Second Edition, Springer Natural Computing Series

[24] Ende, T., *et al.* (2011) *A human-centered approach to robot gesture based communication within collaborative working processes*, IROS 2011

[25] Fastbrick Robotics, *Hadrian X*, `http://fbr.com.au/`, accessed on 23/3/2017, 12:37

[26] Ferrari, F., Paladino, M.P., Jetten, J. (2016) *Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness*, International Journal of Social Robotics 8.2, pp. 287 - 302

[27] Floreano, D. *et al.* (2007) *Evolutionary Conditions for the Emergence of Communication in Robots*, Current Biology 17, pp. 514 - 519

[28] Garrod, S., *et al.* (2007) *Foundations of representation: where might graphical symbol systems come from?*, Cognitive Science 31.6, pp. 961 - 987

[29] Gibbon, D., *et al.* (2003) *A computational model of arm gestures in conversation*, INTERSPEECH 2003

[30] Gibbons, J.D., Chakraborti, S. (2011) *Nonparametric statistical inference*, 4th Edition, Springer Berlin Heidelberg, pp. 168 - 185

[31] Goldin-Meadow, S. (1999) *The role of gesture in communication and thinking,* Trends in cognitive sciences 3.11, pp. 419 - 429

[32] Goodrich, M.A., Schultz, A.C. (2007) *Human–Robot Interaction: A Survey*, Foundations and Trends in Human-Computer Interaction 1.3, pp. 203 - 275

[33] Hanke, T. (2004) *HamNoSys - representing sign language data in language resources and language processing contexts*, LREC 4

[34] Heinerman, J., Rango, M., Eiben, A.E. (2015) *Evolution, individual learning, and social learning in a swarm of real robots*, 2015 IEEE Symposium Series on Computational Intelligence

[35] Hiroshi Ishiguro Laboratories, *Geminoid robots*, `www.geminoid.jp/en/robots.html`, Retrieved 22/3/2017, 9:55h

[36] Hanson Robotics, *Android robots*, `www.hansonrobotics.com/robot-gallery/`, Retrieved 22/3/2017, 9:43h

[37] Honda, *ASIMO, "The World's Most Advanced Humanoid Robot"*, `http://asimo.honda.com/` , Retrieved 23/3/2017, 13:05h

[38] Ikemoto, S., *et al.* (2012) *Physical human-robot interaction: Mutual learning and adaptation*, IEEE robotics & automation magazine 19.4, pp. 24 - 35

[39] ISO, ISO 8373:2012(en) Robots and robotic devices `https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en`

[40] Kennedy, J., Baxter, P., Belpaeme, T. (2016) *Nonverbal Immediacy as a Characterisation of Social Behaviour for Human-Robot Interaction*, International Journal of Social Robotics, pp. 1 - 20

[41] Kim, A., *et al.* (2013) *The Effects of Familiarity and Robot Gesture on User Acceptance of Information*, HRI 2013 Proceedings, pp. 159 - 160

[42] Kirby, S., Cornish, H., Smith, K. (2008) *Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language*, Proceedings of the National Academy of Sciences 105.31, pp. 10681 - 10686.

[43] Kranstedt, A., Kopp, S., Wachsmuth, I. (2002) *Murml: A multimodal utterance representation markup language for conversational agents*, AAMAS'02 Workshop *Embodied conversational agents-let's specify and evaluate them!*

[44] Kruijff-Korbayová, I., *et al.* (2015) *TRADR Project: Long-Term Human-Robot Teaming for Robot Assisted Disaster Response*, KI - Künstliche Intelligenz 29.2, pp. 193 - 201

[45] Leite, I., Carlos M., Paiva, A. (2013) *Social robots for long-term interaction: a survey*, International Journal of Social Robotics 5.2, pp. 291 - 308

[46] Lund, H.H., Hallam, J., Lee, W.P. (1997) *Evolving robot morphology*, IEEE International Conference on Evolutionary Computation

[47] Lund, H., Miglino, O. (1998) *Evolving and breeding robots*, Evolutionary Robotics, Springer Berlin/Heidelberg

[48] Mason, M., Lopes, M.C. (2011) *Robot self-initiative and personalization by learning through repeated interactions*, Proceedings of the 6th international conference on Human-robot interaction, ACM

[49] Marocco, D., Cangelosi, A., Nolfi, S. (2003) *The emergence of communication in evolutionary robots*, Phil. Trans. R. Soc. Lond. A 361, pp. 2397 - 2421

[50] Martell, C.H. (2005) *An extensible, kinematically-based gesture annotation scheme*

[51] McNeill, D. (1985) *So you think gestures are nonverbal?* Psychological review 92.3, pp. 350 - 371

[52] Miller, H. J. *et al.* (2016) "Blissfully Happy" *or* "Ready to Fight"*: Varying Interpretations of Emoji*, 10th International AAAI Conference on Web and Social Media

[53] MIT Media Lab, robot Kismet, `http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html`

[54] MIT Media Lab, robot Nexi, `http://news.mit.edu/2008/nexi-0409`

[55] Mitra, S., Acharya, T. (2007) *Gesture Recognition: A Survey*, IEEE Transactions of Systems, Man, and Cybernetics, Part C: Applications and Reviews, 37.3

[56] Mori, M. (1970) *The Uncanny Valley*, translated by MacDorman, K.F., and Kageki, N., IEEE Robotics & Automation Magazine, June 2012

[57] Nanyang Technological University Singapore, robot Nadine, `http://media.ntu.edu.sg/NewsReleases/Pages/newsdetail.aspx?news=fde9bfb6-ee3f-45f0-8c7b-f08bc1a9a179`

[58] Oxford English Dictionary, `www.oed.com`

[59] PAL, *Personal Assistant for a healthy Lifestyle*, `http://www.pal4u.eu/`, Accessed on 23/3/2017 12:47h

[60] Papadogiorgaki, M., *et al.* (2006) *Gesture synthesis from sign language notation using MPEG-4 humanoid animation parameters and inverse kinematics*, IET, 2nd IET International Conference on Intelligent Environments 1

[61] Powers, A., Kiesler, S. (2006) *The advisor robot: tracing people's mental model from a robot's physical attributes*, Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-Robot Interaction

[62] Quinn, M. (2001) *Evolving communication without dedicated communication channels*, European Conference on Artificial Life, Springer Berlin Heidelberg

[63] Rae, I., Takayama, L., Mutlu, B. (2013) *The influence of height in robot-mediated communication*, Proceedings of the 8th ACM/IEEE international conference on Human-Robot Interaction

[64] Rasmussen, M.K., *et al.* (2012) *Shape-changing interfaces: a review of the design space and open research questions*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM

[65] de Rooij , A., Broekens , J., Lamers , M. H. (2013) *Abstract expressions of affect*, International Journal of Synthetic Emotions 4.1, pp. 1 - 31

[66] Russell, S.J., Norvig, P. (2010) *Artificial Intelligence, A Modern Approach*, Third Edition, Prentice Hall Series in Artificial Intelligence

[67] Saerbeck, M., Bartneck, C. (2010) *Attribution of affect to robot motion*, HRI'10, pp. 53 - 60

[68] Salem, M., *et al.* (2012) *Generation and Evaluation of Communicative Robot Gesture*, International Journal of Social Robotics 4, pp. 201 - 217

[69] Salem, M. *et al.* (2013) *To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability*, International Journal of Social Robotics 5, pp. 313 - 323

[70] Saygin, A.P., *et al.* (2011) *The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions*, Social cognitive and affective neuroscience 7.4, pp. 413 - 422

[71] Scholl, B., Tremoulet, P (2000), *Perceptual causality and animacy*, Trends in cognitive sciences 4.8, pp. 299 - 309

[72] Shead, S. (2017) *Amazon now has 45,000 robots in its warehouses*, `http://www.businessinsider.com/amazons-robot-army-has-grown-by-50-2017-1` Jan. 3, 2017, 8:16 AM, accessed on 23/3/2017, 13:01h

[73] Simbe Robotics, *Tally*, `http://www.simberobotics.com/`, accessed on 23/3/2017, 12:50h

[74] Simon, M. (2017) *Boston Dynamics' New Rolling, Leaping Robot Is an Evolutionary Marvel*, `https://www.wired.com/2017/03/`

boston-dynamics-new-rolling-leaping-robot-evolutionary-marvel/01/03/2017, Retrieved 23/3/2017, 22:27

[75] Sims, K. (1994) *Evolving 3D morphology and behavior by competition*, Artificial Life IV Proceedings, ed. by R. Brooks, P. Maes, MIT Press, pp. 28 - 39

[76] Song, S., Yamada, S. (2017) *Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot*, HRI '17, Vienna, Austria

[77] Spranger, M. (2015) *Incremental Grounded Language Learning in Robot-Robot Interactions - Examples from Spatial Language*, 5th International Conference on Development and Learning and on Epigenetic Robotics

[78] Steels, L. (2001) *Language games for autonomous robots* IEEE Intelligent systems 16.5, pp. 16 - 22

[79] Suga, Y., et al. (2005) *Interactive evolution of human-robot communication in real world*, IROS 2005

[80] Sutton, V. (2014) *Lessons in SignWriting*, 4th edition, 1st edition published in 1990

[81] Takagi, H. (2001) *Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation,* Proceedings of the IEEE 89.9, pp. 1275 - 1296

[82] Trippel, T., *et al.* (2004) *CoGesT: a Formal Transcription System for Conversational Gesture*, LREC 2004

[83] Veruggio, G., Operto, F., and Bekey, G. (2016) *Roboethics: Social and Ethical Implications,* Springer handbook of robotics, Springer International Publishing, pp. 2135 - 2160

[84] Wang, J., Terpenny, J. (2003) *Interactive evolutionary solution synthesis in fuzzy set-based preliminary engineering design*, Journal of intelligent manufacturing 14.2, pp. 153 - 167

[85] Waseda University, KOBIAN, http://www.takanishi.mech.waseda.ac.jp/top/research/kobian/KOBIAN-R/index.htm

[86] Wilson, M. (2002) *Six views of embodied cognition*, Psychonomic Bulletin & Review 9.4, pp. 625 - 636

[87] Xu, J. *et al.* (2015) *Mood contagion of robot body language in human robot interaction*, Auton Agent Multi-Agent Syst 29, pp. 1216 - 1248

[88] Zhou, R., Khemmarat, S., Gao, L. (2010) *The impact of YouTube recommendation system on video views*, Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, ACM

[89] Ziemke, T. (2003) *What's that thing called embodiment?*, in: Alterman, R., Kirsh, D. (eds.) Proceedings of the 25th Annual Conference of the Cognitive Science Society, pp. 1134 - 1139

[90] Ziemke, T., Thill, S. (2014) *Robots are not Embodied! Conceptions of Embodiment and their Implications for Social Human-Robot Interaction*, Seibt, J. *et al. Sociable Robots and the Future of Social Relations*, IOS Press, pp. 49 - 53

Appendix A: Terminology

Robots

The definition of a robot varies widely depending on the source. The Cambridge English dictionary speaks of *"a machine controlled by a computer that is used to perform jobs automatically"* [15] while the Oxford English Dictionary describes a robot as *"A machine capable of automatically carrying out a complex series of movements, esp. one which is programmable"*. They also refer to software robots in the entry: *"A program for automatically performing a task (esp. on the Internet) without continuous human intervention"* [58]. ISO (the International Organization for Standardization) defines a robot as an *"actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks"* [39].

In summary, these definitions describe a robot as a programmable machine that can carry out certain actions (semi-)autonomously, which can move within its environment. This does not describe the actual environment the entity could be in: the definitions could also apply to a virtual space. Internet bots are an example of (ro)bots in virtual space.

We will take a working definition of robots as acting agents. These agents are entities that could in principle exist in physical space, the same space we inhabit as humans, and act on this environment. It is possible and practical to model those (physically realized) robots in virtual space. We acknowledge that this approach will likely result in discrepancies between the virtual model and how it can be realized in practice. However, it can be beneficial to first create a virtual model.

Social robots

Dautenhahn characterizes social robotics by agents that are embodied individuals with histories, that can communicate with other members of the heterogeneous group they belong to. Communication between agents obtains meaning as a result of agents sharing a context, interacting with each other and imitating each other [18].

Bartneck and Forlizzi propose the following definition of social robots: *"A social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact."* They suggest this implies that the social robot should have an embodiment. They propose a framework for classifying social robots along the dimensions form, modality, social norms, autonomy and interactivity. Worth noting is that they write about social norms: *"As social norms can be defined by the interactions between people, we assert that they can be defined by the interactions between people and robots"* [5]. This suggests that norms in social HRI could be different than those that exist between humans. The aspect of norms as described by Bartneck and Forlizzi is approached differently by Dautenhahn. She does not speak of norms, but of *"meaning[, which] is transferred between two agents by sharing the same context"* [18]. Dautenhahn's approach, which allows for meaning to come forth out of a shared contexts, seems preferable to setting behavioural norms preceding the interaction, if a robot is to be developed that can adapt its behaviour to a (new) context. Predefining norms would also require thorough study of every possible (cultural) context that a robot is to be deployed in. If norms can be negotiated between the human and the robotic agent, this would provide more flexibility. Of course, it is likely that people who engage in interaction with a robot, have certain expectations of the subsequent interaction. Violating expectations, for example by very unpredictable, dangerous movement by the robot, is likely to have negative effects on the attitude of the human towards the robot.

Embodiment

We should be careful when talking about robots as embodied systems. Ziemke and Wilson argue that we should disentangle notions around embodied cognition [86, 89]. According to Ziemke, who bases his discussion on Searle's 1980 paper, we should be careful to consider robots as embodied AI, because merely running a computer program in a body does not automatically make the robot behave intentionally, which is characteristic of human (embodied) cognition [90]. In the current paper, the word *embodiment* merely refers to the physical appearance of the robot and is used interchangeably with the term morphology, which describes the entire body plan of a robot, including sensors and motors. We note that sensors and motors are not considered in the experiment that is described later.

Interaction

Interaction in an HRI context implies some form of communication between robots and humans. Goodrich and Schultz classify HRI interactions as either proximate or remote. For remote interactions, we can think of teleoperated robots. Social robots usually take part in proximate interactions with social, cognitive and emotive elements. If we take a programming-based approach, we can say that even fully autonomous systems involve some form of interaction, as humans design algorithms for the robot, as well as reprogram and maintain it if necessary [32]. In the current paper, interaction refers to the situation in which (at least) two agents engaging in communication, thereby affecting one another's actions. The term agent can refer to a human, machine or software program, for instance. An example is the case of an interaction in which one agent sends a signal to the other agent, and the other agent adjusts their behaviour upon reception, as in the work of Steels [78].

Appendix B: Questionnaire

Fake	1	2	3	4	5	Natural
Dead	1	2	3	4	5	Alive
Like	1	2	3	4	5	Dislike
Competent	1	2	3	4	5	Incompetent
Relaxed	1	2	3	4	5	Anxious
Machinelike	1	2	3	4	5	Humanlike
Lively	1	2	3	4	5	Stagnant
Conscious	1	2	3	4	5	Unconscious
Agitated	1	2	3	4	5	Calm
Organic	1	2	3	4	5	Mechanical
Unfriendly	1	2	3	4	5	Friendly
Ignorant	1	2	3	4	5	Knowledgeable
Lifelike	1	2	3	4	5	Artificial
Inert	1	2	3	4	5	Interactive
Foolish	1	2	3	4	5	Sensible
Unkind	1	2	3	4	5	Kind
Moving rigidly	1	2	3	4	5	Moving elegantly
Pleasant	1	2	3	4	5	Unpleasant
Intelligent	1	2	3	4	5	Unintelligent
Apathetic	1	2	3	4	5	Responsive
Awful	1	2	3	4	5	Nice
Responsible	1	2	3	4	5	Irresponsible
Quiescent	1	2	3	4	5	Surprised

1. How did you select gestures? (F.e. based on certain characteristics or certain goals?)

2. Did the quality of the gestures that were presented to you get better or worse, or did it stay the same? (Quality here means that the gesture represented the instruction written at the top of the screen well)

3. Was it difficult to select gestures?

4. Do you have any final remarks?