University of Tartu
Institute of Philosophy and Semiotics
Department of Semiotics

Halil İbrahim Yazıcı

# THE RELEVANCE OF UMWELT THEORY IN EMBODIED ARTIFICIAL INTELLIGENCE RESEARCH

Master's Thesis

Supervisors: Dr. Riin Magnus
Prof. Kalevi Kull

Tartu
2018

I have written the Master Thesis myself independently. All of the other authors' texts, main viewpoints and all the data from other resources have been referred to.

Author: Halil İbrahim Yazıcı

……..

(signature)

Supervisors: Dr. Riin Magnus Prof. Kalevi Kull

..……. ..…….

(signature) (signature)

22.05.2018

**Table of Contents**

## Introduction

Artificial intelligence (henceforth AI) is one of the most frequently debated topics today. It is being introduced into our daily lives at an unprecedented rate, which makes it a target topic both within and outside the academia. While people are not necessarily aware of to what extent they are actually reliant on AI systems, one can say that such awareness is immediately achieved when confronted with the systems that are able to operate autonomously in the physical world (i.e. robots). Consequently, concerns and inquiries pertaining to the capacity of intelligent behavior those systems are able to display have a certain history. One of the theories used to explain how such AI systems perform or should perform is Umwelt theory. It has proven to be an appropriate theory in its capacity to describe how animals perform intelligently in their own physical environments and such descriptions are often generalized by scholars of various backgrounds to physical AI systems as a valid source of inspiration. Despite the fact that Umwelt theory has an undeniable aptitude for this task, to what degree it can be applied to physical AI systems mostly remains an unexplored task that I address in this thesis.

With this work I aim to evaluate the possibility of using Umwelt theory, as it is put forward by the Baltic-German biologist Jakob von Uexküll, as a fundamental basis in AI research. It becomes immediately obvious, however, that AI research is quite diverse in itself and this diversity is also a limiting factor when it comes to applying the theories that either do not directly associate themselves with AI research even if they are situated under the same doctrinal roof or belong to an entirely different field or area of research. Thus, in order for Umwelt theory to have a valid explanatory basis when juxtaposed to theories that essentially belong to AI research, the scope of this thesis has been limited to appropriate fields that find it necessary to emphasize the importance of subject-environment relation, which can be more or less defined under the name of embodied AI.

Another matter of utmost importance about this work is that it is aimed at not only audiences of semiotic interest but also many others such as computer science and philosophy. The thesis, thus, includes perspectives from those respective fields throughout the research in an attempt to demonstrate the fact that regardless of the number of differences between fields, connections can indeed be formed, and it is in fact the ability to cross boundaries rather than

practicing in isolation that lights the spark of creativity and reveals new perspectives simply because “what diverges communicates” (Stengers 2005: 190).

To achieve the aim of this thesis, the research structure has been supplemented with three sub-questions, each corresponding to a chapter, that assist in reaching a conclusion for the main research question. The main research question is formulated as: Does Umwelt theory bear any relevance in the context of embodied AI research? And the three corresponding sub-questions are specified as: (1) In which aspects has Umwelt theory been used in embodied AI research in the previous literature? (2) How did the developments in AI research lead to the adoption of theories (Umwelt theory and other theories pertaining to embodied cognition) between which no connection is apparent? (3) Can Umwelt theory make a substantial contribution to the current embodied AI paradigm?

The thesis starts with an overview of Umwelt theory as its first chapter. The purpose of this chapter is to provide a brief history of biological line of thought preceding and following Umwelt theory and to introduce its defining characteristics in connection with its history. The second chapter corresponds to the first sub-question and explores how scholars of various backgrounds have made use of Umwelt theory with regard to embodied AI research. The analysis mainly follows a chronological order and highlights the concepts of Umwelt theory different scholars have chosen to emphasize in their works. Corresponding to the second sub-question is the third chapter where the history of AI research is discussed. This chapter is structured so that the necessity and emergence of embodiment is made evident through an historical analysis. This analysis also indicates the problems preceding and following the emergence of embodiment to be of semiotic nature. It is also marked as one of the reasons why theories with high emphasis on meaning-making (such as Umwelt theory) were accepted as valid theories of embodiment. The fourth chapter, corresponding to the third sub-question, discusses Umwelt theory’s viability to further the research in embodied AI paradigm through an analysis of its fundamental aspects described in the first chapter. It is proposed in this chapter that the results obtained from the analysis of Umwelt theory’s parts are also represented in Umwelt theory’s totality on larger scale. In the conclusion part I provide a summary of all these chapters with a brief account of my own argument. Finally, I provide a direction where further work in this context might be realized.

## 1. Umwelt Theory, A Brief History and Introduction

The history of biology, or biological line of thought, has seen quite a lot of theories and variations that it is not possible to address them all here in the scope of this research. However, there are certain points in the history of biology that mark the emergence and acknowledgement of different paradigms. One way of locating[1] Umwelt theory among other views can be done by identifying three basic models that correspond to three major paradigms, namely ladder, tree and web. They are briefly described as: (1) ladder in which organisms are depicted as distinct and perfect with no room for evolution, (2) tree, a hierarchical structure where organisms strive for perfection through competition, implying an incompleteness in both themselves and nature, (3) web, a non-hierarchical structure that suggests recognition and interpretation as highly important aspects of the model within which every organism is interconnected and is in a symbiotic relationship (Kull 2003: 592). Ladder was accepted as the prominent model in biology until its replacement by tree model in the 18th century. Tree model, which poses a more significant issue in the Uexkillian context, inherently suggests the existence of a stem from which the species branch off and grow in their diversity. The problem, however, starts with the evident implication of growth and the availability of resources because there are obviously not enough room and resources for every organism to develop freely. This is where the notion of competition enters the scene as both a restrictive and progressive function in the model. It is a decisive means for survival. Darwin's theory of evolution, which has been quite influential even in disciplines other than biology (Dennett 1995: 21), also has in its core the tree model and essentially the idea that natural selection by competition is the primary force in evolution[2]. K.E. von Baer[3], a contemporary of Darwin, opposed his views on the basis that Darwin's theory of evolution focused solely on external mechanical

[1] There are, of course, other ways by which one can situate Umwelt theory among other paradigms in the history of biology (see Kull 2000; Magnus 2008).

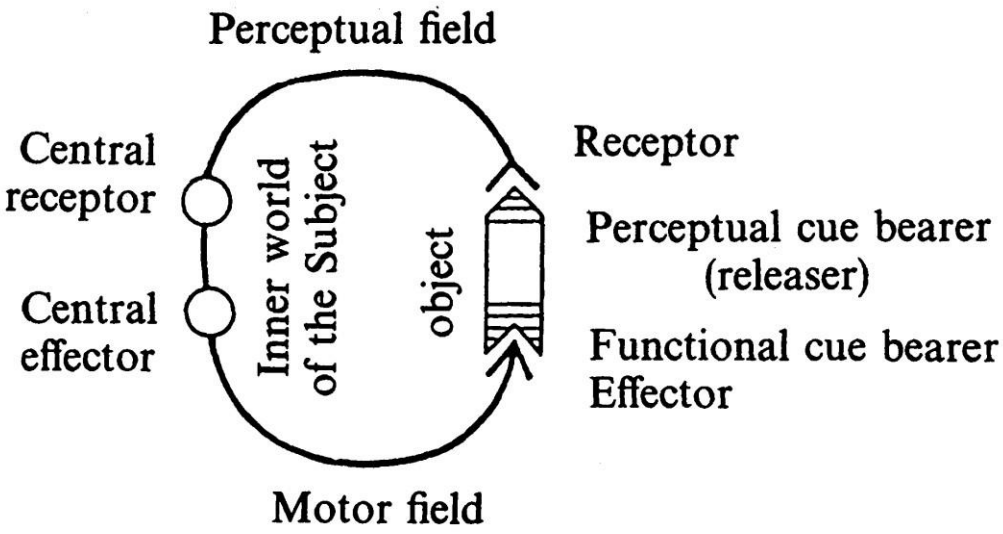
[2] This is not to say Darwin was the first to utilize the tree model for representing phylogenies. In fact, one can see its usage in the works of J.B. Lamarck and even before him in religious contexts (Ruse 1996: 23, 45).

[3] A Baltic-German biologist and embryologist whose works was also greatly valued and accepted. It is important to recognize the influence he had on von Uexküll especially when one considers von Uexküll also studied zoology in University of Dorpat.

forces and random variations, completely ignoring the teleological nature of every organism and organic process (Brentari 2015: 24).

This period in the history of biology, which corresponds to the end of 19th century and the beginning of the 20th century, is also seen as a very productive era for biological thought and it is in this period theoretical biology as a branch of biology was established (Kull 1999: 390). It is also within this period von Uexküll formulated his Umwelt theory[4].

Umwelt, in its essence, can be briefly summarized as the subjective world[5] of the organism within which it sustains its life functions. Every subject lives in their own reality that is in turn determined by their perceptual and operational capabilities, forming a functional cycle. The functional cycle is one of the most essential parts of the Umwelt theory and it is through the intricacies of functional cycle the organism is able to interact with its environment. In a basic functional cycle, the organism, at first, perceives the sensory input through its receptor organs. The received input is then interpreted by its central effector and transformed into action through the effector organs which cause certain changes either the object or environment the organism is interacting with, stimulating the reception of new sensory input as a result.




[4] The term Umwelt first appears in 1909 in his book titled *Umwelt and Innenwelt der Tiere* (The Environment and Inner World of Animals).

[5] An important note has to be made here to acknowledge the impact of Kant on von Uexküll's thought and the direct correlation with his notion of subjective world with Kant's philosophy that considers all reality as subjective appearance (Buchanan 2008: 21).

Figure 1: The Functional Cycle. Adapted from von Uexküll (1992[1934]).

As seen in Figure 1, functional cycle has an emphasis on the coupling of biological processes and they are bound in a continuous loop, explaining the whole process as a unified system. While such depiction might initially be understood as overly-simplistic in its nature, it should be taken into consideration that von Uexküll described the functional cycle at the most basic level to fit into the context of even the simplest animal and stated that more complex animals would have a more complex Umwelten as they would consequently have more functional cycles (Uexküll 1992[1934]: 324). The functional cycle also describes the organism's capacity to make use of signs when interacting with its environment. However, in order for this statement to be understood clearly, the experiential nature of the event for the organism must be highlighted, namely, the object (or the subject) with which the interaction taking place has to have a certain *meaning* for the organism. According to von Uexküll, this meaning cannot be provided by the external forces that claim to shape the organism from the outside[6] (Hoffmeyer 1996: 56), as in Darwin's theory of evolution, but has to be embedded in and provided by their morphology and material structure (Uexküll 1926[1920]: 190). Another point he addresses is that the organism is ever in connection with its environment and cannot be separated from it because Umwelt is only formed when those two are in unison. Neither the organism nor the environment exists without the other. This is an important point in his view because he frequently uses this two-fold layer, also present in the functional cycle as seen in Figure 1., when explaining the relations between two organisms in terms of "likeness". The concept of "likeness" is of great importance to von Uexküll's work. It is ever present in the examples he uses to demonstrate the interconnectedness of the relations both between organisms and their environments. For instance, he gives the example of a spider web and he further elaborates that the spider is able to catch the fly simply because it is flylike and it embodies fly-likeness in its structure which coincidentally applies to the structure of its web. In his words "The web is truly a refined work of art that the spider has painted of the fly" (Uexküll 1982[1940]: 42). It becomes apparent, after going through the examples von Uexküll provides, that there is an inherent mechanism of coupling in his view of biology and, accordingly, life. He introduces the notion of "counterpoints" to explain this mechanism as something that is

[6] Von Uexküll, however, acknowledges the effects of both lifeless things (Uexküll 2001[1937]: 122) and inorganic forces (Uexküll 1926[1920]: 354) on an organism's Umwelt and considers them just as essential.

inherent in all living beings and considers the properties of those units as contrapuntally matched (Uexküll 2001[1937]: 122).

The last notion to be introduced in this chapter as one of the essential parts of the Umwelt theory is the concept of life-task. Life-task is described as the determining factor of the meaning the organism is able to attribute to whatever it encounters in its environment. In von Uexküll's terminology, this concept has more emphasis on the abstract and relational aspects of meaningmaking. Consequently, he states that the meaning-making processes are determined by the organism alone rather than being the result of a cause extrinsic to the organism.

Another thing to note here is that Von Uexküll, when creating his view of biology, required a new terminology which would separate his position from others, especially Darwinians. For this reason, he applied terms from music to explain his view on a grander scale (Kull 1999: 391). He utilized musical terminology in describing the relationship between the basic cells (chime), organs (melody), (within the) organism (symphony), organisms (harmony) and nature (composition). Although it is quite important to remark that von Uexküll himself is a bit hesitant in defining an exact composition of nature as such, he states:

> Nature offers us no theories, so the expression 'a theory of the composition of nature' may be misleading. By such a theory is only meant a generalization of the rules that we believe we have discovered in the study of the composition of nature. (Uexküll 1982[1940]: 52)

To sum up, Umwelt theory stresses the intersubjective and teleological traits of nature and life. And more importantly, it provides us with a model of life, namely the web model. In contrast to the tree model, web model does not have a stem from which it grows. Kull (2003: 594) describes this model as if having no origin[7] and essentially as a communication network where, instead of competition or dominance we observe in the tree model, recognition, co-existence and symbiosis

[7] This statement, combined with Umwelt theory, may be taken as an indication of rejection of evolution in both cases. It should therefore be noted that von Uexküll was not against the idea of evolution as such but only evolution that did not pay any attention to the organism's intersubjective and goal-directed nature, as was the case with Darwin's theory of evolution. Perhaps endosymbiosis (see Margulis 1998) might be considered as a more fitting evolutionary theory candidate.

act as the primary functions. That being said, the web model[8] and von Uexküll's Umwelt theory were only accepted and used by a small group of scholars (from the start of 1930's) for quite some time while Darwinian approach to biology and other disciplines in general enjoyed quite a lot of success. It was not until 1970's von Uexküll's ideas were revived[9] and slowly started to be seen in the works of scholars of various disciplines.

The concepts that are indicated as essential to Umwelt theory will also be extensively discussed in the fourth chapter. Descriptions of these concepts in this chapter should not be taken as final within the scope of this thesis as they remain loyal to the way von Uexküll understood and represented them in his works.

[8] I do not imply here that the web model and Umwelt theory can be used interchangeably or that they are one and the same. This is only to indicate the latter as an account of the former.

[9] Initially, it was T. Sebeok in his various publications and talks that managed to spark interest in von Uexküll again which was followed by Thure von Uexküll and his publications on his father's works (Kull 2001: 11)

## 2. The Reception of Umwelt Theory in Embodied AI Research

In this chapter I will mainly focus on how scholars of various fields used Umwelt theory (or parts of it) in embodied AI research. While trying to follow a chronological order, the focus will first be on individual researchers who frequently used the term and their approach to the issue as they describe it.

### 2.1. Rodney Brooks

One of the first scholars to mention von Uexküll in their works as a computer scientist was Rodney Brooks[10]. Brooks used the term *Merkwelt,* which is an essential part of von Uexküll's functional cycle, first in *Achieving Intelligence Through Building Robots* (1986: 4) mainly as a basis for criticism against the abstraction in AI research which he then claims to be the result of the reductionist mindset common in scientific thought. Another instance of using *Merkwelt* is in his text "Intelligence without representation" (1991: 141) where his approach remains similar. His reason for involving von Uexküll's work is not to suggest a full-scale implementation of Umwelt theory but rather to adopt a certain part of it, namely the perception - action coupling in order for the built robot to be able to map and perform in its environment properly[11]. Brooks claims that the traditional reductionist approach to building robots is not suitable enough when considering the complications, or the possibly complex nature of a given physical environment, as those reduced archetypes cannot account for all objects the robot might encounter in its environment and it is also unrealistic to assume that one can create representations for all of the existing variations. Thus, he suggests not to use a built-in model. Instead, he proposes that the robot models the environment itself according to itself, using its perceptual and actuator units (essentially forming a functional

[10] Rodney Brooks (1954-) is the Panasonic Professor Emeritus of Robotics at MIT. He has actively worked on AI and robotics and is known as one of the initiators of the modern (embodied) approach to AI.

[11] The problem has its origins in the traditional AI research and the symbolic representations those scholars used to create a built-in map which the robot uses when interacting with its environment. I will be addressing this issue more in detail in the third chapter.

cycle) which increases the chances of the robot's functioning properly. In his words: "The world is its own best model" (1990: 5).

In his later works (see Brooks 2017; Brooks et al 1999; Brooks and Stein 1994), he develops and maintains the same idea although the reference to von Uexküll and Umwelt theory is no longer present but inherent.

### 2.2. Claus Emmeche

Claus Emmeche[12], as a semiotician, has mainly focused on artificial life (henceforth AL) and approached the issue from a biological point of view. Nevertheless, he has remarks on AI research (as AL and AI share some fundamental concepts) and uses von Uexküll's works and Umwelt theory quite extensively in a couple of his texts.

The first[13] reference to von Uexküll's work in this context appears in his "Life as an Abstract Phenomenon: Is Artificial Life Possible?" (1992: 472) as an introductory and alternative parallel notion to functionalism that is apparently inherent in strong AL research and cognitive science. The issue addressed here was to provide an example for the argument, namely that cognition is very much dependent on the subject's biological processes which are embedded in its material nature, from the point of view of scholars of phenomenological orientation of which von Uexküll can be associated with. He addressed the possibility of connecting von Uexküll's work with autonomous systems in detail in his later text "Does a Robot have an Umwelt? Reflections on the qualitative biosemiotics of Jakob von Uexküll" (2001). The analysis he provides here is quite extensive in that he first starts from the position of von Uexküll's work in current biology then provides a short history of the research that has been made in autonomous systems and in the conclusion part, he discusses the possibility of a robot's having an Umwelt. The last part bears significance in that Emmeche evaluates this possibility from a biosemiotic point of view. He mentions that every organism that has an Umwelt is able to recognize a *quali sign*, which in

[12] Claus Emmeche (1956-) is an associate professor of theoretical biology at the University of Cophenagen. He has extensively written about biosemiotics, artificial life and theoretical biology.

[13] His text in Danish, however, apparently indicates an earlier connection (see Emmeche 1990).

Peircean terminology refers to the degenerate and simplest form of quality that can be interpreted as a sign, and the inability of recognition in human built artifacts regarding such a sign (Emmeche 2001: 681). This, in turn, has other implications such as the idea that it is the result of the material structure of the artifact or simply the fact that any capacity of recognition is not naturally developed but rather put there by human designers. When considering how much importance subjectivity and the experiential feeling of an activity has in Umwelt theory, he concludes that those artifacts indeed do not possess an Umwelt because they do not have any agency of their own, which every organism can be said to possess. That being said, he states it is nevertheless a possibility to achieve agency in those artifacts in the future.

His works from this point onwards are mainly concerned with the notion of emergence and embodiment of different kinds (see Emmeche 2004, 2007, 2011). Nevertheless, he mentions AL models and their relevance in biosemiotic theory but not in a context that would consider von Uexküll's work especially relevant.

### 2.3. Erich Prem

Another scholar to explicitly make use of von Uexküll's work is Erich Prem[14]. Prem's understanding and application of von Uexküll's theory is rather specific and somewhat has its emphasis more on the functional properties it introduces. Reference to von Uexküll first appears in his "Elements of an Epistemology of Embodied AI" (1996: 98) and then "Epistemic Autonomy in Models of Living Systems" (1997: 6-7), "Semiosis in Embodied Autonomous Systems" (1998: 725). During the course of given texts, he first evaluates the current situation in AI research, particularly emphasizing the effects of Cartesian dualism, and indicates the need to turn towards biology for a more effective model which explains his direct application of von Uexküll's work and functional cycle. Apart from the modeling capacity and some sort of autonomy the application of functional cycle brings, Prem also stresses the notion of anticipation in his texts. Anticipation here is seen as the regulatory driving mechanism for the system in the sense that it provides

[14] Erich Prem is a researcher from the Vienna University of Technology. His involvement with AI research comes from a time he was actively participating in the field.

"objectives" for the system which in turn can be used to provide a certain teleological aspect (Prem 1997: 9), corresponding with von Uexküll's notion of "life-task".

Another aspect that should be mentioned here is the amount of influence the works of Martin Heidegger, who was also influenced (cf. Buchanan 2008) by von Uexküll, has on his philosophical stance. Indeed, he suggests a certain ontological perspective in opposition to the Cartesian understanding in that functions are not merely physical properties of objects, but they also have other qualities that are buried in layer of references to other objects and functions, i.e. a pen refers not only to other functions such as piercing or pointing but also other objects like paper and the hand that holds it. There is also the mention of a social aspect of sign use in the context of robot communication that stresses the purposeful interaction with both the environment and other artifacts, focusing on the dynamic relationship between two parties (Prem 1998: 727).

### 2.4. Andy Clark

Andy Clark[15] is one of the cognitive scientists making explicit use of von Uexküll's work and Umwelt theory. In his book titled *Being There: Putting Brain, Body and World Together Again* (1997), he first introduces the initial robot models that are built without a central planner, rejecting the traditional approach to robotics, and essentially discusses the aspects between the two major approaches to robotics. Von Uexküll's work is introduced as the biological counterpart of the mechanisms that correspond to what is presented as ideal and plausible in robots. He emphasizes the notion of niche and mention that robots created without a central planner exemplify niche dependent sensing which he presents as all too familiar, which also happens to be a core theme of von Uexküll's work and Umwelt theory (Clark 1997: 24). Afterwards he further stresses the importance of mind and body connection with regard to cognition and development.

[15] Andy Clark (1957-) is a professor of philosophy at the University of Edinburgh in Scotland. His work is deeply related with cognitive science and embodiment.

## 2.5. Ricardo Gudwin

The article "Umwelts and Artificial Devices" (1999) by Ricardo Gudwin[16] is not actually concerned directly with von Uexküll or his line of thought. It is actually more like a response to Emmeche's text "Does a Robot have an Umwelt?" (Emmeche 2001) but it is nevertheless important in the context of this work. Gudwin's main argument revolves around his claim that Emmeche fails to make a distinction between a real Umwelt and the internal representation a system has for its Umwelt. He further suggests, in direct opposition to Emmeche, that a robot does indeed have an Umwelt on the basis that the sensory data provided by its sensorimotor units is what constitutes an Umwelt which, according to him, can be reduced to elementary parts so that the robot in question may perform more effectively without having to represent its Umwelt in its entirety but only as a model. In essence, he is suggesting that any system must, to some degree, have internal representations rather than solely relying on the interaction with its environment to achieve a certain degree of intelligence.

## 2.6. Tom Ziemke (and co-authors)

Tom Ziemke[17] is one of the scholars who has extensively discussed von Uexküll's work in relation to robots, computer science and cognitive science. He also expressed the need to explore such relation in its various aspects across various disciplines. I will briefly try to summarize the work done by him and his co-authors according to their chronological order.

Sharkey and Ziemke first use von Uexküll's work and Umwelt theory in their article "Biological and Psychological Foundations of Autonomous Robotics" (1998). Their argument essentially revolves around how the organism in its natural environment is considered to be rooted (different from grounded) and is capable of bodily solidarity and behavioral coherence. They start

[16] Ricardo Gudwin (1967-) is a researcher at the University of Campinas in Brazil. He has written and co-authored several articles about computational semiotics.

[17] Tom Ziemke (1969-) is a professor of cognitive science at the University of Skövde and professor of cognitive systems at Linköping University in Sweden. His work is concerned with topics ranging from human-robot interactions to embodied cognition. His works also deal extensively with Umwelt theory and its relation to embodied AI research.

from the works of Jacques Loeb and then move on to von Uexküll (both of whose works are introduced and analyzed in detail) as a possible source of inspiration for what they at the time call biorobotics while also explaining both the current and the past situation in AI research. Their article "A stroll through the worlds of robots and animals" (2001a) is essentially a revised (see also Sharkey and Ziemke 2000, 2001b) and more detailed (and perhaps a bit more appropriated to fit into a semiotic context) form of their article in 1998. As the name suggests, it also pays a special homage to von Uexküll and his work and the issue of embodiment is elaborated further in relation to the analogy of embodiment between organisms and robots.

Ziemke in his article "Rethinking Grounding" (1999) presents the reader with a brief case of how traditional AI research was conducted and, in the light of the criticism it was subjected to, how the new direction it took was shaped. The new direction is then revealed to be embodiment and physically grounding the symbols in reference to which the robot operates. However, he states that simply building bodies for AI, namely a robot so that it can move about and perform in its immediate environment, is only a partial solution because the task-based nature of the actions it takes still remain determined either by its programming or the limitations of the units that the designers decided to use for one reason or the other. Therefore, it is implied that such an approach can only partially succeed if it does not recognize the non-arbitrary and co-evolutionary relationship between an organism and its environment. It is in this context von Uexküll and Umwelt theory is mentioned as an evidential basis in biology.

Likewise, the next text "The Construction of 'Reality' in the Robot, constructive perspectives on situated artificial intelligence and adaptive robotics" (2001a) discusses the issue of embodiment from a constructive perspective, as the title suggests, using Piaget's and von Uexküll's work as its basis. It becomes evident towards the end, however, that the article is concerned more with situatedness than embodiment because the new paradigm has already successfully (and rightfully) adopted embodiment as its new direction. The former differs from the latter in that it not only implies the existence of a physical body and a functional cycle, but also environment's direct effect on the system's behavior. The final verdict of the paper states that although von Uexküll's work has indeed proved to be most useful for adaptive robotics, his theory remains cumbersome to achieve in many of its aspects and is mostly useful for studying the modeling of constructive processes.

Ziemke's article "Are Robots Embodied?" (2001b) focuses on embodiment types and whether commonly used embodiment, or the common interpretation of embodiment in AI research, is enough for achieving the desired level of intelligence. He classifies types of embodiment and focuses on organismic embodiment as the type of embodiment that is specific to living organisms that currently remain unachievable in robots. He specifically uses von Uexküll's work in further explaining organismic embodiment and indicates that further research is needed in order to somehow realize organismic embodiment level of embodiment in robots. His articles "What's that thing called embodiment?" (2003) and "Embodied AI as Science: Models of Embodied Cognition, Embodied Models of Cognition, or Both?" (2004) are essentially somewhat improved versions of his initial article, the latter being the most elaborate one, with some minor additions.

In "On the epigenesis of meaning in robots and organisms: Could a humanoid robot develop a human(oid) Umwelt?" (2002), he discusses a rather specific case of humanoid (as in built after human appearance) robots' chances of developing a human Umwelt. He first discusses the possibility from the point that if a robot is able to engage in human activities (since it is built that way) long enough, it may in the end develop a human Umwelt. He then discusses *if* the robot would indeed develop an Umwelt, why would we believe it to be human-like when in fact it does not even share the essential features, but only representations of those features, of a human body. The conclusion of the discussion also ends in a similar manner also present in his other articles, namely that robots are not able to achieve intrinsic meaning or autonomy by themselves while living creatures display such features naturally.

Lindblom and Ziemke in "Embodiment and social interaction: A cognitive science perspective" (2007) differentiate between two kinds of embodiment, being simple embodiment and radical embodiment. The former is mainly used to describe the still mechanistic nature envisioned, even if not explicitly, in AI and robotics research where the focus is on the surface level perception-action dependent environment interaction. The latter differs from the former in that it is focused on the autopoietic nature of the organism and its biology as a whole rather than the abstractions of certain functions and their applications. The reference to von Uexküll's work mainly covers the differences between the two kinds of embodiment as his work evidently describes how autonomous subjects and heteronomous mechanisms work differently. The authors

then move on to a more specific case, social interaction, and discuss it in terms of embodiment and human-machine interaction.

Another reference to the works of von Uexküll is found in Sørensen and Ziemke's article "Agents without Agency?" (2007). They first discuss how, in scientific discourse, agency is approached and how the limitations, or even the absence, of agency resulted in reductionistic applications in early, and to some extent contemporary, AI research. Von Uexküll's work is mainly discussed under the topics of self-organization, i.e. how agency is constructed in complex systems, and in providing an alternative basis to reductionist approaches to agency. It is then concluded that the expected level of agency in AI research can so far only be observed in living organisms and it is by studying the model they present agency and other implications of AI research be furthered.

The article "On the role of emotion in biological and robotic autonomy" (2008) offers another approach to embodiment, or more specifically organismic embodiment, from the point of emotions and other regulatory mechanisms, i.e. homeostasis. He mentions that usual adaptations of embodiment in robotics is limited only to physical interaction while the desired level of autonomy requires a more complex structure in its totality. The reference to von Uexküll in this context is used to describe the complex and complete nature of the organisms compared to the artificial creations that do not yet share the notion of life-task. It is implied that notions of embodiment can make progress if there is a homeostatic mechanism regulating the inner states of the robot that could be considered the equal of what is called emotion in organisms, further emphasizing the need to shift from merely physical realizations of embodiment to more "subject" based or organismic versions of embodiment.

His last article about the issue of how types of embodiment differ, "The body of knowledge: on the role of the living body in grounding embodied cognition" (2016), is basically a summary of the differences in what has come to be seen as mainstream embodiment, which more or less reduces the body (and cognition) to physical sensorimotor interaction, and embodiment that is deeply rooted in its underlying biological mechanisms, such as metabolism, emotion, etc., which was previously referred to as organismic embodiment in earlier works.

## 2.7. Winfried Nöth

Another semiotician who also highlighted the connection is Winfried Nöth[18], even if it is limited to a single article. In his text "Semiosis and the Umwelt of a Robot" (2001) Nöth first discusses the similarities between an organism's Umwelt and a robot's probable Umwelt. His argument in this case is quite basic and is as follows: A robot might have an Umwelt because (1) a robot's sensory perception and action units can just as be restrictive and effectively shapes its so-called Umwelt, (2) the variation in the mentioned units attribute to a difference that might correspond to subjectivity in Umwelt theory, (3) the robot's model of environment, or its symbolic representation, corresponds to an organism's *Innenwelt* and (4) the robot possesses both perception and action units for environment interaction which effectively constitutes a functional cycle (Nöth 2001: 696). He then discusses the difference between an organism's and a robot's Umwelt, which is simply expressed with terms of self-referentiality and alloreferentiality. The former indicates the organism's capacity for creating or having meaning for itself while in the latter the reference for meaning making has to be put there from somewhere else. In the robot's case the meaning making, and its task-based nature has to be established by a designer, putting it in the category of an allorefential machine and possibly marking it as not having an Umwelt (Nöth 2001: 697). This discussion is then followed by elaborations of disputes between paradigms and concluding with a possible solution using Peircean terminology.

One thing to note here is that some of Nöth's other works also deal with similar topics such as semiotics in machines or in computer science (see Nöth 1997, 2003). The reason they are not reviewed in this chapter is because of his heavy use of Peircean philosophy rather than mentioning von Uexküll or his work.

## 2.8. Alvaro Moreno and Xabier Barandiaran

[18] Winfried Nöth (1944-) is currently a professor of cognitive semiotics at Sao Paulo Catholic University in Brazil.

Moreno [19] and Barandiaran [20] in their "On What Makes Certain Dynamical Systems Cognitive: A Minimally Cognitive Organization Program" (2006) present us with another form of embodiment, namely emotional embodiment. It is the authors criticism that cognition based on simple behavioral cycles are not plausible and lacking in teleological content in artificial agents. They mention von Uexküll's work along with other authors to emphasize the teleological nature of agents that are autonomous, namely organisms. They suggest that emotional embodiment implies the formation of neurodynamic processes which operate as decoupled from the sensorimotor activity that is happening simultaneously (Moreno and Barandiaran 2006: 180). Such an approach stems from their understanding that cognition alone cannot be explained in terms of behavior alone but must be complemented by an internal structure that adaptively organizes the inner structure of the agent.

To sum up, the scholars who made use of von Uexküll's work has used it in two main categories. The first category can be summarized as forming a valid basis for how things work in biological organisms from which the inspiration is being taken from. The works listed under the authors Brooks, Prem, Clark, Ziemke (and co-authors) and Moreno and Barandiaran make explicit use of von Uexküll's work in this context. The second category of use mainly consists of deriving a valid basis for criticism to use against AI systems with regard to how biological organisms does not work. Emmeche's, Nöth's and Ziemke's (together with his co-authors') works are essentially assuming this position. One important thing to note here, however, is that this list should not be seen as a complete list of works making use of von Uexküll's work in this particular context, but rather as something that is available for development.

[19] Alvaro Moreno is a professor of philosophy of science at the University of Basque Country in Spain. He has written extensively about artificial life, self-organization and complex systems.

[20] Xabier Barandiaran is currently a lecturer at the University of Basque Country in Spain. He has mostly written about situated and embodied cognition.

## 3. AI History, Criticisms and Directions

In this chapter I will first provide a brief account of the history of AI research, emphasizing the points, problems and criticisms that eventually gave rise to the embodiment paradigm, and then discuss why it was thought necessary to adopt theories from disciplines other than computer science or its derivations. By the end of this chapter, I expect to give a thorough explanation of how von Uexküll's work and Umwelt theory is situated among other theories. The analysis will also provide some initial expressions regarding whether Umwelt theory is indeed relevant in the given context.

### 3.1. AI, a brief history

As previously stated, this section will not be covering the totality of the AI research history, as it is virtually impossible to do so within the scope of this work, but rather I will mainly focus on the issues that led to the realization of the requirement of a body for AI. I will try to follow a chronological order. It is, however, important to note that this section will make minimal to no use of the research conducted before the establishment of AI as an individual field since notions such as computability and machine intelligence can be seen in the works of earlier thinkers and scholars.

#### 3.1.1. From 1950's to 1974

What has come to be called AI has its roots in several gatherings that started to gain momentum in 1950's. Dartmouth Summer Project in 1956, which was one of the landmark gatherings in the early AI history, was where John McCarthy persuaded the attendees, and also coined the term, to call their field of research artificial intelligence. McCarthy had several reasons to name the field differently, mainly revolving around the need to separate it from other scientific fields, but what I found most striking was the fact that he wanted to escape any association with

cybernetics [21]. In his words: "Its (cybernetics') concentration on analog feedback seemed misguided and I wished to avoid having either to accept Norbert Wiener as a guru or having to argue with him" (Quoted in Nilsson 2010: 78). This quote actually underlines the general direction AI research would take in years to come.

What AI research of the time (henceforth traditional or classical AI research) concerned itself with was the properties of humans that were thought as "intelligent" as opposed to behaviorist approaches that preceded it[22]. Such properties were listed as features of an intelligent mind and thus the research in this tradition was mainly done in topics such as natural language programming, proving mathematical theorems and so on. One consequence of this line of thought was the computer metaphor for mind, which in turn caused cognition to be seen very much like a computer program. Anything that could be formalized into symbols [23] and calculated accurately in computational terms was introduced as a subject of study. Another implication of such an approach to intelligence, however, was the impression that intelligence could be realized regardless of the hardware (body or brain) it operated on (Pfeifer and Bongard 2007: 27). There are different terms by which this paradigm has come to be called. One of them is cognitivism due to its emphasis on formal representations of an external world's abstract forms and its sole focus on the inner computing mechanism. Another term referring to the traditional approach to AI is GOFAI, short for Good Old Fashioned AI, which was coined by John Haugeland to better indicate the paradigm in light of more recent approaches (cf. Haugeland 1985). His formulation of what is essential to GOFAI presents us with quite a decent summary: (1) "our ability to deal with things intelligently is due to our capacity to think about them reasonably" (including subconscious thinking) and (2) "our capacity to think about things reasonably amounts to a faculty for internal 'automatic' symbol

[21] In a general sense, what cybernetics at the time concerned itself with can be summarized as "causal cycles" or "feedback loops", indicating an action of the system in question causing a certain effect in its environment and the affected environment causing another effect in the system in return. This may be seen as somewhat similar to von Uexküll's functional cycle minus the teleological nature of the organism or the system.

[22] Opposition to behaviorism mainly stemmed from behaviorism's rejection of any theory that could not be observed or tested empirically. "The cognitive revolution", as it came to be called, has its roots in the works of scholars ranging from Turing to Chomsky that made it clear the existence of some underlying mechanism in the human mind.

[23] This is very different from the notion of symbol one would encounter in semiotic context. Symbol here is a basic (and most often direct) representation of an object x that exists in the target system (environment).

manipulation" (Haugeland 1985: 113). The last approach to AI which can be associated with the traditional paradigm is strong AI. It is suggested by this approach that AI should be nothing less than genuine human level intelligence and we are in fact computers ourselves, which is why it is thought human level intelligence can be realized by computation alone.

### 3.1.2. The first AI winter

Traditional AI research, while it had certain achievements, had promised too much, such as the claim AI would be able to do any work a human could do, which they failed to deliver. The achievements of their time were indeed unique and promising, giving way to such optimism regarding the future of the field. However, there were limitations that were brought about either by technology or the way the problems were understood and analyzed. The failure to accomplish the promised goals had consequences which mainly concerned the funding of AI research.

Sir James Lighthill, who was asked to compile a report regarding the progress and adequacy of the AI research, presented in his report (see Lighthill 1973) that while AI research had indeed obtained some positive results (which were nevertheless described as " to a disappointingly smaller extent than had been hoped and expected "), the majority of the research was described as "failed to reach its more grandiose aims" and that there were still problems like combinatorial explosion that remained unsolved. The report had a rather significant impact that not only resulted in limited funding (and funding projects only with an objective in line with the funding organization's intentions i.e. military applications) but also deeply affected the field's public reputation. John McCarthy's response to the report shows both his resentment and acknowledgement in that he accepts there has indeed been some exaggeration in the research conducted. But he states that there has nevertheless been progress and AI research has been moderately successful, which he describes as "perhaps more than social sciences and less than many physical sciences" (McCarthy 1974: 322). Another thing to note here is his description of the field and what should be its object of study: "studying the structure of information and the structure of the problem solving processes independently of applications and independently of its realization in animals or humans" (McCarthy: 317) (which presents us with a better understanding of the mindset) as it was also something Lighthill evaluated in his report.

However, the Lighthill Report was not the only criticism traditional AI had received. Starting from the 60's, various scholars expressed their doubts regarding the claims made by AI researchers.

### 3.1.3. Criticisms of the initial AI research

One of the major criticisms in the traditional AI research was made from within the field. In 1958 Frank Rosenblatt designed one of the earliest neural nets, named 'perceptron', that could work as a binary classifier system (which would eventually be called connectionism as a separate paradigm). It was suggested that given time perceptrons could do much more, but the effort came to an end with the publication of the book *Perceptrons* (1969) by Marvin Minsky and Seymour Papert, which underlined the limitations of the initial perceptron model and suggested that the limitations would exist even in more complex models (Franklin 2014: 19). Even though their suggestion has been mostly proved to be false afterwards, the impact of their work effectively prevented the research in neural networks to get funding (also called the neural net winter) until mid 80's.

Another criticism was from philosophy. Hubert Dreyfus[24] first in his book *Alchemy and Artificial Intelligence* (1965) and later in *What computers can't do: a critique of artificial reason* (1972), where he further improved his ideas, criticized and identified the assumptions beneath the optimism in traditional AI research and suggested an alternative path. His criticism includes four layers. The first layer is the biological assumption that human brain is a general-purpose symbol manipulator and it is therefore possible to achieve human level intelligence by manipulating abstract symbols in the computer. He criticizes the assumption by stating that there is a significant difference between the biological and highly interactive brain and the stagnant and non-interactive machine organization which he claims to be backed by the evidence from biology (Dreyfus 1972: 74). The second layer is the psychological assumption that human minds, and coincidentally human behavior, can be explained by a mechanism that processes a certain set of information inputs. The third layer is the epistemological assumption that all kinds of knowledge can be

[24] Hubert Dreyfus (1929-2017) was a professor of philosophy whose interests included philosophy of AI, existentialism and phenomenology. His exposure to AI research comes from a time when he was teaching at MIT where he also had direct contact with many of the AI researchers.

formalized and turned into rules. The formulated rules then can be used to describe the behavior of the total system. His criticism on the second and the third assumptions stems from the fact that human behavior or the way humans function shows a certain degree of flexibility and the extracted rules cannot be used to generalize the whole behavior of the system (Dreyfus 1972: 114). The last assumption is the ontological one and it is actually the combination of the previous three assumptions. In essence, it is suggested that anything that is of significance to intelligent behavior must, in principle, be analyzable as a set of context free determinate elements (Dreyfus 1972: 68). Dreyfus criticizes this point of view in that the conscious and cognitive process of humans allow them to consider the details and its context simultaneously (what he calls "fringe-consciousness") which is problematic for machines in ambiguous situations (McCorduck 2004: 214).

In his alternative suggestion for AI, Dreyfus highlights the importance of a body which he considers essential for achieving the intended level of intelligence. He draws from the phenomenology of Martin Heidegger and Maurice Merleau-Ponty, especially the former, while formulating his own account. Dreyfus's claim is that AI requires all the data to be formalized, which grows at a rate computers are unable to deal with, and used within certain rules to function. Humans, however, by having bodies, are able to make intelligent decisions not based on prescribed rules but regulations that are also based on bodily needs requiring to be satisfied. He writes:

> This alternative conception of man and his ability to behave intelligently is really an analysis of the way man's skillful bodily activity as he works to satisfy his needs generates the human world. And it is this world which sets up the conditions under which specific facts become accessible to man in indefinite and openended ways, because these facts are originally organized in terms of these needs. (Dreyfus 1972: 193)

Ultimately, it is implied that the formalized data are the result of the abstractions from the experience of interaction between the body and the environment and it is simply not possible to achieve the intelligent behavior without deleting one of the variables, that made the collection of the data possible in the first place, from the equation.

Another well-known criticism of the traditional AI comes from John Searle[25]. Searle in his article "Minds, Brains and Programs" (1980) discusses the case of 'understanding' in a given

[25] John Searle (1932-) currently is a professor emeritus at University of California, Berkeley. He has mainly written about philosophy of mind, philosophy of language and social philosophy. His writings about the capabilities of AI has caused quite the controversy in the field.

computer that is built according to the cognitivist paradigm, or 'strong AI'. He formulates his argument with a thought experiment called the "Chinese Room". It can be basically summarized as follows. A man is locked inside a room surrounded by Chinese speaking people. The man is provided with Chinese symbols which he does not understand and a rule book in English (which he understands). The Chinese speaking people send questions in the room in Chinese and the man is able to answer them by simply referring to the rule book and matching (judging by their shapes) the correct symbols. He then sends out the answers. Searle then supposes the man gets so good at following the rules that the answers he provides are indistinguishable from that of a Chinese native speaker (Searle 1980: 419).

The point of the Chinese room experiment is that even though the people interacting with the computer or AI (the room) think that the AI actually understands their questions and provides answers, the actual situation is quite different. The only thing the room is actually good at is the symbol manipulation. It shows no understanding whatsoever. When it provides an answer, it does not know what the question means, and this issue persists when it formulates an answer. It essentially lacks 'aboutness', it does not know what the symbols refer to (Nilsson 2010: 384). Searle states, when asked whether there can be an understanding based solely on the right kind of program in a computer (and his answer is negative):

> Because the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even symbol manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have only a syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output. (Searle 1980: 427)

While Searle's argument has caused quite a stir in the field and has received a number of responses, there are two that of concern to this research. The first one is the systems reply. It is suggested that while the man in the room does not understand anything, the system as a whole does because the man is a part of the system and understanding in this case is ascribed to the totality of the system the man is part of. Searle apparently thought of this reply as non-sensical as this line of thought would imply that, say, the paper the man would use to calculate a reply to the posed question would have an understanding of Chinese (Searle 1980: 421). The second response is the robot reply. Robot reply implies that the man does not understand Chinese because he is not causally connected

to the environment. Searle interprets this reply as admitting that cognition is not solely based on formal symbol manipulation. He, however, continues by saying that even though we provide the room with perceptual and actuator capacities so as to make it move about or do anything we desire, it would not constitute in any way towards developing an understanding because all that is happening is still happening in the context of symbol manipulation. What previously classified as simple input-output relationship now corresponds to perception (input) and action (output). The room still does not understand what the difference between the two cases is because all it does is formal symbol manipulation (Searle 1980: 423-424).

It can be seen that the first AI winter is caused not only by the inflated optimism within the field but also by the constant criticism the mentioned optimism brought about. Nevertheless, it resulted in both reduced funding for AI research and lowered the level of public and institutional trust and interest in the field for several years. The AI winter would come to an end with the reemergence of connectionism in the 80's.

### 3.1.4. Between 1980 and 1987

After Minsky and Papert's publication (see sub-section 3.1.3.) and its effects, there were little to no research conducted in neural networks save for a few names. The field re-emerged in 80's, due to the works of PDP research group at the University of California (Franklin 2014: 21) and achieved a certain degree of success where their predecessors could not. Briefly connectionism, or the study of AI through neural nets, assumes that cognition is an emergent phenomenon resulting from the interactions between a (not necessarily) large number of simple processing units i.e. neurons (Sun 2014: 109), as opposed to cognitivism where cognition is attributed to a single processing unit.

While the re-emerging interest in AI is partly[26] owed to the success of connectionism, the highlight was on the expert systems. One aspect of success in research was the change in the way

[26] Another reason for the sudden interest in AI after its initial failure to deliver what it promised is due to the competition between countries. In this particular period, it was because the Japanese government decided to invest a rather large amount of money in AI that it was deemed necessary to meet the competition.

knowledge and cognition was understood in the preceding years. While intelligent behavior was understood as stemming from uniform and general knowledge previously, it was suggested by connectionism that such behavior might be the result of using diverse knowledge in differing ways (McCorduck 2004: 298-299). The shift had many implications but perhaps the most important one was the idea to isolate knowledge and create an intelligent program to do a specific task, which was called an expert system afterwards. Some prime examples were MYCIN (a program used to identify bacteria and suggest antibiotics) and XCON (a program to suggest computer components based on the customer's preference). Expert systems proved to be very useful and gained a fair amount of market share with companies constantly investing in the growing industry. In return expert systems saved companies quite a lot of money. It would not, however, last very long. While the expert systems had their uses and were effectively beneficial in some respects, they also were expensive to maintain[27]. The cost of running an expert system soon became unreasonable especially when the general computing power was in constant increase in accordance with Moore's Law[28]. It simply made more sense to wait and acquire systems more capable of general-purpose computing at a cheaper price rather than keep investing in costly, static systems (McCorduck 2004: 435). Another issue about the expert systems was learning and brittleness which became major problems when confronted with systems that could run more applications in software form. In the end, the reign of the expert systems was short lived. They were inflexible, unable to keep up with the dynamic and apparently random situations occasionally encountered in any industrial or working context. The demand gradually came to an end, and with it, the majority of expert systems development.

[27] One thing to note here is the expert systems in the given time period was not only programs that could run on any machine. They came with the machine they operated on which made them much less versatile compared to what we have today.

[28] Moore's Law, named after Gordon Moore (co-founder of Intel), suggests that every two years the number of transistors in a silicon chip would double (see Moore 1965) which would roughly correspond to doubling of the available computing power within the processing unit. While Moore's Law held up, especially in the early decades, the progress has gradually been slowing down in the last decade and has been predicted to come to an end (by Moore himself) in the next decade. Another thing to note about Moore's Law is that it is not a physical law but rather a prediction which became a driving factor for many people in the industry.

### 3.1.5. The second AI winter

In the following years, another pattern similar to the occurrences of the first AI winter repeated itself due to expert systems' not being able to meet the required standards. Cutbacks in funding followed as the main funding government organizations, such as DARPA (Defense Advanced Research Projects Agency) in the US, decided to focus on objectives that either looked more promising or guaranteed immediate results (McCorduck 2004: 430). Connectionism was among the more favored fields and would taste further achievements even though there were only evidence that it could be useful in principle at the given time. Expert systems, which were described as mere 'clever programming', however, were forced to be abandoned as they apparently betrayed the initial (and traditional) goals of AI research (McCorduck 2004: 430).

## 3.2. Nouvelle AI

Having gone through various problems and funding cutbacks, some AI researchers thought that previous AI research had skipped a crucial part in their strict agenda. Approaching intelligence as something that could strictly be formulated into rules and symbols, which were, in turn, used to produce the intelligent behavior in a computer had proven to be very problematic and unfruitful. The problem of achieving intelligence apparently had another issue that laid beneath the mindset of traditional AI research.

When people talk about intelligence or intelligent behavior, they mainly refer to activities or thinking processes such as proving mathematical theorems or playing chess. This has two main implications: (1) that intelligence or intelligent behavior is the sole property of human beings and, (2) that any other behavior that is not accepted as complex is not seen as constitutive of intelligence or intelligent behavior. While these implications were not a secret, as a matter of fact, they were articulated as directly as possible at every chance, they have managed to divert the researchers' attention from what seemed like trivial and daily actions; the actions that everyone took for granted and branded as not requiring intelligence, such as walking, running or acting with common sense. These are the actions that a seven-year-old can perform with ease, yet they happen to be things the million-dollar computers seem unable to achieve as gracefully, even today. The flexible nature of

actions that humans perform daily, regardless of the context, has proven especially hard to be captured in formal and rule-like environments that computers operate in. I think that this issue is reflected perfectly in Moravec's Paradox.

Hans Moravec[29] identifies the issue in artificial intelligence with a dichotomy between higher reasoning, which is described as adult-level problems, and lower level skills such as perception and mobility. He describes the former as comparatively easy to achieve while the latter seems almost impossible to implement, even though it is the set of skills that a one-year-old possesses (Moravec 1988: 15). Moravec's explanation, as to why it is so, appears to be embedded in the biological evolution of our species. He states:

> Since the first multicelled animals appeared about a billion years ago, survival in the fierce competition over such limited resources as space, food, or mates has often been awarded to the animal that could most quickly produce a correct action from inconclusive perceptions. Encoded in the large, highly evolved sensory and motor portions of the human brain is a billion years of experience about the nature of the world and how to survive in it. The deliberate process we call reasoning is, I believe, the thinnest veneer of human thought, effective only because it is supported by this much older and much more powerful, though usually unconscious, sensorimotor knowledge. We are all prodigious olympians in perceptual and motor areas, so good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100 thousand years old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it. (Moravec 1988: 15-16)

While I think Moravec's interpretation is very sufficient and self-explanatory, the idea that higher reasoning capacities of humans is developed on top of apparently more basic and unconscious skills must be stressed further. As a matter of fact, earlier experiments to achieve autonomous behavior in cybernetics focused almost entirely on those basic skills. They did not have any reasoning capacities (as in computation) but relied entirely on analog circuitry. One of the most well-known examples was from William Grey Walter in the early 1950's. Walter designed two electronic tortoises, Elmer and Elsie, and equipped them with a simple artificial nerve system that provided the machines with a basic set of behavior reflexes. A photocell enabled the machines to not only perceive light but also avoid obstacles and bright light, making them attracted only to moderate light under normal conditions. In case their batteries needed re-charging, they instead

[29] Hans Moravec (1948-) currently is a research member at robotics institute at Carnegie Mellon University. He is known for his work on AI, robotics and impacts of technology.

headed towards bright light and re-charged their batteries with an electrical touch sensor. Such behavior is often interpreted to be the counterpart of 'hunger' in biological organisms (Nilsson 2010: 44-45; Sharkey and Ziemke 2001a: 709). Another example of a machine that exhibited similar behavior is the Beast, built by the researchers at John Hopkins University Applied Physics Laboratory in 1960. The Beast, like Walter's mechanical tortoises, did not contain any parts that allowed it to reason but built with on-board electronics that included photocells, a sonar system and a "wall-feeling" arm that enabled it to wander the white colored corridors and look for the dark-covered power plugs. If the system was lacking power and required recharging, the Beast would stop upon finding a power plug and recharge itself. Otherwise, it would keep on patrolling the corridors (Nilsson 2010: 213).

Regarding the difficulties caused by the lack of basic skills, one can see that Minsky has also addressed the problem from a similar perspective. He states:

> In attempting to make our robot work, we found that many everyday problems were much more complicated than the sorts of problems, puzzles, and games adults consider hard. At every point, in that world of blocks, when we were forced to look more carefully than usual, we found an unexpected universe of complication. (Minsky 1986: 29)

His eventual deduction from the problem is that intelligence is the way it is because it is actually an emergent phenomenon from the intricate and unconscious processes of relationships that reside in the body. The main problem here, then, can be identified as the disconnectedness from the real world and lack of a body. Without an actual body through which one can experience the world, it simply is not a plausible argument to assume that human-level intelligence can be developed in artificial agents. Moreover, it is deserving of recognition that the early cybernetic scholars managed to simulate a set of behaviors within such relatively simple systems while the entire cognitivist (and to an extent connectionist[30]) paradigm failed to do so. The reader may observe this as an unfair judgement as the cognitivist paradigm, as stated in earlier sections, was not concerned at all with the body. There exist, however, a few examples where researchers tried to build robots

[30] While connectionism was biologically inspired and had criticized the cognitivist paradigm for its lack of concern regarding the neural organization of a given system, it also disregarded the importance of a body and the systems interaction with the physical world in a similar manner (Sharkey and Ziemke 2001a: 711).

controlled solely by reasoning capacities. One such example is Shakey, a robot built in Stanford Research Institute in 1969, and its reasoning program STRIPS (short for Stanford Research Institute Problem Solver). Shakey was designed to perceive its environment by taking pictures via a television camera and perform actions that would apparently require a certain degree of reasoning and planning before executing them. The important point here, however, is that the creators of Shakey focused less on the environment-agent interaction and very much more on the internal processing and reasoning of its controlling program which, in accordance with the cognitivist paradigm, was based on formalization and rule-like thinking (Pfeifer and Bongard 2007: 40). Consequently, for Shakey to execute any given command, it required a special environment built according to its internal representation system. Moravec (1988: 14) describes Shakey as "an impressive concept but pitiable in action" as the difference between the work put in its reasoning ability and its ability to perform and act was simply too great[31]. This was best observable when Shakey was asked to perform a task that required it to reason *based on* the sensory input it received (for example pushing a block). Performing such tasks always came with a high chance of failure and the completion of them often involved performing them over and over, which would require hours of computation and reasoning, even for such a simple looking task.

Eventually, some researchers realized that in order for AI to perform reliably in real-world scenarios, it would require direct interaction with the physical world, namely a body, and that the focus on formalized internal structures based on symbol manipulation was misguided. Rodney Brooks was one of the first people to claim that such an approach to intelligence was not likely to bear fruit and suggested that instead of building robots with a "top-down" approach in "toy"[32] worlds, the researchers should be more invested in "bottom-up" strategies, where it is possible to deal with the basic set of skills necessary to interact and operate within the "real" world (Brooks 1986: 5-7). It was seen as a necessity for AI to be given a body, through which it can experience the world for itself and be able to exhibit intelligent behavior accordingly. The paradigm, which was called Nouvelle AI by Brooks (1990), would later be named embodied AI, signifying the importance of the body for intelligence to be able to emerge.

[31] It is also noted in Moravec (1988) that the job of developing sensory and mobility capabilities were given to junior researchers as they were not of primary concern for the developers.

[32] Brooks defines "toy" as all existing worlds where it is not up to the AI system itself to do all the understanding for itself, without relying on the human interpreter (Brooks 1986: 5).

While I cannot review the whole body of the embodied paradigm within the scope of this work, there are nevertheless some points that I consider to be useful to be made clear before moving on to theories that were considered appropriate for the embodied paradigm.

### 3.2.1. Symbol grounding problem

Symbol grounding is one of the more essential and challenging problems of AI research, especially for the embodied paradigm. The problem is first articulated by Searle in 1980 (see subsection 3.1.3.) to point out the fact that all meaning making processes that can be said to be taking place in AI is illusory and is the result of the interpreters' own analysis of the situation. The main reason for the lack of understanding in AI is because they do not have 'first-hand semantics' (Sharkey and Ziemke 2001a: 702), or 'intentionality' (Searle 1980: 427), to make something a sign *for* the system rather than making something a sign in the eye of the beholder.

Symbol grounding problem has been addressed by many researchers, but I will mention here some key figures who differed substantially in their views. The first one is Rodney Brooks. Brooks recognized the problem with symbol manipulation and described it as inadequate to represent the world the system was dealing with because it assumed an objective truth about the world which, in turn, was chosen deliberately to make the reasoning process manageable (Brooks 1990: 4-5). He suggested that, in line with Nouvelle AI, representations must be physically grounded in real world rather than be abstract, objectively formalized symbols of the environment the robot is going to operate in. Naturally, this approach also required the system to have a body with actuators and sensors. All the knowledge the system has must be obtained via physical action and all the goals the system has must be expressed with physical action as well. To physically ground the system, Brooks developed a new architecture called "subsumption architecture" which, in very simple terms, allowed the system to prioritize which actions to take based on the sensory input it received from the environment (i.e. if the system is asked to move towards a certain source of sensory input it can detect, the moving action would be subsumed upon detecting the relevant

input to perform another action, such as turning towards the source or avoiding the obstacles in its way) [33].

The second researcher that proposed an alternative way to ground symbols in AI is Stevan Harnad[34]. Harnad approaches the problem directly from Searle's Chinese room thought experiment and, in agreement with Searle, he concludes that the way symbols are used in the cognitivist paradigm are not intrinsic to the system itself, but rather parasitic (that they only have meaning for the interpreter). He suggests that, for a system to be capable of an intrinsically dedicated symbol system, it must be able to distinguish between iconic and categorical representations between which concepts are grounded. The grounded concepts are then fed to a neural network from which new concepts can be recognized by generalizing, based on what the system is already able to identify. An example he provides in this regard is as follows: (1) suppose that the system is able to identify, from learned sensory experience, the concept "horse" by name and iconic or categorical representation, (2) also suppose that the concept "stripes" is similarly grounded, (3) the system would be able to constitute a valid concept of "zebra" by combining and generalizing the elementary categorizations of the first two concepts. In this way "zebra" is equally grounded or, in his words, "inherits the grounding from the initial concepts" (Harnad 1990: 341).

It is important to note here that Harnad's view also requires the behavioral aspect, i.e. a body and learning from interaction with its environment, to be realized in its totality. Another view on symbol grounding is articulated by Sharkey and Ziemke. They, however, approach grounding from a different perspective. It is noted in their work that the approach to grounding meaning in AI research was inspired by biological organisms, which are the only beings in our world to exhibit any kind of intelligent behavior. Thus, the body, generally equipped with sensory and actuator mechanisms, often came to be interpreted as the right way of grounding meaning in the physical world (Sharkey and Ziemke 1998: 381), even for the cognitivist paradigm that simply requires connecting the abstracted symbols to the environment (i.e. Shakey). While this is certainly an improvement over the initial disembodied models of AI, therein lies the potential risk of *wishful*

[33] Subsumption architecture is often criticized for not having the capacity to learn, as the apparent autonomy it provides on the operational level is ultimately pre-determined by the designer.

[34] Stevan Harnad (1945-) is a professor of cognitive science at University of Quebec in Montreal and professor of web science at University of Southampton in UK.

*attribution*. The analogy between a robotic body and the body of a biological organism may seem almost synonymous at first but the difference is actually tremendous. The robot body is just a box, fitted with sensory and mobility units, that is housing its controller program while we observe multi-cellular solidarity in the totality of the biological organisms. The biological organism is a purposeful unit, which is shaped by its bodily structure, and is able to display appropriate behavior. Physically grounding meaning through embodiment and claiming to have achieved the same level of autonomous behavior as in biological organism is thus illusory. This is mainly due to the fact that biological organisms are not only physically grounded and situated in the here and now of their environment but also *rooted* in them which is the result of a long process of mutual interaction and structural coupling (i.e. an Umwelt). Following this argument then, it can be asserted that robots can never be *rooted* in this world simply because they are not of this world and any behavior they display such as goal seeking or obstacle avoiding essentially has no meaning for the robot itself (Sharkey and Ziemke 1998: 383).

### 3.2.2. Turing test

Alan Turing, in his influential paper "Computing Machinery and Intelligence" (1950) proposed a way to evaluate whether the system in question can be counted as intelligent[35]. His original test, which he calls the imitation game, consists of two people getting into different rooms and sending out typewritten answers to some questions that another person asks. This third (who also asks the questions) person then tries to guess who is who based on the answers he receives. Afterwards, Turing asked what would happen if we were to replace one person with an AI and whether, based on the written answer received, the third person could distinguish between the person and AI (Turing 1950: 433-434). The modern version of the Turing test is simply chatting with an AI through an interface and coming to a positive or negative decision. Turing was one of the people whose work implied that a thinking mind could be achieved by computation and accordingly one of the areas to be studied rigorously in the cognitivist paradigm is human language and speech.

[35] There is a continuous debate regarding what Turing actually meant in his paper but for the sake of simplicity, it is assumed in this section that he at least proposed a certain way intelligence should be understood.

While Turing's thesis of computation and Turing Test has received quite a lot of attention[36], there are scholars who think it has either been misinterpreted or is inadequate in evaluating intelligence the way it is proposed and applied.

Daniel Dennett is of the scholars claiming that Turing Test has been misinterpreted. He claims that Turing proposed his test essentially as a conversation stopper and not as a method for confirming or refuting scientific theories. Furthermore, he states his interpretation of what Turing actually proposed as:

> Instead of arguing interminably about the ultimate nature and essence of thinking, why don't we all agree that whatever that nature is, anything that could pass this test would surely have it; then we could turn to asking how or whether some machine could be designed and built that might pass the test fair and square. (Dennett 2004: 296)

I believe his view is, more or less, on par with what Moravec and others have proposed. If it is by language Turing Test evaluates whether a machine can think, then as the only species to have developed this sign system, which presupposes the existence of less complex ones following the history of biological evolution, we serve as our own proof demonstrating the validity of the intended level of intelligence. Thus, by analogy, any other system or organism to have developed a sign system equivalent of our language would have to have followed similar, if not the same, evolutionary steps, marking them as intelligent regardless of their origin.

The problem, however, is that not many people appear to have understood Turing Test in a similar manner Dennett did. Turing Test is often understood very literally and as a certain type of prescribed direction intelligence should take. Such understanding has resulted in what Dennett calls "operationalism"[37] and, accordingly, researchers approached intelligence as something that can be achieved by breaching a barrier. There are also instances of claiming an AI to be intelligent by "fooling" a judge, but fault lies on both sides in such instances as the person evaluating the

[36] A competition, Loebner Prize, organized annually, starting in 1991, has also contributed to the popularization of Turing Test. Initially it was organized to help further the AI research but during the course of years it has also caused controversy in various aspects.

[37] Dennett describes operationalism as "the tactic of defining the presence of some property, for instance, intelligence, as being established once and for all by the passing of some test" (Dennett 2004: 300).

program has to have a certain degree of competence and creativity to force the program to not only imitate human language in specific contexts but also display the capacity to exhibit general intelligence through language. Ultimately, Dennett's interpretation of Turing Test implies that there may be inhuman ways of possessing intelligence and passing a test with the sole objective of passing it to prove something, often leads to illusory results (or is wishful attribution). Stevan Harnad, one of the researchers to claim that Turing Test, with the way it is proposed and applied, is inadequate for evaluating whether a machine is intelligent or is capable of understanding. He states that one of the reasons that Turing made the test commence only on verbal basis was because he did not want the evaluating person to be biased by what the candidate looked like. However, denying an AI the existence of a mind based on what it looks like is not a valid reason. Moreover, the limitation to verbal communication also suggests that language is chosen because: (1) it can be reduced and imitated (to an extent) with symbol manipulation alone and, (2) it is a clear representative of human level intelligence. While being a competent language user can indeed be attributed to possessing intelligence, it is by no means the only way intelligent behavior is performed. Recognizing, identifying and manipulating the objects and the environment is also an indication of intelligence (Harnad 1991: 44). So, Harnad suggests that, instead of relying on the original Turing test which is disembodied and prone to certain errors and limitations, we should start evaluating the systems in real world environments where other intelligent behavior can also be tested. This way the system can be tested against everything a person can do, instead of focusing on an isolated area representing the totality of intelligence. He calls his version of the test "Total Turing Test" and it automatically implies that the system should be a robot.

Perhaps the argument made by Lakoff and Johnson can be used to reinforce what Harnad has proposed (see Lakoff and Johnson 1999). Lakoff and Johnson suggest that our cognition is embodied and, correspondingly, it is reflected in the metaphors we frequently use in our language. Metaphors such as "grasping something" (as in understanding) have their roots in the bodily action of literally grasping an object. Furthermore, we are only able to utilize those metaphors because we have the capacity to perform the action they infer their meaning from in the first place. This view naturally imposes on the original Turing Test the impossibility for any AI system to utilize language in the same way humans do without having access to the features that enable them (i.e. grasping an idea would not emerge as a meaningful concept in an AI system without a unit to perform the grasping action first). If implemented by other and external means (designer) the

criticism made by Searle's Chinese room argument also apply in this context. Evaluation of intelligence through Total Turing Test, thus, provides a more accurate account of assessment than its initial counterpart, especially when the goal is set to be the equivalent of human level intelligence.

### 3.3. Theories of embodiment

The people involved in the emergence of the embodied paradigm often used examples from biological organisms in their arguments to indicate the necessity of the body. Brooks (1986: 7), for instance, referred to insect behavior, which was often conceived as not intelligent. He described them as very robust, being able to perform a variety of tasks in the dynamic world where there also are many obstructive factors (predation etc.). The actions that insects could perform with relative ease were hard to achieve in artificial systems. For this reason, many of the researchers either turned towards the works of authors who already described how biological organisms operate or inspired the emergence of theories that would help realize a similar level of autonomy in artificial systems as observed in biological organisms.

There are three main theories that concern themselves directly with embodiment (or thought have the natural implication of embodiment by the AI researchers). The first one is the enaction theory (see Varela et al. 1991) in cognitive science, inspired heavily by the works of scholars such as Humberto Maturana, Francisco Varela and Evan Thompson. The second one is the dynamical hypothesis of Timothy van Gelder also in cognitive science (see Van Gelder 1995, 1998). And the third one is the Umwelt theory of Jakob von Uexküll as described briefly in the first chapter. I will shortly describe the theories here and move on towards the types of embodiment as it is in those types the theories are applied, bridging the gap between theory and practice and focus on Umwelt theory for the remainder of this thesis.

It is important to note that the researchers differed greatly in to what extent they have utilized those theories in their works. While it can be easily assumed that, through the flow of this thesis, AI researchers have come to depend on those theories when building their own models, it might be the case that they have built their own approaches (and reached the same conclusions)

without any regard to mentioned theories. Hence, the reader should keep it in mind that the theories described in this section provided guidelines, not prescriptions.

Chronologically, the first theory to be the source of inspiration for embodied AI is von Uexküll's Umwelt theory. As was already mentioned before in the previous chapters, it was first utilized by Brooks. Umwelt theory is accepted as a theory of embodiment mainly because of its emphasis on the subject-environment relationship which, through the historical development of AI research, has come to be regarded as the pre-requisite of intelligence within the field. It also has a strong emphasis on the meaning-making processes of the subject but to what extent they are applied or *can be* applied remain questions for the next chapter.

The second theory to be regarded as appropriate in this category is the enaction theory in cognitive science. While often regarded as the modern counterpart[38] of von Uexküll's Umwelt theory, the progenitors of the enaction theory were unfortunately unaware of it. Enaction theory also has a high emphasis on the subject-environment relationship and the concept of subject as something that cannot be explained solely by cause and effect mechanisms. Rather, it is stressed that cognition is a phenomenon highly dependent on the material structure of the subject, which, in turn, determines the world the subject interacts with. Moreover, enaction theory also mentions the notion of autopoiesis as a crucial aspect of the living organisms where, unlike homeostatic systems with a fixed variable such as temperature, the variable to be maintained is the system's own self-organization which results in even more variables in order for the system to self-sustain.

The last theory to be used in embodied AI research is van Gelder's dynamical hypothesis. Dynamical hypothesis is actually a combination of two other theories, dynamic systems theory and the dynamical framework. It is essentially a theory to provide an alternative to the computational models of cognition while being highly mathematical and assuming a quantitative approach when compared to other theories that I briefly described above. Nevertheless, the proper domain of the dynamical hypothesis is described as natural cognitive agents that have a certain evolutionary history with an emphasis on the causal organization of those agents (van Gelder 1998: 619), which makes the dynamical hypothesis a relevant candidate for realizing embodiment in AI systems.

[38] It is important to note here that the works of Maturana and Varela, who initiated enactive movement within the cognitive science, are taken as the primary source of inspiration over other theories when it comes to embodiment. While I am unable to pinpoint the reason behind such logic, I suspect that the publication date and the discourse used in explaining the phenomena play an essential role in this differentiation, especially when compared to von Uexküll's work.

### 3.3.1. Types of embodiment

In this section, I will briefly describe certain types of embodiment as put forward by Wilson (2002) and Ziemke (2003) that bear relevance within the context of this thesis. While both authors essentially suggest that a body to operate within real-world environments is a priori of embodiment, apparently there are researchers who claim otherwise. For example, it is stated by Franklin (1997: 500) that "Software systems with no body in the usual physical sense can be intelligent. But they must be "embodied" in the situated sense of being autonomous agents structurally coupled with their environment". Although this statement strikes me as highly ambiguous and conflicted, the question of how bodies and environments other than physical can be embodied are not within the scope of this thesis and will not be pursued further. Therefore, the types of embodiment discussed below are meant to be compatible with the direction AI research has taken – with implications on physical bodies and real-world environments.

The most relevant argument Wilson makes with respect to the direction embodied AI paradigm is taking is that "cognition is for action". The argument is described as, in Wilson's words: "The function of the mind is to guide action, and cognitive mechanisms such as perception and memory must be understood in terms of their ultimate contribution to situation-appropriate behavior" (Wilson 2002: 626).

The argument stating that "cognition is for action" is also inherent not only in von Uexküll's work, especially when one considers how perception (equivalent of cognition in this argument) gives way to action in the context of functional cycle, but also in the current general direction the embodied AI paradigm is headed towards. An equivalent of Wilsons's "cognition is for action" can be seen in what Ziemke calls "Organismoid" embodiment. It is described as the view:

> […] that at least certain types of organism-like cognition might be limited to organism-like bodies, i.e. physical bodies which at least to some degree have the same or similar form and sensorimotor capacities as living bodies. (Ziemke 2003: 1307)

It is also possible to make the same connection here to von Uexküll's functional cycle and to the direction embodied AI is taking as Wilson's "cognition is for action" argument. Ziemke also makes

the distinction between a type of embodiment that separates it from the previous organism-like bodies from the actual bodies of organisms. "Organismic" embodiment, as he calls it, is described as: "[…] embodiment that holds cognition is not only limited to physical, organism-like bodies, but in fact to organisms, i.e. *living bodies*" (Ziemke 2003: 1308). He mentions that works of authors such as von Uexküll and Maturana are essentially used to describe this type of embodiment and that there is an essential difference between "organismoid" and "organismic" types of embodiment determined by the substance making up those bodies (i.e. living matter vs non-living matter).

Types of embodiment that I briefly described in this section essentially form the very basic structure of every 'autonomous' artificial agent that one can observe today. So, the theories are able to be given practical form with relative ease due to the deceptively simple nature of the mechanisms that lie underneath. The principle of relying on perception and action to act autonomously in a given environment can be seen in the core of many of the artificial agents, ranging from self-driving cars to autonomous delivery vehicles. While von Uexküll's work might be seen as completely compatible with those agents at the first glance, I will argue in the next chapter how such an assumption remains superficial in many aspects.

In short, I have demonstrated the early inclinations of AI research and its failure to deliver what it initially promised, resulting in, also in context with the criticism it was subjected to, the need to look for other ways to achieve intelligent behavior in artificial systems. This search, after stumbling a few more times on the way, has led to the realization that intelligence is not a phenomenon existing solely in the mind as a subject of Cartesian dualism, but rather is very much dependent on the apparently unconscious abilities organisms have developed over time due to evolution. Such an approach eventually made it apparent that to achieve the intended level of intelligence a body that can experience the world was required which is also what made it possible to integrate theories from other disciplines into AI research, one of them being Jakob von Uexküll's Umwelt theory. Those theories have eventually been subsumed by the new paradigm and turned into types of embodiment within which they are represented and realized in various forms such as self-driving cars and autonomous delivery vehicles.

## 4. The Evaluation of Umwelt Theory in Embodied AI and Its Potential Limitations

My hypothesis is that Umwelt theory, with the way it is proposed by von Uexküll, has limits in its usefulness in embodied AI research. In this chapter, I will mainly work towards proving my point through an analysis of Umwelt theory with regards to its essential parts described in the first chapter.

### 4.1. Functional cycle

Functional cycle is one of the essential parts of the Umwelt theory. It not only bridges the perceptual and operational aspects of the physical body, but also connects the inner and outer world of the organism, forming one of the founding pillars of the Umwelt theory. As I have mentioned in the second chapter, the concept of *Merkwelt* was used by Brooks (1986) first and the functional cycle as a whole has been emphasized in the works of Prem (1996, 1997, 1998). The reason functional cycle was used was because the researchers decided to use animals as models since they were intelligent agents and von Uexküll's work in this sense has not only provided the researchers with direction but also indicated the details of the close relationship between perception and action coupling in animal behavior. Consequently, the embodied paradigm came to use the coupling between perception and action in their systems.

Hence, the implementation of a part of Umwelt theory has played a certain role in the development of embodied AI research and caused relevant and permanent changes in the respective paradigm. However, after a certain period of time, it can be seen that the reference to von Uexküll's work has disappeared but the dependence on perception and action as a definitive part of AI has remained still. This issue has several reasons.

The first reason is the habit of reducing every idea into basic and easily applicable notions so as to be able to build upon it with relative ease. This practice is most common in the so-called hard sciences. In case of functional cycle, as soon as it was made intrinsic to the embodied AI paradigm, the initial way it was understood was lost and replaced with some simpler representations of its actual implications. Some authors defined the problem as stemming from the lack of an objective definition of intelligence. Steels (1995: 6), when discussing the issue, states

that as soon as an intelligent behavior is described in terms that make explicit the way it works (i.e. when a chess playing program is described as performing relatively deep searches in search space instead of playing chess), it stops being categorized as intelligent behavior. Thus, the original implications the functional cycle entailed become reduced to mere stimulus and response mechanisms often described in the works of Jacques Loeb (cf. Sharkey and Ziemke 1998, 2001a).

The second reason is that perhaps it was never intended to be used in the same manner as von Uexküll did when he introduced the notion of functional cycle. Brooks states, in his second article with the reference to von Uexküll, that:

> In some circles much credence is given to Heidegger as one who understood the dynamics of existence. Our approach has certain similarities to work inspired by this German philosopher, but our work was not so inspired. It is based purely on engineering considerations. (Brooks 1991: 148)

While it is not possible to simply infer from this statement[39] that Heidegger's work is synonymous with von Uexküll's, it is clear that the latter's work has some bearing on the former (cf. Buchanan 2008). Regardless, the emphasis on the engineering part necessitates a certain level of appropriation of the concepts and their realizations in artificial systems. The process of transferring notions might be seen as a valid case where the loss of meaning appears to be inevitable. Similarly, it is stated by Prem (1997: 8) that the use of functional cycles for engineering concerns might be of use when one is to design a minimalist architecture for environment interactions.

The last reason to be mentioned here is what I call the incompatibility issue, pointing out to a fundamental difference between concepts, or application of those concepts, based on the material composition they are subject to. When von Uexküll introduced the concept of functional cycle, it was meant to describe the unique and subjective Umwelten of each organism (or species) through the sensory experience (perception), which, in turn, gave way to appropriate behavior (operation). The uniqueness part is derived from the differences between the organisms' material structure and furthermore, the organization of the material structure upon which the receptor and

[39] Actually, Brooks here is referring to works of Dreyfus who has used Heidegger's work as his argumentative basis and has criticized the cognitivist paradigm in a similar manner Brooks did. Heidegger is also a more widely referred figure compared to von Uexküll in AI context. Even such notions as Heideggerian AI have been proposed (See Herrera and Sanz 2016).

actuator organs are formed. Consequently, the difference between the structure of organs the organism uses to perceive and act in its environment is what effectively causes any experience to become restricted (subjective) to the organisms with particular receptor and actuator organs. The subjective nature of any experience, therefore, cannot be explained away with mechanistic reductions of the observed behavior and remains inaccessible[40] to those that do not share the required material structure. In a similar vein, Nagel (1974: 437-438) mentions that there are concepts that one cannot understand, "even if the species lasted forever" he adds, simply because one's own structure does not permit him to operate with the concepts of the requisite type.

So, let me put things into perspective. When I talk about my hand and, say, the feeling I get when I move it through a pile of sand, I talk about a particular feeling that is shared among humans, who share the same material structure in their hands, once the mentioned action is performed[41]. Now, let us imagine a robot hand that has the exact size and capacity to move itself in exactly the same manner a human could. When performing the same action, would there be any feeling? Any sign that coming into contact with a particular substance might have some meaning? It is not really possible to reach a positive answer and the reason lies in the material structure of the robot hand. While the robot hand would be made of a material deemed fit by its designer, the skin has evolved in a specific way that involved continuous contact with the mentioned substance and thus has come to be able to detect it and attribute a certain feeling (and meaning) to it when touched.

The realization of apparently the same mechanisms in artificial forms does not mean that it serves exactly the same function. If anything, it makes evident that robots will never have the subjective experience and meaning an organism has, that they will never have an Umwelt. Then, does it make any sense to try and apply Umwelt theory to artificial agents in the first place? It probably does not, other than providing criticism for the claim that an artificial agent has achieved the same level of autonomy as an organism, and this results from the natural incompatibility of material structures and what they entail. At this point, what this claim means appears to be vague and not exactly clear, but such clarity can only be achieved after other examples from von Uexküll's work are demonstrated.

[40] Of course, this is not to say their Umwelten is completely inaccessible for then it would not be possible to model the behavior in question.

[41] My intention is to focus only on the tactile perception and ignore whatever social, cultural or personal connotations moving the hand through a pile of sand might have.

### 4.2. Likeness

As already mentioned in the first chapter, concept of "likeness" has quite a lot of importance in the work of von Uexküll. It is mainly used to indicate a conceptual correlation between either two organisms that interact with each other in some context (such as predator-prey relationship) or a certain aspect of an organism that has come to develop out of interaction with its environment. The concept of "likeness" is most applicable in the context of expert systems. To recapitulate, an expert system is defined as a system that specializes in an isolated domain of knowledge and many embodied expert systems operate within defined environments. While the early expert systems were criticized for being non-dynamic and brittle (they were also disembodied), more recent expert systems can perhaps be evaluated better by the concept of "likeness".

One of the more recent, and rather popular, expert systems are self-driving cars. They have a certain body that is made in a suitable manner for the environments they are going to operate in, namely roads. They are also provided with a general map of how roads connect (the satellite GPS) and various other means for sensory input such as cameras, sonar and LIDAR (light detection and ranging).

While one can easily assume by taking a look at the self-driving cars that they fit into the description of something that may have an Umwelt, it is problematic to do so. Let me start with why it might be assumed so. In Uexküll's work every organism is implied to have co-evolved either with other organism(s) or with its environment. This is also reflected in the respective organisms' material structure (remember also the relation between the fly, spider and its web mentioned in the first chapter). However, before continuing with co-evolution, it is important to point out the difference between von Uexküll's understanding and my own interpretation of his work. Von Uexküll, as a matter of fact, does not discuss co-evolution at all, but, as stated above, he provides examples in terms of "likeness". He also restricts and relates the use of "likeness" with notions such as 'archetypes' and 'meaning-plans', meaning that his understanding of "likeness" has more emphasis on the relational correspondence between entities rather than a direct overlap on the material morphology of the entities in question (von Uexküll 1982[1940]: 43-44). My interpretation of von Uexküll's work, on the other hand, is that there has to be a material correspondence, preceding the conceptual one, before any other semiotic mechanisms can come

into play. While this issue might not be explicitly articulated in von Uexküll's work, it can still be considered as inherent in other parts of his Umwelt theory (especially the functional cycle). Regardless, let me elaborate further as to why it might be thought so. First of all, it is quite important to recall that, in the first chapter, Umwelt theory was introduced with an emphasis on the history of biological thought and was defined as an account of the web model. It was also stated that web model's defining characteristics include recognition, co-existence and symbiosis. Therefore, one can suggest that Umwelt theory, by following the definitive parts of the model it is part of, also has a natural and inherent emphasis on co-evolution. Such generalization might strike one as superficial and it is rather easy to find it lacking in its certain aspects. However, the connection becomes more apparent when one includes in the equation Margulis's work on endosymbiosis. In her work, Margulis defines how organisms are formed by incorporating other organisms, or parts of other organisms. Her theory suggests that evolution has to naturally include not only conflict between organisms but also cooperation following that conflict. As a consequence, this process is highly reflected on the material structures of the organisms. It also enables the researchers to trace back this process, by following the material remnants, back to unicellular organisms and so forth. (Margulis 1998: 55-59).

Again, one might object to forming such relation either by stating co-evolution might not exist in the same manner in Umwelt theory or by simply stating that those are two different theories. However, von Uexküll states that:

> To be 'fly-like' means that the body structure of the spider has taken on certain characteristics – not from a specific fly, but rather from the fly's archetype. To express it more accurately, the spider's 'fly-likeness' comes about when its body structure has adopted certain themes from the fly's melody. (von Uexküll 1982[1940]: 66)

One can infer from this quotation that even though von Uexküll does not necessarily define a coevolutionary process as something prevalent in the material structure of the organism in question, he nevertheless concedes that it is in the material morphology of the organism that the concept of "likeness" is realized. If it is according to a building-plan the spider is able to weave its web for the fly, even before it encounters one, then this naturally brings about the question of how these building-plans are formed. Perhaps the most encompassing answer for this question would

be that the building-plan of the organism is encoded in its genes, which are also subject to certain evolutionary processes. However, von Uexküll does not explore the question in such depth. This is where Margulis's work enter the scene as it suggests that genes are the primary indicator of a co-evolutionary process through endosymbiosis which is evidently demonstrated by the organism's material structure. Therefore, Umwelt theory can be claimed to have an inherent but unarticulated emphasis on co-evolution.

Continuing with the self-driving cars, or any other expert-system that is embodied in a similar way, it can be claimed that by having the knowledge on how to operate in traffic, various forms of perception and the ability to act in a certain way, it possesses a certain "likeness" to the environment it is coupled with. One can also claim that there exists a certain form of evolution in the form of machine learning that is realized by repeating the action over time and achieving a somewhat smoother behavioral capacity.

Now, let us investigate why such expert systems are actually "not like" either their environments or the objects they interact within those environments and do not have their own Umwelten. It has already been stated that co-evolution has a high emphasis on the material structure of the system. If the self-driving car were to be able to co-evolve, it would inherently recognize and adapt accordingly to the environment it is supposed to operate in without the need to be directed or maintained by a human designer for human intentions. The claim that it may show the capacity for evolution in the form of machine learning is also a counterintuitive argument in itself because what is actually evolving in this case is the controlling program(s) of the self-driving car and not the actual body of the car, as is observed in the case of organisms with all the parts it uses for displaying intelligent behavior. Hence, such systems are described as allopoietic (as opposed to autopoietic) meaning that the level of self-organization in artificial systems and their bodies do not work the same way as it does in organisms, which is described as the result of a continuous relationship with its environment, but rather is the consequence of processes that are independent of the organization of the machine in its totality (i.e. parts are produced separately and then combined to build the car) (Sharkey and Ziemke 2001a: 733). Being an allopoietic system effectively prevents a system from showing the expected degree of "likeness" as the material structure of interacting parties has not co-evolved (or is lacking a certain history of interaction that would make their relationship meaningful) and unable to exhibit naturally inherent recognition.

A recent example of this issue can be seen in a report stating that slight changes to a street sign, for example attaching a couple of stickers to a stop sign, causes the mechanism behind its perception system to classify it as something else (see Ackerman 2017). While this is not an indication that the mentioned perception systems do not really work, it certainly points out to a limitation in their capacity to work properly. The problem here stems from the fact that there were no examples of the particular instance (stickers on a stop sign) in the data base the system uses to ground its evaluation and the impossibility of adding every possible variation of modification, to the mentioned data base, one can make on a street sign. In terms of "likeness", this amounts to the fact that self-driving cars do not bear the material and conceptual relation to their environment. Organisms, however, simply because they were co-evolved in this world and display "likeness" at a very fundamental and material level, are able adapt to situations that they have not encountered before with relative ease compared to artificial systems.

Incompatibility here, then, again results from the impossibility of achieving the intended level of intelligent behavior because the web model presenting the co-evolutionary processes in biological organisms, and the theory of von Uexküll relying on the aspects of this model, presupposes a certain material composition both within the organism and its environment.

### 4.3. Life task

The last part of the Umwelt theory that will be discussed in this chapter is the concept of life task. Von Uexküll describes life task as something that "consists of utilizing the meaningcarriers and the meaning-factors, respectively, according to their particular building-plan" (Uexküll 1982[1940]: 36). An organism's life task is essentially the combination of signs that the organism is able to recognize and interpret, demonstrated by its behavioral capacity accordingly. All of those, in turn, are determined by its material nature.

Artificial systems do not have the capacity to inherently recognize something as a sign. This issue has already been addressed by Searle (see section 3.1.3.) as a problem of "intentionality" and as a problem of "first-hand semantics" by Sharkey and Ziemke (see section 3.2.1.). All interpretation capacity in an artificial system is put there by its designers. This has two implications with regard to the current argument: (1) that all artificial systems are heteronomous (as opposed to

autonomous), i.e. that signs do not have any meaning *for* the artificial system but for the creators of those systems as outside observers and (2) that those systems actually act as extensions of the Umwelten of their creators (see also von Uexküll 1992[1957]: 388-390).

Self-driving cars can again be taken as examples. The functions one tends to ascribe to a self-driving car, such as driving, actually have no inherent meaning for the system itself. Driving is a human activity and all the semiotic meaning the activity has is due to a history of interactions that formed the environment those systems are currently operating in. What the artificial system is doing, when it looks like it is driving in the sense that a human might, is actually just moving from one place to another using certain rules and regulations telling it what to do. It has no inherent understanding of a human, a traffic sign or any other part of traffic that might turn this moving process into driving. While a counterargument of such a claim may be that this is exactly what humans do, one should bear in mind the case of modified traffic signs in the previous section and the fact that the presented issue has no effect on human drivers because it is a sign for humans in a system built for humans.

One might also argue that the case presented above is not strictly applicable to all the artificial systems because it is strictly limited to human connotations and does not necessarily put an emphasis on the material structure of the systems in question. A simpler, and perhaps a more accurate one as well, example would be the artificial bees (see Wyss Institute 2017). Regardless of the reason they are developed and being used, those systems represent the human understanding of a bee which, in most contexts, would mean that it is a means of natural pollination. However, it is only an effect humans observe from their restricted points of view, and perhaps not something bees themselves pay deliberate attention to. What is happening here is that the meaningful existence of a bee is being reduced to a single and extrinsic function, also without paying any attention to how bees are situated in their own ecosystems (i.e. a bee might be a food source for another organism as a result of the co-evolutionary process described in the previous section). As a result, those artificial bees do not have any meaning for absolutely anything in their environment and the only function they can perform is put there by human designers for human ends. I have briefly mentioned the difference between the concepts of grounding and being rooted in section 3.2.1. The concept of life task can also be used to demonstrate that artificial systems can be grounded (that they can operate dependent on the physical input) but not rooted. What being rooted essentially means is that a system is embedded in a given environment and is not just able to

perform pre-given tasks. Being rooted in this sense also involves that all the meaning the system infers from in its environment is *for* itself and not just to serve some external end. This self-serving nature is due to the system's own material structure which has co-evolved in correspondence with its environment and thus enables it to create meaning for its own needs.

This is also another instance of incompatibility as we cannot make the physical artificial systems rooted in our environments simply because it is not possible to simulate physical evolution and somehow bring them up to speed, matching the biological organisms.

### 4.4. Incompatibility

I believe I have demonstrated, with examples, that the issue of incompatibility is essentially a problem of morphological and material structure and Umwelt theory especially provides us with insight when it comes to the differences those material structures causes (i.e. differing Umwelten) even when the material structure of all the organisms is the same at the most basic level (i.e. all intelligent life on the Earth is carbon based).

A brief example from the field of artificial life might be used to illustrate this point better[42]. Chris Langton states, when elaborating on the nature of life, that: "Life is a property of *form*, not *matter;* a result of the organization of the matter rather than something that inheres in the matter itself" (Langton 1989: 41). Claus Emmeche's response to his view is that form and matter are interdependent and not medium independent (Emmeche 1992: 473) and following this line of thought one can claim that there is an incompatibility between the physical and essentially "dead material" that is being used to build robots and the "living material", which is essentially a combination[43] of cells every single one of which is said to be capable of interpretation (Hoffmeyer 2013: 151-153), that constitutes the living bodies. In a similar vein, perhaps it would be better to mention Hoffmeyer's depiction of the body as a swarming body. He suggests that, since every cell

[42] Perhaps one important note here is that such approaches in AL was also adopted to AI (see Brooks and Steels 1995).

[43] von Uexküll also describes the interconnectedness between the cells using musical terminology (see first chapter).

is capable of interpretation and therefore recognition, there should not be a separation between body and mind as the body is the true mind (Hoffmeyer 1996: 115, 1997: 939; Emmeche et. al. 2002: 19). Consequently, such an approach classifies the brain as an organ where consciousness is realized as a result of the need to reach a consensus between the swarms of swarms that operate individually (but in a unified manner) in the body. However, it should be noted that while Hoffmeyer was familiar with von Uexküll's work and used it in his works when formulating his own theories, there are other scholars who reached similar conclusions, although by different means[44]. One such scholar is Michael Gazzaniga who, in his book *The Social Brain*, describes brain as a confederation where many mental systems co-exist (Gazzaniga 1985: 6). Similarly, Minsky also describes the brain as composed of plenty other unconscious mechanisms (Minsky 1986: 322). Regardless of how one reaches the conclusion, the emphasis on the brain as the prime organ for intelligence appears to be in decline in all cases while other structures that form the body are highlighted as the source of intelligent behavior. The effects of the difference between materials also becomes most evident in cases where it is expected, out of functional similarity, that the robot perform as intelligently in real life situations and in physical environments, which is a behavior displayed all the time almost effortlessly by any biological organism. Thus, the incompatibility between the natures of those materials has a direct effect on the capacity to display intelligent behavior in environments when the material in question does not have a native evolutionary history.

The implications of such incompatibility, however, is also reflected in the limits of the theory that is being used to explain how intelligent behavior can never be intrinsic to the systems that are designed to operate in real world environments. Since Umwelt theory suggests that: (1) intelligent behavior is the result of co-evolution that is both made available by and is the result of the material structure the system shares also both with other organisms and their environments and that (2) artificial systems, as they are essentially alien systems trying to operate on a physical reality without bearing any natural and material connection to it, are doomed to fail in this regard. The viability of Umwelt theory then, with respect to how one might realize the intended quality of

[44] The difference is caused by whether the researcher attributes the capacity to interpret to the smaller units (i.e. cells) that cause the function in question in the body/brain. Hoffmeyer does indeed attribute such capacity to those units while Minsky approaches them (and the functions that they make up) as unconscious.

intelligent behavior in artificial systems, becomes questionable because it suggests an inevitable end without providing any alternatives. In a similar context, Daniel Dennett states that:

> "It is rather as if philosophers were to proclaim themselves expert explainers of the methods of stage magicians, and then, when we ask how the magician does the sawing-the lady-in-half trick, they explain that it is really quite obvious: the magician doesn't really saw her in half; he simply makes it appear that he does.
>
> "But how does he do *that*?" we ask. "Not our department," say the philosophers." (Dennett 1984: 131).

To appropriate Dennett's statement into my argument, I can claim that application of Umwelt theory to AI research works in a similar way. It provides a method to refute any claim that artificial systems can behave as intelligently and autonomously as biological organisms. It does not just state what those systems appear to do but also gives valid ground to criticism by explaining how things work in biological organisms and how things do not work the same way in artificial systems. What it does not say, however, is how to proceed forward because there exists a certain barrier, in the heart of which lies the material nature of systems that I have discussed earlier, that prevents it from doing so.

If the implications of Umwelt theory are in essence unrealistic to achieve in physical artificial systems, then it is only natural that researchers either turn to other sources of inspiration or reduce those implications (such as the functional cycle) to basic and simpler realizations. This is also the reason why in the third chapter the direction AI research took was explained in such length. Regardless of the inspiration it could have or have taken from von Uexküll's work, it was already headed towards a certain direction that was outlined by Moravec, Minsky and others whose work was not so inspired by Umwelt theory. Another indication of such limitation is the amount of research that is concerned with building robots and the amount of references indicating a connection to von Uexküll's work in those research. I have not been able to find any indicating that connection[45].

To sum up, the issue of incompatibility that is seen in the material level is also reflected in the theory level which proves to have similar limitations. Thus, Umwelt theory, as proposed and

[45] Of course, one can say that Brooks' work was indicative of this connection but it is clear that he is more fascinated with organisms (i.e. insects) and transforming the qualities of those organisms for engineering concerns rather than von Uexküll's work itself and its direct applications.

discussed by von Uexküll, does not present a way to further the research on physical artificial systems (robots) and is bound to remain incompatible with its current limitations.

## Conclusion

With the growing interest towards AI research, the application of Umwelt theory has found its way into the relevant literature to explain the behavior of such systems better. However, the utilization of Umwelt theory mostly remained a theoretical concern, implicating a practical difficulty in its application to the systems in question which this work indicated. In this thesis, three different sub-questions were focused upon to be able to answer the question whether Jakob von Uexküll's Umwelt theory bears any relevance with respect to embodied AI research. While answering the first sub-question – "In which aspects has Umwelt theory been used in embodied AI research in previous literature?" –, it was established that Umwelt theory, or parts of Umwelt theory, have been used for mainly two purposes. In the first case, Umwelt theory was used by various scholars (e.g. Brooks, Prem and Clark) to describe how organisms function since they tried to base the intelligent behavior on the organisms from which the inspiration was being taken. Concepts such as the functional cycle became the main element to be incorporated into the embodied AI paradigm. Other concepts such as *Merkwelt, Innenwelt* and the fact that the organism occupies its own ecological niche has also been articulated and pointed out as defining aspects of how organisms function. In the second case, Umwelt theory was mainly utilized (e.g. Ziemke and co-authors and Emmeche) to provide a valid basis for criticism against the scholars who have claimed to have achieved the same level of intelligence one can observe in organisms. Such use mainly revolved around the fact that those realizations of so-called intelligent behavior in robots are actually not as intelligent as their biological counterparts and that Umwelt theory is an indicator of such a gap.

Regarding the second sub-question – "How did the developments in AI research lead to the adoption of theories (Umwelt theory and other theories pertaining to embodied cognition) between which no connection is apparent?" – it was demonstrated how, in its earlier years, AI research was focused on achieving human-level intelligence by manipulating formal symbols in rule-like environments and was thought that such intelligence could be achieved independent of the platform it was operating on. While such an approach proved some immediate results, which caused the

researchers to make grand claims like they would soon achieve human-level intelligence, it was soon proved that intelligence could not be as easily realized. Criticisms, along with limitations in funding, followed and the amount of research in the field was significantly decreased, which was called the AI winter afterwards. Most note-worthy figures who have criticized the field in this era are Hubert Dreyfus and John Searle and their criticisms appear to be valid even today. Research in AI repeated the same scenario after the first AI winter and it was only after the second AI winter that some researchers realized the need for a new direction if the field was to survive. Nouvelle AI, which was later named as embodied AI, was formulated by the works of various researchers such as Brooks, Moravec and Minsky as an alternative approach to the traditional approach to intelligence. The distinguishing part of this paradigm was that it no longer described intelligence as a feature solely belonging to the mind but rather accepted that intelligence is an emergent phenomenon resulting from the constant interaction the subject performs with its environment. This approach also indicated that the subject should be a robot. Therefore, some researchers turned to theories where interaction with the environment by means of perception-action coupling are highlighted. Among the theories was von Uexküll's Umwelt theory as it provided a detailed account of the required type of interaction. Consequently, its enabling aspects were assimilated into the embodied paradigm and represented within certain types of embodiment that can be seen as simpler models of the advanced autonomous systems we observe today.

The third sub-question – "Can Umwelt theory make a substantial contribution to the current embodied AI paradigm?" – found its answer in the discussion of three distinctive parts of Umwelt theory which were identified to be of utmost importance for AI research, specifically the concepts of functional cycle, likeness and life-task. Starting from the functional cycle, it was argued, for each of the above-mentioned concepts, that there is a material composition underlying their realizations in organisms. The possibility of applying and realizing these concepts in physical artificial systems such as self-driving cars have been discussed in comparison to organisms, and with respect to concepts such as co-evolution, and claimed as impossible to achieve or dissimilar in their realizations. The reason for the difference can be rationalized with the concept of incompatibility which presupposed the existence of a certain material composition in the substance that made up the agents in question, be it biological or artificial, before any of the previously stated concepts can be realized. Consequently, the issue of incompatibility presented while applying the parts of Umwelt theory to physical artificial systems, is similarly reflected in Umwelt theory's

totality which marked its limitation and inability to make a substantial contribution to the current embodied AI paradigm.

Therefore, regarding the main research question of the thesis – "Does umwelt theory bear any relevance in the context of embodied AI research?"–, the conclusion was negative. This is due to the fact that Umwelt theory is essentially a theory to explain the teleological nature of behavior the organisms display with strong emphasis on their material structures. Such theory becomes immediately incompatible when applied to physical artificial systems because those systems do not share the material nature of the environments they are operating in. Consequently, it makes no sense to build a system based on Umwelt theory and then inevitably criticize it because it is unable to achieve the intelligent and autonomous behavior of the biological organisms it was initially based on. Although none of the researchers who pointed out to a connection between Umwelt theory and embodied AI research have directly articulated an incompatibility caused by the Umwelt theory itself, it becomes evident after one takes a look at the research where a robot is built – there are no cases of using Umwelt theory in such contexts. Hence, Umwelt theory is revealed to be limited in its usefulness, which is also a conclusion that distinguishes this thesis from previous works on the relations between embodied AI research and Umwelt theory.

In light of what has been said so far, it is plausible to argue that Umwelt theory is a specific theory that is tailored for biological organisms which eliminates the artificial systems as a natural candidate to which it can be applied in a constructive way. As a result, AI research has found its own way to the embodied paradigm of AI and will probably continue to do so without the help of Umwelt theory.

As for the future prospects, I would commit to the same fallacy I have criticized Umwelt theory for if I did not provide a vision for potential endeavors and just described an impossibility. Today there are disembodied programs that perform intelligently, and those programs have surpassed human understanding in the areas they are operating in. One such example is Google's AlphaGo, a program to have defeated the top human players in the game of go. Umwelt theory can be applied to analyze how those apparently disembodied programs are able to demonstrate intelligent behavior *if* it is modified to transcend the boundaries of physical materiality it is built upon. While the original theory is tailored to fit the organisms with physical bodies in real world environments, the object of study of the modified theory would be the programs with digital bodies in virtual environments where mathematics act as the medium. I believe both semiotics and

Umwelt research would benefit greatly from such perspectives into rather overlooked aspects of the possibility of alternative realities.

**References**


Ackerman, E. 2017. *Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms.* Website: IEEE Spectrum. Available https://spectrum.ieee.org/cars-thatthink/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learningalgorithms, 2017, August 4. Last visited 20th of April, 2018.

Barandiaran, X.; Moreno A. 2006. On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior* 14(2): 171-185.

Brentari, C. 2015. *Jakob von Uexküll: The Discovery of the Umwelt between Biosemiotics and Theoretical Biology*. Springer, Dordrecht.

Brooks, R. 1986. *Achieving Artificial Intelligence through Building Robots*. (Technical Report Memo 899.) Cambridge, MA: MIT AI Lab.

— 1990. Elephants Don't Play Chess. *Robotics and Autonomous Systems* 6: 3-15.

— 1991. Intelligence without representation. *Artificial Intelligence* 47: 139-159.

— 2017. What Is It Like to Be A Robot? Blog: Rodney Brooks: Robots, AI and other stuff. Available http://rodneybrooks.com/what-is-it-like-to-be-a-robot/, 2017, March 18. Last visited 20th of April, 2018.

Brooks, R.; Stein, L. A. 1994. Building Brains for Bodies. *Autonomous Robots* 1: 7-25

Brooks, R.; Steels, L. 1995. *The Artificial Life Route to Artificial Intelligence*. Lawrence Erlbaum Associates Publishers, New Jersey.

Brooks, R.; Breazeal, C.; Marjanovic, M.; Scassellati, B.; Williamson, M. M. 1999. The Cog Project: Building a Humanoid Robot. In Nehaniv, C. L. (ed.) *Computation for Metaphors, Analogy, and Agents.* New York: Springer. pp: 52-87.

Buchanan, B. 2008. *Onto-Ethologies: The Animal Environments of Uexküll, Heidegger, MerleauPonty and Deleuze.* State University of New York Press, Albany.

Clark, A. 1997. *Being There*. Cambridge, MA: MIT Press.

Dennett, D. C. 1984. Cognitive wheels: the frame problem of AI. In Hookway, C. (ed.) Minds, *Machines and Evolution: Philosophical Studies*. Cambridge University Press. pp: 129-151.

— 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* London, Allen Lane.

— 2004. Can Machines Think? In Teuscher, C. (ed.) *Alan Turing: Life and Legacy of a Great Thinker.* Springer. Berlin, Heidelberg. pp: 295-316.

Dreyfus, H. L. 1972. *What Computers Can't Do: A Critique of Artificial Reason.* Harper & Row, New York.

Emmeche, C. 1990. Kognition og omverden - om Jakob von Uexküll og hans bidrag til kognitionsforskningen. *Almen Semiotik* 2: 52-67.

— Life as an abstract phenomenon: Is artificial life possible? In Varela, F. J.; Bourgine, P. (eds.) T*oward a Practice of Autonomous Systems - Proceedings of the First European Conference on Artificial Life.* Cambridge, MA: MIT Press. pp: 466-474.

— 2001. Does a robot have an Umwelt? *Semiotica* 134 (1/4): 653-693.

— 2004. A-life, organism and body: The semiotics of emergent levels. In Bedeau, M.; Husbands, P.; Hutton, T.; Kumar, S.; Suzuki, H. (eds.) *Workshop and Tutorial proceedings.*

*Ninth International Conference on the Simulation and Synthesis of Living Systems*. Boston, Massachusetts. pp: 117-124.

— 2007. A biosemiotic note on organisms, animals, machines, cyborgs, and the quasiautonomy of robots. *Pragmatics & Cognition* 15 (3): 455-483.

— 2011. Organism and Body: The Semiotics of Emergent Levels of Life. In Emmeche, C.; Kull, K. (eds.), *Towards a Semiotic Biology. Life is the Action of Signs*. Imperial College Press, London. pp: 91-112.

Emmeche, C.; Kull, K.; Stjernfelt, F., 2002. *Reading Hoffmeyer, Rethinking Biology*. Tartu University Press.

Franklin, S. 1997. Autonomous agents as embodied AI. *Cybernetics and Systems* 25(8): 499-520.

— 2014. History, motivations and core themes. In Frankish, K.; Ramsey, W. M. (eds.), *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press. pp: 15-34.

Gazzaniga, M. S. 1985. *The Social Brain.* Basic Books, New York.

Gudwin, R. 1999. Umwelts and artificial devices. In da Costa R.; Santaella L. (eds.) *Seminário Avançado de Comunicacao e Semiotica: Novos Modelos de Representação: Vida Artificial e Inteligência Artificial*. São Paulo, Pontifícia Universidade Católica. pp: 51-56.

Harnad, S. 1990. The Symbol Grounding Problem. *Physica D* 42: 335-346.

— 1991. Other Bodies, Other Minds: A machine Incarnation of an Old Philosophical Problem. *Minds and Machines* 1: 43-54.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. The MIT Press.

Herrera, C.; Sanz, R. 2016. Heideggerian AI and the being of robots. In Müller, V. C. (ed.) *Fundamental Issues of Artificial Intelligence.* Springer. pp: 497-517.

Hoffmeyer, J. 1996. *Signs of Meaning in the Universe*. Indiana University Press.

— 1997. The swarming body. In Rauch, I.; Carr, G. F. (eds.) *Semiotics around the World: Synthesis in Diversity*. *Proceedings of the Fifth Congress of the International Association for Semiotic Studies, Berkeley 1994*. Mouton de Gruyter, Berlin. pp: 937-940.

— 2013. Why do we need a semiotic understanding of life? In Henning, B.; Scarfe, A. (eds.), *Beyond Mechanism. Putting Life Back into Biology*. Lexington Books. pp: 147-168

Lakoff, G.; Johnson, M. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books, New York.

Kull, K. 1999. Biosemiotics in the twentieth century: a view from biology. *Semiotica* 127(1/4): 385-414.

— 2000. Trends in theoretical biology: the 20th century. *Aquinas* 43(2), 235–249.

— 2001. Jakob von Uexkull: An introduction. *Semiotica* 134(1/4): 1-59.

— 2003. Ladder, tree, web: The ages of biological understanding. S*ign Systems Studies* 31.2: 589-603.

Langton, C. G. 1989. Artificial Life. In Langton, C., G. (ed.), *Artificial Life*. Addison-Wesley. pp: 1-47.

Lighthill, J. 1973. Artificial Intelligence: A General Survey. In *Lighthill Report: Artificial Intelligence: a paper symposium*. Science Research Council.

Lindblom, J.; Ziemke, T. 2007. Embodiment and social interaction: A cognitive science perspective. In Ziemke, T.; Zlatev, J.; Frank, R. M.(eds.) *Body, Language and Mind Volume 1: Embodiment.* Mouton de Gruyter. pp: 129-166.

Magnus, R. 2008. Biosemiotics Within and Without Biological Holism: A Semio-historical Analysis. *Biosemiotics* 1: 379-396.

Margulis, L. 1998. *The Symbiotic Planet: A New Look at Evolution.* Weidenfeld & Nicolson, London.

McCarthy, J. 1974. Review of "Artificial Intelligence: A General Survey". *Artificial Intelligence* 5 (3): 317-322.

McCorduck, P. 2004. *Machines Who Think. A personal Inquiry into the History and Prospects of Artificial Intelligence*. A K Peters, Natick, Massachusetts.

Minsky, M. 1986. *The Society of Mind*. Simon & Schuster, New York.

Moore, G. 1965. Cramming more components onto integrated circuits. *Electronics* 38 (8): 114117.

Moravec, H. 1988. *Mind Children. The future of Robot and Human Intelligence*. Harvard University Press.

Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* Vol.83 (4): 435-450.

Nilsson, N. J. 2010. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press.

Nöth, W. 1997. Representation in semiotics and in computer science. *Semiotica* 115(3/4): 203213.

— 2001.Semiosis and the Umwelt of a robot. *Semiotica* 134(1/4): 695-699.

— 2003. Semiotic Machines. *S.E.E.D. Journal* (3): 81-99.

Pfeifer, R.; Bongard, J. 2007. *How the Body Shapes the Way We Think*. The MIT Press.

Prem, E. 1996. Elements of an Epistemology of Embodied AI. In *Proceedings of the AAAI Fall Symposium on Embodied Cognition and Action.* AAAI Press. Menlo Park, CA. pp: 97-101.

— 1997. Epistemic autonomy in models of living systems. In Husbands, P.; Harvey, I. (eds.) *Fourth European Conference on Artificial Life (ECAL 97)*. Cambridge, MA: MIT Press. pp: 2-9.

— 1998. Semiosis in embodied autonomous systems. In *Proceedings of the IEEE International Symposium on Intelligent Control*. Piscataway, NJ: IEEE. pp: 724-729.

Ruse, M. 1996. *Monad to Man: The Concept of Progress in Evolutionary Biology*. Harvard University Press.

Searle, J. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3 (3): 417-457.

Sharkey, N. E.; Ziemke, T. 1998. A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science* 10 (3-4): 361-391.

— 2000. Life, Mind, and Robots: The Ins and Outs of Embodied Cognition. In Wermter, S.; Sun, R. (eds.) *Hybrid Neural Systems*. LNAI 1778. Springer Verlag, Berlin, Heidelberg. pp: 313-332.

—

2001a. A stroll through the worlds of robots and animals. *Semiotica* 134 (1/4): 701-746.

— 2001b. Mechanistic vs. Phenomenal Embodiment: Can Robot Embodiment Lead to Strong AI? *Cognitive Systems Research* 2(4): 251-262.

Steels, L. 1995. When are robots intelligent autonomous systems? *Robotics and Autonomous Systems* 15: 3-9.

Stengers, I. 2005. Introductory notes on an ecology of practices. *Cultural Studies Review* Vol 11 (1): 183-196.

Sun, R. 2014. Connectionism and neural networks. In Frankish, K.; Ramsey, W. M. (eds.), *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press. pp: 108-128.

Sørensen, M. H.; Ziemke, T. 2007. Agents without Agency? *Cognitive Semiotics* 0 (1): 102-124.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 49: 433-460.

Uexküll, Jakob von 1926[1920]. *Theoretical Biology*. New York: Harcourt, Brace

— 1982 [1940]. The Theory of Meaning. *Semiotica* 42(1): 25-82.

— 1992[1957]. A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica* 89 (4): 319-391.

— 2001 [1937]. The new concept of Umwelt: A link between science andhumanities. *Semiotica* 134(1/4): 111-123.

Van Gelder, T. 1995. What might cognition be, if not computation? *Journal of Philosophy* 91:

—
345-381.
1998. The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences* 21: 615-665.

Wilson, M. 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review* 9 (4): 625636.

Wyss Institute 2017. *Autonomous Flying Microrobots (Robobees).* Available https://wyss.harvard.edu/technology/autonomous-flying-microrobots-robobees/. Last visited 20th of April, 2018.

Varela, F. J.; Thompson, E.; Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press.

Ziemke, T. 1999. Rethinking grounding. In Riegler, A.; Peschl, M.; von Stein, A. (eds.) *Understanding Representation in the Cognitive Sciences*. Plenum Press, New York. pp: 177-190.

— 2001a. The Construction of 'Reality' in the Robot. *Foundations of Science* 6(1): 163-233

— 2001b. Are Robots Embodied? In Balkenius, C.; Zlatev, J.; Brezeal, C.; Dautenhahn, K.; Kozima, H. (eds.) *Proceedings of the First International Workshop on Epigenetic Robotics: Modelling Cognitive Development in Robotic Systems.* Lund University Cognitive Studies, vol. 85, Lund, Sweden. pp: 75-83.

— 2002. On the epigenesis of meaning in robots and organisms: Could a humanoid robot develop a human(oid) Umwelt? *Sign Systems Studies* 30.1: 101-111.

—

— 2003. What's that thing called embodiment? In Alterman, R.; Kirsh, D. (eds.) *Proceedings of the 25th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum. pp. 1305-1310.

2004. Embodied AI as Science: Models of Embodied Cognition, Embodied Models of Cognition, or Both? In Iida, F.; Pfeifer, R.; Steels, L.; Kuniyoshi, Y. (eds.) *Embodied Artificial Intelligence.* Heidelberg: Springer. pp. 27-36.

— 2008. On the role of emotion in biological and robotic autonomy. *BioSystems* 91: 401-408.

— 2016. The body of knowledge: On the role of the living body in grounding embodied cognition. *BioSystems* 148: 4-11.

## Kokkuvõte

### Omailmateooria asjakohasus kehastunud tehisintellekti uuringutes

Käesoleva magistritöö eesmärk on uurida, kas Jakob von Uexkülli omailmateooria omab tähtsust kehastunud tehisintellekti uuringutes. Töö keskendub omailmateooria ja kehastunud tehisintellekti uurimuste seostele, kuna mõlemad valdkonnad peavad subjekti ja keskkonna interaktsioone intellekti võtmeaspektideks. Töös antakse esmalt lühike ülevaade omailmateooriast ning selle kesksetest teemadest, paigutades selle ühtlasi teiste bioloogia teooriate taustale. Seejärel vaadeldakse, kuidas erineva taustaga uurijad on mõtestanud omailmateooriat seoses kehastunud tehisintellekti uuringutega, ning selles kontekstis tõstetakse esile omailmateooria kaks põhilist funktsiooni. Selleks et põhjendada, miks kehastunud tehisintellekti uuringud üldse tekkisid, analüüsitakse tehisintellekti uuringute ajalugu kehastumise mõiste kujunemise taustal, arvestades ühtlasi tehisintellekti uuringute suhtes tehtud kriitikat ja selle valdkonna arenguid. Seejärel näidatakse, kuidas kehastunud tehisintellekti uuringutes võeti kasutusele omailmateooria. Töö lõpuosas väidetakse omailmateooria põhiseisukohtadele tuginedes, et organismid ja tehislikud agendid on ühildamatud. Näidatakse, et ühildamatuse probleem eksisteerib kõige fundamentaalsemal ja materiaalsel tasandil. Sellest järeldub, et omailmateooria rakendamine kehastunud tehisintellekti uuringutes on piirangutega ja ühildamatuse kaudu on need piirid näha ka omailmateoorias eneses, muutes selle kehastunud tehisintellekti uuringute jaoks mitte eriti oluliseks. Seda näitab ka kehastunud tehisintellekti ajalooline areng.

**Summary**

**The Relevance of Umwelt Theory in Embodied Artificial Intelligence Research.**

This thesis aims to evaluate whether Umwelt theory of Jakob von Uexküll has any relevance in embodied artificial intelligence (AI) research. The scope of the thesis is limited to Umwelt theory's relation with embodied AI research because they both has a natural emphasis on the subject's interaction with its environment as one of the core aspects of intelligence. The thesis starts with a brief introduction of Umwelt theory and its core aspects while also historically locating it among other biological theories. Afterwards it is demonstrated how scholars of various backgrounds have used Umwelt theory in relation to embodied AI research and it is revealed to have two basic functions in this context. In order to establish a stable position for how the embodied paradigm in AI research came about, the history of AI research is analyzed with respect to the emergence of the notion of embodiment along with the developments and criticisms AI research was subjected to. It is then exhibited how Umwelt theory came to be used in embodied AI research. In the last part, it is argued, through the core aspects of Umwelt theory, that there exists a certain incompatibility between biological organisms and the artificial agents through which embodiment is realized. The incompatibility issue is then indicated to exist at the most basic and material level. Consequently, it is claimed that Umwelt theory, within the way it is proposed and applied, has certain limits when it comes to furthering the research in embodied AI and, through the incompatibility demonstrated previously, such limits are reflected in the theory itself, marking it as not especially relevant as it is exhibited by the historical development of embodied AI.

Non-exclusive license to reproduce thesis and make thesis public

I, Halil İbrahim Yazıcı

_____,

(author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive license) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

The Relevance of Umwelt Theory in Embodied Artificial Intelligence Research

_____,

(title of thesis)

supervised by Dr. Riin Magnus

_____

Prof. Kalevi Kull

_____,

(supervisor's name)

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive license does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 22.05.2018