Article

Experiments with a First Prototype of a Spatial Model of Cultural Meaning through Natural-Language Human-Robot Interaction [†]

Oliver Schürer [1],*, Benjamin Stangl [1], Christoph Hubatschke [2] and Christoph Müller [1]

[1] Department for Architectural Theory and Philosophy of Technics, Vienna University of Technology, 1040 Vienna, Austria; benjamin@stangl.eu (B.S.); mueller@a-theory.tuwien.ac.at (C.M.)

[2] Department of Philosophy, University of Vienna, 1010 Vienna, Austria; christoph.hubatschke@univie.ac.at

* Correspondence: schuerer@tuwien.ac.at; Tel.: +43-(1)-588-012-5122

† This paper is an extended version of our paper published in Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA), Island of Rhodes, Greece, 21–23 June 2017.

Abstract: When using assistive systems, the consideration of individual and cultural meaning is crucial for the utility and acceptance of technology. Orientation, communication and interaction are rooted in perception and therefore always happen in material space. We understand that a major problem lies in the difference between human and technical perception of space. Cultural policies are based on meanings including their spatial situation and their rich relationships. Therefore, we have developed an approach where the different perception systems share a hybrid spatial model that is generated by artificial intelligence—a joint effort by humans and assistive systems. The aim of our project is to create a spatial model of cultural meaning based on interaction between humans and robots. We define the role of humanoid robots as becoming our companions. This calls for technical systems to include still inconceivable human and cultural agendas for the perception of space. In two experiments, we tested a first prototype of the communication module that allows a humanoid to learn cultural meanings through a machine learning system. Interaction is achieved by non-verbal and natural-language communication between humanoids and test persons. This helps us to better understand how a spatial model of cultural meaning can be developed.

1. Introduction

Our everyday life is increasingly enriched by a variety of technologies. Artificial intelligence (A. I.) is already applied in areas such as the stock market, face detection or in controlling mobile phones. A fusion of A. I. with robotics will provide services for everyday life which will not only satisfy individual desires but also solve social problems. These assistive systems are being developed for areas such as marketing, entertainment, communication, health care and care for the elderly. In the field of caregiving, our aging society is increasingly counting on help by humanoid robots. Anthropomorphic machines are more likely to support humans and build confidence when they can move "naturally" in the intimate physical and social spaces where caregiving and assistance actually take place. Their physical humanoid shape does not require any major changes to our human environment. What is crucial for assistive robots to be accepted is that they independently learn from their environment and to interact by means of social skills [1,2].

2. Perception—Space Relation

In contrast to existing domestic engineering or ubiquitous computing systems, social robots are mobile and partly autonomously operating technologies. We are currently living in spaces permeated by technology, and the influence of technology is constantly increasing. In modern buildings, basic utility and domestic engineering systems are already in effect for electricity, wireless networks, and ventilation/lighting systems. In this paper we refer to robots which resemble humans in form and which are built and used for that very reason, namely to humanoid robots. We understand humanoid robots as artificial agents resembling human bodies, perceiving input from their environment, possessing action capabilities and having some aspects of autonomy.

The innovative functions of humanoid robots range from having access to physical objects and digital objects from other technical systems, to their movement in a room and social interaction with humans. Mobile robots introduce new possibilities where fixed technologies reach their limits. They do not only move in physical space but also operate in virtual space through their digital capabilities. They represent an interface in a network between different types of spaces—a mobile and autonomous interface.

In the development of contemporary technology, two approaches complement each other. So-called ubiquitous computing, conceptualized by Mark Weiser in the early nineties [3], and the development of robots [4] are two related strands of 21st century technological achievements. Weiser's approach assumes that technology is woven into everyday life and eventually becomes indistinguishable. While some technologies discreetly infiltrate our living environment, we perceive social robots as material entities. Since they simulate human skills and human appearance, we see humanoids as the most expressive form of robots, about to operate in our intimate spaces.

A perception system allows for the active reception and coordination of information received through one or several sensory systems. The purpose of a perception system is to let us behave effectively within a certain environment. In this case, technical perception is understood as the transformation of sensor information into data models to be processed algorithmically by software. A data model is understood as an abstract formalization of objects and relations that occur in a specific set of environments, in our case, in living spaces. In both developments, technical performance is achieved through a network of automated operating sensors and the algorithmic processing of data in models by means of software. The collaboration of these components leads to technical perception systems. It is state of the art that technical perception systems express objects and users in numbers, register spaces geometrically and present them visually. For instance, the tasks of this technology can range from the trivial regulation of room temperature to enabling the autonomous behaviour of hyper-complex machines such as humanoid robots. These examples show the expediency of different technical perception systems, whose different objectives produce different forms of spatial perception. While in the first example, air temperature is simply connected with heating, the second example constitutes "the world" of an artificial intelligence system. The complexity and unpredictability of the everyday world in which we humans move effortlessly presents a major challenge for hyper-complex technical systems. All technical perception systems are fundamentally different from human perception systems. Their quantitatively formal and abstract parameters contrast with our personal parameters that are embedded in a cultural context. Humans and humanoid assistants act according to very different perceptions of living space.

Technological developments are currently faced with a paradox: some robots are developed for intimate areas of human existence, yet they cannot participate in the space we humans perceive. Artificial intelligence is limited by its technical perception systems and data models. Data are only recorded according to their perception systems and processed according to the model. This defines the performance characteristics of technical systems. Thus, technical systems exist in different spaces of perception than humans. Technical perception systems are fundamentally different from the human system since they conceive space through geometries, numbers and images. For humans, space is only marginally defined by formal-abstract factors and mostly through personal experience,

association and habits, all rooted in cultural meaning that is not absolutely defined. Instead, there is permanent negotiation between inter-subjective meaning and individual meaning—negotiation among individuals and between individuals, objects and processes. Even though robots are equipped with very advanced, sophisticated technical perception systems, they are unable to grasp human meaning or to cope with human perception of space. That is where the problems culminate that motivate our research.

For instance, the state of research in this field is defined by concepts such as "object recognition" [5], "semantic scene labeling" [6] and "intention recognition" [7]. Object detection allows for a technical system to recognize material objects which consequently are labeled by means of object recognition or scene labeling, however this is only possible for objects which do not deviate too much from the form that is typical for this class of objects. Furthermore, intention recognition considers a functional connection between recognized objects and their related actions. These existing concepts share the fact that they solve the given problem according to technical possibilities. Thereby the cultural meaning of objects is marginalized, individual meanings cannot be introduced and tend to disappear within the formal representation of data.

2.1. Theory of Relationships between Human and Humanoids

The term language game [8], coined by Ludwig Wittgenstein, implies that every verbal expression is rooted in human life, not only because the various human language games make sense there, but also because language is woven into human action. Every word, every term and every sentence has a meaning which depends on the context of the action and situation in which it is uttered. Language games also involve mathematics and formal logic. Referring to the philosophical language- game, we conceive a space game that also makes use of relations between verbal expressions and human practices. The concept is extended by the fact that human action is woven into space (see details on space in Section 2.2). In contrast to the language game, the space game aims to negotiate cultural meaning between humans and humanoids to constitute a shared cultural meaning of space. Through natural language communication between humans and machines, a mutual construction of space and meaning becomes possible with dialogue. Thanks to its humanoid shape, the social robot can imitate human gestures and postures to enrich interaction with non-verbal communication. Using this interaction, humanoids can perceive space which was formed interactively from meanings, hence exceeding mere technical parameters.

It was a relatively simple and everyday situation from which the French philosopher Jacques Derrida developed his complex thoughts and profound reflections. One day when Derrida stepped out of the shower, he realized that he was standing naked in front of his cat which had sneaked into the bathroom. What fascinated him about the situation was the simple observation that he had felt ashamed to be exposed to his cat's glance. Although the cat certainly did not have any idea what nudity meant and was most likely not interested in his nudity, Derrida still felt observed by the gaze of his cat. "How can an animal look you in the face?" Derrida asked in the lecture, "The Animal That Therefore I Am", in which he describes this very situation [9]. Humanoid robots for domestic needs are built to look us in the face. It almost seems as if this skill were their main purpose, even one of the reasons why they were invented. These robots have a face we can look at, a face that can look back at us and is able to return our gaze [10–12]. With cameras as their eyes, humanoid robots do not only have the ability to "see", but they can also film. Additionally, they possess a number of other sensors which constantly monitor and evaluate their environment. Furthermore, they are connected with numerous other helping and assistive technical artifacts. All this generates a completely different perception of space, time and actions than that of humans who are socioculturally conditioned and influenced by architecture. Now that humanoid robots are advancing more and more into our private and intimate spaces and are thereby able to look at us—should we feel ashamed? Or are these robots simply an accumulation of technically based offers in a system they represent, merely robotic servants

whose glances are not touching us? Or are they companions, hybrid creatures who share our intimate spaces while potentially being able to share what they see worldwide via the internet?

Which principles make sense for creating a relationship between unique, human individuals and some kind of serial, distorted, robotic "mirror image of a human", namely a humanoid? All too often the relationship between humans and robots is understood as one of many different variations of master and servant, different yet always hierarchical relationships. However, we have developed a very different concept for the relationship by integrating its spatial context.

In her book "When Species Meet" [13], the American philosopher and theorist Donna Haraway criticizes Derrida's approach because he did not take the possibilities of the cat's gaze seriously and overlooked the various opportunities and meanings of the cat's glance. This gaze of the cat is—according to Haraway—an invitation to a "becoming with", what she calls "companion species", because each glance is always a reciprocal process, each instance of living with such a creature is a becoming with. According to Haraway this becoming with requires a different approach to ethics, an ethics of shared responsibility. Just as a guide dog is responsible for its owner, the owner is likewise responsible for the dog. A life together is based on a shared responsibility. This is a principle which is not only valid for relationships between humans and dogs: "Responsibility is a relationship crafted in intra-action through which entities, subjects and objects, come into being" [13].

Based on Haraway's approach, we understand the humanoid robot as some kind of "companion (species)" to the humans who live closely with these technical devices. What can a shared responsibility between humans and robots look like? Based on a similar understanding of shared responsibility, we want to reflect upon ethical, political and spatial consequences and develop a technical system in which humans take care of their robots which likewise support humans in everyday life. Furthermore, it always has to be considered in which power structure these humanoids are developed, explored and applied. As Langdon Winner asked in his essay, "Do Artifacts have Politics?" [14], we also want to question which cultural, ethical and political conceptions should be installed in such robots, especially related to architectural spaces for living that constitute our residential environment.

Technical perception systems express objects and humans by means of abstract representations. However, they are not capable of associating, varying and contextualizing meaning. Hence, they are not capable of placing those objects in a meaningful relation to each other. In contrast to this we have developed an approach in which the creation of a spatial model of cultural meaning is based on interaction between human and robot. The machine is thereby considered a "companion".

2.2. Theory of Cultural Space Production

The human lifeworld is essentially spatial. Thus, space is mainly a holistic substrate of meanings and their positioning from which all relations discussed in social and cultural space theories emerge [15,16]. Architecture devotes itself especially to aspects of the space in which we live. Changes lie ahead of the domain that will transform behaviour concerning the creation and use of space that has been culturally conserved for a long time. Increasingly, buildings as we know them will be altered by mobile parts and sophisticated control systems. We are standing on the cusp of a new development. In the near future, interaction between technical artefacts such as houses and robots will become part of the common understanding of space.

Space, the basis of the relational lifeworld of humans, is fed into the managing algorithms of specific technical systems by technical perception systems through numbers, geometries and images. However, they are neither able to detect the meanings, nor to put these parameters of our surroundings into culturally meaningful mutual relationships. Thus, technical systems exist in other spaces of perception than humans. The powers of each space of perception compete for the interpretational sovereignty concerning the actual everyday world. This structural otherness of human and technical realties contains potential conflicts and risks. Often users feel intimidated or pressured by assistive systems and therefore alienated and denaturalized in a lifeworld that is no longer theirs and thus are therefore rejected.

The social and cultural spaces of humans are diverse. We understand culture in this context as the values and knowledge that inform society as the realm of communication, activities and organizations of humans. Humanoid robots can be employed in private households, on semipublic premises or in public malls. Social space is not a permanently fixed space, but is in fact constantly "produced" [17] and is therefore in a permanent state of change. The production of space is therefore never a neutral process, but always determined by power structures, economic interests and cultural hegemonies. We call such an understanding of space a cultural concept of space. Just as space technologies should not be considered neutral because even those power structures are interwoven, our project tries to address those questions of power and also wants to support the empowerment of users in relation to artificial intelligence. "Technical objects" [18] are regarded as active protagonists. Humanoid robots are not only simple objects or neutral actors, but must be regarded and examined as active designers of such spaces.

How spaces are designed, modified and produced and how we move and interact with each other in these spaces is mainly influenced by perception. What we see, hear and feel is co-determined by our environment. However, the way in which objects, spaces and humans are seen, felt and heard is significantly influenced by cultural factors. It is also constantly re-negotiated and transformed what cultural meanings the spaces and objects have. Hence, we see humanoids as co-producers of social spaces and feel that they interact equally with the space and the humans in it. For these machines to perform tasks such as caregiving, selling products or other assistive services, and also to interact appropriately with humans, they must learn to cope with these cultural meanings, i.e., with this cultural concept of space. Cultural meaning cannot simply be programmed. Not only because meanings change, but also because they constantly must be re-negotiated and can differ greatly from one individual to the next. Thus, humanoid robots must not only learn to recognize these cultural meanings, but also to co-create them. They will occupy a new position between existing technologies and humans.

With this approach, technical systems must be enabled to develop a concept of space through the interactive requisition of cultural meanings; humanoid robots and artificial intelligence are being domesticated. The intended spatial model of cultural meaning is understood as an adaptive system. Through interaction it adapts to its environment and to different individual and spatial conditions.

3. Hypothesis and Questions

Summarizing Section 2, the overall context of our research questions is how to develop a spatial model of cultural meaning in a machine learning system as the operational basis for human as well as humanoid perception systems. As a hybrid spatial model, it is conceived to mediate between humans and humanoids by integrating cultural meanings into human-robot interaction (HRI) on the inter-subjective and the individual level. Hence, a kind of medium will be created that is informed by the different perception systems. Thus, the humanoid robot and its artificial intelligence can be involved in a shared context of meaning with humans. For this, we use the skills of a humanoid robot for human-like interaction as an approach for data acquisition and for model generation. In the spatial model discussed here, the meanings of objects and their relations are determined through verbal interaction and interconnected use of the algorithms of machine learning. Through the gathering and connecting of located meanings a spatial model is developed, which is no longer reducible to technical parameters or to human perception. Instead, a hybrid spatial model based on interaction is created.

Our approach follows the hypothesis that in natural language communication between humans and robots and with the help of the particular features of a humanoid robot (body language, autonomous movement, having a very intuitively designed conversation), a spatial model of cultural meaning can be developed. We decided to use and combine existing robust technical solutions to set up the system, to create a proof of concept. This model is neither reducible to technical quantifiability nor to human spatial perceptions, but it introduces a third hybrid interpretation of space based on the cultural meaning of objects and spatial features.

The most important research questions for this first stage of development of our prototype are:

1. Can the interactive creation of a spatial model of cultural meaning be technically robust, even in a bustling environment, and can it be successful in real-time?
2. Will users accept the generated model as representative? (narrative interviews)
3. Will users consider this representation as a relevant abstraction of individual and inter-subjective meaning of space and objects? (narrative interviews)
4. How do users react to the humanoid as a dialogue partner? (Kind of acceptance, e.g., terms used, usage and modulation of speech such as adult versus child, dominance versus subordination by observation and filming).

4. Research Methodology Based on the Arts

These questions will be asked using trans-disciplinary methods known from artistic research and will be reflected philosophically; thus, new trans-disciplinary approaches and methods ought to be developed and interpolated in the current debates and discussions concerning robot ethics [19].

Research on humanoid robots as a possible future "companion species" also means to integrate the above questions concerning new kinds of perception. Attributing meanings to objects, rooms, components, people and machines is a network-type reference system consisting of geometric, figurative, social and cultural relations. Thus, the meanings of single parts and their relationships constitute each other mutually and are constantly transformed dynamically in negotiations. Integrating/using sensors, which technically produce completely different images of the environment than the human perception system, trans-disciplinary artistic research can especially help to make these types of perception visible and can also show how interaction between humanoids and the human perception system can be interwoven to generate a cultural model of lifeworld from individual and inter-subjective meanings: A spatial model which is neither reducible to the human nor to the humanoid perspective.

Research on problems of the lifeworld requires a methodical course of action which integrates approaches from different disciplines. The transdisciplinary approach to this project requires merging expertise from architecture, automation technology and philosophy. Synergies develop through the overlapping and sharpening of questions such as ours in search of a spatial model of cultural meaning. In addition, we use methods of artistic research to widen our approach. Artistic research is already being used in HRI studies as in the case of the Theatrical Robot [2]. There, the artistic implementation of human-robot interaction is applied in a performance for the benefit of studies in social science. In contrast, the artistic research we use in our project is not an artistic tool for social science, but a contribution to research methodology.

Our complex lifeworld actually consists of a surprisingly large number of individual objects and processes. With teleological methods we can only grasp some of these objects and processes in our everyday discourse. Artistic research methods do not follow explicit targets but are developed on an open-ended and yet goal-oriented scale. They do not necessarily need to be developed along cause-effect relations and can apply non-causal effects. They explicitly apply vagueness, intentional misreading and superpositions of established concepts. Artistic research can integrate references and sources from different academic disciplines, but also from non-scientific bodies of human knowledge. Original phantasies and deductive argumentation are both considered: "[. . .] where logical thinking is naturally intertwined with associative and intuitive conceptualization" [20]. For the present project, it is necessary to work with artistic research methodology since it does not focus on efficiency, but on the aesthetics of interaction, not on reproducibility, but on the variability of interaction, not on precision, but on cultural depth of interaction. Artistic research in conjunction with social scientific methods enables us to develop, evaluate and reflect upon questions concerning the cultural meaning of space, the different kinds of spatial perception, and spatial relationships between robots and non-robots. These specific kinds of interactive and reflective processes [21–23] are applied to the actualization of a spatial model of cultural meaning. The findings and insights will not be rated with social, technical

and technological targets in a cause-effect relation, but will be evaluated in the form of open results in context of the lifeworld.

5. Prototype, Verbal and Non-Verbal Communication between Humanoid Robots and Humans to Learn Meanings, Using a Machine Learning System

5.1. Experimental Settings

As a first attempt of implementation, we developed a prototype, to be tested in two different experimental settings. A first experiment was held at "The Long Night of the Robots", an event open to the public at the Technical University of Vienna, which we used to involve different people of all ages. We had a variety of participants reaching from those unfamiliar with machine learning and robots, to interested experts from various different domains, including but not limited to philosophy, linguistics and electrical engineering. The second experiment was held with a sample of voluntary participants of architecture students in a controlled laboratory setting. The results of the experiments are expected to produce considerations on how the prototype of a spatial model of cultural meaning can be further developed, and which problems and questions still have to be solved.

5.1.1. Setting 1: "Long Night of the Robots"

We were invited to contribute to the "Long Night of the Robots", which took place in Vienna on 25 November 2016 as part of the "European Robotics Week". We took the opportunity for a first experiment with our first prototype of verbal and non-verbal communication between the humanoid and the visitors to generate cultural meaning in the machine learning system. We expected the visitors to have different degrees of prior knowledge, different social and cultural backgrounds and interests, and that they would be mostly interested in robotics as lay persons. Our contribution was to offer visitors a "conversation" with a humanoid robot: a natural-verbal interaction supported by gestures about the space and the meanings of some surrounding objects.

The principle elements of the setting to generate a spatial model of cultural meaning consist of the following: (a) humanoid robot, (b) machine learning system, (c) test persons, (d) objects and areas in the room.

(a)　Humanoid robot: A stationary humanoid robot, the model called "Romeo" by the firm of Softbank Robotics, is equipped with our machine learning system. Besides speech input and output, a crucial ability of the robot is to extend verbal communication by using various gestures. We consider it of crucial importance to create a communication experience, similar to a conversation among humans. The robot was trained to use various non-verbal gestures and common-sense phrases of small talk, to enrich the interaction. In order to clearly communicate the gestures of the robot in the turbulent and completely crowded room, the robot pointed at the objects with a flashlight to ask for their name, function and meaning (Figure 1).

(b)　Machine learning system: Through the cloud-based speech recognition of Google, the voice input is converted into text. The access to Wordnet through LibLeipzig enables the definition of a particular grammatical element and thus a first categorization of each term by its word class. The specific categorization of the entire input is made by the library of machine learning by Orange in Python. The machine learning system contextualizes the meanings generated through interaction with certain areas of the room in relation to the position of the robot. Terms were categorized and reduced to their basic form. The answers by the visitors were itemized for each of the six objects in the categories of name, function and features (Figure 2). Our machine learning system generates a hierarchy of meanings, but without letting the individual meanings vanish. On the contrary, the individual meanings are especially used as part of the further exchange by an individual user with the machine. An important aim is to correspond also to the individual range of meanings of the users of the assistive technology. With this prototype, individual meanings

are captured and counted, but they are not yet set in relation, in order to accurately develop the accumulation of inter-subjective meanings.

Figure 1. Robot Romeo in the experimental setting at the "Long Night of the Robots" at Vienna University of Technology.
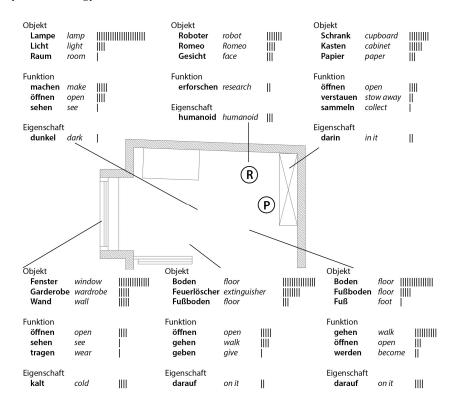

Figure 2. Floor plan of the experimental area with the location of the six objects, translated into English: the robot (R) and the interview partner (P). Naming of object, function attributed to object, property associated to function and object. Bars indicate how often names and functions were mentioned.

Through the different conversations the machine learning system culminates in a variety of terms for each item requested by the robot, and also for all the other items which might have been used. These terms are evaluated by the artificial intelligence, depending on frequency.

(c)   Test persons: To make the technical system transparent to the visitors, the inclusion of human statements and the placing of meanings are presented to the visitors on a screen. Since both happen in real time, the visitors can follow how their stated meanings are included in the system. Since a German-speaking public in Vienna had been invited to this public event, the robot communicated in German. Modifying the speech recognition from German to English and changing the LibLeipzig library to TextBlob would make the experiment reproducible in English.

(d)   The setting of the interaction was a 60 m^2 room in the robotics laboratory, divided into an experiment area and a visitor area that included several objects and parts of the room: a cabinet, an overhead light, a fire extinguisher, the floor, a window and the robot Romeo itself. The objects were spread out through the room.

5.1.2. Experimental Process 1

The robot points at objects and parts of the room and asks the visitors about personal meaning (Figure 1). Four types of questions in different variations were asked about six predefined objects in the room. Types of questions were: name, property, function and meaning of an object. Questions were asked in natural language with common-sense phrasing and accompanied by distinct non-verbal gestures. The responses of the visitors were registered by the machine-learning system. Parallel to communication with the robot, visitors could also observe on a screen how their spoken answers were allocated to a specific location within a floor plan of the room and weighted by previous answers (Figure 2).

5.1.3. Setting 2: "Young Architects"

The principal elements for the setting to generate a spatial model of cultural meaning were as mentioned above (Section 5.1.1).

In the second setting (Figures 3 and 4), we expanded the system to another robot developed especially for social interaction, the Pepper model by Softbank Robotics. Based on our experience from the first experiment, we adapted the following points:

1.   Less ambiguity in the non-verbal gestures for separate single objects in the room.
2.   Questioning an individual person rather than individuals from among a group.
3.   To evaluate the accuracy of speech recognition, we also transcribed the user input manually, parallel to autonomous speech recognition.
4.   Larger number of persons and more intensive questioning.

Figure 3. Robot Pepper in the experimental setting at the "Young Architects" experiment, at Vienna University of Technology. The questioned objects in this picture are: robot, chair, backpack and waste bin; a lamp is not visible.
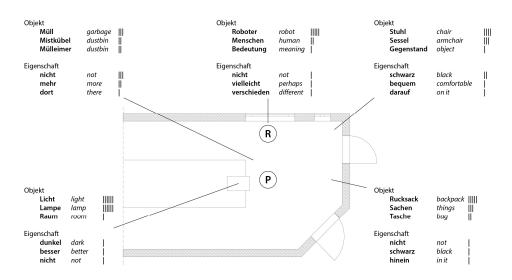
Objekt		
Müll	*garbage*	|||
Mistkübel	*dustbin*	||
Mülleimer	*dustbin*	||

Eigenschaft		
nicht	*not*	|||
mehr	*more*	||
dort	*there*		

Objekt		
Roboter	*robot*	|||||
Menschen	*human*	||
Bedeutung	*meaning*		

Eigenschaft		
nicht	*not*		
vielleicht	*perhaps*		
verschieden	*different*		

Objekt		
Stuhl	*chair*	||||
Sessel	*armchair*	|||
Gegenstand	*object*		

Eigenschaft		
schwarz	*black*	||
bequem	*comfortable*		
darauf	*on it*		

Objekt		
Licht	*light*	||||||
Lampe	*lamp*	||||||
Raum	*room*		

Eigenschaft		
dunkel	*dark*		
besser	*better*		
nicht	*not*		

Objekt		
Rucksack	*backpack*	|||||
Sachen	*things*	|||
Tasche	*bag*	||

Eigenschaft		
nicht	*not*		
schwarz	*black*		
hinein	*in it*		

Figure 4. Floor plan of the experimental area with location of the five objects, translated into English, the robot (R) and the interview partner (P). Naming of object, function attributed to object, property associated to function and object. Bars indicate how often names and functions were mentioned.

After interacting with the robot, the participants completed a brief quantitative questionnaire. It asked for general assessments of the robot, the setting and the experiment, and also about experience with robotics. In addition, some of the participants were asked for a narrative individual interview. The 17 interviews were recorded with a video camera and generally lasted 15 to 25 min. They were guideline-supported narrative interviews. The participants had an opportunity to speak freely about the experiment itself, about their general opinion regarding robots in the private and the public sphere, and their relevant expectations and emotions. Another goal of the narrative interviews was to find out how the participants perceived the experiment, what the participants thought was being researched with the experiment, and whether they feel that generating a spatial model of cultural meaning is possible and useful. Another goal was to improve interaction with the robot as indicated by the feedback in the first experiment.

5.1.4. Experimental Process 2

Showing and questioning was conducted as described in Section 5.1.2. In the first experiment, many people found it necessary to formulate extensive descriptions for the objects, and these were partly imprecise due to their length. It was shown in the first experiment that the answers pertaining to name and characteristic sometimes led to shorter answers to questions about the name and the meaning of the objects, which were of greater importance for the project. Therefore, in the second experiment, the type of questions was limited to those about the name and the meaning of an object. To make the dialogue more diversified, there were five versions for each of these questions (Figures 5–7). To make the questioning process more accessible to visitors, we asked questions in two parts. In the first part, the robot asked for the name, in the second part for the meaning of objects (Supplementary Material Code S1).

In total, 60 persons in four groups were questioned about five objects each. This means that there was a total of 600 interactions. The average duration per session was between 3 and 5 min.

```
...
# Vordefinierte Fragen
Level_0 = [
"Was ist das? Kannst du das beschreiben?",
"Super! Und was ist das? Darüber würde ich gern mehr wissen.",
"Was ist das? Kannst du mir das erklären",
"Ok, Was ist das? Das interessiert mich auch.",
"Was bin ich?"
]


Level_1 = [
"Vielen Dank, dass du das beantwortet hast! Und was bedeutet dieser Gegenstand fuer euch?",
"Gut zu wissen! Und was ist die Bedeutung dieses Objekts?",
"Ok, und welche Bedeutung hat dieser Gegenstand für euch Nicht-Roboter?",
"Aha, kannst du mir die Bedeutung dieses Gegenstandes verraten?",
"Ok, kannst mir auch meine Bedeutung nennen?"
]
...
```

Figure 5. Excerpt of variations of the humanoid robot dialogue (Examples of questions).

```
...
# Predefined movements
def b0():
    names = list()
    times = list()
    keys = list()

    names.append("HeadPitch")
    times.append([1.96])
    keys.append([[-0.21936, [3, -0.666667, 0], [3, 0, 0]]])

    names.append("HeadYaw")
    times.append([1.96])
    keys.append([[0, [3, -0.666667, 0], [3, 0, 0]]])

    names.append("HipPitch")
    times.append([1.96])
    keys.append([[-0.0352817, [3, -0.666667, 0], [3, 0, 0]]])
...
```

Figure 6. Script of spatial model of cultural meaning. Only the angles of the different joints of the robot were predefined as the starting point.

```
...
# Basic structure of objects
Object class ():
    List = []


    def __init__(self, Id, movement):
        self.Id = Id
        self.level = -1
        self.question = 0


        self.movement = movement


        Object.list.append(self)


# space-related movement
neues_Objekt = Objekt(0, z2)   ??? übersetzen?
neues_Objekt = Objekt(1, z4)
neues_Objekt = Objekt(2, z5)
neues_Objekt = Objekt(3, z3)
neues_Objekt = Objekt(4, z0)
...
```

Figure 7. Script of spatial model of cultural meaning. For starting the script at the beginning of the experiment, only one class and five empty instances were predefined.

5.2. Results of the Experiment

Evaluation of research question 1: As part of the "European Robotics Week", we invited the public, and there were other appeals in the form of press releases inviting people to the "Long Night of the Robots" at Vienna University of Technology. The organizers were overwhelmed by the response (more than 1000 visitors). This large number of visitors exceeded our expectations and greatly strained our resources. We had an opportunity to do narrative interviews with mainly lay people, but also with a philosopher specializing in Philosophy of Technology, a psychoanalyst, and several researchers in the field of robotics. However, a total of 258 interactions between visitors and the robot Romeo could be recorded.

In the "Young Architects" experiment, the environmental conditions were exactly controlled. The result was that 603 interactions between robots and participants were recorded. We did 60 interviews using questionnaires and 17 narrative video interviews.

Evaluation of research question 2: In the narrative interviews that we conducted with some of the visitors, nearly everyone instantly accepted the representation of the model. The reason was familiarity with a floor plan and the fact that the text was shown on the screen. This made up for the strangeness of the humanoid and the abstractness of the invisible machine learning system. Hence it was mostly understood as relevant.

We did not show the floor plan of the room to the architecture students, only the script for the machine learning system. The script was an unfamiliar abstraction for the students, but they understood it for the most part.

Evaluation of research question 3: The level of abstraction was understood and considered relevant for matters of the everyday lifeworld.

In the second experiment the level of abstraction was not understood but considered relevant for matters of the everyday lifeworld.

Evaluation of research question 4: In the first experiment, the robot was set up to ask questions and store the answers in a machine learning system. The participants reacted with different modes of speech to the robot and his questions: some spoke slowly and noticeably more articulated, almost as if they were speaking to a toddler. Others used only single words. Another group of visitors spoke as if they were answering the question of another human.

For the most part, the architecture students spoke slowly and in very articulated form.

Both groups were aware that the robot could not reply in context, nevertheless many still tried to turn the question and answer dialogue into a complex conversation. Counter-questions were often asked, and the robot was asked whether it had understood what was said. Many people tried different modes of speech in their conversation with the robot, and therefore changed their modes of speech as the conversation developed (Supplementary Material Table S1).

5.3. Experimental Results, Evaluation of Prototype

Our system achieved reliable terms for most of the objects and areas of the room. In Experiment 1, however, the objects "fire extinguisher" and "floor" where mentioned more often as a result of equivocal pointing and its interpretations. In this case the attribution of "fire extinguisher" was succeeded by "floor".

Additionally, some unexpected connections occurred. As an example, "wardrobe" was associated with the feature "cold". "Light" was associated with "darkness" in both experiments. This was not obvious at the beginning, exemplifying therefore the before mentioned constant negotiation process. The same occurred with terms like "cold" to express the season. The object "cabinet" was connected to the feature that something is "in it", just like the floor is a part of the room "on which" something is situated. The examples "wardrobe" and "floor" show that inherent features and concepts from the verbal analysis were filtered out correctly. In the machine learning system, our space game maps cognitive connections between objects. For example, the term wardrobe is directly associated with the terms "wear" and "cold". Furthermore "trash can" is connected to topics like environment and the waste of things. The development of these connections was confirmed in the second experiment. An interesting detail was found in relation to the term "robot". The machine learning system showed relations to the term "help", indicating at a general acceptance of a humanoid robot as an assistive system.

We learned that our system has to solve the challenge to separate the context and to bring terms into context (future work 2), furthermore we have to strengthen the filter of the system, as for example the often-used word "not", in sentences as "I'm not sure" or "I don't know exactly" was categorized as a function in the second experiment.

With both experiments we found that names of objects were resolved with a high likelihood, followed by function and property (Figures 2 and 4). Meanings on the other hand where found to vary greatly, but at the same time we found that core descriptions started to emerge. Here, experiments with bigger samples of participants need to be conducted. A combination with the field of Scene Labeling might be promising. Speech recognition was not satisfying even in the controlled environment of the second experiment. We used a special voice-optimized microphone in both experiments and transcribed the input of participants by-hand to learn about error sources. This is a technical problem to be solved in both hardware and software.

Most importantly, our expectation was that inter-subjective and individual meanings could be mapped and related with a non-verbal and natural-language dialogue. In this respect, the prototype worked. Indeed, a spatial model of cultural meaning started to develop.

6. Conclusions

The two experiments are only the first for building and constantly improving our first prototype of a cultural spatial model. These experiments and the generated first prototype are still very basic and therefore in no way able to compete with state of the art technical perception systems. As space

is constantly produced (and by no means is this production a neutral process), we could already see in this basic stage of the cultural spatial model, that even basic questions on everyday objects and room elements in humanoid-human-interaction opened up moments of reflection and irritation. Speaking with a humanoid about space forced some of the participants, especially the architecture students, to rethink how to explain objects and their functions. This basic act of reflecting on everyday objects is already part of the process we described in the first part of the paper. Therefore, playing the architecture space game is not only for producing data for the machine learning system, but also to start processes of reflection and to add new viewpoints regarding the question of space in human participants. As we have also argued in the theoretical part, it is essential for our space game to talk about artifacts with a humanoid robot in order to produce space, to follow the robot's gaze, to follow his arm movements in natural language and in non-verbal interaction.

Although most of the conversation between the robot and the participants was about concrete objects, questions about their function concerned their spatiality, i.e., their position in the room, their relation to other objects, their function in the room, their collective and individual meanings, etc. In future experiments, we hope not only to constantly expand the number of objects to be talked about, but also to introduce more elements of the room as a container, such as the floor, the windows, the doors, and so on. The aim of the cultural spatial model is to produce a hybrid spatial model, combining technical and formal perceptions of space with the cultural and subjective viewpoints of the participants. In the long run, we expect the emergence of a dense, hybrid and shared model of space that does not depend on the actual object in a certain room but on what these objects mean in both contexts—in the technical as well as the human perception system.

7. Future Work

1. Inputs into the machine learning system will be correlated to each other and categorized. This will be done with the classification and regression algorithms by Orange in Python, as for example with the simple method of "k-nearest-neighbors".

After evaluation of the interaction and the thus generated spatial model, the results will be applied to the development of further modules in several prototypes, as mentioned above:

2. Relationship between the meaning of objects and the space they are in.
3. Using ultrasound measurement by expanding the robots' sensors to help the robot to recognize objects and building parts and simultaneously to locate specific meanings by combining the robot's pointing direction.
4. Generation of answers by means of the machine learning system.
5. Processing inter-subjective as well as individual meanings with the machine learning system.

Once a robust prototype of the three modules has been generated as a spatial model of cultural meaning, further goals are as follows:

6. Ethical problems of privacy on at least two levels: usage of meanings in the system in different everyday situations and usage of individual data in the machine leaning system.
7. The introduction of time as a dimension in the model. The planned approaches consider time as the fourth dimension or cinematically as the third dimension in a two-dimensional space.
8. After the assessment, interconnection and categorization of the meanings generated, we will test how to place whole networks of meanings and their categories.
9. Applying the meanings and the various concepts of space in different languages/cultures, for instance in Japanese.
10. Prediction of meanings using the classification and regression methods mentioned above.
11. Extending the natural-verbal communication with non-verbal elements such as human gestures in connection with facial expressions and brainwaves of human moods.

References

1. Duffy, B.R. Anthropomorphism and the social robot. *Robot. Auton. Syst.* **2003**, *42*, 177–190. [CrossRef]
2. Dautenhahn, K. Methodology and themes of human-robot interaction: A growing research field. *Int. J. Adv. Robot. Syst.* **2007**, *4*. [CrossRef]
3. Weiser, M. Ubiquitos Computing. *Computer* **1993**, *10*, 71–72. [CrossRef]
4. Brooks, R.A. Intelligence without representation. *Artif. Intell.* **1991**, *47*, 139–159. [CrossRef]
5. Varvadoukas, T.; Giannakidou, E.; Gomez, J.V.; Mavridis, N. Indoor Furniture and Room Recognition for a Robot using Internet-derived Models and Object Context. In Proceedings of the 10th International Conference on Frontiers of Information Technology, Islamabad, India, 17–19 December 2012; pp. 122–128.
6. Koppula, H.S.; Anand, A.; Joachims, T.; Saxena, A. Semantic Labeling of 3D Point Clouds for Indoor Scenes. In *Advances in Neural Information Processing Systems 11, Proceedings of 24th Annual Conference on Neural Information Processing Systems, Granada, Spain, 12–15 December 2011*; Curran Associates, Inc.: New York, NY, USA, 2012; ISBN 9781618395993.
7. Krauthausen, P. *Learning Dynamic Systems for Intention Recognition in Human-Robot-Cooperation*; KIT Scientific Publishing: Karlsruhe, Germany, 2013.
8. Wittgenstein, L. *Philosophische Untersuchungen*; Wissenschaftliche Buchgesellschaft: Frankfurt, Germany, 2001.
9. Derrida, J. The Animal That Therefore I Am (More to Follow). *Crit. Inq.* **2002**, *28*, 369–418. [CrossRef]
10. Coeckelbergh, M. Humans, Animals, and Robots: A Phenomenological Approach to Human-Robot Relations. *Int. J. Soc. Robot.* **2011**, *3*, 197–204. [CrossRef]
11. Coeckelbergh, M. *Growing Moral Relations: Critique of Moral Status Ascription*; Palgrave Macmillan: Basingstoke, UK, 2012.
12. Coeckelbergh, M.; Gunkel, D.J. Facing Animals: A Relational, Other-Oriented Approach to Moral Standing. *J. Agric. Environ. Ethics* **2014**, *27*, 715–733. [CrossRef]
13. Haraway, D. *When Species Meet*; University of Minnesota Press: Minneapolis, MN, USA, 2008.
14. Winner, L. Do Artifacts have Politics? In *The Whale and the Reactor: A Search for Limits in an Age of High Technology*; University of Chicago Press: Chicago, IL, USA, 1987.
15. Dünne, J. *Raumtheorie: Grundlagentexte aus Philosophie und Kulturwissenschaften (Suhrkamp Taschenbuch Wissenschaft)*; Suhrkamp: Berlin, Germany, 2006. (In German)
16. Günzel, S. *Raumwissenschaften (Suhrkamp Taschenbuch Wissenschaft)*; Suhrkamp: Berlin, Germany, 2008. (In German)
17. Lefebvre, H. *The Production of Space*; Blackwell: Oxford, UK, 1991.
18. Simondon, G. *Die Existenzweise Technischer Objekte*; Diaphanes: Zürich, Switzerlands, 2012. (In German)
19. Lin, P.; Abney, K.; Bekey, G.A. *Robot Ethics*; MIT-Press: Cambridge, MA, USA, 2015.
20. Dyrssen, C. Navigating in Heterogeneity: Architectural Thinking and Art-based Research. In *The Routledge Companion to Research in the Arts*; Biggs, H., Karlsson, M., Eds.; Routledge: Oxford, UK, 2011.
21. Klein, J. Was ist Künstlerische Forschung. Kunsttexte.de/Auditive Perspektiven, Nr. 2/2011. Available online: www.kunsttexte.de (accessed on 10 January 2018). (In German)
22. Rey, A.; Schöbi, S. *Künstlerische Forschung—Positionen und Perspektiven*; Museum für Gestaltung Zürich: Zürich, Switzerlands, 2009. (In German)
23. Badura, J.; Dubach, S.; Haarmann, A.; Mersch, D.; Rey, A.; Schenker, C.; Toro, G. *Künstlerische Forschung: Ein Handbuch*; Diaphanes: Zürich, Switzerlands, 2015. (In German)